Bootstrapping factor-augmented regression models

Sílvia Gonçalves and Benoit Perron*

Département de sciences économiques, CIREQ and CIRANO, Université de Montréal

March 20, 2013

Abstract

This paper proposes and theoretically justifies bootstrap methods for regressions where some of the regressors are factors estimated from a large panel of data. We derive our results under the assumption that $\sqrt{T}/N \to c$, where $0 \leq c < \infty$ (N and T are the cross-sectional and the time series dimensions, respectively), thus allowing for the possibility that factor estimation error enters the limiting distribution of the OLS estimator as an asymptotic bias term (as was recently discussed by Ludvigson and Ng (2011)). We consider general residual-based bootstrap methods and provide a set of high-level conditions on the bootstrap residuals and on the idiosyncratic errors such that the bootstrap distribution of a rotated OLS estimator is consistent. We subsequently verify these conditions for a simple wild bootstrap residual-based procedure.

Keywords: factor model, bootstrap, asymptotic bias.

1 Introduction

Factor-augmented regressions where some of the regressors are estimated from a large set of data are increasingly popular in empirical work. Inference in these models is complicated by the fact that the regressors are estimated and thus measured with error. Recently, Bai and Ng (2006) derived the asymptotic distribution of the OLS estimator in this case under a set of standard regularity conditions. In particular, they show that the asymptotic distribution of the OLS estimator is unaffected by the estimation of the factors when $\sqrt{T}/N \to 0$, where N and T are the cross-sectional and the time series dimensions, respectively. While their simulation study does not consider inference on the coefficients themselves (they look at the conditional mean and forecast), they report noticeable size distortions in some situations.

*This paper is dedicated to the memory of Hal White, who was a great econometrician and a great mentor. Special thanks to the guest editor Norm Swanson and three anonymous referees for helpful comments. We also want to thank participants at the Conference in honor of Halbert White (San Diego, May 2011), the Conference in honor of Ron Gallant (Toulouse, May 2011), the Conference in honor of Hashem Pesaran (Cambridge, July 2011), the Fifth CIREQ time series conference (Montreal, May 2011), the 17th International Panel Data conference (Montreal, July 2011), the NBER/NSF Time Series conference at MSU (East Lansing, September 2011), and the Canadian Econometric Study Group (Toronto, October 2011) as well as seminar participants at HEC Montreal, Syracuse University, Georgetown University, Tilburg University, University of Rochester, Cornell University, University of British Columbia, University of Victoria, Texas A&M, Rice University and University of Pennsylvania. We also thank Valentina Corradi, Nikolay Gospodinov, Guido Kuersteiner and Serena Ng for very useful comments and discussions. Gonçalves acknowledges financial support from the NSERC and MITACS whereas Perron acknowledges financial support from the SSHRC and MITACS.

The main contribution of this paper is to propose and theoretically justify bootstrap methods for inference in the context of the factor-augmented regression model. A few other contributions consider the validity of the bootstrap in this context. Corradi and Swanson (2011) consider the bootstrap for forecast stability tests with factors, while Yamamoto (2011) considers the bootstrap for factor-augmented vector autoregressions (FAVAR) proposed by Boivin and Bernanke (2003). Finally, Shintani and Guo (2011) prove the validity of the bootstrap to carry out inference on the persistence of factors. Recent empirical applications of the bootstrap include Ludvigson and Ng (2007, 2009,2011) and Gospodinov and Ng (2011), where the bootstrap has been used in the context of predictability tests based on factor-augmented regressions without theoretical justification.

Our main contribution is to establish the first order asymptotic validity of the bootstrap for factor-augmented regression models under assumptions similar to those of Bai and Ng (2006) but without the condition that $\sqrt{T}/N \to 0$. Specifically, we assume that $\sqrt{T}/N \to c$, where $0 \leq c < \infty$, thus allowing for the possibility that factor estimation error affects the limiting distribution of the OLS estimator. As it turns out, when $c > 0$, an asymptotic bias term appears in the distribution, reflecting the contribution of factors estimation uncertainty. This bias problem was recently discussed by Ludvigson and Ng (2011), who proposed an analytical bias correction procedure. Instead, here we focus on the bootstrap and provide a set of conditions under which it can reproduce the limiting distribution of the OLS estimator, including the bias term.

The bootstrap method we propose is made up of two main steps. In a first step, we obtain a bootstrap panel data set from which we estimate the bootstrap factors by principal components. The bootstrap panel observations are generated by adding the estimated common components from the original panel and bootstrap idiosyncratic residuals. In a second step, we generate a bootstrap version of the response variable by again relying on a residual-based bootstrap where the bootstrap observations of the dependent variable are obtained by summing the estimated regression mean and a bootstrap regression residual. We provide a set of high level conditions on these bootstrap residuals and idiosyncratic errors that allow us to characterize the limiting distribution of the bootstrap OLS estimator under the assumption that $\sqrt{T}/N \to c$, where $0 \leq c < \infty$. These high level conditions essentially require that the bootstrap idiosyncratic errors be weakly dependent across individuals and over time and that the bootstrap regression scores satisfy a central limit theorem. We then verify these high level conditions for a residual-based wild bootstrap scheme under a stronger set of conditions.

A crucial result in proving the first order asymptotic validity of the bootstrap is the consistency of the bootstrap principal component estimator. Given our residual-based bootstrap, the "latent" factors underlying the bootstrap data generating process (DGP) are given by the estimated factors. These are not identified by the bootstrap principal component estimator due to the well-known identification problem of factor models. However, contrary to the rotation indeterminacy problem that affects the principal component estimator, this indeterminacy is easily resolved in the bootstrap world, where the bootstrap rotation matrix depends on bootstrap population values that are functions of the original data. As a consequence, to bootstrap the distribution of OLS estimator, our proposal is to rotate the bootstrap OLS estimator using the feasible bootstrap rotation matrix. This amounts to sign-adjusting the bootstrap OLS regression estimates asymptotically.

The rest of the paper is organized as follows. In Section 2, we first describe the setup and the assumptions, and then derive the asymptotic theory of the OLS estimator when $\sqrt{T}/N \rightarrow c$. In Section 3, we introduce the residual-based bootstrap method and characterize a set of conditions under which the bootstrap distribution consistency follows. Section 4 proposes a wild bootstrap implementation of the residual-based bootstrap and proves its consistency. Section 5 discusses the Monte Carlo results and Section 6 concludes. Three mathematical appendices are included. Appendix A contains the proofs of the results in Section 2, Appendix B the proofs of the results in Section 3, and Appendix C the proofs of the results in Section 4.

2 Asymptotic theory when $\sqrt{T}/N \rightarrow c$, where $0 \leq c < \infty$

We consider the following regression model

$$y_{t+h} = \alpha' F_t + \beta' W_t + \varepsilon_{t+h}, \ t = 1, \ldots, T - h, \tag{1}$$

where $h \geq 0$. The q observed regressors are contained in W_t. The r unobserved regressors F_t are the common factors in the following panel factor model,

$$X_{it} = \lambda_i' F_t + e_{it}, \quad i = 1, \ldots, N, \ t = 1, \ldots, T, \tag{2}$$

where the $r \times 1$ vector λ_i contains the factor loadings and e_{it} is an idiosyncratic error term. In matrix form, we can write (2) as

$$X = F\Lambda' + e,$$

where X is a $T \times N$ matrix of stationary data, $F = (F_1, \ldots, F_T)'$ is $T \times r$, r is the number of common factors, $\Lambda = (\lambda_1, \ldots, \lambda_N)'$ is $N \times r$, and e is $T \times N$.

The factor-augmented regression model described in (1) and (2) has recently attracted a lot of attention in econometrics. One of the first papers to discuss this model in the forecasting context was Stock and Watson (2002). Recent empirical applications include Ludvigson and Ng (2007) who consider predictive regressions of excess stock returns and augment the usual set of predictors by including estimated factors from a large panel of macro and financial variables, Ludvigson and Ng (2009,2011) who consider this approach in the context of predictive regressions of bond excess returns, Gospodinov and Ng (2011) who study predictive regressions for inflation using principal components from a panel of commodity convenience yields, and Eichengreen, Mody, Nedeljkovic, and Sarno (2012) who use common factors extracted from credit default swap (CDS) spreads during the recent financial crisis to look at spillovers across banks.

Estimation proceeds in two steps. Given X, we estimate F and Λ with the method of principal components. In particular, F is estimated with the $T \times r$ matrix $\tilde{F} = \begin{pmatrix} \tilde{F}_1 & \ldots & \tilde{F}_T \end{pmatrix}'$ composed of $\sqrt{T}$ times the eigenvectors corresponding to the r largest eigenvalues of of XX'/TN (arranged in decreasing order), where the normalization $\frac{\tilde{F}'\tilde{F}}{T} = I_r$ is used. The matrix containing the estimated loadings is then $\tilde{\Lambda} = \left(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_N\right)' = X'\tilde{F}\left(\tilde{F}'\tilde{F}\right)^{-1} = X'\tilde{F}/T$.

In the second step, we run an OLS regression of y_{t+h} on $\hat{z}_t = \begin{pmatrix} \tilde{F}_t' & W_t' \end{pmatrix}'$, i.e. we compute

$$\hat{\delta} \equiv \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \left(\sum_{t=1}^{T-h} \hat{z}_t \hat{z}_t'\right)^{-1} \sum_{t=1}^{T-h} \hat{z}_t y_{t+h}, \tag{3}$$

where $\hat{\delta}$ is $p \times 1$ with $p = r + q$.

As is well known in this literature, the principal components $\tilde{F}_t$ can only consistently estimate a transformation of the true factors F_t, given by HF_t, where H is a rotation matrix defined as

$$H = \tilde{V}^{-1} \frac{\tilde{F}'F}{T} \frac{\Lambda'\Lambda}{N}, \tag{4}$$

where $\tilde{V}$ is the $r \times r$ diagonal matrix containing on the main diagonal the r largest eigenvalues of XX'/NT, in decreasing order, see Bai (2003).

One important implication is that $\hat{\delta}$ consistently estimates $\delta \equiv \begin{pmatrix} \alpha' H^{-1} & \beta' \end{pmatrix}'$, and not $\begin{pmatrix} \alpha' & \beta' \end{pmatrix}'$. In particular, given (1), adding and subtracting appropriately yields

$$y_{t+h} = \underbrace{\begin{pmatrix} \alpha' H^{-1} & \beta' \end{pmatrix}}_{=\delta'} \underbrace{\begin{pmatrix} \tilde{F}_t \\ W_t \end{pmatrix}}_{=\hat{z}_t} + \alpha' H^{-1}\left(HF_t - \tilde{F}_t\right) + \varepsilon_{t+h},$$

or, equivalently,

$$y_{t+h} = \hat{z}_t'\delta + \alpha' H^{-1}\left(HF_t - \tilde{F}_t\right) + \varepsilon_{t+h}, \tag{5}$$

where the second term represents the contribution from estimating the factors.

Recently, Bai and Ng (2006) derived the asymptotic distribution of $\sqrt{T}\left(\hat{\delta} - \delta\right)$ under a set of regularity conditions and the assumption that $\sqrt{T}/N \to 0$. Our goal in this section is to derive the limiting distribution of $\hat{\delta}$ under the assumption that $\sqrt{T}/N \to c$, where c is not necessarily zero. We use the following assumptions, which are similar to Bai's (2003) assumptions and slightly weaker than the Bai and Ng (2006) assumptions. We let $z_t = \begin{pmatrix} F_t' & W_t' \end{pmatrix}'$, where z_t is $p \times 1$, with $p = r + q$.

Assumption 1 - Factors and factor loadings

(a) $E\|F_t\|^4 \leq M$ and $\frac{1}{T}\sum_{t=1}^{T} F_t F_t' \to^P \Sigma_F > 0$, where Σ_F is a non-random $r \times r$ matrix.

(b) The loadings λ_i are either deterministic such that $\|\lambda_i\| \leq M$, or stochastic such that $E\|\lambda_i\|^4 \leq M$. In either case, $\Lambda'\Lambda/N \to^P \Sigma_\Lambda > 0$, where Σ_Λ is a non-random matrix.

(c) The eigenvalues of the $r \times r$ matrix $(\Sigma_\Lambda \Sigma_F)$ are distinct.

Assumption 2 - Time and cross section dependence and heteroskedasticity

(a) $E(e_{it}) = 0$, $E|e_{it}|^8 \leq M$.

(b) $E(e_{it}e_{js}) = \sigma_{ij,ts}$, $|\sigma_{ij,ts}| \leq \bar{\sigma}_{ij}$ for all (t,s) and $|\sigma_{ij,ts}| \leq \tau_{ts}$ for all (i,j) such that $\frac{1}{N}\sum_{i,j=1}^{N} \bar{\sigma}_{ij} \leq M$, $\frac{1}{T}\sum_{t,s=1}^{T} \tau_{ts} \leq M$, and $\frac{1}{NT}\sum_{t,s,i,j} |\sigma_{ij,ts}| \leq M$.

(c) For every (t,s), $E\left|N^{-1/2}\sum_{i=1}^{N}(e_{it}e_{is} - E(e_{it}e_{is}))\right|^4 \leq M$.

Assumption 3 - Moments and weak dependence among $\{z_t\}$, $\{\lambda_i\}$ and $\{e_{it}\}$.

(a) $E\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_t e_{it}\right\|^2\right) \leq M$, where $E(F_t e_{it}) = 0$ for all (i,t).

(b) For each t, $E\left\|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T}\sum_{i=1}^{N} z_s (e_{it}e_{is} - E(e_{it}e_{is}))\right\|^2 \leq M$, where $z_s = \begin{pmatrix} F_s' & W_s' \end{pmatrix}'$.

(c) $E\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T} z_t e_t' \Lambda\right\|^2 \leq M$, where $E(z_t \lambda_i' e_{it}) = 0$ for all (i,t).

(d) $E\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \lambda_i e_{it}\right\|^2\right) \leq M$, where $E(\lambda_i e_{it}) = 0$ for all (i,t).

(e) As $N, T \to \infty$, $\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j' e_{it} e_{jt} - \Gamma \to^P 0$, where $\Gamma \equiv \lim_{N,T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \Gamma_t > 0$, and $\Gamma_t \equiv Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \lambda_i e_{it}\right)$.

Assumption 4 - weak dependence between idiosyncratic errors and regression errors

(a) For each t and $h \geq 0$, $E\left|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T-h}\sum_{i=1}^{N} \varepsilon_{s+h}(e_{it}e_{is} - E(e_{it}e_{is}))\right|^2 \leq M$.

(b) $E\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-h}\sum_{i=1}^{N} \lambda_i e_{it} \varepsilon_{t+h}\right\|^2 \leq M$, where $E(\lambda_i e_{it} \varepsilon_{t+h}) = 0$ for all (i,t).

Assumption 5 - moments and CLT for the score vector

(a) $E(\varepsilon_{t+h}) = 0$ and $E|\varepsilon_{t+h}|^2 < M$.

(b) $E\|z_t\|^4 \leq M$ and $\frac{1}{T}\sum_{t=1}^{T} z_t z_t' \to^P \Sigma_{zz} > 0$.

(c) As $T \to \infty$, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h} z_t \varepsilon_{t+h} \to^d N(0, \Omega)$, where $E\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h} z_t \varepsilon_{t+h}\right\|^2 < M$, and $\Omega \equiv \lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h} z_t \varepsilon_{t+h}\right) > 0$.

Assumption 1(a) imposes the assumption that factors are non-degenerate. Assumption 1(b) ensures that each factor contributes non-trivially to the variance of X_t, i.e. the factors are pervasive and affect all cross sectional units. These assumptions ensure that there are r identifiable factors in the model. Recently, Onatski (2011) considers a class of "weak" factor models, where the factor loadings are modeled as local to zero. Under this assumption, the estimated factors are no longer consistent for the unobserved (rotated) factors. In this paper, we do not consider this possibility. Assumption 1(c) ensures that $Q \equiv p\lim\left(\tilde{F}'F/T\right)$ is unique. Without this assumption, Q is only determined up to orthogonal transformations. See Bai (2003, proof of his Proposition 1).

Assumption 2 imposes weak cross-sectional and serial dependence conditions on the idiosyncratic error term e_{it}. In particular, we allow for the possibility that e_{it} is dependent across individual units and over time, but we require that the degree of dependence decreases as the time and the cross sectional distance (regardless of how it is defined) between observations increases. This assumption is compatible with the approximate factor model of Chamberlain and Rothschild (1983) and Connor and Korajczyk (1986, 1993), in which cross-sectional units are weakly correlated. Assumption 2 allows for heteroskedasticity in both dimensions and requires the idiosyncratic error term to have finite eighth moments.

Assumption 3 restricts the degree of dependence among the vector of regressors $\{z_t\}$, the factor loadings $\{\lambda_i\}$ and the idiosyncratic error terms $\{e_{it}\}$. If we assume that $\{z_t\}$, $\{\lambda_i\}$ and $\{e_{it}\}$ are mutually independent (as in Bai and Ng (2006)), then Assumptions 3(a), 3(c) with $z_t = F_t$ and 3(d) are implied by Assumptions 1 and 2. Similarly, Assumption 3(b) holds if we assume that $\{z_t\}$ and $\{e_{it}\}$ are independent and the following weak dependence condition on $\{e_{it}\}$ holds. For each t,

$$\frac{1}{T^2 N} \sum_{i=1}^{N} \sum_{s,q=1}^{T} \left| Cov\left(e_{it} e_{is}, e_{it} e_{iq}\right)\right| \leq M. \tag{6}$$

For a similar assumption, see Bai (2009, Assumption C.4). This assumption holds if e_{it} is i.i.d. over i and t and $E\left(e_{it}^4\right) < M$. Assumptions 3(a)-3(c) are equivalent to Assumptions D, F1 and F2 of Bai (2003) when $z_t = F_t$.

To describe Assumption 3(e), for each t, let

$$\phi_t \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{it}, \text{ and } \Gamma_t \equiv Var\left(\phi_t\right) = E\left(\phi_t \phi_t'\right),$$

since we assume that $E\left(\lambda_i e_{it}\right) = 0$ for all (i,t). Assumption 3(e) requires that $\frac{1}{T}\sum_{t=1}^{T}\left[\phi_t\phi_t' - E\left(\phi_t\phi_t'\right)\right]$ converges in probability to zero. This follows under weak dependence conditions on $\{\lambda_i e_{it}\}$ over (i,t).

Assumption 4 imposes weak dependence between the idiosyncratic errors and the regression errors. Part (a) holds if $\{e_{it}\}$ is independent of $\{\varepsilon_t\}$ and the weak dependence condition (6) holds. Similarly, part (b) holds if $\{\lambda_i\}, \{e_{is}\}$ and $\{\varepsilon_t\}$ are three mutually independent groups of random variables and Assumption 2 holds.

