# Bootstrapping factor-augmented regression models 

Ślvia Gonçalves and Benoit Perron*<br>Département de sciences économiques, CIREQ and CIRANO, Université de Montréal

November 22, 2010


#### Abstract

The main contribution of this paper is to propose and theoretically justify bootstrap methods for factor-augmented regressions where some of the regressors are factors estimated from a large panel of data. We consider a residual-based bootstrap method where in a first step bootstrap panel observations are generated by adding the estimated common components to bootstrap residuals. Similarly, in a second step, we generate the bootstrap observations on the dependent variable by adding the estimated regression mean and the bootstrap regression residuals. To mimic the fact that in the original regression model the true factors are latent and need to be estimated, we regress the bootstrap dependent variable on the bootstrap estimated factors (the remaining regressors are kept fixed). This produces a bootstrap OLS estimator whose distribution can be used to compute the quantiles of the distribution of the OLS estimator.

We first provide a set of high level conditions on the bootstrap residuals and on the idiosyncratic errors such that the bootstrap distribution is consistent. We subsequently verify these conditions for a simple wild bootstrap residual-based procedure. Although this method generates bootstrap idiosyncratic errors that are independent (but possibly heteroskedastic) in both dimensions, its validity holds under the general approximate factor model of Bai and Ng (2006) provided $\sqrt{T} / N \rightarrow$ 0 . Our Monte Carlo simulation results confirm the superior finite sample properies of the wild bootstrap over the normal approximation even when there is serial dependence in the idiosyncratic error term.


## 1 Introduction

Factor-augmented regressions where some of the regressors, called factors, are estimated from a large set of data are increasingly popular in empirical work. Inference in these models is complicated by the problem of generated regressors analyzed by Pagan (1984). Recently, Bai and Ng (2006) derive the asymptotic distribution of the OLS estimator in this case under a set of standard regularity conditions. In particular, they show that the asymptotic covariance matrix is unaffected by the estimation of the factors when $\sqrt{T} / N \rightarrow 0$, 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), they report noticeable size distortions in some situations.

[^0]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. Recent empirical applications of the bootstrap in this context include Ludvigson and Ng (2007, 2009a,b) and Gospodinov and Ng (2010), where the bootstrap has been used in the context of predictability tests based on factor-augmented regressions without theoretical justification. Here we establish the first order asymptotic validity of the bootstrap for factor-augmented regression models under a set of regularity conditions similar to those used by Bai and Ng (2006).

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. To mimic the fact that in the original regression model the true factors are latent and need to be estimated, we regress the bootstrap response variable on the estimated bootstrap factors. This produces a bootstrap OLS estimator whose bootstrap distribution can be used to replicate the distribution of the OLS estimator.

A crucial result in proving the first order asymptotic validity of the bootstrap in this context 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. Nevertheless, these are not identified by the bootstrap principal component estimator due to the well-known identification problem of factor models. By relying on results of Bai and Ng (2010) (see also Stock and Watson (2002)), we show that the bootstrap estimated factors identify the estimated factors up to a change of sign. Contrary to the rotation indeterminacy problem that affects the principal component estimator, this sign indetermination 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 $t$-statistics, we should center the bootstrap OLS regression estimates around the sign-adjusted estimated regression coefficients.

We provide a set of high level conditions on the bootstrap residuals and idiosyncratic errors such that the bootstrap principal components have the appropriate rate of convergence. These high level conditions essentially require that the bootstrap idiosyncratic errors be weakly dependent across individuals and over time (so that the bootstrap factor estimation error is appropriately controlled), and that the bootstrap regression scores satisfy a central limit theorem. We show that these high level conditions are satisfied for a residual-based wild bootstrap scheme, where the wild bootstrap is used to generate the bootstrap idiosyncratic error term in the first step, and also in the second step when generating the regression residuals. The two steps are performed independently of each other.

Although the wild bootstrap idiosyncratic errors are independent over time and across individuals
by construction (but are heteroskedastic), the first order asymptotic validity of the wild residualbased bootstrap method holds under the general conditions of Bai and Ng (2006), which allow for weak time series and cross sectional dependence in the idiosyncratic error term. The main reason is that we maintain the condition that $\sqrt{T} / N \rightarrow 0$ so that the asymptotic covariance matrix of the OLS estimator does not depend to first order on the dependence structure of the idiosyncratic errors. The main motivation for using the wild bootstrap in the second step of our residual-based bootstrap is that we follow Bai and $\operatorname{Ng}$ (2006) and assume that the regression errors are a possibly heteroskedastic martingale difference sequence. Under a more general dependence assumption, the wild bootstrap would not be appropriate and we should instead consider a block bootstrap. We do not pursue this possibility here but note that our bootstrap high level conditions would be useful in establishing the validity of the block bootstrap in this context as well.

The rest of the paper is organized as follows. In Section 2, we describe the setup and review the assumptions and the asymptotic theory derived in Bai and Ng (2006). In Section 3, we introduce the residual-based bootstrap method and characterize a set of high level 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 3. Appendix B provides a set of high level conditions on the bootstrap residuals and idiosyncratic error terms such that the bootstrap factor estimation error satisfies the appropriate rates of convergence. These high level conditions can be verified for any residual-based bootstrap method. Appendix C verifies them for the wild bootstrap method we consider in Section 4.

A word on notation. As usual in the bootstrap literature, we use $P^{*}$ to denote the bootstrap probability measure, conditional on a given sample. For any bootstrap statistic $T_{N T}^{*}$, we write $T_{N T}^{*}=o_{P^{*}}(1)$, in probability, or $T_{N T}^{*} \rightarrow P^{P^{*}} 0$, in probability, when for any $\delta>0, P^{*}\left(\left|T_{N T}^{*}\right|>\delta\right)=$ $o_{P}(1)$. We write $T_{N T}^{*}=O_{P^{*}}(1)$, in probability, when for all $\delta>0$ there exists $M_{\delta}<\infty$ such that $\lim _{N, T \rightarrow \infty} P\left[P^{*}\left(\left|T_{N T}^{*}\right|>M_{\delta}\right)>\delta\right]=0$. Finally, we write $T_{N T}^{*} \rightarrow d^{d^{*}} D$, in probability, if conditional on a sample with probability that converges to one, $T_{N T}^{*}$ weakly converges to the distribution $D$ under $P^{*}$, i.e. $E^{*}\left(f\left(T_{N T}^{*}\right)\right) \rightarrow^{P} E(f(D))$ for all bounded and uniformly continuous functions $f$.

## 2 Setup, assumptions and estimation

### 2.1 Setup

We consider the following regression model

$$
\begin{equation*}
y_{t+h}=\alpha^{\prime} F_{t}+\beta^{\prime} W_{t}+\varepsilon_{t+h}, t=1, \ldots, T-h, \tag{1}
\end{equation*}
$$

where $h \geq 0$. This model is known in the literature as a factor-augmented regression model due to the presence of $F_{t}$, which is a vector containing $r$ latent common factors. The observable regressors are
contained in $W_{t}$. The unobserved regressors $F_{t}$ are the common factors in the following panel factor model,

$$
\begin{equation*}
X_{i t}=\lambda_{i}^{\prime} F_{t}+e_{i t}, \quad i=1, \ldots, N, t=1, \ldots, T \tag{2}
\end{equation*}
$$

where the $r \times 1$ vector $\lambda_{i}$ contains the factor loadings and $e_{i t}$ is an idiosyncratic error term. None of $\left\{\lambda_{i}\right\},\left\{F_{t}\right\}$ and $\left\{e_{i t}\right\}$ are observed, but they can be estimated from the observed panel data set $\left\{X_{i t}: i=1, \ldots, N, t=1, \ldots, T\right\}$. In particular, we can obtain estimated factors $\tilde{F}_{t}$ and run a standard regression of $y_{t+h}$ on $\tilde{F}_{t}$ and $W_{t}$ to estimate $\alpha$ and $\beta$. We will review in more detail below the estimation methods proposed in this context.