Assumption 5 imposes moment conditions on $\{\varepsilon_{t+h}\}$, on $\{z_t\}$ and on the score vector $\{z_t \varepsilon_{t+h}\}$. Part b) requires $\{z_t z_t'\}$ to satisfy a law of large numbers. Part c) requires the score to satisfy a central limit theorem, where Ω denotes the limiting variance of the scaled average of the scores. The dependence structure of the scores $\{z_t \varepsilon_{t+h}\}$ dictates the form of the covariance matrix estimator to be used for inference on δ. For instance, Bai and Ng (2006) assume that $\Omega = \lim \frac{1}{T}\sum_{t=1}^{T-h} E\left(z_t z_t' \varepsilon_{t+h}^2\right)$, and the appropriate covariance matrix estimator is a heteroskedasticity robust variance estimator. As we will see later, the form of Ω will also dictate the type of bootstrap we should use.

Given Assumptions 1-5, we can state our main result as follows. We introduce the following notation:

$$\begin{aligned} H_0 &\equiv p\lim H = V^{-1} Q \Sigma_\Lambda, \ \ Q \equiv p\lim \left(\frac{\tilde{F}' F}{T} \right), \ \ V \equiv p\lim \tilde{V}, \quad \text{and} \\ \Phi_0 &\equiv diag\left(H_0, I_q\right). \end{aligned}$$

Additionally, we let $\Sigma_{WF} = E\left(W_t F_t'\right)$.

Theorem 2.1 *Let Assumptions 1-5 hold. If $\sqrt{T}/N \to c$, with $0 \le c < \infty$, then*

$$\sqrt{T}\left(\hat{\delta} - \delta\right) \to^d N\left(-c\Delta_\delta, \Sigma_\delta\right),$$

where $\Sigma_\delta = \left(\Phi_0'\right)^{-1} \Sigma_{zz}^{-1} \Omega \Sigma_{zz}^{-1} \Phi_0^{-1}$, and

$$\begin{aligned} \Delta_\delta &\equiv \begin{pmatrix} \Delta_\alpha \\ \Delta_\beta \end{pmatrix} = \left(\Phi_0 \Sigma_{zz} \Phi_0'\right)^{-1} \begin{pmatrix} \Sigma_{\tilde{F}} + V \Sigma_{\tilde{F}} V^{-1} \\ \Sigma_{W\tilde{F}} V \Sigma_{\tilde{F}} V^{-1} \end{pmatrix} \left(H_0^{-1}\right)' \alpha, \ \ \textit{with} \\ \Sigma_{\tilde{F}} &\equiv V^{-1}\left(Q \Gamma Q'\right) V^{-1}, \\ \Sigma_{W\tilde{F}} &\equiv p\lim \left(\frac{W' \tilde{F}}{T} \right) = \Sigma_{WF} H_0'. \end{aligned}$$

If $\Sigma_{WF} = 0$, the asymptotic bias is equal to

$$-c\Delta_\delta = -c \begin{pmatrix} \left[\Sigma_{\tilde{F}} + V \Sigma_{\tilde{F}} V^{-1}\right] \left(H_0^{-1}\right)' \alpha \\ 0 \end{pmatrix}.$$

Theorem 2.1 gives the asymptotic distribution of $\sqrt{T}\left(\hat{\delta} - \delta\right)$ under the condition that $\sqrt{T}/N \to c$, where $0 \le c < \infty$. When $c = 0$, we obtain the same limiting distribution as in Bai and Ng (2006) under a set of assumptions that is weaker than theirs, as we discussed above. As in Bai and Ng (2006), factors estimation error does not contribute to the asymptotic distribution when $c = 0$. This is no longer the case when $c > 0$. Under this alternative condition, an asymptotic bias appears, as was recently discussed by Ludvigson and Ng (2011) in the context of a simpler regression model without observed regressors W_t. Our Theorem 2.1 complements their results by providing an expression for the bias of $\hat{\delta}$ when the factor-augmented regression model includes also observed regressors in addition to the unobserved factors F_t.

Several remarks follow. First, the expression for Δ_δ is proportional to $\left(H_0^{-1}\right)' \alpha = p\lim \hat{\alpha}$, implying that when $\alpha = 0$, no asymptotic bias exists independently of the value of c. Second, the asymptotic bias for both $\hat{\alpha}$ and $\hat{\beta}$ is a function of

$$\Sigma_{\tilde{F}} \equiv V^{-1} Q \Gamma Q' V^{-1} = \lim_{N,T \to \infty} \frac{1}{T} \sum_{t=1}^{T} V^{-1} Q \Gamma_t Q' V^{-1},$$

where $V^{-1} Q \Gamma_t Q' V^{-1}$ is the asymptotic variance of $\sqrt{N}\left(\tilde{F}_t - H F_t\right)$ (see Bai (2003)). Thus, the bias depends on the sampling variance-covariance matrix of the estimation error incurred by the principal components estimator $\tilde{F}_t$, averaged over time. This variance matrix depends on the cross sectional dependence of $\{e_{it}\}$ via $\Gamma_t \equiv Var\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_i\right)$. As we will see next, the main implication for the validity of the bootstrap is that it needs to reproduce this cross sectional dependence when $c \neq 0$ but not otherwise. Third, the existence of measurement error in $\tilde{F}_t$ contaminates the estimators of the remaining parameters β, i.e. $\hat{\beta}$ is asymptotically biased due to the measurement error in $\tilde{F}_t$. The asymptotic bias associated with $\hat{\beta}$ will be zero only in the special case when the observed regressors and the factors are not correlated (i.e. $\Sigma_{WF} = 0$) (or when $\alpha = 0$).

3 A general residual-based bootstrap

The main contribution of this section is to propose a general residual-based bootstrap method and discuss its consistency for factor-augmented regression models under a set of sufficient high-level conditions on the bootstrap residuals. These high level conditions can be verified for any bootstrap scheme that resamples residuals. We verify these conditions for a two-step wild bootstrap scheme in Section 4.

3.1 Bootstrap data generating process and estimation

Let $\{e_t^* = (e_{1t}^*, \ldots, e_{Nt}^*)'\}$ denote a bootstrap sample from $\left\{\tilde{e}_t = X_t - \tilde{\Lambda}\tilde{F}_t\right\}$ and $\{\varepsilon_{t+h}^*\}$ a bootstrap sample from $\left\{\hat{\varepsilon}_{t+h} = y_{t+h} - \hat{\alpha}'\tilde{F}_t - \hat{\beta}' W_t\right\}$. We consider the following bootstrap DGP:

$$y_{t+h}^* = \hat{\alpha}'\tilde{F}_t + \hat{\beta}' W_t + \varepsilon_{t+h}^*, \quad t = 1, \ldots, T-h, \tag{7}$$

$$X_t^* = \tilde{\Lambda}\tilde{F}_t + e_t^*, \quad t = 1, \ldots, T. \tag{8}$$

Estimation proceeds in two stages. First, we estimate the factors by the method of principal components using the bootstrap panel data set $\{X_t^*\}$. Second, we run a regression of y_{t+h}^* on the bootstrap *estimated* factors and on the fixed observed regressors W_t.

Because the bootstrap scheme used to generate y_{t+h}^* is residual-based, we fix the observed regressors W_t in the bootstrap regression. We replace $\tilde{F}_t$ with $\tilde{F}_t^*$ to mimic the fact that in the original regression model the factors F_t are latent and need to be estimated with $\tilde{F}_t$. This yields the bootstrap OLS estimator

$$\hat{\delta}^* = \left(\frac{1}{T}\sum_{t=1}^{T-h} \hat{z}_t^* \hat{z}_t^{*\prime}\right)^{-1} \frac{1}{T}\sum_{t=1}^{T-h} \hat{z}_t^* y_{t+h}^*. \tag{9}$$

$\hat{\delta}^*$ is the bootstrap analogue of $\hat{\delta}$, the OLS estimator based on the original sample.

3.2 Bootstrap high level conditions

In this section, we provide a set of high level conditions on $\{e_{it}^*\}$ and $\{\varepsilon_{t+h}^*\}$ that will allow us to characterize the bootstrap distribution of $\hat{\delta}^*$.

A word on notation. As usual in the bootstrap literature, we use P^* to denote the bootstrap probability measure, conditional on the original sample (defined on a given probability space $(\Omega, \mathcal{F}, P)$). Because the sample depends on N and T, as well as on the given sample realization ω, P^* is a random measure that depends on N, T and ω and we should write $P^*_{NT,\omega}$. However, for simplicity, we omit the indices on P^*. Similarly, we omit the indices NT when referring to the bootstrap samples $\{e^*_{it}, \varepsilon^*_{t+h}\}$. For any bootstrap statistic T^*_{NT}, we write $T^*_{NT} = o_{P^*}(1)$, in probability, or $T^*_{NT} \to^{P^*} 0$, in probability, when for any $\delta > 0$, $P^*(|T^*_{NT}| > \delta) = o_P(1)$. We write $T^*_{NT} = O_{P^*}(1)$, in probability, when for all $\delta > 0$ there exists $M_\delta < \infty$ such that $\lim_{N,T\to\infty} P[P^*(|T^*_{NT}| > M_\delta) > \delta] = 0$. Finally, we write $T^*_{NT} \to^{d^*} D$, in probability, if conditional on a sample with probability that converges to one, T^*_{NT} weakly converges to the distribution D under P^*, i.e. $E^*(f(T^*_{NT})) \to^P E(f(D))$ for all bounded and uniformly continuous functions f.

Condition A* (weak time series and cross section dependence in e^*_{it})

(a) $E^*(e^*_{it}) = 0$ for all (i,t).

(b) $\frac{1}{T}\sum_{t=1}^T \sum_{s=1}^T |\gamma^*_{st}|^2 = O_P(1)$, where $\gamma^*_{st} = E^*\left(\frac{1}{N}\sum_{i=1}^N e^*_{it}e^*_{is}\right)$.

(c) $\frac{1}{T^2}\sum_{t=1}^T \sum_{s=1}^T E^*\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N \left(e^*_{it}e^*_{is} - E^*(e^*_{it}e^*_{is})\right)\right|^2 = O_P(1)$.

Condition B* (weak dependence among $\hat{z}_t, \tilde{\lambda}_i$, and e^*_{it})

(a) $\frac{1}{T}\sum_{t=1}^T \sum_{s=1}^T \tilde{F}_s \tilde{F}'_t \gamma^*_{st} = O_P(1)$.

(b) $\frac{1}{T}\sum_{t=1}^T E^*\left\|\frac{1}{\sqrt{TN}}\sum_{s=1}^T \sum_{i=1}^N \hat{z}_s\left(e^*_{it}e^*_{is} - E^*(e^*_{it}e^*_{is})\right)\right\|^2 = O_P(1)$, where $\hat{z}_s = \begin{pmatrix} \tilde{F}'_s & W'_s \end{pmatrix}'$.

(c) $E^*\left\|\frac{1}{\sqrt{TN}}\sum_{t=1}^T \sum_{i=1}^N \hat{z}_t \tilde{\lambda}'_i e^*_{it}\right\|^2 = O_P(1)$.

(d) $\frac{1}{T}\sum_{t=1}^T E^*\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N \tilde{\lambda}_i e^*_{it}\right\|^2 = O_P(1)$.

(e) $\frac{1}{T}\sum_{t=1}^T \left(\frac{\tilde{\Lambda}' e^*_t}{\sqrt{N}}\right)\left(\frac{e^{*\prime}_t \tilde{\Lambda}}{\sqrt{N}}\right) - \Gamma^* = o_{P^*}(1)$, in probability, where $\Gamma^* \equiv \frac{1}{T}\sum_{t=1}^T Var^*\left(\frac{1}{\sqrt{N}}\tilde{\Lambda}' e^*_t\right) > 0$.

Condition C* (weak dependence between e^*_{it} and ε^*_{t+h})

(a) $\frac{1}{T}\sum_{t=1}^T E^*\left|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T-h} \sum_{i=1}^N \varepsilon^*_{s+h}\left(e^*_{it}e^*_{is} - E^*(e^*_{it}e^*_{is})\right)\right|^2 = O_P(1)$.

(b) $E^*\left\|\frac{1}{\sqrt{TN}}\sum_{t=1}^{T-h} \sum_{i=1}^N \tilde{\lambda}_i e^*_{it}\varepsilon^*_{t+h}\right\|^2 = O_P(1)$, where $E\left(e^*_{it}\varepsilon^*_{t+h}\right) = 0$ for all (i,t).

(c) $\frac{1}{T}\sum_{t=1}^{T-h} \sum_{s=1}^T \tilde{F}_s \varepsilon^*_{t+h} \gamma^*_{st} = O_{P^*}(1)$, in probability.

Condition D* (bootstrap CLT)

(a) $E^*\left(\varepsilon^*_{t+h}\right) = 0$ and $\frac{1}{T}\sum_{t=1}^{T-h} E^*\left|\varepsilon^*_{t+h}\right|^2 = O_P(1)$.

(b) $\Omega^{*-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h} \hat{z}_t \varepsilon^*_{t+h} \to^{d^*} N(0, I_p)$, in probability, where $E^*\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h} \hat{z}_t \varepsilon^*_{t+h}\right\|^2 = O_P(1)$, and $\Omega^* \equiv Var^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h} \hat{z}_t \varepsilon^*_{t+h}\right) > 0$.

Conditions A*-D* are the bootstrap analogue of Assumptions 1 through 5. However, contrary to Assumptions 1-5, which pertain to the data generating process and cannot be verified in practice, Conditions A*-D* can be verified for any particular bootstrap algorithm used to generate the bootstrap residuals and idiosyncratic error terms. More importantly, we can devise bootstrap schemes to verify these conditions and hence ensure bootstrap validity. For instance, part (a) of Condition A* requires the bootstrap mean of e_{it}^* to be zero for all (i,t) whereas part (a) of Condition D* requires that the same is true for ε_t^*. The practical implication is that we should make sure to construct bootstrap residuals with mean zero, e.g. to recenter residuals before applying a nonparametric bootstrap method. Parts b) and c) of Condition A* impose weak dependence conditions on $\{e_{it}^*\}$ over (i,t). For instance, these conditions are satisfied if we resample $\{e_{it}^*\}$ in an i.i.d. fashion over the two indices (i,t). Condition B* imposes further restrictions on the dependence among $\hat{z}_t$, $\tilde{\lambda}_i$ and the idiosyncratic errors e_{it}^*. Since $\hat{z}_t$ and $\tilde{\lambda}_i$ are fixed in the bootstrap world, Condition B* is implied by appropriately restricting the dependence of e_{it}^* over (i,t). Similarly, Condition C* restricts the amount of dependence between $\{e_{is}^*\}$ and $\left\{\varepsilon_{t+h}^*\right\}$. If these two sets of bootstrap innovations are independent of one another, then weak dependence on $\{e_{is}^*\}$ suffices for Condition C* to hold. Finally, Condition D* requires the bootstrap regression scores $\hat{z}_t\varepsilon_{t+h}^*$ to obey a central limit theorem in the bootstrap world.

3.3 Bootstrap results

Under Conditions A*-D*, we can show the consistency of the bootstrap principal component estimator $\tilde{F}^*$ for a rotated version of the true "latent" bootstrap factors $\tilde{F}$, a crucial result in proving the first order asymptotic validity of the bootstrap in this context.

More specifically, according to (7)-(8), the common factors underlying the bootstrap panel data $\{X_t^*\}$ are given by $\tilde{F}_t$ (with $\tilde{\Lambda}$ as factor loadings). Nevertheless, just as $\tilde{F}_t$ estimates a rotation of F_t, the estimated bootstrap factors $\tilde{F}_t^*$ estimate $H^*\tilde{F}_t$, where H^* is the bootstrap analogue of the rotation matrix H defined in (4), i.e.

$$H^* = \tilde{V}^{*-1}\frac{\tilde{F}^{*\prime}\tilde{F}}{T}\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}, \tag{10}$$

where $\tilde{V}^*$ is the $r \times r$ diagonal matrix containing on the main diagonal the r largest eigenvalues of $X^*X^{*\prime}/NT$, in decreasing order.

Lemma 3.1 *Let Assumptions 1-5 hold and suppose we generate bootstrap data* $\left\{y_{t+h}^*, X_t^*\right\}$ *according to the residual-based bootstrap DGP (7) and (8) by relying on bootstrap residuals* $\left\{\varepsilon_{t+h}^*\right\}$ *and* $\{e_t^*\}$ *such that Conditions A*-D* are satisfied. Then, as* $N,T \to \infty$,

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t^* - H^*\tilde{F}_t\right\|^2 = O_{P^*}\left(\delta_{NT}^{-2}\right),$$

in probability, where $\delta_{NT} = \min\left(\sqrt{N}, \sqrt{T}\right)$.

According to Lemma 3.1, the time average of the squared deviations between the *estimated* bootstrap factors $\tilde{F}_t^*$ and a rotation of the "latent" bootstrap factors given by $H^*\tilde{F}_t$ vanishes in probability under the bootstrap measure P^* as $N, T \to \infty$, conditional on a sample which lies in a set with probability P converging to one. Contrary to H, H^* does not depend on population values and can be computed for any bootstrap sample, given the original sample. Hence, rotation indeterminacy is not a problem in the bootstrap world. Because the bootstrap factor DGP (8) satisfies the constraints that $\tilde{F}'\tilde{F}/T = I_r$ and $\tilde{\Lambda}'\tilde{\Lambda}$ is a diagonal matrix, we can actually show that H^* is asymptotically (as $N, T \to \infty$) equal to $H_0^* = diag\,(\pm 1)$, a diagonal matrix with diagonal elements equal to ± 1, where the sign of is determined by the sign of $\tilde{F}^{*\prime}\tilde{F}/T$ (see Lemma B.1; the proof follows by arguments similar to those used in Bai and Ng (2011) and Stock and Watson (2002)). Thus, the bootstrap factors are identified up to a change of sign.

The main implication from Lemma 3.1 is that the bootstrap OLS estimator that one obtains from regressing y_{t+h}^* on $\hat{z}_t^*$ estimates a rotated version of $\hat{\delta}$, given by $\delta^* \equiv \begin{pmatrix} \hat{\alpha}' H^{*-1} & \hat{\beta}' \end{pmatrix}' = \Phi^{*\prime -1}\hat{\delta}$, where $\Phi^* = diag\,(H^*, I_q)$. Asymptotically, δ^* is equal to $\Phi_0^{*\prime -1}\hat{\delta}$, where $\Phi^* = diag\,(H_0^*, I_q)$, which can be interpreted as a sign-adjusted version of $\hat{\delta}$.

Our next result characterizes the asymptotic bootstrap distribution of $\sqrt{T}\left(\hat{\delta}^* - \delta^*\right)$ when $\sqrt{T}/N \to c$, with $0 \leq c < \infty$. We add the two following conditions.

Condition E*. $p\lim \Omega^* = \Phi_0 \Omega \Phi_0'$.

Condition F*. $p\lim \Gamma^* = Q\Gamma Q'$.

Ω^* is the bootstrap variance of the scaled average of the bootstrap regression scores $\hat{z}_t \varepsilon_{t+h}^*$ (as defined in Condition D*(b)). Since $\tilde{F}_t$ estimates a rotated version of the latent factors given by $H_0 F_t$, $\hat{z}_t$ estimates a rotated version of z_t given by $\Phi_0 z_t$, and therefore Ω^* is the sample analog of $\Phi_0 \Omega \Phi_0'$ provided we choose ε_{t+h}^* to mimic the time series dependence of ε_{t+h}. Condition E* imposes formally this condition. Similarly, by Condition B*(e), $\Gamma^* \equiv \frac{1}{T}\sum_{t=1}^{T} Var^*\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\lambda}_i e_{it}^*\right)$. Because $\tilde{\lambda}_i$ estimates $Q\lambda_i$, Γ^* is the sample analogue of $Q\Gamma Q'$ if we choose e_{it}^* to mimic the cross sectional dependence of e_{it} (interestingly enough, mimicking the time series dependence of e_{it} is not relevant). Condition F* formalizes this requirement.