The factor-augmented regression model described in (11) and (2) has recently attracted a lot of attention in econometrics. One of the attractive features of this model is the fact that a large amount of information (the $X_{i t}$ 's) can be summarized in a few estimated factors (the $\tilde{F}_{t}$ ). By including the estimated factors as regressors in an otherwise standard regression model, we can effectively take into account a large number of predictors without running into the curse of dimensionality that would arise if we tried to include them all in the first place. One of the first papers to discuss this model in the forecasting context was Stock and Watson (2002), who studied the consistency of feasible forecasts of $y_{T+h}$ based on the estimated conditional mean of (1) when $F_{t}$ is replaced with an estimate $\tilde{F}_{t}$. Bai and Ng (2006) derived an asymptotic distribution theory for the estimated regression parameters of (11) as well as for the feasible forecasts based on these parameters and the estimated factors. Recent empirical applications include Ludvigson and $\operatorname{Ng}$ (2007, 2009a,b) and Gospodinov and Ng (2010). Ludvigson and Ng (2007) 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 $\mathrm{Ng}(2009 \mathrm{a}, \mathrm{b})$ consider this approach in the context of predictive regressions of bond excess returns. Gospodinov and Ng (2010) study predictive regressions for inflation using principal components from a panel of commodity convenience yields, while Eichengreen, Mody, Nedeljkovic, and Sarno (2009) use common factors extracted from credit default swap (CDS) spreads during the recent financial crisis to look at spillovers across banks.

### 2.2 Assumptions

We follow Bai and $\operatorname{Ng}$ (2006) (see also Bai and $\operatorname{Ng}$ (2002) and Bai (2003)) and rely on the following set of assumptions. Throughout, $M$ is a generic finite constant.

Assumption $\mathbf{A}$ (common factors) $E\left\|F_{t}\right\|^{4} \leq M$ and $\frac{1}{T} \sum_{t=1}^{T} F_{t} F_{t}^{\prime} \rightarrow^{P} \Sigma_{F}>0$, where $\Sigma_{F}$ is a non-random $r \times r$ matrix.

Assumption B (heterogeneous factor loadings) The loadings $\lambda_{i}$ are either deterministic such that $\left\|\lambda_{i}\right\| \leq M$, or stochastic such that $E\left\|\lambda_{i}\right\|^{4} \leq M$. In either case, $\Lambda^{\prime} \Lambda / N \rightarrow^{P} \Sigma_{\Lambda}>0$, where $\Sigma_{\Lambda}$ is a non-random matrix.

## Assumption C (time and cross-section weak dependence and heteroskedasticity)

1. $E\left(e_{i t}\right)=0, E\left|e_{i t}\right|^{8} \leq M$.
2. $E\left(e_{i t} e_{j s}\right)=\sigma_{i j, t s},\left|\sigma_{i j, t s}\right| \leq \bar{\sigma}_{i j}$ for all $(t, s)$ and $\left|\sigma_{i j, t s}\right| \leq \tau_{t s}$ for all $(i, j)$ such that

$$
\frac{1}{N} \sum_{i, j=1}^{N} \bar{\sigma}_{i j} \leq M, \quad \frac{1}{T} \sum_{t, s=1}^{T} \tau_{t s} \leq M, \text { and } \frac{1}{N T} \sum_{t, s, i, j}\left|\sigma_{i j, t s}\right| \leq M
$$

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

Assumption $\mathbf{D}\left\{\lambda_{i}\right\},\left\{F_{t}\right\}$, and $\left\{e_{i t}\right\}$ are three mutually independent groups. Dependence within each group is allowed.
Assumption E For each $t, E\left\|\frac{1}{\sqrt{T N}} \sum_{s=1}^{T} \sum_{i=1}^{N} F_{s}\left(e_{i t} e_{i s}-E\left(e_{i t} e_{i s}\right)\right)\right\|^{2} \leq M$.
Assumption A imposes the assumption that factors are non-degenerate. Assumption 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 (2009b) 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 C imposes weak cross-sectional and serial dependence conditions in the idiosyncratic error term $e_{i t}$. In particular, we allow for the possibility that $e_{i t}$ 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 section units are weakly correlated. Assumption C allows for heteroskedasticity in both dimensions and requires the idiosyncratic error term to have finite eighth moments.

According to Assumption D, the factor loadings, the factors and the idiosyncratic error terms are three mutually independent groups of random variables. This assumption is standard in classic factor analysis. Bai (2003) relaxes the independence assumption between factors and idiosyncratic errors by allowing for weak dependence between the two sets. For simplicity, we will maintain the independence assumption throughout. When factors and idiosyncratic errors are independent, Assumption E is implied by the following more primitive condition

$$
\frac{1}{N T^{2}} \sum_{t, s, l, u=1}^{T} \sum_{i, j=1}^{N}\left|\operatorname{Cov}\left(e_{i t} e_{i s}, e_{j l} e_{j u}\right)\right| \leq M,
$$

which holds under suitable weak dependence conditions in $\left\{e_{i t}\right\}$ across $i$ and over $t$. Bai (2009) relies on this condition (part 1 of his Assumption C.4) to establish the asymptotic properties of the interactive
effects estimator.
Our next assumption is a high level assumption on the regressors $z_{t}=\left(F_{t}^{\prime}, W_{t}^{\prime}\right)^{\prime}$, the error term $\varepsilon_{t+h}$, and the scores $z_{t} \varepsilon_{t+h}$ of the regression model (11). Bai and $\operatorname{Ng}$ (2006) rely on this assumption to derive the asymptotic distribution of the ordinary least squares estimator of the parameters in (1). See also Stock and Watson (2002) for a similar set of assumptions.

Assumption $\mathbf{F}$ Let $z_{t}=\left(F_{t}^{\prime}, W_{t}^{\prime}\right)^{\prime}$.

1. $E\left\|z_{t}\right\|^{4} \leq M$.
2. $E\left(\varepsilon_{t+h} \mid y_{t}, z_{t}, y_{t-1}, z_{t-1}, \ldots\right)=0$, for any $h \geq 0$, and $z_{t}$ and $\varepsilon_{t}$ are independent of $e_{i s}$ for all $(i, s, t)$.
3. $\frac{1}{T} \sum_{t=1}^{T} z_{t} z_{t}^{\prime} \rightarrow^{P} \Sigma_{z z}>0$.
4. $\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} z_{t} \varepsilon_{t+h} \rightarrow^{d} N\left(0, \Sigma_{z z . \varepsilon}\right)$, where $\Sigma_{z z . \varepsilon} \equiv E\left(z_{t} z_{t}^{\prime} \varepsilon_{t+h}^{2}\right)=p \lim \frac{1}{T} \sum_{t=1}^{T-h} z_{t} z_{t}^{\prime} \varepsilon_{t+h}^{2}>0$.

Assumption F. 1 requires the existence of finite fourth moments for the elements of $z_{t}$. Assumption F. 2 imposes a martingale difference condition on the regression errors $\varepsilon_{t+h}$. Under this condition, $\varepsilon_{t+h}$ is serially uncorrelated but possibly heteroskedastic. Assumption F. 2 also imposes an independence condition between $\left(z_{t}, \varepsilon_{t}\right)$ and the idiosyncratic errors $e_{i s}$. In particular, $\varepsilon_{t}$ is independent of $e_{i s}$ for all $(i, s, t)$. Assumption F. 3 is the standard absence of multicollinearity condition and ensures that the regression coefficients are well identified. Assumption F. 4 is a high level central limit theorem condition on the scores of the regression model. Under the martingale difference condition F.2, the scores are serially uncorrelated but possibly heteroskedastic. The form of the asymptotic covariance matrix $\Sigma_{z z . \varepsilon} \equiv E\left(z_{t} z_{t}^{\prime} \varepsilon_{t+h}^{2}\right)$ reflects this assumption.

### 2.3 Estimation

The factor-augmented regression model (11) cannot be directly estimated because the common factors $F_{t}$ are not observed. Given the factor panel model (2), the idea is to first use the panel $\left\{X_{i t}\right\}$ to estimate $F_{t}$, and then run a regression of $y_{t+h}$ on the estimated factors $\tilde{F}_{t}$ and $W_{t}$.

In matrix form, we can write (21) as

$$
X=F \Lambda^{\prime}+e
$$

where $X$ is a $T \times N$ matrix of stationary data, $F=\left(F_{1}, \ldots, F_{T}\right)^{\prime}$ is $T \times r, r$ is the number of common factors, $\Lambda=\left(\lambda_{1}, \ldots, \lambda_{N}\right)^{\prime}$ is $N \times r$, and $e$ is $T \times N$. We assume throughout that the number of factors, $r$, is known. One could proceed by estimating the number of factors using the information criteria of Bai and Ng (2002), Groen and Kapetanios (2009), or the test of Onatski (2009a).

Given $X$, we estimate $F$ and $\Lambda$ with the method of principal components. The objective function that we wish to minimize is given by

$$
V(F, \Lambda)=\frac{1}{T N} \sum_{i=1}^{N} \sum_{t=1}^{T}\left(X_{i t}-\lambda_{i}^{\prime} F_{t}\right)^{2}=\frac{1}{T N} \sum_{i=1}^{N}\left(X_{i}-F \lambda_{i}\right)^{\prime}\left(X_{i}-F \lambda_{i}\right),
$$

where $X_{i}=\left(X_{i 1}, \ldots, X_{i T}\right)^{\prime}$ is $T \times 1$. Some normalization has to be imposed. The two most common is that either $\frac{F^{\prime} F}{T}=I_{r}$ or $\frac{\Lambda^{\prime} \Lambda}{N}=I_{r}$. In the case where the normalization $\frac{F^{\prime} F}{T}=I_{r}$ is used, the solution is the $T \times r$ matrix $\tilde{F}=\left(\begin{array}{lll}\tilde{F}_{1} & \ldots & \tilde{F}_{T}\end{array}\right)^{\prime}$ composed of $\sqrt{T}$ times the eigenvectors corresponding to the $r$ largest eigenvalues of of $X X^{\prime} / T N$ (arranged in decreasing order). The matrix containing the estimated loadings is then $\tilde{\Lambda}=\left(\tilde{\lambda}_{1}, \ldots, \tilde{\lambda}_{N}\right)^{\prime}=X^{\prime} \tilde{F}\left(\tilde{F}^{\prime} \tilde{F}\right)^{-1}=X^{\prime} \tilde{F} / T$.

A distinctive feature of factor models is that $F$ and $\Lambda$ are not separately identified. For any invertible matrix $A$, we can write $F \Lambda^{\prime}=F A A^{-1} \Lambda^{\prime} \equiv G \Gamma^{\prime}$, implying that $X=F \Lambda^{\prime}+e$ is observationally equivalent to $X=G \Gamma^{\prime}+e$. For this reason, $\tilde{F}$ can only consistently estimate the space spanned by F. In particular, under Assumptions A-D, Bai and Ng (2002) show that

$$
\begin{equation*}
\frac{1}{T} \sum_{t=1}^{T}\left\|\tilde{F}_{t}-H \tilde{F}_{t}\right\|^{2}=O_{P}\left(\delta_{N T}^{-2}\right) \tag{3}
\end{equation*}
$$

where $\delta_{N T}=\min (\sqrt{N}, \sqrt{T})$. The matrix $H$ is defined as

$$
\begin{equation*}
H=\tilde{V}^{-1} \frac{\tilde{F}^{\prime} F}{T} \frac{\Lambda^{\prime} \Lambda}{N}, \tag{4}
\end{equation*}
$$

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

Consider now the factor-augmented regression of $y_{t+h}$ on $\hat{z}_{t}=\left(\tilde{F}_{t}^{\prime}, W_{t}^{\prime}\right)^{\prime}$. Given (11), adding and subtracting appropriately yields

$$
y_{t+h}=\underbrace{\left(\begin{array}{ll}
\alpha^{\prime} H^{-1} & \beta^{\prime}
\end{array}\right)}_{=\delta^{\prime}} \underbrace{\binom{\tilde{F}_{t}}{W_{t}}}_{=\hat{z}_{t}}+\alpha^{\prime} H^{-1}\left(H F_{t}-\tilde{F}_{t}\right)+\varepsilon_{t+h},
$$

or, equivalently,

$$
\begin{equation*}
y_{t+h}=\hat{z}_{t}^{\prime} \delta+\alpha^{\prime} H^{-1}\left(H F_{t}-\tilde{F}_{t}\right)+\varepsilon_{t+h} . \tag{5}
\end{equation*}
$$

The least squares estimator of $\delta$ is

$$
\hat{\delta} \equiv\binom{\hat{\alpha}}{\hat{\beta}}=\left(\sum_{t=1}^{T-h} \hat{z}_{t} \hat{z}_{t}^{\prime}\right)^{-1} \sum_{t=1}^{T-h} \hat{z}_{t} y_{t+h} .
$$