Theorem 3.1 *Let Assumptions 1-5 hold and consider any residual-based bootstrap scheme for which Conditions A*-D* are verified. Suppose $\sqrt{T}/N \to c$, where $0 \leq c < \infty$. If in addition the two following conditions hold: (1) Condition E* is verified, and (2) $c = 0$ or Condition F* is verified; then as $N, T \to \infty$,*

$$\sqrt{T}\left(\hat{\delta}^* - \delta^*\right) \to^{d^*} N\left(-c\left(\Phi_0^{*\prime}\right)^{-1}\Delta_\delta, \left(\Phi_0^{*\prime}\right)^{-1}\Sigma_\delta \Phi_0^{*-1}\right),$$

in probability, where $\delta^* \equiv \left(\Phi_0^{*-1}\right)'\hat{\delta}$, with $\Phi_0^* = p\lim \Phi^* = diag\,(H_0^*, I_q)$ a diagonal matrix with ± 1 in the main diagonal, and Δ_δ and Σ_δ are as defined in Theorem 2.1.

According to Theorem 3.1, $\sqrt{T}\left(\hat{\delta}^* - \delta^*\right)$ is asymptotically distributed as a normal random vector with mean equal to $-c\left(\Phi_0^{*\prime}\right)^{-1}\Delta_\delta$. Just as the asymptotic bias of $\sqrt{T}\left(\hat{\alpha} - \left(H_0^{-1}\right)'\alpha\right)$ is proportional to $\left(H_0^{-1}\right)'\alpha$, the bootstrap asymptotic bias is proportional to $\left(H_0^{*-1}\right)'\hat{\alpha}$. Since $\hat{\alpha}$ converges in probability to $\left(H_0^{-1}\right)'\alpha$, the bootstrap bias of $\hat{\alpha}^*$ converges to $-c\left(H_0^{*\prime}\right)^{-1}\Delta_\alpha$ provided we ensure that Condition F* is satisfied. It is interesting that the bootstrap bias of $\hat{\beta}^*$ is unaffected by the rotation problem. Since the bootstrap analogue of $\Sigma_{W\tilde{F}}$ is $p\lim \frac{W'\tilde{F}^*}{T}$, which converges to $\Sigma_{W\tilde{F}}H_0^{*\prime}$, the rotation matrix $H_0^{*\prime}$ "cancels out" with $\left(H_0^{*-1}\right)'\hat{\alpha}$ (Lemma B.4 formalizes this argument). Similarly, the asymptotic variance-covariance matrix of $\hat{\delta}^*$ is equal to $\left(\Phi_0^{*\prime}\right)^{-1}\Sigma_\delta\Phi_0^{*-1}$ provided we choose ε_{t+h}^* so as to verify Condition E*.

For bootstrap consistency, we need the bootstrap bias and variance to match the bias and variance of the limiting distribution of the original OLS estimator. Since H_0^* (hence Φ_0^*) is not necessarily equal to the identity matrix, Theorem 3.1 shows that this is not the case. Hence, the bootstrap distribution of $\sqrt{T}\left(\hat{\delta}^* - \delta^*\right)$ is not a consistent estimator of the sampling distribution of $\sqrt{T}\left(\hat{\delta} - \delta\right)$ in general. This is true even if we choose ε_{t+h}^* and e_{it}^* such that Conditions E* and F* are satisfied. The reason is that the bootstrap factors are not identified. In particular, because the bootstrap principal components estimator does not necessarily identify the sign of the bootstrap factors, the mean of each element of $\sqrt{T}\left(\hat{\delta}^* - \delta^*\right)$ corresponding to the coefficients associated with the latent factors may have the "wrong" sign even asymptotically. The same "sign" problem will affect the off-diagonal elements of the bootstrap covariance matrix asymptotically (although not the main diagonal elements). As we mentioned above, the coefficients associated with W_t are correctly identified in the bootstrap world as well as in the original sample and therefore this sign problem does not affect these coefficients.

In order to obtain a consistent estimator of the distribution of $\sqrt{T}\left(\hat{\delta} - \delta\right)$, our proposal is to consider the bootstrap distribution of the rotated version of $\sqrt{T}\left(\hat{\delta}^* - \delta^*\right)$ given by $\sqrt{T}\left(\Phi^{*\prime}\hat{\delta}^* - \hat{\delta}\right)$. This rotation is feasible because Φ^* does not depend on any population quantities and can be computed for any bootstrap and original samples. Since Φ^* is asymptotically equal to $\Phi_0^* = diag\left(\pm 1, I_q\right) = diag\left(sign\left(\tilde{F}^{*\prime}\tilde{F}\right), I_q\right)$, $\Phi^{*\prime}\hat{\delta}^*$ is asymptotically equal to a sign-adjusted version of $\hat{\delta}^*$. The following result is an immediate corollary to Theorems 2.1 and 3.1.

Corollary 3.1 *Under the conditions of Theorem 3.1, if $\sqrt{T}/N \to c$, where $0 \leq c < \infty$, as $N, T \to \infty$, then* $\sup_{x\in\mathbb{R}^p}\left|P^*\left(\sqrt{T}\left(\Phi^{*\prime}\hat{\delta}^* - \hat{\delta}\right) \leq x\right) - P\left(\sqrt{T}\left(\hat{\delta} - \delta\right) \leq x\right)\right| \to^P 0.$

Corollary 3.1 justifies the use of a residual-based bootstrap method for constructing bootstrap percentile confidence intervals for the elements of δ. When $c = 0$, the crucial condition for bootstrap validity is Condition E*, which requires $\{\varepsilon_t^*\}$ to be chosen so as to mimic the dependence structure of the scores $z_t\varepsilon_{t+h}$. This condition ensures that the bootstrap variance-covariance matrix of $\Phi^{*\prime}\hat{\delta}^*$ is correct asymptotically. When $c > 0$, Condition F* is also crucial to ensure that the bootstrap distribution correctly captures the bias. When both Conditions E* and F* are satisfied, the bootstrap contains a built-in bias correction term that is absent in the Bai and Ng (2006) asymptotic normal distribution, and we might expect it to outperform the normal approximation. A bootstrap method that does not involve factor estimation in the bootstrap world will not contain this bias correction and will not be valid in this context.

4 Wild bootstrap

In this section we illustrate the use of the high-level conditions above and propose a particular bootstrap method for generating $\left\{\varepsilon_{t+h}^*\right\}$ and $\{e_{it}^*\}$ when $h=1$.

Bootstrap algorithm

1. For $t = 1, \ldots, T$, let
$$X_t^* = \tilde{\Lambda}\tilde{F}_t + e_t^*,$$
where $\{e_t^* = (e_{1t}^*, \ldots, e_{Nt}^*)\}$ is such that
$$e_{it}^* = \tilde{e}_{it}\eta_{it},$$
is a resampled version of $\left\{\tilde{e}_{it} = X_{it} - \tilde{\lambda}_i'\tilde{F}_t\right\}$ obtained with the wild bootstrap. The external random variables η_{it} are i.i.d. across (i,t) and have mean zero and variance one.

2. Estimate the bootstrap factors $\tilde{F}^*$ and the bootstrap loadings $\tilde{\Lambda}^*$ using X^*.

3. For $t = 1, \ldots, T-1$, let
$$y_{t+1}^* = \hat{\alpha}'\tilde{F}_t + \hat{\beta}'W_t + \varepsilon_{t+1}^*,$$
where the error term ε_{t+1}^* is a wild bootstrap resampled version of $\hat{\varepsilon}_{t+1}$, i.e.
$$\varepsilon_{t+1}^* = \hat{\varepsilon}_{t+1}v_{t+1},$$
where the external random variable v_{t+1} is i.i.d. $(0,1)$ and is independent of η_{it}.

4. Regress y_{t+1}^* generated in 3. on the estimated bootstrap factors and the fixed regressors $\hat{z}_t^* = \left(\tilde{F}_t^{*\prime}, W_t'\right)'$. This yields the bootstrap OLS estimators
$$\hat{\delta}^* = \left(\sum_{t=1}^{T-1} \hat{z}_t^* \hat{z}_t^{*\prime}\right)^{-1} \sum_{t=1}^{T-1} \hat{z}_t^* y_{t+1}^*.$$

To prove the validity of this residual-based wild bootstrap, we add the following assumptions[1].

Assumption 6. λ_i are either deterministic such that $\|\lambda_i\| \leq M < \infty$, or stochastic such that $E\|\lambda_i\|^{12} \leq M < \infty$ for all i; $E\|F_t\|^{12} \leq M < \infty$; $E|e_{it}|^{12} \leq M < \infty$, for all (i,t); and for some $q > 1$, $E|\varepsilon_{t+1}|^{4q} \leq M < \infty$, for all t.

[1] Under Assumptions 6-8, some of Assumptions 1-5 simplify. In particular, we can show that Assumption 2(c) and Assumptions 3(a)-(d) and 4 are implied by Assumptions 1,2, 6-8 if we impose in addition the mutual independence among $\{F_t\}, \{\lambda_i\}$ and $\{e_{is}\}$ and require condition (6).

Assumption 7. $E(\varepsilon_{t+1}|y_t, F_t, y_{t-1}, F_{t-1}, \ldots) = 0$, and F_t and ε_t are independent of the idiosyncratic errors e_{is} for all (i, s, t).

Assumption 8. $E(e_{it}e_{js}) = 0$ if $i \neq j$.

Assumption 6 strengthens the moment conditions assumed in Assumption 1.b), 2.a), and 5.a), respectively. The moment conditions on λ_i, F_t and e_{it} suffice to show that $E\left|\lambda_i' F_s e_{it}\right|^4 < M$ (while maintaining that $E|e_{it}|^8 < M$). If we assume that the three groups of random variables $\{F_t\}$, $\{e_{it}\}$ and $\{\lambda_i\}$ are mutually independent (as in Bai and Ng (2006)), then it suffices that $E\|\lambda_i\|^4 \leq M < \infty$, $E\|F_t\|^4 \leq M < \infty$ (and $E|e_{it}|^8 < M$).

Assumption 7 imposes a martingale difference condition on the regression errors ε_{t+1}, implying that these are serially uncorrelated but possibly heteroskedastic. In addition, ε_t is independent of e_{is} for all (i, t, s). Under Assumption 7,

$$\Omega = \lim Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1} z_t \varepsilon_{t+1}\right) = \lim \frac{1}{T}\sum_{t=1}^{T-1} E\left(z_t z_t' \varepsilon_{t+1}^2\right),$$

which motivates using a wild bootstrap to generate ε_{t+1}^*. For this bootstrap scheme,

$$\Omega^* = \frac{1}{T}\sum_{t=1}^{T-1} \hat{z}_t \hat{z}_t' \hat{\varepsilon}_{t+1}^2,$$

is consistent for $\Phi_0 \Omega \Phi_0'$ under Assumptions 1-7. For $h > 1$, ε_{t+h} will be serially correlated, and some block bootstrap based method is required in this case. Corradi and Swanson (2011) have established the validity of the m out of n bootstrap when ε_{t+h} is not a martingale difference sequence when $c = 0$. We assume independence between e_{it} and ε_{t+1} and generate ε_{t+1}^* independently of e_{it}^*, but we conjecture that our results will be valid under weak forms of correlation between the two sets of errors because the limiting distribution of the OLS estimator remains unchanged under Assumptions 1-5, which allow for dependence between e_{it} and ε_{t+1}, as we proved in Theorem 2.1.

Assumption 8 assumes cross section independence in $\{e_{it}\}$, but allows for serial correlation and heteroskedasticity in both directions. This assumption motivates the use of a wild bootstrap to generate $\{e_{it}^*\}$. For this bootstrap scheme, we can show that

$$\Gamma^* = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{N}\sum_{i=1}^{N} \tilde{\lambda}_i \tilde{\lambda}_i' \tilde{e}_{it}^2 \equiv \frac{1}{T}\sum_{t=1}^{T} \tilde{\Gamma}_t,$$

where $\tilde{\Gamma}_t$ corresponds to estimator (5a) in Bai and Ng (2006, p. 1140). As shown by Bai and Ng, this estimator is consistent for $Q\Gamma Q'$ under cross section independence (but potential heteroskedasticity). Assumption 8 assumes this is the case and thus justifies Condition F* in this context. As we discussed in the previous section, Condition F* is not needed if $c = 0$. Thus, a wild bootstrap is still asymptotically valid if the idiosyncratic errors are cross sectionally (and serially) dependent when $\sqrt{T}/N$ converges to zero (as assumed in Bai and Ng (2006)).

Our main result is as follows.

Theorem 4.1 *Suppose that a residual-based wild bootstrap is used to generate* $\{e^*_{it}\}$ *and* $\{\varepsilon^*_{t+1}\}$ *with* $E^* |\eta_{it}|^4 < C$ *for all* (i,t) *and* $E^* |v_{t+1}|^{4q} < C$ *for all* t*, for some* $q > 1$*. Under Assumptions 1-7, if* $\sqrt{T}/N \to c$*, where* $0 \leq c < \infty$*, and either Assumption 8 holds or* $c = 0$*, the conclusions of Corollary 3.1 follow.*

5 Monte Carlo results

In this section, we report results from a simulation experiment that documents the properties of our bootstrap procedure in factor-augmented regressions.

The data-generating process (DGP) is similar to the one used in Ludvigson and Ng (2011) to analyze bias. We consider the single factor model:

$$y_{t+1} = \alpha F_t + \varepsilon_{t+1}, \tag{11}$$

where F_t is drawn from a standard normal distribution independently over time. The regression error ε_{t+1} will either be standard normal or heteroskedastic over time as specified below. The $(T \times N)$ matrix of panel variables is generated as:

$$X_{it} = \lambda_i F_t + e_{it}, \tag{12}$$

where λ_i is drawn from a $U[0,1]$ distribution (independent across i) and the properties of e_{it} will be discussed below. The only difference with Ludvigson and Ng (2011) is that they draw the loadings from a standard normal distribution. The use of a uniform distribution increases the cross-correlations and leads to larger biases without having to set the idiosyncratic variance to large values (they set it to 16 in one experiment). Note that this DGP satisfies the conditions PC1 in Bai and Ng (2011) which implies that $H_0 = \pm 1$.

We consider six different scenarios outlined in the table below. We consider two values for the coefficient, either $\alpha = 0$ or 1. When $\alpha = 0$, the OLS estimator is unbiased, and the properties of the idiosyncratic components do not matter asymptotically. This leads us to consider only one scenario with $\alpha = 0$. The other five scenarios have $\alpha = 1$ but differ according to the properties of the regression error, ε_t, and of the idiosyncratic error, e_{it}.

DGP	α	ε_{t+1}	e_{it}
1 (homo-homo)	0	$N(0,1)$	$N(0,1)$
2 (homo-homo)	1	$N(0,1)$	$N(0,1)$
3 (hetero-homo)	1	$N\left(0, \frac{F_t^2}{3}\right)$	$N(0,1)$
4 (hetero-hetero)	1	$N\left(0, \frac{F_t^2}{3}\right)$	$N\left(0, \sigma_i^2\right)$
5 (hetero-AR)	1	$N\left(0, \frac{F_t^2}{3}\right)$	$AR + N\left(0, \sigma_i^2\right)$
6 (hetero-CS)	1	$N\left(0, \frac{F_t^2}{3}\right)$	$CS + N(0,1)$

DGP 1 is the simplest case with $\alpha = 0$ and both error terms i.i.d. standard normal in both dimensions. DGP 2 is the same but with $\alpha = 1$. This allows us to isolate the effect of a non-zero coefficient on bias and inference while keeping everything else the same. The third experiment introduces conditional heteroskedasticity in the regression error. We do so by making the variance of ε_t depend on the factor and scale so that the asymptotic variance of $\hat{\alpha}$, Σ_α, is 1 in all experiments.

The fourth DGP adds heteroskedasticity to the idiosyncratic error. The variance of e_{it} is drawn from $U[.5, 1.5]$ so that the average variance is the same as the homoskedastic case. The fifth DGP introduces serial correlation in the idiosyncratic error term with autoregressive parameter equal to 0.5. The innovations are scaled by $\left(1 - .5^2\right)^{1/2}$ to preserve the variance of the idiosyncratic errors. Finally, the last experiment introduces cross-sectional dependence among idiosyncratic errors. The design is similar to the one in Bai and Ng (2006): the correlation between e_{it} and e_{jt} is $0.5^{|i-j|}$ if $|i-j| \leq 5$. We rescale e_{it} so that Γ is the same as for the other cases.

We concentrate on inference about the parameter α in (11). We consider asymptotic and bootstrap symmetric percentile t confidence intervals at a nominal level of 95%. We report experiments based on 5000 replications with $B = 399$ bootstrap repetitions. We consider three values for N (50, 100, and 200) and T (50, 100, and 200).

We tailor our inference procedures to the properties of the error terms. When ε_t is homoskedastic (DGP 1 and 2), we use the variance estimator under homoskedasticity:

$$\hat{\Sigma}_\alpha = \hat{\sigma}_\varepsilon^2 \left(\frac{1}{T} \sum_{t=1}^{T-1} \tilde{F}_t^2 \right)^{-1} \tag{13}$$

whereas we use the heteroskedastic-robust version for DGPs 3-6:

$$\hat{\Sigma}_\alpha = \left(\frac{1}{T} \sum_{t=1}^{T-1} \tilde{F}_t^2 \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T-1} \tilde{F}_t^2 \hat{\varepsilon}_{t+1}^2 \right) \left(\frac{1}{T} \sum_{t=1}^{T-1} \tilde{F}_t^2 \right)^{-1}. \tag{14}$$

Similarly, we use the homoskedastic estimator of Γ for cases 1-3, the heteroskedasticity-robust estimator for cases 4-5, and the CS-HAC estimator of Bai and Ng (2006) in case 6 with the window size n equal to $\min\left(\sqrt{N}, \sqrt{T}\right)$. We consider the wild residual-based bootstrap described in Section 4 with the two external variables η_{it} and v_t both drawn from i.i.d. $N(0, 1)$.

We report two sets of results. The first set of results is the bias of the OLS estimator. Because the OLS estimator does not converge to α but to $H^{-1\prime}\alpha$ (and H converges to $+1$ or -1) and because its bias is proportional to this, the bias will be positive for some replications and negative for others. Reporting the average bias over replications is therefore misleading in this situation. To circumvent this, we report the bias of the *rotated* OLS coefficient, $H'\hat{\alpha}$. This rotated coefficient converges to α in all replications. Note that this rotation is not possible in the real world since the matrix H depends on population parameters. Note also that this rotation is possible in the bootstrap world (and indeed necessary to obtain consistent inference of the entire coefficient vector, see Corollary 3.1). For the bootstrap world, we report the average of $H'H^{*\prime}\hat{\alpha}^* - H'\hat{\alpha}$, again to ensure that the sign of this bias is always the same. Secondly, we present coverage rates of the associated confidence intervals. For comparison, we also include results for the case where factors do not need to be estimated and are treated as observed (row labeled "true factor" in the tables). This quantifies the loss from having to estimate the factors.