Bai and Ng (2006) derived the asymptotic distribution of $\hat{\delta}$ under Assumptions A-F and the condition that $\sqrt{T} / N \rightarrow 0$. Their Theorem 1 shows that

$$
\sqrt{T}(\hat{\delta}-\delta) \rightarrow^{d} N\left(0, \Sigma_{\delta}\right)
$$

where

$$
\Sigma_{\delta}=\Phi_{0}^{\prime-1} \Sigma_{z z}^{-1} \Sigma_{z z, \varepsilon} \Sigma_{z z}^{-1} \Phi_{0}^{-1}
$$

where $\Sigma_{z z . \varepsilon}$ is defined in Assumption F. 4 and $\Phi_{0}=p \lim \left[\begin{array}{cc}H & 0 \\ 0 & I\end{array}\right]$.
A consistent estimator of $\Sigma_{\delta}$ is

$$
\begin{equation*}
\hat{\Sigma}_{\delta}=\left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{z}_{t} \hat{z}_{t}^{\prime}\right)^{-1}\left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{z}_{t} \hat{z}_{t}^{\prime} \hat{\varepsilon}_{t+h}^{2}\right)\left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{z}_{t} \hat{z}_{t}^{\prime}\right)^{-1}, \tag{6}
\end{equation*}
$$

where $\hat{\varepsilon}_{t+h}=y_{t+h}-\hat{z}_{t}^{\prime} \hat{\delta}$ are the regression residuals in (5). This is equation (3) in Bai and $\operatorname{Ng}$ (2006).
Some comments are in order. First, because the factor model is unidentified, the least squares estimator $\hat{\delta}$ is consistent for $\delta=\left(\alpha^{\prime} H^{-1}, \beta^{\prime}\right)^{\prime}$, and not for $\left(\alpha^{\prime}, \beta^{\prime}\right)^{\prime}$. In particular, the OLS regression coefficients associated with $\tilde{F}_{t}$ are consistent only for a rotation of the true parameters $\alpha$, where the rotation is determined by the matrix $H$ defined above. Thus, the estimated parameters do not necessarily have a structural interpretation. An exception is when $\alpha=0$, in which case $\delta=\left(0, \beta^{\prime}\right)^{\prime}=$ $\left(\alpha^{\prime}, \beta^{\prime}\right)^{\prime}$. Another exception is when the data generating process (DGP) is such that we can identify $H$. This has been recently studied by Bai and Ng (2010), who provide sufficient conditions for $H$ to be asymptotically equal to a diagonal matrix with $\pm 1$ in the main diagonal. For instance, this will be true if the DGP happens to satisfy the normalization conditions $F^{\prime} F / T=I_{r}$ and $\Lambda^{\prime} \Lambda$ happens to be equal to a diagonal matrix. See Bai and Ng (2010) for two other identification schemes.

The second comment is that if $\sqrt{T} / N \rightarrow 0$, estimation of the true factors does not impact the estimation of the asymptotic covariance matrix of $\hat{\delta}$. In particular, the covariance matrix estimator $\hat{\Sigma}_{\delta}$ given in (6) is exactly the same we would compute if the true factors were observed. There is no need to adjust the standard errors to take into account the presence of generated regressors. This is in contrast to the usual results (see Pagan (1984)) that show that in a standard regression model the asymptotic variance of the OLS estimator changes when we replace unobserved regressors by a first-step estimated version of these. In the factor-augmented regression model, Bai and Ng (2006) show that there is no such effect as long as $\sqrt{T} / N \rightarrow 0$.

## 3 A general residual-based bootstrap

In this section we consider a general residual-based bootstrap for factor-augmented regressions and provide high level conditions under which this method is first order asymptotically valid.

Suppose the bootstrap DGP generates $\left(y_{t+h}^{*}, X_{t}^{*}\right)$ using an estimated version of the factor model representation (11) and (2) given by

$$
\begin{align*}
y_{t+h}^{*} & =\hat{\alpha}^{\prime} \tilde{F}_{t}+\hat{\beta}^{\prime} W_{t}+\varepsilon_{t+h}^{*}, \quad t=1, \ldots, T-h ;  \tag{7}\\
X_{t}^{*} & =\tilde{\Lambda} \tilde{F}_{t}+e_{t}^{*}, \quad t=1, \ldots, T \tag{8}
\end{align*}
$$

where $\left\{\varepsilon_{t+h}^{*}\right\}$ is a bootstrap sample generated from the regression residuals $\left\{\hat{\varepsilon}_{t+h}=y_{t+h}-\hat{\alpha}^{\prime} \tilde{F}_{t}-\hat{\beta}^{\prime} W_{t}\right\}$
and $\left\{e_{t}^{*}=\left(e_{1 t}^{*}, \ldots, e_{N t}^{*}\right)^{\prime}\right\}$ is a bootstrap sample from $\left\{\hat{e}_{t}=X_{t}-\tilde{\Lambda} \tilde{F}_{t}\right\}$. Here, $X_{t}=\left(X_{1 t}, \ldots, X_{N t}\right)^{\prime}$ and $X_{t}^{*}=\left(X_{1 t}^{*}, \ldots, X_{N t}^{*}\right)^{\prime}$. At this point, we are not specific about which bootstrap methods we use to generate these residuals and idiosyncratic error terms. Instead, our goal is to provide a set of high level conditions under which any bootstrap scheme generated according to (77) and (8) will be asymptotically valid.

Estimation proceeds in two stages. First, we estimate the factors by the method of principal components using the bootstrap panel data set $\left\{X_{t}^{*}\right\}$. Second, we run a regression of $y_{t+h}^{*}$ on the bootstrap estimated factors and the fixed observed regressors $W_{t}$. Under Bai and Ng's (2006) assumptions, a simple residual-based bootstrap method that does not take into account the factor estimation uncertainty in the bootstrap samples (i.e. a bootstrap method based only on the second step of our proposed method) is asymptotically valid. We do not consider this possibility here because factor estimation uncertainty has an impact in finite samples and therefore estimating the factors in the bootstrap world is important to improve finite sample accuracy. Yamamoto (2009) compares bootstrap methods with and without factor estimation for the factor-augmented vector autoregression (FAVAR) model. He concludes that the latter is worse than the first in terms of finite sample accuracy.

Given $\left\{X_{t}^{*}\right\}$, we estimate the bootstrap factor loadings and the bootstrap factors by minimizing the bootstrap objective function

$$
V^{*}(F, \Lambda)=\frac{1}{T N} \sum_{t=1}^{T} \sum_{i=1}^{N}\left(X_{i t}^{*}-\lambda_{i}^{\prime} F_{t}\right)^{2}
$$

subject to the normalization constraint that $F^{\prime} F / T=I_{r}$. The $T \times r$ matrix containing the estimated bootstrap factors is denoted by $\tilde{F}^{*}=\left(\tilde{F}_{1}^{*}, \ldots, \tilde{F}_{T}^{*}\right)^{\prime}$ and it is equal to the $r$ eigenvectors of $X^{*} X^{* \prime} / N T$ (multiplied by $\sqrt{T}$ ) corresponding to the $r$ largest eigenvalues. The $N \times r$ matrix of estimated bootstrap loadings is given by $\tilde{\Lambda}^{*}=\left(\tilde{\lambda}_{1}^{*}, \ldots, \tilde{\lambda}_{N}^{*}\right)^{\prime}=X^{* \prime} \tilde{F}^{*} / T$.