Table 1 provides results for the first two DGPs and illustrates our results. For each DGP, the top panel gives the bias associated with the OLS estimator as well as the plug-in and bootstrap estimates. The plug-in estimate is obtained by replacing the unknown quantities in Theorem 2.1 by sample analogues. The second panel for each design provides the coverage rate of intervals based on asymptotic theory, either using the OLS estimator or its plug-in bias-corrected version, based on OLS using true factors, and based on the wild bootstrap. From table 1, we see that, as expected, bias is nil when $\alpha = 0$ (case 1). When $\alpha \neq 0$, a negative bias appears. DGP 2 shows that the magnitude of this bias is decreasing in N (and T). The plug-in estimate of this quantity provides a reasonable approximation to it. However, the bootstrap captures the behavior of the bias well as predicted by theory and better than the plug-in estimate.

Coverage rates are consistent with these findings. When $\alpha = 0$ (DGP 1), asymptotic theory is nearly perfect and matches closely the results based on the observed factors. In case 2, with $\alpha = 1$, OLS inference suffers from noticeable distortions for all sample sizes. This is because the estimator is biased and the associated t-statistic is not centered at 0. Plug-in bias correction corrects most of these distortions. The bootstrap provides even better inference and is quite accurate for $N \geq 100$. The loss in accuracy in inference is due to the estimation of the factors as illustrated by the results with the true factors.

Tables 2 and 3 provide results for the other DGPs and show the robustness of the results to the presence of heteroskedasticity in both errors (DGPs 3 and 4), and serial correlation in the idiosyncratic errors (DGP 5). The bias results in table 2 are very similar to those in table 1, although coverage rates reported in table 3 deteriorate relative to the simpler homoskedastic case. The presence of cross-sectional dependence (DGP 6) is interesting. The wild bootstrap is theoretically not valid since it does not replicate the cross-sectional dependence. Indeed, we see that, contrary to the other cases, the plug-in estimate of the bias is often better than the bootstrap, especially with $N = 50$.

6 Conclusion

The main contribution of this paper is to give a set of sufficient high-level conditions under which any residual-based bootstrap method is valid in the context of the factor-augmented regression model in cases where $\sqrt{T}/N \to c$, $0 \leq c < \infty$. Our results show that two crucial conditions for bootstrap validity in this context are that the bootstrap regression scores replicate the time series dependence of the true regression scores, and that either $c = 0$ or the bootstrap replicates also the cross-sectional dependence of the idiosyncratic error terms.

Our high-level conditions can be checked for any implementation of the bootstrap in this context. We verify them for a particular scheme based on a two-step application of the wild bootstrap. Although the wild bootstrap preserves heteroskedasticity, its validity depends on a martingale difference condition on the regression errors and on cross-sectional independence of the idiosyncratic errors when $c \neq 0$.

Although our general results in Sections 2 and 3 allow for serial correlation in the scores (see in particular our Assumption 5(b)), the particular implementation of the two-step residual-based wild bootstrap we consider in Section 4 is not robust to serial dependence. A block bootstrap based method is required in this case. We plan on investigating the validity of such a method in future work, and our high-level conditions will be useful in establishing this result.

A second important extension of the results in this paper is to propose a bootstrap scheme that is able to replicate the cross-sectional dependence of the idiosyncratic error term. As our results show, this is crucial for capturing the bias when $c \neq 0$. Our wild bootstrap based method does not allow for cross-sectional dependence. Because there is no natural cross-sectional ordering, devising a nonparametric bootstrap method that is robust to cross-sectional dependence of unknown form is a challenging task.

Another important extension of the results in this paper is the construction of interval forecasts, which we are currently investigating.

A Appendix A: Proofs of results in Section 2

We rely on the following lemmas to prove Theorem 2.1.

Lemma A.1 *Let Assumptions 1-5 hold. Then,* $\frac{1}{T}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\varepsilon_{t+h} = O_P\left(\frac{1}{\delta_{NT}\sqrt{T}}\right)$, *where* $\delta_{NT} = \min\left(\sqrt{N}, \sqrt{T}\right)$.

Lemma A.2 *Let Assumptions 1-5 hold. Then, if* $\sqrt{T}/N \to c$, *where* $0 \leq c < \infty$, *for any fixed* $h \geq 0$,

a) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\left(\tilde{F}_t - HF_t\right)' = cV^{-1}Q\Gamma Q'V^{-1} + o_P(1)$.

b) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}HF_t\left(\tilde{F}_t - HF_t\right)' = cQ\Gamma Q'V^{-2} + o_P(1)$.

c) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}W_t\left(\tilde{F}_t - HF_t\right)' = c\Sigma_{WF}H_0'Q\Gamma Q'V^{-2} + o_P(1)$.

d) *Letting* $\Sigma_{\tilde{F}} \equiv V^{-1}Q\Gamma Q'V^{-1}$, *we have that*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\tilde{F}_t\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\alpha = c\underbrace{\left(\Sigma_{\tilde{F}} + V\Sigma_{\tilde{F}}V^{-1}\right)p\lim\left(\hat{\alpha}\right)}_{\equiv B_\alpha} + o_P(1), \tag{15}$$

and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}W_t\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\alpha = c\underbrace{\left(\Sigma_{W\tilde{F}}V\Sigma_{\tilde{F}}V^{-1}\right)p\lim\left(\hat{\alpha}\right)}_{\equiv B_\beta} + o_P(1), \tag{16}$$

where $\Sigma_{W\tilde{F}} = p\lim\left(\frac{W'\tilde{F}}{T}\right) = \Sigma_{WF}H_0'$, *with* $\Sigma_{WF} \equiv E\left(W_tF_t'\right)$.

Proof of Theorem 2.1. Write

$$\sqrt{T}\left(\hat{\delta} - \delta\right) = \left(\frac{1}{T}\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t'\right)^{-1}\left\{\begin{array}{c}\underbrace{\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\begin{pmatrix}HF_t\\W_t\end{pmatrix}\varepsilon_{t+h}}_{\equiv A} + \underbrace{\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\begin{pmatrix}\tilde{F}_t - HF_t\\0\end{pmatrix}\varepsilon_{t+h}}_{\equiv B}\\ -\underbrace{\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\hat{z}_t\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\alpha}_{\equiv C}\end{array}\right\}. \tag{17}$$

By Assumption 5 and given the definition of $\Phi_0 = diag\left(p\lim H, I_q\right)$,

$$A = \begin{pmatrix}H & 0\\0 & I_q\end{pmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}z_t\varepsilon_{t+h} \to^d N\left(0, \Phi_0\Omega\Phi_0'\right).$$

By Lemma A.1, $B \to^P 0$ and by Lemma A.2d),

$$C = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\begin{pmatrix}\tilde{F}_t\\W_t\end{pmatrix}\left(\tilde{F}_t - HF_t\right)'\left(H^{-1}\right)'\alpha = -c\begin{pmatrix}B_\alpha\\B_\beta\end{pmatrix} + o_P(1),$$

where B_α and B_β are as defined in (15) and (16). Given Assumptions 1-5,

$$\frac{1}{T}\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t' = \Phi_0\left(\frac{1}{T}\sum_{t=1}^{T-h}z_tz_t'\right)\Phi_0' + o_P(1) = \Phi_0\Sigma_{zz}\Phi_0' + o_P(1).$$

This implies that if $\sqrt{T}/N \to c$, $\sqrt{T}\left(\hat{\delta}-\delta\right) \to^d N\left(-c\Delta_\delta, \Sigma_\delta\right)$, with $\Sigma_\delta = \Phi_0'^{-1}\Sigma_{zz}^{-1}\Omega\Sigma_{zz}^{-1}\Phi_0^{-1}$, and

$$\Delta_\delta \equiv \begin{pmatrix}\Delta_\alpha \\ \Delta_\beta\end{pmatrix} = \left(\Phi_0\Sigma_{zz}\Phi_0'\right)^{-1}\begin{pmatrix}\Sigma_{\tilde{F}} + V\Sigma_{\tilde{F}}V^{-1} \\ \Sigma_{W\tilde{F}}V\Sigma_{\tilde{F}}V^{-1}\end{pmatrix} p\lim\left(\hat{\alpha}\right).$$

When $\Sigma_{WF} \equiv E\left(W_tF_t'\right) = 0$, we have that $\Sigma_{W\tilde{F}} = 0$, implying that

$$\Delta_\delta = \begin{pmatrix} H_0'^{-1}\Sigma_F^{-1}H_0^{-1} & 0 \\ 0 & \Sigma_W^{-1}\end{pmatrix}\begin{pmatrix}\Sigma_{\tilde{F}} + V\Sigma_{\tilde{F}}V^{-1} \\ 0\times V\Sigma_{\tilde{F}}V^{-1}\end{pmatrix} p\lim\left(\hat{\alpha}\right) = \begin{pmatrix}\Sigma_{\tilde{F}} + V\Sigma_{\tilde{F}}V^{-1} \\ 0\end{pmatrix} p\lim\left(\hat{\alpha}\right),$$

since $H_0'^{-1}\Sigma_F^{-1}H_0^{-1} = I_r$ given that we can show that $H_0\Sigma_F = Q = \left(H_0'\right)^{-1}$.

Proof of Lemma A.1. The proof is based on the following identity:

$$\begin{aligned}\tilde{F}_t - HF_t &= \tilde{V}^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st} + \frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st} + \frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s\eta_{st} + \frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right) \\ &\equiv \tilde{V}^{-1}\left(A_{1t} + A_{2t} + A_{3t} + A_{4t}\right), \qquad (18)\end{aligned}$$

where $\gamma_{st} = E\left(\frac{1}{N}\sum_{i=1}^{N}e_{is}e_{it}\right)$, $\zeta_{st} = \frac{1}{N}\sum_{i=1}^{N}\left(e_{is}e_{it} - E\left(e_{is}e_{it}\right)\right)$, $\eta_{st} = \frac{1}{N}\sum_{i=1}^{N}\lambda_i'F_se_{it} = F_s'\frac{\Lambda'e_t}{N}$ and $\xi_{st} = F_t'\frac{\Lambda'e_s}{N} = \eta_{ts}$. Using the identity (18), we have that

$$\frac{1}{T}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\varepsilon_{t+h} = \tilde{V}^{-1}\left(I + II + III + IV\right),$$

where $I = T^{-2}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}\varepsilon_{t+h}$, and

$$II = T^{-2}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\tilde{F}_s\zeta_{st}\varepsilon_{t+h},\ III = T^{-2}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\tilde{F}_s\eta_{st}\varepsilon_{t+h};\text{ and } IV = T^{-2}\sum_{t=1}^{T-h}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\varepsilon_{t+h}.$$

Since $\tilde{V}^{-1} = O_P(1)$ (see Lemma A.3 of Bai (2003), which shows that $\tilde{V} \to^P V > 0$), we can ignore $\tilde{V}^{-1}$. $I = O_P\left(T^{-1/2}\delta_{NT}^{-1}\right)$ by the same argument as given in the proof of Lemma A1 of Bai and Ng (2006) (this uses the fact that $\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$ under our assumptions; see Lemma A.1 of Bai (2003)). For II, we have

$$\|II\| \le \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\left|\frac{1}{T}\sum_{t=1}^{T-h}\zeta_{st}\varepsilon_{t+h}\right|^2\right)^{1/2} = O_P\left(\frac{1}{\sqrt{NT}}\right) = O_P\left(\frac{1}{\sqrt{T}\delta_{NT}}\right),$$

given that $\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s\right\|^2 \le \frac{2}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2 + \frac{2}{T}\|H\|^2\sum_{s=1}^{T}\|F_s\|^2 = O_P(1)$, and

$$\frac{1}{T}\sum_{s=1}^{T}E\left|\frac{1}{T}\sum_{t=1}^{T-h}\zeta_{st}\varepsilon_{t+h}\right|^2 = \frac{1}{TN}\frac{1}{T}\sum_{s=1}^{T}E\left|\frac{1}{\sqrt{TN}}\sum_{t=1}^{T-h}\sum_{i=1}^{N}\left(e_{is}e_{it} - E\left(e_{is}e_{it}\right)\right)\varepsilon_{t+h}\right|^2 = O\left(\frac{1}{TN}\right),$$

by Assumption 4(a). For III, a similar argument yields

$$III \leq \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\left|\frac{1}{T}\sum_{t=1}^{T-h}\eta_{st}\varepsilon_{t+h}\right|^2\right)^{1/2} = O_P\left(\frac{1}{\sqrt{NT}}\right),$$

since

$$\frac{1}{T}\sum_{s=1}^{T}\left(\frac{1}{T}\sum_{t=1}^{T-h}\eta_{st}\varepsilon_{t+h}\right)^2 \leq \underbrace{\frac{1}{T}\sum_{s=1}^{T}\|F_s\|^2}_{=O_P(1)}\cdot\frac{1}{TN}\underbrace{\left\|\frac{1}{\sqrt{TN}}\sum_{t=1}^{T-h}\Lambda' e_t\varepsilon_{t+h}\right\|^2}_{=O_P(1)\text{ by Assumption 4(b)}} = O_P\left(\frac{1}{TN}\right).$$

Finally, for IV, we have that

$$\|IV\| \leq \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\left(\frac{1}{T}\sum_{t=1}^{T-h}\xi_{st}\varepsilon_{t+h}\right)^2\right)^{1/2} = O_P\left(\frac{1}{\sqrt{NT}}\right),$$

since

$$\frac{1}{T}\sum_{s=1}^{T}\left(\frac{1}{T}\sum_{t=1}^{T-h}\xi_{st}\varepsilon_{t+h}\right)^2 \leq \frac{1}{TN}\underbrace{\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{e_s'\Lambda}{\sqrt{N}}\right\|^2\right)}_{=O_P(1)\text{ by Assumption 3(d)}}\underbrace{\left(\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}F_t\varepsilon_{t+h}\right\|^2\right)}_{=O_P(1)\text{ by Assumption 5(c)}} = O_P\left(\frac{1}{TN}\right).$$

Proof of Lemma A.2. Proof of part a) Using the identity (18), we can write

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\left(\tilde{F}_t - HF_t\right)\left(\tilde{F}_t - HF_t\right)' = \tilde{V}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\left(A_{1t}+A_{2t}+A_{3t}+A_{4t}\right)\left(A_{1t}+A_{2t}+A_{3t}+A_{4t}\right)'\tilde{V}^{-1},$$

where A_{it} $(i = 1, \ldots, 4)$ are defined in (18). We analyze each term separately (ignoring $\tilde{V}^{-1}$). We can show that $\frac{1}{T}\sum_{t=1}^{T-h}A_{1t}A_{1t}' = O_P\left(T^{-1}\right)$, implying that $\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}A_{1t}A_{1t}' = O_P\left(T^{-1/2}\right) = o_P(1)$. Indeed, by Cauchy-Schwartz

$$\left\|\frac{1}{T}\sum_{t=1}^{T-h}A_{1t}A_{1t}'\right\| \leq \frac{1}{T}\sum_{t=1}^{T-h}\|A_{1t}\|^2 \leq \frac{1}{T}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s\right\|^2\right)\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma_{st}^2\right) = O_P\left(\frac{1}{T}\right).$$

We can show that $\frac{1}{T}\sum_{t=1}^{T-h}A_{2t}A_{2t}' = O_P\left((NT)^{-1}\right) + O_P\left(N^{-1}\delta_{NT}^{-2}\right)$, implying that $\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}A_{2t}A_{2t}' = O_P\left(\frac{1}{\sqrt{T}N}\right) + O_P\left(\frac{\sqrt{T}}{N}\delta_{NT}^{-2}\right) = o_P(1)$, if $\sqrt{T}/N \to c < \infty$. Indeed,

$$\begin{aligned}\left\|\frac{1}{T}\sum_{t=1}^{T-h}A_{2t}A_{2t}'\right\| &\leq \frac{1}{T}\sum_{t=1}^{T-h}\|A_{2t}\|^2 = \frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s + HF_s\right)\zeta_{st}\right\|^2 \\ &\leq \frac{1}{T}\sum_{t=1}^{T-h}\left\|\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\zeta_{st}\right\|^2 + \frac{1}{T}\sum_{t=1}^{T-h}\left\|H\frac{1}{T}\sum_{s=1}^{T}F_s\zeta_{st}\right\|^2 \equiv a_{22.1} + a_{22.2},\end{aligned}$$

where

$$a_{22.1} \equiv \frac{1}{T}\sum_{t=1}^{T-h}\left\|\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\zeta_{st}\right\|^2 \leq \underbrace{\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\right)}_{=O_P\left(\delta_{NT}^{-2}\right)}\underbrace{\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|\zeta_{st}\right|^2\right)}_{=O_P\left(N^{-1}\right)\text{ by Assumption 2(c)}} = O_P\left(\frac{1}{N\delta_{NT}^2}\right),$$

and

$$a_{22.2} \leq \|H\|^2\frac{1}{T}\sum_{t=1}^{T-h}\left\|\frac{1}{T}\sum_{s=1}^{T}F_s\zeta_{st}\right\|^2 = \|H\|^2\frac{1}{TN}\underbrace{\frac{1}{T}\sum_{t=1}^{T-h}\left\|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s\left(e_{is}e_{it} - E\left(e_{is}e_{it}\right)\right)\right\|^2}_{=O_P(1)\text{ by Assumption 3(b)}} = O_P\left(\frac{1}{TN}\right).$$