According to (8), the common factors underlying the bootstrap panel data $\left\{X_{t}^{*}\right\}$ are given by $\tilde{F}_{t}$ (with $\tilde{\Lambda}$ as factor loadings). Nevertheless, and as noted in the previous section, the factor model is unidentified and therefore the estimated bootstrap factors $\tilde{F}_{t}^{*}$ do not identify $\tilde{F}_{t}$. Instead, $\tilde{F}_{t}^{*}$ estimates $H^{*} \tilde{F}_{t}$, where $H^{*}$ is the bootstrap analogue of the rotation matrix $H$ defined in (44), i.e.

$$
\begin{equation*}
H^{*}=\tilde{V}^{*-1} \frac{\tilde{F}^{* \prime} \tilde{F}}{T} \frac{\tilde{\Lambda}^{\prime} \tilde{\Lambda}}{N} \tag{9}
\end{equation*}
$$

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

Contrary to $H, H^{*}$ does not depend on population values and can be computed. Hence, the rotation indeterminacy problem is not as acute in the bootstrap world. In particular, because the bootstrap factor model (8) satisfies the constraints that $\tilde{F}^{\prime} \tilde{F} / T=I_{r}$ and $\tilde{\Lambda}^{\prime} \tilde{\Lambda}$ is a diagonal matrix, it turns out that $H^{*}$ is asymptotically (as $N, T \rightarrow \infty$ ) a diagonal matrix with diagonal elements equal
to $\pm 1$. Specifically, we can show that

$$
H^{*}=H_{0}^{*}+O_{P^{*}}\left(\delta_{N T}^{-2}\right),
$$

in probability, where $H_{0}^{*}=\operatorname{diag}( \pm 1)$, by relying on the arguments of Bai and Ng (2010) (see also Stock and Watson (2002)). Therefore, the bootstrap factors are identified up to a change of sign.

Given the estimated bootstrap factors $\tilde{F}_{t}^{*}$, the second estimation stage is to regress $y_{t+h}^{*}$ on $\hat{z}_{t}^{*} \equiv$ $\left(\tilde{F}_{t}^{* \prime}, W_{t}^{\prime}\right)^{\prime}$. 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 (11) the factors $F_{t}$ are latent and need to be estimated with $\tilde{F}_{t}$. This yields the bootstrap OLS estimator

$$
\begin{equation*}
\hat{\delta}^{*} \equiv\binom{\hat{\alpha}^{*}}{\hat{\beta}^{*}}=\left(\sum_{t=1}^{T-h} \hat{z}_{t}^{*} \hat{z}_{t}^{* \prime}\right)^{-1} \sum_{t=1}^{T-h} \hat{z}_{t}^{*} y_{t+h}^{*} . \tag{10}
\end{equation*}
$$

From (7), by adding and subtracting appropriately, we have that

$$
y_{t+h}^{*}=\underbrace{\left(\begin{array}{ll}
\hat{\alpha}^{\prime} H^{*-1} & \hat{\beta}^{\prime} \tag{11}
\end{array}\right)}_{=\delta^{* \prime}} \underbrace{\binom{\tilde{F}_{t}^{*}}{W_{t}}}_{=\hat{z}_{t}^{*}}+\hat{\alpha}^{\prime} H^{*-1}\left(H^{*} \tilde{F}_{t}-\tilde{F}_{t}^{*}\right)+\varepsilon_{t+h}^{*} .
$$

Here, $\delta^{*}$ is the set of "parameters" that the bootstrap OLS estimator $\hat{\delta}^{*}$ identifies provided we appropriately control the bootstrap factor estimation uncertainty (as captured by the second term). The following condition provides a set of high level conditions under which this is the case.

Condition $\mathbf{A}^{*}$ Let $\delta_{N T}=\min (\sqrt{N}, \sqrt{T})$ and suppose the following conditions hold in probability, as $N, T \rightarrow \infty$,

1. $\frac{1}{T} \sum_{t=1}^{T}\left\|\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}\right\|^{2}=O_{P^{*}}\left(\delta_{N T}^{-2}\right)$;
2. $\frac{1}{T} \sum_{t=1}^{T}\left(\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}\right) \hat{z}_{t}^{\prime}=O_{P^{*}}\left(\delta_{N T}^{-2}\right)$, where $\hat{z}_{t}=\left(\tilde{F}_{t}^{\prime}, W_{t}^{\prime}\right)^{\prime}$;
3. $\frac{1}{T} \sum_{t=1}^{T}\left(\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}\right) \hat{z}_{t}^{* \prime}=O_{P^{*}}\left(\delta_{N T}^{-2}\right)$, where $\hat{z}_{t}^{*}=\left(\tilde{F}_{t}^{* \prime}, W_{t}^{\prime}\right)^{\prime}$;
4. $\frac{1}{T} \sum_{t=1}^{T}\left(\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}\right) \varepsilon_{t+h}^{*}=O_{P^{*}}\left(\delta_{N T}^{-2}\right)$; and
5. $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{z}_{t} \varepsilon_{t+h}^{*} \rightarrow d^{d^{*}} N\left(0, \Phi_{0} \Sigma_{z z, \varepsilon} \Phi_{0}^{\prime}\right)$, where $\Phi_{0}=\operatorname{diag}(p \lim H, I)$ and $\Sigma_{z z, \varepsilon}=E\left(z_{t} z_{t}^{\prime} \varepsilon_{t+h}^{2}\right)$.

With Condition A*, we can show the validity of the residual-based bootstrap. Appendix B contains primitive conditions on the bootstrap residuals $\left\{\varepsilon_{t+h}^{*}\right\}$ and on the bootstrap idiosyncratic error terms $\left\{e_{i t}^{*}\right\}$ such that (the first four parts of) Condition $\mathrm{A}^{*}$ hold. In the next section we will propose a wild bootstrap method for generating $\left\{\varepsilon_{t+h}^{*}\right\}$ and $\left\{e_{i t}^{*}\right\}$ such that Condition A* is verified.

Theorem 3.1 Assume Assumptions $A-F$ 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 Condition $A^{*}$ is satisfied. Then, as $N, T \rightarrow \infty$ with $\sqrt{T} / N \rightarrow 0$,
a) With probability approaching one,

$$
\sqrt{T}\left(\hat{\delta}^{*}-\delta^{*}\right) \rightarrow^{d^{*}} N\left(0, \Sigma_{\delta}^{*}\right)
$$

where $\delta^{*}=\Phi^{*-1} \hat{\delta}$, with $\Phi^{*}=\operatorname{diag}\left(H^{*}, I\right)$, and $\Sigma_{\delta}^{*}=\left(\Phi_{0}^{* \prime}\right)^{-1} \Sigma_{\delta}\left(\Phi_{0}^{*}\right)^{-1}$, where $\Phi_{0}^{*}$ is a diagonal matrix with $\pm 1$ as main diagonal elements.
b) For each individual element $j$,

$$
\sup _{x \in \mathbb{R}}\left|P^{*}\left(\sqrt{T}\left(\hat{\delta}_{j}^{*}-\delta_{j}^{*}\right) \leq x\right)-P\left(\sqrt{T}\left(\hat{\delta}_{j}-\delta_{j}\right) \leq x\right)\right| \xrightarrow{P} 0 .
$$

Part a) of Theorem 3.1 shows that the bootstrap asymptotic distribution of $\sqrt{T}\left(\hat{\delta}^{*}-\delta^{*}\right)$ is $N\left(0, \Sigma_{\delta}^{*}\right)$ provided the bootstrap samples $\left\{\varepsilon_{t+h}^{*}\right\}$ and $\left\{e_{i t}^{*}\right\}$ are such that Condition A* is satisfied and Bai and Ng's (2006) conditions hold.