We can show that if $\sqrt{T}/N \to c$, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}A_{3t}A_{3t}' = cQ\Gamma Q' + o_P(1)$. Indeed,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T-h}A_{3t}A_{3t}' &= \frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s + HF_s\right)\eta_{st}\right)\left(\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s + HF_s\right)\eta_{st}\right)' \\ &\equiv a_{33.1} + a_{33.2} + a_{33.2}' + a_{33.3},\end{aligned}$$

where $\|a_{33.1}\| \leq \frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2\frac{1}{T^2}\sum_{t=1}^{T-h}\sum_{s=1}^{T}|\eta_{st}|^2 = O_P\left(\frac{1}{N\delta_{NT}^2}\right)$ under our assumptions (in particular, Assumption 3(d) is useful here). Thus, $\sqrt{T}a_{33.1} = O_P\left(\frac{\sqrt{T}}{N}\frac{1}{\delta_{NT}^2}\right) = o_P(1)$, if $\sqrt{T}/N \to c$. Using similar arguments, we can show that $\sqrt{T}a_{33.2} = O_P\left(\frac{\sqrt{T}}{N}\frac{1}{\delta_{NT}}\right) = o_P(1)$. For $a_{33.3}$, write

$$\begin{aligned}a_{33.3} &= \frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{1}{T}\sum_{s=1}^{T}HF_s\eta_{st}\right)\left(\frac{1}{T}\sum_{s=1}^{T}HF_s\eta_{st}\right)' = H\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{1}{T}\sum_{s=1}^{T}F_sF_s'\frac{\Lambda' e_t}{N}\right)\left(\frac{1}{T}\sum_{s=1}^{T}\frac{e_t'\Lambda}{N}F_sF_s'\right)H' \\ &= H\left(\frac{F'F}{T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{\Lambda' e_t}{N}\right)\left(\frac{e_t'\Lambda}{N}\right)\left(\frac{F'F}{T}\right)H' = O_P\left(\frac{1}{N}\right),\end{aligned}$$

given Assumption 3(e). When multiplied by $\sqrt{T}$, this term is of order $O_P\left(\frac{\sqrt{T}}{N}\right)$ and therefore it will not converge to zero in probability when $c \neq 0$. To compute its probability limit, note that

$$\frac{HF'F}{T} = \frac{\left(FH' - \tilde{F} + \tilde{F}\right)'F}{T} = -\frac{\left(\tilde{F} - FH'\right)'F}{T} + \frac{\tilde{F}'F}{T} = -Q + o_P(1)$$

since $\frac{\left(\tilde{F}-FH'\right)'F}{T} = O_P\left(\delta_{NT}^{-2}\right) = o_P(1)$ (see Lemma B.2 of Bai (2003)) and $p\lim\frac{\tilde{F}'F}{T} = Q$. Thus,

$$\sqrt{T}a_{33.3} = \frac{\sqrt{T}}{N}H\left(\frac{F'F}{T}\right)\left\{\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{\Lambda' e_t}{\sqrt{N}}\right)\left(\frac{e_t'\Lambda}{\sqrt{N}}\right)\right\}\left(\frac{F'F}{T}\right)H' = cQ\Gamma Q' + o_P(1),$$

where we have used Assumption 3(e) and the fact that $\sqrt{T}/N = c + o(1)$. To complete the proof, we show that the remaining terms are asymptotically negligible. We can show that $\frac{1}{T}\sum_{t=1}^{T-h}A_{4t}A_{4t}' = O_P\left(\frac{1}{N\delta_{NT}^2}\right) + O_P\left(\frac{1}{N\delta_{NT}}\right) + O_P\left(\frac{1}{TN}\right)$, which implies that if $\sqrt{T}/N \to c$, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}A_{4t}A_{4t}' = O_P\left(\frac{\sqrt{T}}{N\delta_{NT}^2}\right)+$

$O_P\left(\frac{\sqrt{T}}{N\delta_{NT}}\right)+O_P\left(\frac{1}{\sqrt{TN}}\right)=o_P(1)$. Indeed, given the definition of A_{4t},

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T-h}A_{4t}A_{4t}' &= \frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s-HF_s+HF_s\right)\xi_{st}\right)\left(\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s-HF_s+HF_s\right)\xi_{st}\right)' \\ &\equiv a_{44.1}+a_{44.2}+a_{44.2}'+a_{44.3},\end{aligned}$$

where by Cauchy-Schwartz and Assumption 3(d),

$$\|a_{44.1}\|\leq\frac{1}{T}\sum_{t=1}^{T-h}\left\|\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s-HF_s\right)\xi_{st}\right\|^2\leq\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s-HF_s\right\|^2\frac{1}{T^2}\sum_{t=1}^{T-h}\sum_{s=1}^{T}|\xi_{st}|^2=O_P\left(\delta_{NT}^{-2}\right)O_P\left(N^{-1}\right),$$

and

$$\|a_{44.2}\|\leq\left(\frac{1}{T}\sum_{t=1}^{T-h}\left\|\frac{1}{T}\sum_{s=1}^{T}HF_s\xi_{st}\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T-h}\left\|\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s-HF_s\right)\xi_{st}\right\|^2\right)^{1/2}=O_P\left(N^{-1/2}\right)O_P\left(N^{-1/2}\delta_{NT}^{-1}\right).$$

Thus, $\sqrt{T}a_{44.1}=O_P\left(\delta_{NT}^{-2}\right)O_P\left(\sqrt{T}N^{-1}\right)=o_P(1)$ and $\sqrt{T}a_{44.2}=O_P\left(\sqrt{T}N^{-1}\right)O_P\left(\delta_{NT}^{-1}\right)=o_P(1)$ if $\sqrt{T}/N\to c$. Finally, for $a_{44.3}$, by Assumption 3(c),

$$\begin{aligned}a_{44.3} &= H\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{1}{T}\sum_{s=1}^{T}F_s\xi_{st}\right)\left(\frac{1}{T}\sum_{s=1}^{T}\xi_{st}F_s'\right)H' \\ &= H\left(\frac{1}{T}\sum_{s=1}^{T}F_s\frac{e_s'\Lambda}{N}\right)\frac{1}{T}\sum_{t=1}^{T-h}F_tF_t'\left(\frac{1}{T}\sum_{s=1}^{T}\frac{\Lambda'e_s}{N}F_s'\right)H'=O_P\left((NT)^{-1/2}\right)O_P(1)\,O_P\left((NT)^{-1/2}\right),\end{aligned}$$

implying that $\sqrt{T}a_{44.3}=o_P(1)$. We now deal with the cross terms. From Cauchy-Schwartz and the fact that $T^{-1}\sum_{t=1}^{T-h}\|A_{1t}\|^2=O_P\left(T^{-1}\right)$ and $T^{-1}\sum_{t=1}^{T-h}\|A_{2t}\|^2=O_P\left(N^{-1}\delta_{NT}^{-2}\right)$, it follows that $T^{-1}\sum_{t=1}^{T-h}A_{1t}A_{2t}'=O_P\left(\delta_{NT}^{-1}(TN)^{-1/2}\right)$, which implies that $T^{-1/2}\sum_{t=1}^{T-h}A_{1t}A_{2t}'=o_P(1)$. Similarly, we can show that $T^{-1}\sum_{t=1}^{T-h}A_{1t}A_{3t}'=O_P\left((TN)^{-1/2}\right)$ since $T^{-1}\sum_{t=1}^{T-h}\|A_{3t}\|^2=O_P\left(N^{-1}\right)$; $T^{-1}\sum_{t=1}^{T-h}A_{1t}A_{4t}'=O_P\left((TN)^{-1/2}\right)$, given that $T^{-1}\sum_{t=1}^{T-h}\|A_{4t}\|^2=O_P\left(N^{-1}\right)$; $T^{-1}\sum_{t=1}^{T-h}A_{2t}A_{3t}'=O_P\left(N^{-1}\delta_{NT}^{-1}\right)$; $T^{-1}\sum_{t=1}^{T-h}A_{2t}A_{4t}'=O_P\left(N^{-1}\delta_{NT}^{-1}\right)$; and $T^{-1}\sum_{t=1}^{T-h}A_{3t}A_{4t}'=O_P\left(N^{-1}\delta_{NT}^{-1}\right)$. For this last term, an application of Cauchy-Schwartz inequality is not enough since it would imply that this term is of order $O_P\left(N^{-1}\right)$, which when multiplied by $\sqrt{T}$ would not go to zero if $\lim\sqrt{T}/N=c\neq0$. Therefore, a different argument is required. In particular, we write

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T-h}A_{3t}A_{4t}' &= \frac{1}{T}\sum_{t=1}^{T-h}A_{3t}\left(\frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s\xi_{st}\right)'=\frac{1}{T}\sum_{t=1}^{T-h}A_{3t}\left(\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s-HF_s\right)\xi_{st}+H\frac{1}{T}\sum_{s=1}^{T}F_s\xi_{st}\right)' \\ &= \frac{1}{T}\sum_{t=1}^{T-h}A_{3t}\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s-HF_s\right)'\xi_{st}+H\frac{1}{T^2}\sum_{t=1}^{T-h}A_{3t}\left(\sum_{s=1}^{T}F_s\xi_{st}\right)'.\end{aligned}$$

The first term is of order $O_P\left(N^{-1}\delta_{NT}^{-1}\right)$ by Cauchy-Schwartz inequality. For the second term, we decompose A_{3t} as follows,

$$A_{3t} = \frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s\eta_{st} = \frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\eta_{st} + H\frac{1}{T}\sum_{s=1}^{T}F_s\eta_{st}.$$

Thus, we get that

$$\frac{1}{T^2}\sum_{t=1}^{T-h}A_{3t}\sum_{s=1}^{T}F_s\xi_{st} = \frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\eta_{st}\right)\left(\frac{1}{T}\sum_{s=1}^{T}F_s\xi_{st}\right)' + H\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{1}{T}\sum_{s=1}^{T}F_s\eta_{st}\right)\left(\frac{1}{T}\sum_{s=1}^{T}F_s\xi_{st}\right)'.$$

The first term is of order $O_P\left(N^{-1}\delta_{NT}^{-1}\right)$ using Cauchy-Schwartz and the bounds found before. For the second term, we use the definitions of η_{st} and ξ_{st} to write

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{1}{T}\sum_{s=1}^{T}F_s\eta_{st}\right)\left(\frac{1}{T}\sum_{s=1}^{T}F_s\xi_{st}\right)' &= \frac{1}{NT}\left(\frac{F'F}{T}\right)\left(\frac{1}{\sqrt{TN}}\sum_{t=1}^{T-h}\Lambda' e_t F_t'\right)\left(\frac{1}{\sqrt{TN}}\sum_{s=1}^{T-h}\Lambda' e_s F_s'\right) \\
&= \frac{1}{NT}O_P(1)\,O_P(1)\,O_P(1) = O_P\left(\frac{1}{TN}\right),
\end{aligned}$$

given Assumption 3(c). This completes the proof of part a) since it shows that all the terms except the term that depends on $\frac{1}{T}\sum_{t=1}^{T-h}A_{3t}A_{3t}'$ are asymptotically negligible when $\sqrt{T}/N \to c$.

Proof of part b). Replacing $\tilde{F}_t - HF_t = \tilde{V}^{-1}\left(A_{1t} + A_{2t} + A_{3t} + A_{4t}\right)$ yields

$$H\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}F_t\left(\tilde{F}_t - HF_t\right)' = H\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}F_t\left(A_{1t} + A_{2t} + A_{3t} + A_{4t}\right)'\tilde{V}^{-1} \equiv \sqrt{T}H\left(b_{f1} + b_{f2} + b_{f3} + b_{f4}\right)\tilde{V}^{-1},$$

where A_{it} are as defined previously. Again, we consider each term separately. We can show that $T^{-1}\sum_{t=1}^{T-h}F_tA_{1t}' = O_P\left(\delta_{NT}^{-1}T^{-1/2}\right)$, which implies that $\sqrt{T}Hb_{f1}\tilde{V}^{-1} = o_P(1)$ under our assumptions. Indeed, using the decomposition of A_{1t},

$$\begin{aligned}
b_{f1} &\equiv \frac{1}{T}\sum_{t=1}^{T-h}F_tA_{1t}' = \frac{1}{T}\sum_{t=1}^{T-h}F_t\left(\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)'\gamma_{st}\right) + \frac{1}{T}\sum_{t=1}^{T-h}F_t\left(\frac{1}{T}\sum_{s=1}^{T}F_s'\gamma_{st}\right)H' \\
&\equiv b_{f1.1} + b_{f1.2},
\end{aligned}$$

where $\|b_{f1.1}\| \leq O_P(1)\,O_P\left(\delta_{NT}^{-1}T^{-1/2}\right)$ by an application of Cauchy-Schwartz inequality and the fact that $T^{-1}\sum_{t=1}^{T-h}\|F_t\|^2 = O_P(1)$ and $T^{-1}\sum_{t=1}^{T-h}\left\|\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)'\gamma_{st}\right\|^2 = O_P\left(\delta_{NT}^{-2}T^{-1}\right)$. For $b_{f1.2}$, we have

$$b_{f1.2} \equiv \frac{1}{T}\sum_{t=1}^{T-h}F_t\left(\frac{1}{T}\sum_{s=1}^{T}F_s'\gamma_{st}\right)H' = \frac{1}{T}\left(\frac{1}{T}\sum_{t=1}^{T-h}\sum_{s=1}^{T}F_tF_s'\gamma_{st}\right)H' = O_P\left(T^{-1}\right),$$

since

$$E\left\|\frac{1}{T}\sum_{t=1}^{T-h}\sum_{s=1}^{T}F_tF_s'\gamma_{st}\right\| \leq \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left\|F_tF_s'\gamma_{st}\right\| \leq \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}|\gamma_{st}|\left(E\|F_t\|^2\right)^{1/2}\left(E\|F_s\|^2\right)^{1/2} = O(1),$$

given Assumptions 1 and 2. Thus, $b_{f1} = O_P\left(\delta_{NT}^{-1}T^{-1/2}\right)$. Similarly, we can show that $b_{f2} \equiv T^{-1}\sum_{t=1}^{T-h} F_t A_{2t}' = O_P\left((TN)^{-1/2}\right)$ since by definition of A_{2t},

$$\|b_{f2}\| = \left\|\frac{1}{T^2}\sum_{t=1}^{T-h}\sum_{s=1}^{T} F_t\zeta_{st}\tilde{F}_s'\right\| \leq \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_s\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{T}\sum_{t=1}^{T-h}F_t\zeta_{st}\right\|^2\right)^{1/2} = O_P(1)\,O_P\left(\frac{1}{\sqrt{TN}}\right),$$

where Assumption 3(b) is used to bound the second term. The same exact reasoning applies to show that $b_{f3} \equiv T^{-1}\sum_{t=1}^{T-h} F_t A_{3t}' = O_P\left((TN)^{-1/2}\right)$, with the difference that we use Assumption 3(c) to conclude that $T^{-1}\sum_{s=1}^{T}\left\|T^{-1}\sum_{t=1}^{T-h}F_t\eta_{st}\right\|^2 = O_P\left((TN)^{-1/2}\right)$. To end the proof, we show that $\sqrt{T}b_{f4} = c\Sigma_F\Gamma Q'V^{-1} + o_P(1)$. Replacing A_{4t} with its definition (and decomposing it into two terms) yields

$$b_{f4} \equiv \frac{1}{T}\sum_{t=1}^{T-h}F_tA_{4t}' = \frac{1}{T}\sum_{t=1}^{T-h}F_t\left(\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)'\xi_{st}\right) + \frac{1}{T}\sum_{t=1}^{T-h}F_t\left(\frac{1}{T}\sum_{s=1}^{T}(HF_s)'\xi_{st}\right) \equiv b_{f4.1} + b_{f4.2}.$$

Starting with $b_{f4.2}$, and given that $\xi_{st} = F_t'\frac{\Lambda' e_s}{N}$, we have that

$$b_{f4.2} = \frac{1}{\sqrt{TN}}\left(\frac{1}{T}\sum_{t=1}^{T-h}F_tF_t'\right)\left(\frac{1}{\sqrt{TN}}\sum_{s=1}^{T}\Lambda' e_sF_s'\right)H' = O\left(\frac{1}{\sqrt{TN}}\right),$$

given Assumptions 1, 3(c) and the fact that $H = O_P(1)$. Thus, $\sqrt{T}b_{4f.2} = o_P(1)$. Next, consider $b_{f4.1}$. We have that

$$b_{f4.1} = \frac{1}{T}\sum_{t=1}^{T-h}F_t\frac{1}{T}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)'\left(F_t'\frac{\Lambda' e_s}{N}\right) = \left(\frac{1}{T}\sum_{t=1}^{T-h}F_tF_t'\right)\left(\frac{1}{T}\sum_{s=1}^{T}\frac{\Lambda' e_s}{N}\left(\tilde{F}_s - HF_s\right)'\right),$$

where the first term is $O_P(1)$ and the second term can be shown to be $O_P\left(N^{-1}\right)$. Thus, we will get a non negligible contribution from $b_{f4.1}$ when multiplying by $\sqrt{T}$. Specifically, using the usual decomposition for $\tilde{F}_s - HF_s$, we have that

$$\frac{1}{T}\sum_{s=1}^{T}\frac{\Lambda' e_s}{N}\left(\tilde{F}_s - HF_s\right)' = \frac{1}{T}\sum_{s=1}^{T}\frac{\Lambda' e_s}{N}\left(A_{1s} + A_{2s} + A_{3s} + A_{4s}\right)'\tilde{V}^{-1}. \tag{19}$$

We will show that the only non-negligible term is the term involving A_{3s}. The first term is $O_P\left((NT)^{-1/2}\right)$ by an application of the Cauchy-Schwartz inequality, given Assumption 3(d) and the fact that $T^{-1}\sum_{s=1}^{T}\|A_{1s}\|^2 = O_P\left(T^{-1}\right)$; so this term is $o_P(1)$ when multiplied by $\sqrt{T}$. Similarly, the second term is $O_P\left(N^{-1}\delta_{NT}^{-1}\right)$, given Assumption 3(d) and the fact that $T^{-1}\sum_{s=1}^{T}\|A_{2s}\|^2 = O_P\left(N^{-1}\delta_{NT}^{-2}\right)$. For the last term in (19), a more careful analysis is required. We have that

$$\frac{1}{T}\sum_{s=1}^{T}\frac{\Lambda' e_s}{N}A_{4s}' = \frac{1}{T}\sum_{s=1}^{T}\frac{\Lambda' e_s}{N}\frac{1}{T}\sum_{t=1}^{T}\left(\tilde{F}_t - HF_t\right)'\xi_{ts} + H\frac{1}{T}\sum_{s=1}^{T}\frac{\Lambda' e_s}{N}\frac{1}{T}\sum_{t=1}^{T}F_t'\xi_{ts}. \tag{20}$$

By Cauchy-Schwartz inequality, we can bound the first term in (20) by

$$\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t - HF_t\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{T}\sum_{s=1}^{T}\frac{\Lambda' e_s}{N}\xi_{ts}\right\|^2\right)^{1/2} = O_P\left(\frac{1}{\delta_{NT}}\right)O_P\left(\frac{1}{N}\right) = O_P\left(\frac{1}{N\delta_{NT}}\right),$$

since $\xi_{st} = \eta_{st}$, implying that

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{T}\sum_{s=1}^{T}\frac{\Lambda' e_s}{N}\xi_{ts}\right\|^2 = \frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{T}\sum_{s=1}^{T}\frac{\Lambda' e_s}{N}\eta_{st}\right\|^2 \leq \frac{1}{T}\sum_{s=1}^{T}\left\|\frac{\Lambda' e_s}{N}\right\|^2\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}|\eta_{st}|^2 = O_P\left(\frac{1}{N^2}\right).$$

The second term in (20) can be shown to be $O_P\left((TN)^{-1}\right)$ given Assumption 3(c). Thus, (20) is $O_P\left(N^{-1}\delta_{NT}^{-1}\right)$, which is $o_P(1)$ when multiplied by $\sqrt{T}$. To complete the proof, we analyze the dominant term in $b_{f4.1}$ which comes from the contribution involving A_{3s}. Using the definition of A_{3s}, it follows that

$$\frac{1}{T}\sum_{s=1}^{T}\frac{\Lambda' e_s}{N}A_{3s}'\tilde{V}^{-1} = \frac{1}{N}\frac{1}{T}\sum_{s=1}^{T}\left(\frac{\Lambda' e_s}{\sqrt{N}}\right)\left(\frac{e_s'\Lambda}{\sqrt{N}}\right)\frac{F'\tilde{F}}{T}\tilde{V}^{-1} = \frac{1}{N}\left(\Gamma + o_P(1)\right)Q'V^{-1},$$

given Assumption 3(e) and the fact that $\frac{F'\tilde{F}}{T} = Q' + o_P(1)$ and $\tilde{V} \to^P V$. Thus,

$$\sqrt{T}b_{f4.1} = \left(\frac{1}{T}\sum_{t=1}^{T-h}F_tF_t'\right)\left(\frac{\sqrt{T}}{N}\left(\Gamma + o_P(1)\right)Q'V^{-1}\right) = c\Sigma_F\Gamma Q'V^{-1} + o_P(1).$$

This implies that $\sqrt{T}Hb_{f4.1}\tilde{V}^{-1} = cH_0\Sigma_F\Gamma Q'V^{-1}V^{-1} + o_P(1) = cQ\Gamma Q'V^{-2} + o_P(1)$, because $H_0 = p\lim H$ is such that $H_0\Sigma_F = Q$.

Proof of part c). The proof follows closely the proof of part b) by relying on moment and dependence conditions involving the extra regressors W_t. In particular, writing

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}W_t\left(\tilde{F}_t - HF_t\right)' = \frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}W_t\left(A_{1t} + A_{2t} + A_{3t} + A_{4t}\right)'\tilde{V}^{-1} \equiv \sqrt{T}\left(d_1 + d_2 + d_3 + d_4\right)\tilde{V}^{-1},$$

we verify that $\sqrt{T}d_i = o_P(1)$ for $i = 1, 2, 3$ by using the same arguments as for b_{f1}, b_{f2} and b_{f3}. The only term that has a nonzero contribution is d_4. Following the same arguments as for b_{f4},

$$\begin{aligned}
\sqrt{T}d_4\tilde{V}^{-1} &= \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T-h}W_tF_t'\right)\left(\frac{1}{T}\sum_{s=1}^{T}\frac{\Lambda' e_s}{N}\left(\tilde{F}_s - HF_s\right)'\right)\tilde{V}^{-1} = c\Sigma_{WF}\Gamma Q'V^{-2} + o_P(1) \\
&= c\Sigma_{WF}H_0'V\left(V^{-1}Q\Gamma Q'V^{-1}\right)V^{-1} + o_P(1), \text{ since } H_0'Q = I_r \\
&= c\Sigma_{W\tilde{F}}V\Sigma_{\tilde{F}}V^{-1} + o_P(1), \text{ since } \Sigma_{W\tilde{F}} = \Sigma_{WF}H_0' \text{ and } \Sigma_{\tilde{F}} = V^{-1}Q\Gamma Q'V^{-1}.
\end{aligned}$$

Proof of part d). This follows immediately from parts a), b) and c) of this Lemma.

B Appendix B: Proofs of results in Section 3

This Appendix is organized as follows. First, we provide some auxiliary lemmas, then we prove the results in Section 3, and finally we prove the auxiliary lemmas. In the proofs we will repeatedly use the fact that $O_{P^*}(1) O_P(1) = O_{P^*}(1) O_{P^*}(1) = O_{P^*}(1)$ in probability and $O_{P^*}(1) o_P(1) = O_{P^*}(1) o_{P^*}(1) = o_{P^*}(1)$ in probability. For a proof of these properties as well as other properties that justify the transition between the bootstrap stochastic orders and the stochastic orders for the original sample see Cheng and Huang (2010, Lemma 3) or Chang and Park (2003, Lemma 1).

Lemma B.1 *Let* $H^* = \tilde{V}^{*-1} \frac{\tilde{F}^{*\prime} \tilde{F}}{T} \frac{\tilde{\Lambda}' \tilde{\Lambda}}{N}$. *Under Conditions A*-D*, we have that if* $\delta_{NT} = \min\left(\sqrt{N}, \sqrt{T}\right)$,

a) $H^* H^{*\prime} = I_r + O_{P^*}\left(\delta_{NT}^{-2}\right)$, *in probability, i.e.* H^* *is asymptotically an orthogonal matrix.*

b) $H^* = H_0^* + O_{P^*}\left(\delta_{NT}^{-2}\right)$, *in probability, where* H_0^* *is a diagonal matrix with* ± 1 *on the main diagonal.*

c) $\tilde{V}^* = H^* \tilde{V} H^{*\prime} + O_{P^*}\left(\delta_{NT}^{-2}\right) = \tilde{V} + O_{P^*}\left(\delta_{NT}^{-2}\right)$, *in probability.*

Lemma B.2 *Let Assumptions 1-5 hold and suppose we generate bootstrap data* $\left\{y_{t+h}^*, X_t^*\right\}$ *according to the residual-based bootstrap DGP (7) and (8) by relying on bootstrap residuals* $\left\{\varepsilon_{t+h}^*\right\}$ *and* $\left\{e_t^*\right\}$ *such that Conditions A*-D* are satisfied. Then, as* $N, T \to \infty$,

$$\frac{1}{T} \sum_{t=1}^{T-h} \left(\tilde{F}_t^* - H^* \tilde{F}_t\right) \varepsilon_{t+h}^* = O_{P^*}\left(\frac{1}{\delta_{NT} \sqrt{T}}\right),$$

in probability, for $h \geq 0$.

Lemma B.3 *Suppose conditions A*-D* hold. Then,the following statements hold in probability, as* $N, T \to \infty$,

a)
$$\frac{1}{T} \sum_{t=1}^{T-h} \left(\tilde{F}_t^* - H^* \tilde{F}_t\right) \left(\tilde{F}_t^* - H^* \tilde{F}_t\right)' = \frac{1}{N} \tilde{V}^{*-1} H^* \left[\frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{\tilde{\Lambda}' e_t^*}{\sqrt{N}}\right) \left(\frac{e_t^{*\prime} \tilde{\Lambda}}{\sqrt{N}}\right)\right] H^{*\prime} \tilde{V}^{*-1} + O_{P^*}\left(\frac{1}{T}\right) + O_{P^*}\left(\frac{1}{N \delta_{NT}}\right) + O_{P^*}\left(\frac{1}{\sqrt{NT}}\right).$$

b)
$$\frac{1}{T} \sum_{t=1}^{T-h} H^* \tilde{F}_t \left(\tilde{F}_t^* - H^* \tilde{F}_t\right)' = H^* \frac{1}{N} \left(\frac{1}{T} \sum_{t=1}^{T-h} \tilde{F}_t \tilde{F}_t'\right) \left[\frac{1}{T} \sum_{s=1}^{T} \left(\frac{\tilde{\Lambda}' e_s^*}{\sqrt{N}}\right) \left(\frac{e_s^{*\prime} \tilde{\Lambda}}{\sqrt{N}}\right)\right] \left(\frac{\tilde{F}' \tilde{F}^*}{T}\right) \tilde{V}^{*-2} + O_{P^*}\left(\frac{1}{\delta_{NT} \sqrt{T}}\right) + O_{P^*}\left(\frac{1}{N \delta_{NT}}\right).$$

c)
$$\frac{1}{T} \sum_{t=1}^{T-h} W_t \left(\tilde{F}_t^* - H^* \tilde{F}_t\right)' = \frac{1}{N} \left(\frac{1}{T} \sum_{t=1}^{T-h} W_t \tilde{F}_t'\right) \left[\frac{1}{T} \sum_{s=1}^{T} \left(\frac{\tilde{\Lambda}' e_s^*}{\sqrt{N}}\right) \left(\frac{e_s^{*\prime} \tilde{\Lambda}}{\sqrt{N}}\right)\right] \left(\frac{\tilde{F}' \tilde{F}^*}{T}\right) \tilde{V}^{*-2} + O_{P^*}\left(\frac{1}{\delta_{NT} \sqrt{T}}\right) + O_{P^*}\left(\frac{1}{N \delta_{NT}}\right).$$

Lemma B.4 *Suppose conditions A*-D* hold. For any $h \geq 0$, if $\sqrt{T}/N \to c$, $0 \leq c < \infty$, then*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \tilde{F}_t^* \left(\tilde{F}_t^* - H^* \tilde{F}_t \right)' \left(H^{*\prime} \right)^{-1} \hat{\alpha} = c \left(H_0^{*\prime} \right)^{-1} \underbrace{\left[\tilde{V}^{-1} \Gamma^* \tilde{V}^{-1} + \Gamma^* \tilde{V}^{-2} \right] \hat{\alpha}}_{\equiv B_\alpha^*} + o_{P^*}(1),$$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} W_t \left(\tilde{F}_t^* - H^* \tilde{F}_t \right)' \left(H^{*\prime} \right)^{-1} \hat{\alpha} = \underbrace{c \left[\tilde{\Sigma}_{W\tilde{F}} \Gamma^* \tilde{V}^{-2} \right] \hat{\alpha}}_{\equiv B_\beta^*} + o_{P^*}(1),$$

in probability, where $\tilde{\Sigma}_{W\tilde{F}} = \frac{1}{T} \sum_{t=1}^{T-h} W_t \tilde{F}_t'$.

Proof of Lemma 3.1. The proof is based on the following identity:

$$\tilde{F}_t^* - H^* \tilde{F}_t = \tilde{V}^{*-1} \left(\underbrace{\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_s^* \gamma_{st}^*}_{\equiv A_{1t}^*} + \underbrace{\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_s^* \zeta_{st}^*}_{\equiv A_{2t}^*} + \underbrace{\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_s^* \eta_{st}^*}_{\equiv A_{3t}^*} + \underbrace{\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_s^* \xi_{st}^*}_{\equiv A_{4t}^*} \right),$$

where $\gamma_{st}^* = E^* \left(\frac{1}{N} \sum_{i=1}^N e_{is}^* e_{it}^* \right)$, $\zeta_{st}^* = \frac{1}{N} \sum_{i=1}^N \left(e_{is}^* e_{it}^* - E^* \left(e_{is}^* e_{it}^* \right) \right)$, $\eta_{st}^* = \frac{1}{N} \sum_{i=1}^N \tilde{\lambda}_i' \tilde{F}_s e_{it}^* = \tilde{F}_s' \frac{\tilde{\Lambda}' e_t^*}{N}$ and $\xi_{st}^* = \frac{1}{N} \sum_{i=1}^N \tilde{\lambda}_i' \tilde{F}_t e_{is}^* = \eta_{ts}^*$. Ignoring $\tilde{V}^{*-1}$ (which is $O_{P^*}(1)$), it follows that

$$\frac{1}{T} \sum_{t=1}^{T} \left\| \tilde{F}_t^* - H^* \tilde{F}_t \right\|^2 \leq \frac{4}{T} \sum_{t=1}^{T} \left(\|A_{1t}^*\|^2 + \|A_{2t}^*\|^2 + \|A_{3t}^*\|^2 + \|A_{4t}^*\|^2 \right),$$

By the Cauchy-Schwartz inequality, $\left\| \sum_{s=1}^T \tilde{F}_s^* \gamma_{st}^* \right\|^2 \leq \left(\sum_{s=1}^T \left\| \tilde{F}_s^* \right\|^2 \right) \left(\sum_{s=1}^T \gamma_{st}^{*2} \right)$, implying that

$$\frac{1}{T} \sum_{t=1}^{T} \|A_{1t}^*\|^2 \leq \frac{1}{T} \underbrace{\left(\frac{1}{T} \sum_{s=1}^{T} \left\| \tilde{F}_s^* \right\|^2 \right)}_{= \frac{\|\tilde{F}^*\|^2}{T} = r \text{ because } \frac{\tilde{F}^{*\prime} \tilde{F}^*}{T} = I_r} \underbrace{\left(\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \gamma_{st}^{*2} \right)}_{= O_P(1) \text{ by Condition A*(b).}} = O_P \left(\frac{1}{T} \right).$$

For the second term, we have that $\frac{1}{T} \sum_{t=1}^T \|A_{2t}^*\|^2 \leq \left(\frac{1}{T} \sum_{s=1}^T \left\| \tilde{F}_s^* \right\|^2 \right) \left(\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T |\zeta_{st}^*|^2 \right) = O_{P^*} \left(\frac{1}{N} \right)$. In particular, by Condition A*(c), we can show that

$$\frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} E^* |\zeta_{st}^*|^2 = \frac{1}{N} \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} E^* \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(e_{is}^* e_{it}^* - E^* \left(e_{is}^* e_{it}^* \right) \right) \right|^2 = O_P \left(\frac{1}{N} \right),$$

which explains why the second term is $O_{P^*} \left(\frac{1}{N} \right)$. For the third term,

$$\frac{1}{T} \sum_{t=1}^{T} \|A_{3t}^*\|^2 = \frac{1}{T} \sum_{t=1}^{T} T^{-2} \left\| \sum_{s=1}^{T} \tilde{F}_s^* \tilde{F}_s' \frac{\tilde{\Lambda}' e_t^*}{N} \right\|^2 \leq \frac{1}{T} \sum_{t=1}^{T} \left\| \frac{\tilde{\Lambda}' e_t^*}{N} \right\|^2 \left\| T^{-1} \sum_{s=1}^{T} \tilde{F}_s^* \tilde{F}_s' \right\|^2 = O_{P^*} \left(\frac{1}{N} \right),$$

since $\left\| T^{-1} \sum_{s=1}^T \tilde{F}_s^* \tilde{F}_s' \right\|^2 \leq r^2$, whereas by Condition B*(d) and Markov's inequality, $\frac{1}{T} \sum_{t=1}^T \left\| \frac{\tilde{\Lambda}' e_t^*}{N} \right\|^2 = O_{P^*} \left(\frac{1}{N} \right)$. The fourth term follows by the same arguments.

Proof of Theorem 3.1. The proof follows the proof of Theorem 2.1. In particular, we can write the bootstrap analogue of (17), viz $\sqrt{T}\left(\hat{\delta}^* - \delta^*\right) = \left(\frac{1}{T}\sum_{t=1}^{T-h}\hat{z}_t^*\hat{z}_t^{*\prime}\right)^{-1}(A^* + B^* + C^*)$, where conditional on the original data, with probability converging to one, we have that

$$A^* = \Phi^* \frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\hat{z}_t\varepsilon_{t+h}^* \to^{d^*} N\left(0, \Phi_0^*\left(\Phi_0\Omega\Phi_0'\right)\Phi_0^{*\prime}\right),$$

given Conditions D*(b) and E*, and given that $\Phi_0^* \equiv p\lim\Phi^*$. In addition, $B^* = \frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\left(\tilde{F}_t^* - H^*\tilde{F}_t\right)\varepsilon_{t+h}^* = o_{P^*}(1)$ (given Lemma B.2); and

$$C^* = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\hat{z}_t^*\left(\tilde{F}_t^* - H^*\tilde{F}_t\right)'\left(H^{*-1}\right)'\hat{\alpha} \to^{P^*} -c\left(\Phi_0^{*\prime}\right)^{-1}\begin{pmatrix} B_\alpha^* \\ B_\beta^* \end{pmatrix},$$

where B_α^* and B_β^* are defined in Lemma B.4. Under Assumptions 1-5, $p\lim\tilde{V} = V$, $p\lim\hat{\alpha} = \left(H_0'\right)^{-1}\alpha$, $p\lim\tilde{\Sigma}_{W\tilde{F}} = \Sigma_{W\tilde{F}}$, and $p\lim\Gamma^* = Q\Gamma Q'$ by Condition F*, which implies that $B_\alpha^* \to^P B_\alpha$ and $B_\beta^* \to^P B_\beta$; and finally $\frac{1}{T}\sum_{t=1}^{T-h}\hat{z}_t^*\hat{z}_t^{*\prime} = \Phi_0^*\left(\Phi_0\Sigma_{zz}\Phi_0'\right)\Phi_0^{*\prime} + o_{P^*}(1)$. This implies that $\sqrt{T}\left(\hat{\delta}^* - \delta^*\right) \to^{d^*} N\left(-c\left(\Phi_0^{*\prime}\right)^{-1}\Delta_\delta, \left(\Phi_0^{*\prime}\right)^{-1}\Sigma_\delta\left(\Phi_0^*\right)^{-1}\right)$, in probability.

Proof of Corollary 3.1. By Theorem 2.1, under Assumptions 1-5, and if $\sqrt{T}/N \to c$, $0 \le c < \infty$, $\sqrt{T}\left(\hat{\delta} - \delta\right) \to^d Z \sim N\left(-c\Delta_\delta, \Sigma_\delta\right)$. Thus, from a multivariate version of Polya's Theorem (cf. Bhattacharya and Rao (1986)), it follows that $\sup_x\left|P\left(\sqrt{T}\left(\hat{\delta} - \delta\right) \le x\right) - \Phi\left(x; -c\Delta_\delta, \Sigma_\delta\right)\right| = o(1)$, where $\Phi\left(x; -c\Delta_\delta, \Sigma_\delta\right)$ denotes the distribution function of Z. Then, the result follows if we show that

$$\sup_x\left|P\left(\sqrt{T}\left(\Phi^{*\prime}\hat{\delta}^* - \hat{\delta}\right) \le x\right) - \Phi\left(x; -c\Delta_\delta, \Sigma_\delta\right)\right| = o_P(1). \tag{21}$$

Under the stated assumptions, from Theorem 3.1 we have that $\sqrt{T}\left(\Phi^{*\prime}\hat{\delta}^* - \hat{\delta}\right) \to^{d^*} N\left(-c\Delta_\delta, \Sigma_\delta\right)$, in probability. The result (21) follows from Polya's Theorem.

Proof of Lemma B.1. Part a) follows from part b). For part b), by definition, $H^* = \tilde{V}^{*-1}\left(\frac{\tilde{F}^{*\prime}\tilde{F}}{T}\right)\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N} = \tilde{V}^{*-1}H^*\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N} + O_{P^*}\left(\delta_{NT}^{-2}\right)$, given that $H^* = \tilde{F}^{*\prime}\tilde{F}/T + O_{P^*}\left(\delta_{NT}^{-2}\right)$. Left multiplying both sides by $\tilde{V}^*$ yields

$$\tilde{V}^*H^* = H^*\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N} + O_{P^*}\left(\delta_{NT}^{-2}\right) = H^*\tilde{V} + O_{P^*}\left(\delta_{NT}^{-2}\right), \tag{22}$$

since $\tilde{V} = \frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}$ by construction of the principal components. By Lemma A.3 of Bai (2003), $\tilde{V} \to^P V > 0$ and therefore we can write that $\tilde{V}^*H^* = H^*V + o_{P^*}(1)$, or transposing, that $VH^{*\prime} = H^{*\prime}\tilde{V}^* + o_{P^*}(1)$. Thus, $H^{*\prime}$ is (for large N and T) the matrix of eigenvectors of V. Since V is a diagonal matrix, $H^{*\prime}$ is also diagonal, asymptotically. Moreover, because V has distinct eigenvalues (by assumption), it follows that its eigenvectors have only one nonzero value and this is +1 or -1 (because H^* is orthogonal). Therefore $H^{*\prime}$ is for large N and T a diagonal matrix with ± 1 in the main diagonal (in particular, $H^* = diag(sign(\tilde{F}^{*\prime}\tilde{F})))$. Part c) follows from (22) by right multiplying by $H^{*\prime}$ and using parts a) and b).

Proof of Lemma B.2. The proof follows exactly as the proof of Lemma A.1 using conditions A*-D* instead of Assumptions 1-5.