The first feature to notice is that the bootstrap estimated coefficients $\hat{\delta}^{*}$ are centered around

$$
\delta^{*}=\Phi^{* \prime-1} \hat{\delta}
$$

where $\Phi^{*}=\operatorname{diag}\left(H^{*}, I\right)$, and not around $\hat{\delta}$. Because $H^{*}$ is asymptotically equal to a diagonal matrix with $\pm 1$ on the main diagonal, $\delta^{*}$ is asymptotically equal to $\hat{\delta}$, up to a sign change, i.e.

$$
\delta^{*}=\delta_{0}^{*}+O_{P *}\left(\delta_{N T}^{-2}\right)
$$

where $\delta_{0}^{*} \equiv \Phi_{0}^{* /-1} \hat{\delta}$ and $\Phi_{0}^{*}=\operatorname{diag}(\operatorname{diag}( \pm 1), I)$ is a diagonal matrix where the first $r$ main diagonal elements are $\pm 1$. Thus, $\hat{\delta}^{*}$ only identifies the sign-adjusted coefficients associated with the estimated factors. Stock and Watson (2002) and Bai and Ng (2010) also discuss the sign indeterminacy.

The second feature to notice is that the asymptotic bootstrap covariance matrix of $\hat{\delta}^{*}$ is $\Sigma_{\delta}^{*}=$ $\left(\Phi_{0}^{* \prime}\right)^{-1} \Sigma_{\delta}\left(\Phi_{0}^{*}\right)^{-1}$. For first order asymptotic bootstrap validity, we need $\Sigma_{\delta}^{*}=\Sigma_{\delta}$, where $\Sigma_{\delta}$ is the asymptotic covariance matrix of $\hat{\delta}$ (as shown by Bai and $\operatorname{Ng}(2006)$ ). Since $\Phi_{0}^{*}=\operatorname{diag}(\operatorname{diag}( \pm 1), I)$, this equality does not hold and therefore the residual-based bootstrap does not consistently estimate the distribution of the vector $\sqrt{T}(\hat{\delta}-\delta)$. Nevertheless, because $\Phi_{0}^{*}$ is a diagonal matrix with $\pm 1$ as the first $r$ main diagonal elements (and +1 for the remaining ones), the main diagonal elements of $\Sigma_{\delta}^{*}$ are the same as the main diagonal elements of $\Sigma_{\delta}$. This implies that the bootstrap variances of each individual element of $\hat{\delta}_{j}^{*}$ coincide asymptotically with the variances of $\hat{\delta}_{j}$. Consequently, under the conditions of Theorem 3.1 the bootstrap distribution of $\sqrt{T}\left(\hat{\delta}_{j}^{*}-\delta_{j}^{*}\right)$ is consistent for the distribution of $\sqrt{T}\left(\hat{\delta}_{j}-\delta_{j}\right)$. This is the content of part b), which justifies using the residual-based bootstrap for constructing bootstrap percentile-type confidence intervals for the individual elements of $\delta$.

When interest focuses on inference involving the entire vector $\delta$ (or a subvector containing more
than one of the coefficients associated with the estimated factors), we need to modify the bootstrap procedure described above to ensure that the asymptotic bootstrap covariance matrix is equal to $\Sigma_{\delta}$. One easy modification is to consider the bootstrap distribution of a sign-adjusted vector of bootstrap estimates given by

$$
\tilde{\delta}^{*}=\Phi_{0}^{* /} \hat{\delta}^{*}
$$

The asymptotic bootstrap covariance matrix of $\tilde{\delta}^{*}$ is equal to

$$
\operatorname{Var}^{*}\left(\sqrt{T} \tilde{\delta}^{*}\right)=\Phi_{0}^{* \prime} \operatorname{Var}^{*}\left(\sqrt{T} \hat{\delta}^{*}\right) \Phi_{0}^{*}=\Phi_{0}^{* \prime}\left(\Phi_{0}^{* \prime}\right)^{-1} \Sigma_{\delta}\left(\Phi_{0}^{*}\right)^{-1} \Phi_{0}^{*}=\Sigma_{\delta}
$$

Sign-adjusting the OLS bootstrap estimates $\hat{\delta}^{*}$ not only delivers the correct asymptotic covariance matrix for $\tilde{\delta}^{*}$ but also implies that we can center $\tilde{\delta}^{*}$ around $\hat{\delta}$ (instead of $\delta^{*}$ ). Notice that sign-adjusting $\hat{\delta}^{*}$ is exactly equivalent to sign-adjusting the factors $\tilde{F}^{*}$, i.e. $\tilde{\delta}^{*}$ is equal to the OLS estimator from the regression of $y_{t+h}^{*}$ on $H_{0}^{*} \tilde{F}_{t}^{*}$ and $W_{t}$.

The following result provides the consistency of the bootstrap distribution of $\sqrt{T}\left(\tilde{\delta}^{*}-\hat{\delta}\right)$ as an estimator of the distribution of $\sqrt{T}(\hat{\delta}-\delta)$.

Corollary 3.1 Under the assumptions of Theorem 3.1, as $N, T \rightarrow \infty$,
a) With probability approaching one,

$$
\sqrt{T}\left(\tilde{\delta}^{*}-\hat{\delta}\right) \rightarrow^{d^{*}} N\left(0, \Sigma_{\delta}\right)
$$

b)

$$
\sup _{x \in \mathbb{R}^{\mathrm{dim}(\delta)}}\left|P^{*}\left(\sqrt{T}\left(\tilde{\delta}^{*}-\hat{\delta}\right) \leq x\right)-P(\sqrt{T}(\hat{\delta}-\delta) \leq x)\right| \xrightarrow{P} 0
$$

## 4 A residual-based wild bootstrap

Both Theorem 3.1 and Corollary 3.1 justify the use of a residual-based bootstrap method for constructing bootstrap percentile confidence intervals for the elements of $\delta$ provided we choose the bootstrap innovations $\left\{\varepsilon_{t+h}^{*}\right\}$ and $\left\{e_{i t}^{*}\right\}$ such that Condition $A^{*}$ holds.

In this section we propose a particular bootstrap method for generating $\left\{\varepsilon_{t+h}^{*}\right\}$ and $\left\{e_{i t}^{*}\right\}$ and show its first-order asymptotic validity under a set of primitive conditions.

## Bootstrap algorithm

1. For $t=1, \ldots, T$, let

$$
X_{t}^{*}=\tilde{\Lambda} \tilde{F}_{t}+e_{t}^{*}
$$

where $\left\{e_{t}^{*}=\left(e_{1 t}^{*}, \ldots, e_{N t}^{*}\right)\right\}$ is such that

$$
e_{i t}^{*}=\tilde{e}_{i t} \eta_{i t}
$$

is a resampled version of $\left\{\tilde{e}_{i t}=X_{i t}-\tilde{\lambda}_{i}^{\prime} \tilde{F}_{t}\right\}$ obtained with the wild bootstrap. The external random variables $\eta_{i t}$ 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 $\tilde{\Lambda}^{*} u \operatorname{sing} X^{*}$.
3. For $t=1, \ldots, T-h$, let

$$
y_{t+h}^{*}=\hat{\alpha}^{\prime} \tilde{F}_{t}+\hat{\beta}^{\prime} W_{t}+\varepsilon_{t+h}^{*},
$$

where the regressors are kept fixed and the error term $\varepsilon_{t+h}^{*}$ is a wild bootstrap resampled version of $\hat{\varepsilon}_{t+h}$, i.e.

$$
\varepsilon_{t+h}^{*}=\hat{\varepsilon}_{t+h} v_{t+h}
$$

where the external random variable $v_{t+h}$ is i.i.d. $(0,1)$ and is independent of $\eta_{i t}$.
4. Regress $y_{t+h}^{*}$ generated in 3. on the fixed regressors $W_{t}$ and on the estimated bootstrap factors $\tilde{F}_{t}^{*}$. This yields the bootstrap OLS estimators

$$
\hat{\delta}^{*}=\left(\sum_{t=1}^{T-h} \hat{z}_{t}^{*} \hat{z}_{t}^{* \prime}\right)^{-1} \sum_{t=1}^{T-h} \hat{z}_{t}^{*} y_{t+h}^{*}
$$

where $\hat{z}_{t}^{*}=\left(\tilde{F}_{t}^{* \prime}, W_{t}^{\prime}\right)^{\prime}$.
Given Theorem 3.1 and Corollary 3.1, the wild residual-based bootstrap method is first order asymptotically valid provided it satisfies Condition $A^{*}$. The following assumption (together with Assumptions A-F) ensures that this is the case.

Assumption WB For some $q>1$,

1. $\lambda_{i}$ are either deterministic such that $\left\|\lambda_{i}\right\| \leq M<\infty$, or stochastic such that $E\left\|\lambda_{i}\right\|^{4 q} \leq M<\infty$ for all $i$.
2. $E\left|e_{i t}\right|^{8 q} \leq M<\infty$, for all $(i, t)$.
3. $E\left\|z_{t}\right\|^{4 q} \leq M<\infty$, for all $t$, where $z_{t}=\left(F_{t}^{\prime}, W_{t}^{\prime}\right)^{\prime}$.
4. $E\left|y_{t+h}\right|^{4 q} \leq M<\infty$, for all $t, h$.

Assumption WB. 1 strengthens Assumption B when the factor loadings are stochastic by requiring that $\lambda_{i}$ have uniformly bounded moments of order slightly larger than 4. Similarly, Assumptions WB. 2 and WB. 3 slightly strengthen the moment conditions of Assumptions C. 1 and E.1, respectively. Assumption WB. 4 is new; together with Assumption WB. 3 it ensures that $E\left|\varepsilon_{t+h}\right|^{4 q} \leq M<\infty$.

The following auxiliary result shows that Condition $A^{*}$ is satisfied under Assumptions A-F and WB.

Lemma 4.1 Under Assumptions $A-F$ and $W B$, Condition $A^{*}$ holds for the wild bootstrap provided $E^{*}\left|\eta_{i t}\right|^{4}<C$ and $E^{*}\left|v_{t+h}\right|^{4 q}<C$, for some $q>1$.

Our main result is as follows.

Theorem 4.1 Suppose Assumptions A-F and WB hold. Then the conclusions of Theorem 3.1 and Corollary 3.1 hold for the wild bootstrap provided $\sqrt{T} / N \rightarrow 0$ and $E^{*}\left|\eta_{i t}\right|^{4}<C$ for all $(i, t)$ and $E^{*}\left|v_{t+h}\right|^{4 q}<C$ for all $t$, for some $q>1$.

Although the wild bootstrap idiosyncratic errors $\left\{e_{i t}^{*}=\hat{e}_{i t} \eta_{i t}\right\}$ are independent (but possibly heteroskedastic) along both dimensions, the wild bootstrap residual-based method is asymptotically valid under Assumptions A-F and WB. These conditions allow for weak dependence across $(i, t)$. The validity of the wild bootstrap in the first step depends crucially on the assumption that $\sqrt{T} / N \rightarrow 0$ since under this assumption the covariance matrix of $\hat{\delta}$ does not depend on the dependence structure of $e_{i t}$.

The main motivation for using the wild bootstrap to generate $\varepsilon_{t+h}^{*}$ in the second step is that under Assumption F.2, $\varepsilon_{t+h}$ is a possibly heteroskedastic martingale difference sequence, so the wild bootstrap applied to the regression residuals is a natural choice. Under more general dependence conditions on $\varepsilon_{t+h}$ that would induce serially correlated scores, a wild bootstrap would not be valid. In this case, block bootstrap methods would be appropriate. Condition A* could still be used to establish the validity of the block bootstrap although we do not pursue this possibility here.

## 5 Monte Carlo results

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

The DGP follows Bai and $\operatorname{Ng}$ (2006) closely. The dependent variable is generated as

$$
\begin{equation*}
y_{t}=\beta+\alpha_{1} f_{1, t}+\alpha_{2} f_{2, t}+\varepsilon_{t}, \tag{12}
\end{equation*}
$$

where $\varepsilon_{t}$ is independent over time but possibly heteroskedastic. The factors are generated as

$$
\begin{equation*}
f_{j, t}=.8^{j} f_{j, t-1}+\left(1-.8^{2 j}\right)^{\frac{1}{2}} v_{j, t}, \tag{13}
\end{equation*}
$$

where $v_{t}=\binom{v_{1, t}}{v_{2, t}} \sim$ i.i.d. $N\left(0, I_{2}\right)$.
The $(T \times N)$ matrix of panel variables $X$ is generated as

$$
\begin{equation*}
X=F \Lambda^{\prime}+e, \tag{14}
\end{equation*}
$$

where

$$
\underset{(T \times 2)}{F}=\left(\begin{array}{cc}
f_{1,1} & f_{2,1} \\
\vdots & \vdots \\
f_{1, T} & f_{2, T}
\end{array}\right), \underset{(n \times 2)}{\Lambda}=\left(\begin{array}{cc}
\lambda_{1,1} & \lambda_{1,2} \\
\vdots & \vdots \\
f_{n, 1} & \lambda_{n, 2}
\end{array}\right)
$$

The loadings are drawn from $U[0,1]$ once and for all. We consider three values for $N(50,100$, and 200) and two values for $T$ ( 50 and 100). We also consider two values for the coefficients on the factors, either $\alpha=(0,0)^{\prime}$ or $\alpha=(1,1)^{\prime}$. The constant $\beta$ is set to unity throughout.

We consider three scenarios for the two error terms:

| DGP | $\varepsilon_{t}$ | $e_{i t}$ |
| :--- | :--- | :--- |
| 1 (homo-homo) | $N(0,1)$ | $N(0,1)$ |
| 2 (hetero-hetero) | $N\left(0, f_{1 t}^{2}\right)$ | $N\left(0, \sigma_{i}^{2}\right)$ |
| 3 (hetero-AR) | $N\left(0, f_{1 t}^{2}\right)$ | $\operatorname{AR}(1)+N\left(0, \sigma_{i}^{2}\right)$ |

DGP 1 is the simplest case where all the errors are independent in both dimensions and identically distributed. DGP 2 introduces heteroskedasticity in both error terms, while DGP 3 adds serial correlation to the idiosyncratic errors. In DGP 3, the autoregressive parameter is set to 0.5 . When $e_{i t}$ is heteroskedastic (DGP 2 and 3), its variance is drawn from $U[.5,1.5]$.

We concentrate on inference about the parameters in (12) . For this purpose, we use the heteroskedasticityrobust version of the covariance matrix of $\hat{\delta}=\left(\hat{\alpha}^{\prime}, \hat{\beta}^{\prime}\right)^{\prime}$ which is equation (6).