Proof of Lemma B.3. *Proof of part a).* We follow closely the proof of Lemma A.2. Specifically, we analyze each of the terms in

$$\frac{1}{T}\sum_{t=1}^{T-h}\left(\tilde{F}_t^* - H^*\tilde{F}_t\right)\left(\tilde{F}_t^* - H^*\tilde{F}_t\right)' = \tilde{V}^{*-1}\frac{1}{T}\sum_{t=1}^{T-h}\left(A_{1t}^* + A_{2t}^* + A_{3t}^* + A_{4t}^*\right)\left(A_{1t}^* + A_{2t}^* + A_{3t}^* + A_{4t}^*\right)'\tilde{V}^{*-1}.$$

In particular, by Lemma 3.1, and the appropriate bootstrap high level conditions, we can show that: $\frac{1}{T}\sum_{t=1}^{T-h} A_{1t}^* A_{1t}^{*\prime} = O_{P^*}\left(\frac{1}{T}\right)$, using Condition A*(b); $\frac{1}{T}\sum_{t=1}^{T-h} A_{2t}^* A_{2t}^{*\prime} = O_{P^*}\left(\frac{1}{N\delta_{NT}^2}\right)$, using Conditions A*(c) and B*(b); $\frac{1}{T}\sum_{t=1}^{T-h} A_{3t}^* A_{3t}^{*\prime} = \frac{1}{N}H^*\left(\frac{\tilde{F}'\tilde{F}}{T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{\tilde{\Lambda}' e_t^*}{\sqrt{N}}\right)\left(\frac{e_t^{*\prime}\tilde{\Lambda}}{\sqrt{N}}\right)\left(\frac{\tilde{F}'\tilde{F}}{T}\right)H^{*\prime} + O_{P^*}\left(\frac{1}{N\delta_{NT}}\right)$, using Condition B*(d); each of $\frac{1}{T}\sum_{t=1}^{T-h} A_{4t}^* A_{4t}^{*\prime}$, $\frac{1}{T}\sum_{t=1}^{T-h} A_{2t}^* A_{3t}^{*\prime}$, $\frac{1}{T}\sum_{t=1}^{T-h} A_{2t}^* A_{4t}^{*\prime}$ and $\frac{1}{T}\sum_{t=1}^{T-h} A_{3t}^* A_{4t}^{*\prime}$ is $O_{P^*}\left(\frac{1}{N\delta_{NT}}\right)$, using condition B*(d); $\frac{1}{T}\sum_{t=1}^{T-h} A_{1t}^* A_{2t}^{*\prime} = O_{P^*}\left(\frac{1}{\delta_{NT}\sqrt{NT}}\right)$; $\frac{1}{T}\sum_{t=1}^{T-h} A_{1t}^* A_{3t}^{*\prime}$ and $\frac{1}{T}\sum_{t=1}^{T-h} A_{1t}^* A_{4t}^{*\prime}$ are $O_{P^*}\left(\frac{1}{\sqrt{NT}}\right)$, in probability. This implies the result. *Proof of part b).* We analyze each of the terms in

$$H^*\frac{1}{T}\sum_{t=1}^{T-h}\tilde{F}_t\left(A_{1t}^* + A_{2t}^* + A_{3t}^* + A_{4t}^*\right)'\tilde{V}^{*-1} = H^*\left(b_{f1}^* + b_{f2}^* + b_{f3}^* + b_{f4}^*\right)\tilde{V}^{*-1},$$

where we let $b_{fi}^* \equiv \frac{1}{T}\sum_{t=1}^{T-h}\tilde{F}_t A_{it}^{*\prime}$, for $j = 1, \ldots, 4$. Following exactly the same steps as in the proof of part b) of Lemma A.2, we can show that: $b_{f1}^* = O_{P^*}\left(\frac{1}{\delta_{NT}\sqrt{T}}\right)$, in probability, given Condition B*(a) and Lemma 3.1; $b_{f2}^* = O_{P^*}\left(\frac{1}{\sqrt{TN}}\right)$, given Condition B*(b) and Lemma 3.1; $b_{f3}^* = O_{P^*}\left(\frac{1}{\sqrt{TN}}\right)$, given Condition B*(c) and Lemma 3.1; $b_{f4}^* = \frac{1}{N}\left(\frac{1}{T}\sum_{t=1}^{T-h}\tilde{F}_t\tilde{F}_t'\right)\left[\frac{1}{T}\sum_{s=1}^{T}\left(\frac{\tilde{\Lambda}' e_s^*}{\sqrt{N}}\right)\left(\frac{e_s^{*\prime}\tilde{\Lambda}}{\sqrt{N}}\right)\right]\left(\frac{\tilde{F}'\tilde{F}^*}{T}\right)\tilde{V}^{*-1} + O_{P^*}\left(\frac{1}{\sqrt{TN}}\right) + O_{P^*}\left(\frac{1}{N\delta_{NT}}\right)$, given Condition B*(c) and Lemma 3.1. *Proof of part c).* This follows by the same arguments used in the proof of Lemma A.1.c).

Proof of Lemma B.4. Part a) follows from Lemma B.3.a) and Lemma B.1.c), given in particular the assumption that $\sqrt{T}/N \to c$, the fact that $H_0^* = diag(\pm 1)$, and given Condition B*(e). Similarly, part b) follows from part b) of Lemma B.3 and part c) of Lemma B.1, given Condition B*(e) and the fact that $\frac{1}{T}\sum_{t=1}^{T-h}\tilde{F}_t\tilde{F}_t' = I_r + o_P(1)$, that $\frac{\tilde{F}'\tilde{F}^*}{T}\tilde{V}^{*-1} = \tilde{V}^{-1}H^{*\prime}$, and that $H^{*-1}\left(H^{*\prime}\right)^{-1} = I_r + o_{P^*}(1)$, in probability. Part c) follows similarly using part c) of Lemma B.3.

C Appendix C: Proofs of results in Section 4

First, we state an auxiliary result and its proof. Then we prove Theorem 4.1.

Lemma C.1 *Suppose Assumptions 1-5 hold. If in addition either: (1) $\{F_s\}$, $\{\lambda_i\}$ and $\{e_{it}\}$ are mutually independent and for some $p \geq 2$, $E\left|e_{it}\right|^{2p} \leq M < \infty$, $E\left\|\lambda_i\right\|^p \leq M < \infty$ and $E\left\|F_t\right\|^p \leq M < \infty$, or (2) for some $p \geq 2$, $E\left|e_{it}\right|^{3p} \leq M < \infty$, $E\left\|\lambda_i\right\|^{3p} \leq M < \infty$ and $E\left\|F_t\right\|^{3p} \leq M < \infty$, it follows that (i) $\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t - HF_t\right\|^p = O_P(1)$; (ii) $\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_i - H^{-1\prime}\lambda_i\right\|^p = O_P(1)$; and (iii) $\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\tilde{e}_{it}^p = O_P(1)$.*

Proof of Lemma C.1. Proof of (i). We rely on the following identity (see Bai and Ng (2002), proof of Theorem 1): $\tilde{F}_t - HF_t = \tilde{V}^{-1}\left(\frac{1}{T}\sum_{s=1}^T \tilde{F}_s\psi_{st} + \frac{1}{T}\sum_{s=1}^T \tilde{F}_s\eta_{st} + \frac{1}{T}\sum_{s=1}^T \tilde{F}_s\xi_{st}\right)$, where $\psi_{st} = \frac{1}{N}\sum_{i=1}^N e_{is}e_{it}$; $\eta_{st} = \frac{1}{N}\sum_{i=1}^N \lambda_i' F_s e_{it}$; and $\xi_{st} = \eta_{ts}$. It follows that by the c_r inequality,

$$\frac{1}{T}\sum_{t=1}^T \left\|\tilde{F}_t - HF_t\right\|^p \leq 3^{p-1}\left\|\tilde{V}^{-1}\right\|^p \left(\frac{1}{T}\sum_{t=1}^T a_t + \frac{1}{T}\sum_{t=1}^T b_t + \frac{1}{T}\sum_{t=1}^T c_t\right),$$

where $a_t = \frac{1}{T^p}\left\|\sum_{s=1}^T \tilde{F}_s\psi_{st}\right\|^p$; $b_t = \frac{1}{T^p}\left\|\sum_{s=1}^T \tilde{F}_s\eta_{st}\right\|^p$; and $c_t = \frac{1}{T^p}\left\|\sum_{s=1}^T \tilde{F}_s\xi_{st}\right\|^p$. Let χ_{st} denote either ψ_{st}, η_{st} or ξ_{st}. We can write

$$\left\|\sum_{s=1}^T \tilde{F}_s\chi_{st}\right\|^p = \left(\left\|\sum_{s=1}^T \tilde{F}_s\chi_{st}\right\|^2\right)^{p/2} \leq \left(\sum_{s=1}^T \left\|\tilde{F}_s\right\|^2 \sum_{s=1}^T |\chi_{st}|^2\right)^{p/2}.$$

It follows that

$$\frac{1}{T}\sum_{t=1}^T \frac{1}{T^p}\left\|\sum_{s=1}^T \tilde{F}_s\chi_{st}\right\|^p \leq r^{p/2}\frac{1}{T}\sum_{t=1}^T\left(\frac{1}{T}\sum_{s=1}^T |\chi_{st}|^2\right)^{p/2} \leq r^{p/2}\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T |\chi_{st}|^p,$$

given that $\frac{1}{T}\sum_{s=1}^T \left\|\tilde{F}_s\right\|^2 = r$ and where the last inequality follows again by the c_r inequality. Thus it suffices to show that $E|\chi_{st}|^p \leq M < \infty$ to prove that the above term is $O_P(1)$. Starting with $\chi_{st} = \psi_{st}$,

$$E|\psi_{st}|^p = E\left|\frac{1}{N}\sum_{i=1}^N e_{it}e_{is}\right|^p \leq \frac{1}{N}\sum_{i=1}^N E|e_{it}e_{is}|^p \leq \frac{1}{N}\sum_{i=1}^N\left(E|e_{it}|^{2p}\right)^{1/2}\left(E|e_{is}|^{2p}\right)^{1/2} \leq M < \infty,$$

given that we assume $E|e_{it}|^{2p} \leq M < \infty$. When $\chi_{st} = \eta_{st}$, we have that

$$E\left|\lambda_i' F_s e_{it}\right|^p \leq \left(E\|\lambda_i e_{it}\|^{\frac{3p}{2}}\right)^{2/3}\left(E\|F_s\|^{3p}\right)^{1/3} \leq \left(E\|\lambda_i\|^{3p}E|e_{it}|^{3p}\right)^{1/3}\left(E\|F_s\|^{3p}\right)^{1/3} \leq M,$$

which implies $E|\eta_{st}|^p \leq M$. Note that if we assume that $\{\lambda_i\}$, $\{F_s\}$ and $\{e_{it}\}$ are three mutually independent groups of random variables, then it suffices that $E\|\lambda_i\|^p \leq M$, $E\|F_s\|^p \leq M$, and $E|e_{it}|^p \leq M$ to bound $E\left|\lambda_i' F_s e_{it}\right|^p$. The term that depends on $\chi_{st} = \xi_{st}$ can be dealt with similarly.

Proof of (ii). Note that $\tilde{\Lambda} = \frac{X'\tilde{F}}{T}$. Since $X = F\Lambda' + e$, it follows that $\tilde{\Lambda}' = \frac{\tilde{F}'F}{T}\Lambda' + \frac{\tilde{F}'e}{T}$, thus implying that $\tilde{\lambda}_i = \frac{\tilde{F}'F}{T}\lambda_i + \frac{\tilde{F}'e_i}{T}$, where $e_i = (e_{i1}, \ldots, e_{iT})'$. We can write

$$\tilde{\lambda}_i = \frac{\tilde{F}'FH'}{T}H'^{-1}\lambda_i + \frac{\tilde{F}'e_i}{T} = H'^{-1}\lambda_i - T^{-1}\tilde{F}'\left(\tilde{F} - FH'\right)H'^{-1}\lambda_i + T^{-1}\left(\tilde{F} - FH'\right)'e_i + T^{-1}\left(FH'\right)'e_i.$$

Thus,

$$\frac{1}{N}\sum_{i=1}^N \left\|\tilde{\lambda}_i - H^{-1\prime}\lambda_i\right\|^p \leq 3^{p-1}\left(\begin{array}{c}\frac{1}{N}\sum_{i=1}^N \left\|T^{-1}\tilde{F}'\left(\tilde{F} - FH'\right)H'^{-1}\lambda_i\right\|^p \\ +\frac{1}{N}\sum_{i=1}^N \left\|T^{-1}\left(\tilde{F} - FH'\right)'e_i\right\|^p + \frac{1}{N}\sum_{i=1}^N \left\|T^{-1}\left(FH'\right)'e_i\right\|^p\end{array}\right).$$

For the first term, we have that

$$\frac{1}{N}\sum_{i=1}^{N}\left\|T^{-1}\tilde{F}'\left(\tilde{F}-FH'\right)H'^{-1}\lambda_i\right\|^p \leq \left\|T^{-1/2}\tilde{F}\right\|^p \left\|T^{-1/2}\left(\tilde{F}-FH'\right)\right\|^p \left\|H'^{-1}\right\|^p \frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^p,$$

where $\left\|T^{-1/2}\tilde{F}\right\|^p = r^{p/2}$, $\left\|T^{-1/2}\left(\tilde{F}-FH'\right)\right\|^p = O_P\left(\delta_{NT}^{-p}\right) = O_P(1)$, $\left\|H'^{-1}\right\|^p = O_P(1)$, and the last factor is $O_P(1)$ given that $E\|\lambda_i\|^p \leq M < \infty$. For the second term, $\frac{1}{N}\sum_{t=1}^{N}\left\|T^{-1}\left(\tilde{F}-FH'\right)' e_i\right\|^p \leq \left\|T^{-1/2}\left(\tilde{F}-FH'\right)\right\|^p \frac{1}{N}\sum_{t=1}^{N}\left\|T^{-1/2}e_i\right\|^p$, where the first factor is $O_P(1)$ and the second factor is dominated by

$$\frac{1}{N}\sum_{i=1}^{N}\left(\left\|T^{-1/2}e_i\right\|^2\right)^{p/2} = \frac{1}{N}\sum_{t=1}^{N}\left(T^{-1}e_i'e_i\right)^{p/2} = \frac{1}{N}\sum_{t=1}^{N}\left(T^{-1}\sum_{t=1}^{T}e_{it}^2\right)^{p/2} \leq \frac{1}{NT}\sum_{t=1}^{N}\sum_{t=1}^{T}e_{it}^p,$$

which is $O_P(1)$ given the assumption that $E|e_{it}|^p \leq M$. The third term can be bounded similarly using in particular the fact that $E\|F_t\|^2 \leq M < \infty$.

Proof of (iii). We can write $\tilde{e}_{it} = e_{it} - \lambda_i' H^{-1}\left(\tilde{F}_t - HF_t\right) - \left(\tilde{\lambda}_i - H^{-1\prime}\lambda_i\right)'\tilde{F}_t$, which implies that

$$\begin{aligned}\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}|\tilde{e}_{it}|^p \quad &\leq \quad 3^{p-1}\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}|e_{it}|^p + 3^{p-1}\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^p\left\|H^{-1}\right\|^p\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t - HF_t\right\|^p \\ &\quad +3^{p-1}\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_i - H^{-1\prime}\lambda_i\right\|^p\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t\right\|^p.\end{aligned}$$

The first term is $O_P(1)$ given that $E|e_{it}|^p = O(1)$; the second term is $O_P(1)$ since $E\|\lambda_i\|^p = O(1)$ and given part (i); and the third term is $O_P(1)$ given parts (ii) and (iii), since in particular

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t\right\|^p \leq \frac{1}{T}\sum_{t=1}^{T}\left\|HF_t + \left(\tilde{F}_t - HF_t\right)\right\|^p \leq 2^{p-1}\left(\|H\|^p\frac{1}{T}\sum_{t=1}^{T}\|F_t\|^p + \frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t - HF_t\right\|^p\right) = O_P(1).$$

Proof of Theorem 4.1. We verify Condition A*-F*. We start with Condition A*. Since $e_{it}^* = \tilde{e}_{it}\eta_{it}$ where η_{it} is i.i.d.$(0,1)$ across (i,t), part a) follows immediately. For part b), note that $\gamma_{st}^* = \frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{it}\tilde{e}_{is}1(t=s)$, which implies that $\frac{1}{T}\sum_{t,s}^{T}\gamma_{st}^{*2} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{it}^2\right)^2$. This expression is bounded by $\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{it}^4$, which is $O_P(1)$ under our assumptions by an application of Lemma C.1 (iii) with $p=4$. For c), note that for any (t,s),

$$E^*\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(e_{it}^*e_{is}^* - E^*\left(e_{it}^*e_{is}^*\right)\right)\right|^2 = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}Cov^*\left(e_{it}^*e_{is}^*, e_{jt}^*e_{js}^*\right) = \frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{it}^2\tilde{e}_{is}^2 Var\left(\eta_{it}\eta_{is}\right),$$

given that $Cov^*\left(e_{it}^*e_{is}^*, e_{jt}^*e_{js}^*\right) = 1(i=j)\tilde{e}_{it}^2\tilde{e}_{is}^2 Var\left(\eta_{it}\eta_{is}\right)$. Thus, condition A*(c) becomes

$$\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{it}^2\tilde{e}_{is}^2\underbrace{Var\left(\eta_{it}\eta_{is}\right)}_{\leq\bar{\eta}} \leq \bar{\eta}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{e}_{it}^2\right)^2 \leq \bar{\eta}C\frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{e}_{it}^4 = O_P(1),$$

for some constants $\bar{\eta}$ and C, which holds given Lemma C.1 (iii) with $p = 4$. For Condition B*(a), using the bootstrap time series independence, we have that $\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{F}_s\tilde{F}_t'\gamma_{st}^* = \frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t\tilde{F}_t'\gamma_{tt}^* = \frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t\tilde{F}_t'\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{it}^2\right)$, which is bounded by

$$\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t\tilde{F}_t'\right\|^2\right)^{1/2}\left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{it}^2\right)^2\right]^{1/2} \leq \left(\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t\right\|^4\right)^{1/2}\left[\frac{1}{T}\frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{N}\tilde{e}_{it}^4\right]^{1/2} = O_P(1).$$

For Condition B*(b), we have that

$$\frac{1}{T}\sum_{t=1}^{T}E^*\left\|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T}\sum_{i=1}^{N}\hat{z}_s\left(e_{it}^*e_{is}^* - E^*\left(e_{it}^*e_{is}^*\right)\right)\right\|^2 = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{T}\sum_{s=1}^{T}\sum_{l=1}^{T}\hat{z}_s'\hat{z}_l\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}Cov^*\left(e_{it}^*e_{is}^*, e_{jt}^*e_{jl}^*\right),$$

where by the bootstrap cross sectional independence, we can show that $Cov^*\left(e_{it}^*e_{is}^*, e_{jt}^*e_{jl}^*\right) = 0$ when $i \neq j$ for any t, s, l, and when $i = j$, $Cov^*\left(e_{it}^*e_{is}^*, e_{it}^*e_{il}^*\right) = 1(s = l)\tilde{e}_{it}^2\tilde{e}_{is}^2 Var^*\left(\eta_{it}\eta_{is}\right)$. It follows that

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}E^*\left\|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T}\sum_{i=1}^{N}\hat{z}_s\left(e_{it}^*e_{is}^* - E^*\left(e_{it}^*e_{is}^*\right)\right)\right\|^2 &\leq \bar{\eta}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{e}_{it}^2\right)\left(\frac{1}{T}\sum_{s=1}^{T}\hat{z}_s'\hat{z}_s\tilde{e}_{is}^2\right)\\
\leq\ \bar{\eta}\left[\frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{e}_{it}^4\right]^{1/2}\left[\frac{1}{T}\sum_{s=1}^{T}\|\hat{z}_s\|^4\frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{s=1}^{T}\tilde{e}_{is}^4\right]^{1/2} &= O_P(1).
\end{aligned}$$

Next consider Condition B*(c). We can show that

$$E^*\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\tilde{\Lambda}'e_t^*}{\sqrt{N}}\hat{z}_t'\right\|^2 = \frac{1}{T}\sum_{t=1}^{T}\|\hat{z}_t\|^2\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_i\right\|^2\tilde{e}_{it}^2 \leq \left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{z}_t\|^4\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_i\right\|^2\tilde{e}_{it}^2\right)^2\right)^{1/2},$$

where by Cauchy-Schwartz,

$$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_i\right\|^2\tilde{e}_{it}^2\right)^2 \leq \frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_i\right\|^4\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{it}^4 = O_P(1).$$

In particular, $\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_i\right\|^4 \leq 2^3\left(\frac{1}{N}\sum_{i=1}^{N}\left\|H^{-1\prime}\lambda_i\right\|^4 + \frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_i - H^{-1\prime}\lambda_i\right\|^4\right) = O_P(1)$ by Lemma C.1(ii) with $p = 4$. For Condition B*(d), we have that

$$\frac{1}{T}\sum_{t=1}^{T}E^*\left\|\frac{\tilde{\Lambda}'e_t^*}{\sqrt{N}}\right\|^2 = \frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_i\right\|^2\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{e}_{it}^2\right) \leq \left(\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_i\right\|^4\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\tilde{e}_{it}^4\right)^{1/2} = O_P(1).$$

For Condition B*(e), we show that $A^* = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\tilde{\lambda}_i\tilde{\lambda}_j'\left(e_{it}^*e_{jt}^* - E^*\left(e_{it}^*e_{jt}^*\right)\right) = o_P(1)$. This expression has mean zero under the bootstrap measure by construction. So, it suffices to show that its variance tends to zero in probability. Take the case where the number of factors r is equal to 1, for simplicity. Then,

$$Var^*\left(A^*\right) = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{N^2}\sum_{i,j,k,l}^{N}\tilde{\lambda}_i\tilde{\lambda}_j\tilde{\lambda}_l\tilde{\lambda}_k Cov^*\left(e_{it}^*e_{jt}^*, e_{ls}^*e_{ks}^*\right).$$

Using the properties of the wild bootstrap, we can show that $Cov^*\left(e_{it}^* e_{jt}^*, e_{ls}^* e_{ks}^*\right) = 0$ if $t \neq s$, for any (i, j, k, l), and

$$Cov^*\left(e_{it}^* e_{jt}^*, e_{lt}^* e_{kt}^*\right) = \begin{cases} \tilde{e}_{it}^4 Var^*\left(\eta_{it}^2\right) \text{ if } i = j = k = l \\ \tilde{e}_{it}^2 \tilde{e}_{jt}^2, \text{ if } i = k \neq j = l \\ 0, \text{ otherwise} \end{cases},$$

which implies that $Var^*(A^*) \leq \bar{\eta} \frac{1}{T} \frac{1}{N} \sum_{i=1}^N \tilde{\lambda}_i^4 \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \tilde{e}_{it}^4 = O_P\left(\frac{1}{T}\right) = o_P(1)$, given that $\frac{1}{N} \sum_{i=1}^N \tilde{\lambda}_i^4 = O_P(1)$ and $\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \tilde{e}_{it}^4 = O_P(1)$ by an application of lemma C.1 with $p = 4$. Thus, $\Gamma^* = \frac{1}{T} \sum_{t=1}^T \frac{1}{N} \sum_{i=1}^N \tilde{\lambda}_i^2 \tilde{e}_{it}^2$ for the wild bootstrap. Condition F* is satisfied because by Bai and Ng (2006), $\Gamma^* \to^P Q \Gamma Q'$. Next, we verify Condition C*. For $t = 1, \ldots, T-h$, let $\varepsilon_{t+h}^* = \hat{\varepsilon}_{t+h} v_{t+h}$, where $v_{t+h} \sim$i.i.d.$(0,1)$. Part a) follows as Condition B*(b), using the independence between ε_{t+h}^* and e_{it}^*. For part b), we have that

$$E^*\left\|\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \frac{\tilde{\Lambda}' e_t^*}{\sqrt{N}} \varepsilon_{t+h}^*\right\|^2 = \frac{1}{TN} E^*\left\{\left(\sum_{t=1}^{T-h} \sum_{s=1}^{T-h} \varepsilon_{t+h}^* \varepsilon_{s+h}^* \left(\sum_{i=1}^N \tilde{\lambda}_i' e_{it}^*\right)\left(\sum_{j=1}^N \tilde{\lambda}_j e_{js}^*\right)\right)\right\}$$

$$= \frac{1}{T} \sum_{t=1}^{T-h} \hat{\varepsilon}_{t+h}^2 \frac{1}{N} \sum_{i=1}^N \left\|\tilde{\lambda}_i\right\|^2 \tilde{e}_{it}^2 \leq \left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{\varepsilon}_{t+h}^4\right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{1}{N} \sum_{i=1}^N \left\|\tilde{\lambda}_i\right\|^2 \tilde{e}_{it}^2\right)^2\right)^{1/2} = O_P(1),$$

where the second equality uses the independence between $\{e_{it}^*\}$ and $\{\varepsilon_s^*\}$ and the fact that $E^*\left(\varepsilon_{t+h}^* \varepsilon_{s+h}^*\right) = 0$ if $t \neq s$ and $E^*\left(e_{it}^* e_{js}^*\right) = 1(i = j \text{ and } t = s)\, \tilde{e}_{it}^2$. Part c) follows because $\gamma_{st}^* = 0$ for $t \neq s$ and by repeated application of Cauchy-Schwartz inequality, we have that

$$\frac{1}{T} \sum_{t=1}^T \tilde{F}_t \varepsilon_{t+h}^* \gamma_{tt}^* \leq \left(\frac{1}{T} \sum_{t=1}^T \left\|\tilde{F}_t\right\|^4 \frac{1}{T} \sum_{t=1}^{T-h} \left\|\varepsilon_{t+h}^*\right\|^4\right)^{1/4} \left(\frac{1}{T} \frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N \tilde{e}_{it}^4\right)^{1/2},$$

which is $O_{P^*}(1)$ provided $\frac{1}{T} \sum_{t=1}^{T-h} \varepsilon_{t+h}^{*4} = O_{P^*}(1)$ in probability. For this, it suffices that $\frac{1}{T} \sum_{t=1}^{T-h} E^*\left(\varepsilon_{t+h}^{*4}\right) = O_P(1)$. But by the properties of the wild bootstrap on ε_{t+h}^*, we have that for some constant $C < \infty$. $\frac{1}{T} \sum_{t=1}^{T-h} E^*\left(\varepsilon_{t+h}^{*4}\right) = \frac{1}{T} \sum_{t=1}^{T-h} \hat{\varepsilon}_{t+h}^4 E^*\left(v_{t+h}^4\right) \leq C \frac{1}{T} \sum_{t=1}^{T-h} \hat{\varepsilon}_{t+h}^4 = O_P(1)$. Finally. we verify Condition D*. $E^*\left(\varepsilon_{t+h}^*\right) = 0$ by construction. Moreover, we have that $\frac{1}{T} \sum_{t=1}^{T-h} E^*\left|\varepsilon_{t+h}^*\right|^2 = \frac{1}{T} \sum_{t=1}^{T-h} \hat{\varepsilon}_{t+h}^2 \leq \left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{\varepsilon}_{t+h}^4\right)^{1/2} = O_P(1)$. So, part a) is verified. For part b), let $w_t^* \equiv (\Omega^*)^{-1/2} \hat{z}_t \varepsilon_{t+h}^*$ and note that w_t^* is an heterogeneous array of independent random vectors (given that ε_{t+h}^* is conditionally independent but heteroskedastic). Thus, we apply a CLT for heterogeneous independent vectors (see e.g. Proposition 2.27 of van der Vaart (1998)). Since $E^*(w_t^*) = 0$ and $Var^*\left(T^{-1/2} \sum_{t=1}^{T-h} w_t^*\right) = I$, it suffices to verify Lyapunov's condition (a sufficient condition for Lindeberg's condition). In particular, we can show that for some $d > 1$, $T^{-d} \sum_{t=1}^{T-h} E^*\left\|w_t^*\right\|^{2d} = O_P\left(T^{1-d}\right) = o_P(1)$. Condition E* is satisfied since Ω^* converges to $\Phi_0 \Omega \Phi_0' > 0$, by Bai and Ng (2006) and Condition F* was verified in the main text.

References

[1] Bai, J., 2003. "Inferential theory for factor models of large dimensions," *Econometrica*, 71, 135-172.

[2] Bai, J., 2009. "Panel data models with interactive fixed effects," *Econometrica*, 77, 1229-1279.

[3] Bai, J. and S. Ng, 2002. "Determining the number of factors in approximate factor models," *Econometrica*, 70, 191-221.

[4] Bai, J. and S. Ng, 2006. "Confidence intervals for diffusion index forecasts and inference with factor-augmented regressions," *Econometrica*, 74, 1133-1150.

[5] Bai, J. and S. Ng, 2011. "Principal Components Estimation and Identification of the Factors", manuscript, Columbia University.

[6] Bhattacharya, R. and R. Rao, 1986. Normal Approximation and Asymptotic Expansions, Krieger.

[7] Boivin, J. and B. S. Bernanke, 2003, "Monetary policy in a data-rich environment", *Journal of Monetary Economics*, 50, 525-546.

[8] Chang, Y. and J. Park, 2003. "A sieve bootstrap for the test of a unit root", *Journal of Time Series Analysis*, 24, 379-400.

[9] Chamberlain, G. and M. Rothschild, 1983. "Arbitrage, factor structure and mean-variance analysis in large asset markets," *Econometrica*, 51, 1305-1324.

[10] Cheng, G. and J.Z. Huang, 2010. "Bootstrap consistency for general semiparametric M-estimation", *Annals of Statistics*, 38, 2884-2915.

[11] Connor, G. and R. A. Korajczyk, 1986. "Performance measurement with the arbitrage pricing theory," *Journal of Financial Economics*, 15, 373-394.

[12] Connor, G. and R. A. Korajczyk, 1993. "A test for the number of factors in an approximate factor model," *Journal of Finance*, XLVIII, 1263-1291.

[13] Corradi, V. and N. Swanson, 2011. "Testing for Structural Stability of Factor Augmented Forecasting Models", manuscript.

[14] Eichengreen, B. A. Mody, M. Nedeljkovic, and L. Sarno, 2012. "How the Subprime Crisis Went Global: Evidence from Bank Credit Default Swap Spreads", *Journal of International Money and Finance*, 31, 1299-1318

[15] Gospodinov, N. and S. Ng, 2011. "Commodity prices, convenience yields and inflation," *Review of Economics and Statistics*, forthcoming.

[16] Ludvigson, S. and S. Ng, 2007. "The empirical risk return relation: a factor analysis approach," *Journal of Financial Economics*, 83, 171-222.

[17] Ludvigson, S. and S. Ng, 2009. "Macro factors in bond risk premia," *Review of Financial Studies*, 22, 5027-5067.

[18] Ludvigson, S. and S. Ng, 2011. "A factor analysis of bond risk premia," in Ullah, A. and D. Giles eds., Handbook of Empirical Economics and Finance, Chapman and Hall, 313-372,

[19] Onatski, A., 2011. "Asymptotics of the principal components estimator of large factor models with weakly influential factors," manuscript, University of Cambridge.

[20] Shintani, M. and Z-Y Guo, 2011. "Finite sample performance of principal components estimators for dynamic factor models: asymptotic vs. bootstrap approximations," manuscript, Vanderbilt University.

[21] Stock, J. H. and M. Watson, 2002. "Forecasting using principal components from a large number of predictors," *Journal of the American Statistical Association*, 97, 1167-1179.

[22] van der Vaart, A.W., 1998. Asymptotics Statistics, Cambridge University Press.

[23] Yamamoto, Y., 2011. "Bootstrap inference for impulse response functions in factor-augmented vector autoregressions," manuscript, University of Alberta.

Table 1: Bias and coverage rate of 95% CIs for delta - Homoskedastic cases

		N = 50			N = 100			N = 200		
		T = 50	T = 100	T = 200	T = 50	T = 100	T = 200	T = 50	T = 100	T = 200
						Bias				
DGP 1 alpha = 0 homo, homo	bias	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	plug-in	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	WB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
						Coverage rate				
	OLS	95.0	94.8	95.0	94.2	95.3	94.9	94.6	95.1	94.6
	plug-in	92.8	92.1	92.3	92.8	94.1	93.8	93.8	94.6	93.8
	True factor	95.1	94.6	95.2	94.4	95.1	94.9	94.5	95.4	94.4
	WB	97.3	96.3	96.0	96.0	95.8	95.1	96.0	95.5	94.8
						Bias				
DGP 2 alpha = 1 homo, homo	bias	-0.17	-0.14	-0.12	-0.11	-0.08	-0.07	-0.09	-0.06	-0.04
	plug-in	-0.09	-0.09	-0.10	-0.05	-0.05	-0.05	-0.03	-0.03	-0.03
	WB	-0.12	-0.10	-0.10	-0.09	-0.07	-0.06	-0.07	-0.05	-0.04
						Coverage rate				
	OLS	71.5	65.2	52.0	84.6	84.2	80.3	89.1	90.1	89.0
	plug-in	84.4	85.8	86.4	88.8	91.2	91.9	90.6	92.6	92.5
	True factor	95.1	94.6	95.2	94.4	95.1	94.9	94.5	95.4	94.4
	WB	92.2	90.9	90.4	93.4	94.0	94.0	94.1	94.9	94.2

Each part of the table reports estimates of the bias in the estimation of α and the associated coverage rate for the OLS estimator, the bias-corrected estimator that corrects bias by plugging in sample analogues of the quantities in Theorem 3.1, and the wild bootstrap respectively. For reference, we also report coverage rate for the OLS estimator that uses the true factors instead of factors estimated by principal components. All results are based on 5000 replications and B=399 bootstraps.

Table 2. Bias in estimation of alpha - More general cases

		N = 50			N = 100			N = 200		
		T = 50	T = 100	T = 200	T = 50	T = 100	T = 200	T = 50	T = 100	T = 200
DGP 3	bias	-0.17	-0.14	-0.13	-0.12	-0.09	-0.07	-0.09	-0.06	-0.04
alpha = 1	plug-in	-0.09	-0.09	-0.10	-0.05	-0.05	-0.05	-0.03	-0.03	-0.03
hetero, homo	WB	-0.12	-0.10	-0.10	-0.09	-0.07	-0.06	-0.07	-0.05	-0.04
DGP 4	bias	-0.19	-0.15	-0.14	-0.12	-0.10	-0.08	-0.10	-0.06	-0.05
alpha = 1	plug-in	-0.10	-0.10	-0.10	-0.05	-0.06	-0.06	-0.03	-0.03	-0.03
hetero, hetero	WB	-0.13	-0.12	-0.11	-0.10	-0.08	-0.07	-0.08	-0.06	-0.04
DGP 5	bias	-0.20	-0.16	-0.14	-0.14	-0.10	-0.08	-0.10	-0.07	-0.05
alpha = 1	plug-in	-0.09	-0.10	-0.10	-0.05	-0.06	-0.06	-0.03	-0.03	-0.03
hetero, AR+hetero	WB	-0.12	-0.12	-0.11	-0.09	-0.08	-0.07	-0.07	-0.06	-0.04
DGP 6	bias	-0.14	-0.12	-0.12	-0.08	-0.07	-0.07	-0.05	-0.04	-0.03
alpha = 1	plug-in	-0.07	-0.07	-0.07	-0.04	-0.04	-0.04	-0.02	-0.02	-0.02
hetero, CS+homo	WB	-0.05	-0.05	-0.04	-0.04	-0.03	-0.03	-0.03	-0.02	-0.02

The table reports estimates of the bias in the estimation of α for the OLS estimator, the bias-corrected estimator that corrects bias by plugging in sample analogues of the quantities in Theorem 3.1, and the wild bootstrap respectively. All results are based on 5000 replications and B=399 bootstraps.

Table 3: Coverage rate of 95% CIs for delta - More general cases

		N = 50			N = 100			N = 200		
		T = 50	T = 100	T = 200	T = 50	T = 100	T = 200	T = 50	T = 100	T = 200
DGP 3	OLS	60.3	57.2	46.5	75.1	78.2	75.7	81.4	86.1	87.8
alpha = 1	plug-in	77.3	81.5	84.7	81.3	87.7	89.1	85.2	89.5	91.7
hetero, homo	True factor	91.0	92.3	93.9	90.8	93.4	93.1	90.7	93.2	94.1
	WB	91.4	91.8	92.3	91.9	93.9	93.1	93.4	94.1	94.7
DGP 4	OLS	56.2	52.9	39.8	72.3	75.5	72.8	80.4	85.4	86.4
alpha = 1	plug-in	75.8	80.2	82.6	80.8	86.2	89.4	84.1	88.9	91.5
hetero, hetero	True factor	90.6	92.6	94.2	90.3	93.4	94.4	90.6	92.8	94.1
	WB	91.7	92.3	91.9	92.3	93.5	93.8	92.9	94.1	94.6
DGP 5	OLS	50.5	50.0	39.4	69.6	74.3	71.8	77.8	84.2	86.0
alpha = 1	plug-in	70.4	78.1	81.8	78.4	86.1	88.2	82.5	88.6	91.6
hetero, AR + hetero	True factor	91.0	92.3	93.9	90.8	93.4	93.1	90.7	93.2	94.1
	WB	88.6	91.0	92.0	90.2	93.2	92.7	92.3	93.6	94.3
DGP 6	OLS	70.6	64.8	53.1	81.5	83.2	80.2	87.1	89.1	90.1
alpha = 1	plug-in	80.1	81.2	80.2	84.7	88.8	88.7	88.3	90.8	92.4
hetero, CS + homo	True factor	91.0	92.3	93.9	90.8	93.4	93.1	90.7	93.2	94.1
	WB	83.0	78.9	70.8	88.2	89.0	85.6	92.2	92.4	92.2

Each part of the table reports estimates of the bias in the estimation of α and the associated coverage rate for the OLS estimator, the bias-corrected estimator that corrects bias by plugging in sample analogues of the quantities in Theorem 3.1, and the wild bootstrap respectively. For reference, we also report coverage rate for the OLS estimator that uses the true factors instead of factors estimated by principal components. All results are based on 5000 replications and B=399 bootstraps.