We consider the wild residual-based bootstrap described in Section 4. The two external variables $\eta_{i t}$ and $v_{t}$ are both i.i.d. $N(0,1)$. Since the properties of the idiosyncratic errors do not enter the asymptotic distribution of $\hat{\delta}$ to first-order, we also consider the i.i.d. bootstrap of the idiosyncratic errors. We have not verified Condition A* for this case, but it is clear that it will hold under an appropriate set of assumptions similar to those used for the wild bootstrap. For the first DGP (which assumes homoskedastic regression errors), we also consider the i.i.d. bootstrap for the regression error $\varepsilon_{t}$.

We consider two types of confidence intervals, symmetric percentile- $t$ and symmetric percentile, although our theory only covers the percentile case. The nominal level is $95 \%$. We report experiments based on 1000 replications with $B=399$ bootstrap repetitions.

The results are in Tables 1-3. Each cell has three columns corresponding to the constant, the first (dominant) factor, and the second factor, respectively. Each bootstrap method has two lines, the first one is the coverage rate for the percentile interval, while the second one is for the percentile- $t$ interval.

The performance of asymptotic theory depends heavily on the true value of the parameters. When the two coefficients on the factors are 0 , asymptotic theory is generally reliable in all three designs, although some mild size distortions appear for the smaller sample sizes. For non-zero values of $\alpha$, however, there are large size distortions for the factor coefficients, especially for the first one. These distortions are reduced as $N$ increases, but they remain important even with $N=200$. Behavior for the constant is very good for all configurations. The undercoverage for non-zero coefficients are worst in DGP 3 with serial correlation in the idiosyncratic error where it is as low as $47 \%$ and $49.9 \%$ for $N=T=50$.

To understand the source of these large distortions when the coefficients are not 0 , notice that we
can rewrite $\hat{\delta}$ (properly centered and scaled) as

$$
\begin{equation*}
\sqrt{T}(\hat{\delta}-\delta)=\left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{z}_{t} \hat{z}_{t}^{\prime}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \hat{z}_{t} \varepsilon_{t+h}+R_{1, N T} \tag{15}
\end{equation*}
$$

where

$$
R_{1, N T} \equiv \sqrt{T}\left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{z}_{t} \hat{z}_{t}^{\prime}\right)^{-1} \frac{1}{T} \sum_{t=1}^{T-h} \hat{z}_{t}\left(H F_{t}-\tilde{F}_{t}\right)^{\prime}\left(\alpha^{\prime} H^{-1}\right)^{\prime}
$$

reflects the factors estimation uncertainty. When the true value of the coefficients is 0 , this remainder is identically 0 , and the resulting asymptotic distribution is a good approximation to the finite-sample distribution of the first term in expression (15). When $\alpha \neq 0$, however, the remainder is not 0 , and while it disappears asymptotically under $\sqrt{T} / N \rightarrow 0$ and Assumptions A-F, it plays a large role in finite samples. In particular, our simulations revealed that it increases the variance of the estimator in finite samples. In other words, the asymptotic distribution has too small a variance relative to the actual behavior of the estimator, leading to confidence intervals that are too narrow relative to those that would be need to obtain correct coverage. See Ludvigson and $\operatorname{Ng}(2009 b)$ for more on the impact of relaxing the condition that $\sqrt{T} / N \rightarrow 0$.

Our wild residual-based bootstrap is quite successful in correcting these distortions. For example, when $N=50$ and $T=50$, the coverage rate of the confidence intervals in DGP 1 (where both errors are homoskedastic) go from 77.2 and $86.3 \%$ to 93.1 and $97.4 \%$. In general, we do have a coverage rate for the second factor that is higher than $95 \%$, but overall, the bootstrap seems to perform well in providing reliable inference for the first two designs. Percentile-t intervals have slightly higher coverage rates than the percentile intervals and this exacerbates the over-coverage problem for the second factor. For the third design, while the bootstrap does correct some of the large distortions, it is not completely accurate. For example with $N=50$ and $T=100$, the coverage rate of the asymptotic intervals is $51.8 \%$ and $81.4 \%$ respectively. The wild bootstrap improves these numbers to $80.5 \%$ and $91.0 \%$.

If we compare the bootstrap methods, there is little to choose from between the iid and the wild bootstrap of the idiosyncratic errors. If anything, the iid bootstrap performs slightly better in a few cases. There is more of a difference for DGP 1 where both sets of errors are indeed iid. In that case, using the iid bootstrap in both steps improves the coverage rate of the confidence intervals.

## 6 Conclusion

In this paper, we have given conditions under which a bootstrap scheme is valid in factor-augmented regressions under similar regularity conditions as Bai and Ng (2006). We have suggested a scheme that satisfies these conditions based on the wild bootstrap and documented the performance of this algorithm in a simulation experiment.

Extension of the present work to cases where the panel dimensions are not well-approximated by
the condition $\sqrt{T} / N \rightarrow 0$ (longer and narrower panels) seems important. Large distortions appeared in cases characterized by a non-negligible contribution of the remainder of (15) which is important when this condition is not satisfied.

A second important extension is the case of factor-augmented vector autoregressions (FAVAR) first suggested by Boivin and Bernanke (2002). This case has recently been analyzed by Yamamoto (2009), who proposes a bootstrap scheme that exploits the VAR structure in the factors and the panel variables.

## A Appendix A. Proofs of results in Section 3

Proof of Theorem 3.1, Part a). From (11), we can write

$$
\sqrt{T}\left(\hat{\delta}^{*}-\delta^{*}\right)=\left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{z}_{t}^{*} \hat{z}_{t}^{* \prime}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \hat{z}_{t}^{*} \varepsilon_{t+h}^{*}+R_{1, N T}^{*},
$$

where

$$
R_{1, N T}^{*} \equiv \sqrt{T}\left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{z}_{t}^{*} \hat{z}_{t}^{* \prime}\right)^{-1} \frac{1}{T} \sum_{t=1}^{T-h} \hat{z}_{t}^{*}\left(H^{*} \tilde{F}_{t}-\tilde{F}_{t}^{*}\right)^{\prime}\left(\hat{\alpha}^{\prime} H^{*-1}\right)^{\prime}
$$

reflects the contribution of factor estimation uncertainty to the stochastic expansion of the bootstrap OLS estimator. By Condition $A^{*} 3$., $R_{1, N T}^{*}=O_{P^{*}}(\sqrt{T} / N)=o_{P^{*}}(1)$, in probability, if $\sqrt{T} / N \rightarrow 0$. Similarly, we can write

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \hat{z}_{t}^{*} \varepsilon_{t+h}^{*}=\Phi^{*} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \hat{z}_{t} \varepsilon_{t+h}^{*}+R_{2, N T}^{*},
$$

where

$$
\Phi^{*} \equiv\left(\begin{array}{cc}
H^{*} & 0 \\
0 & I
\end{array}\right)=\Phi_{0}^{*}+O_{P^{*}}\left(\delta_{N T}^{-2}\right),
$$

in probability, where $\Phi_{0}^{*}=\operatorname{diag}( \pm 1)$, and where

$$
R_{2, N T}^{*} \equiv \sqrt{T} \frac{1}{T} \sum_{t=1}^{T-h}\binom{\tilde{F}_{t}^{*}-H^{*} F_{t}^{*}}{0} \varepsilon_{t+h}^{*} .
$$

By Condition $A^{*} 4$, we can show that $R_{2, N T}^{*}=O_{P^{*}}(\sqrt{T} / N)=o_{P^{*}}(1)$ if $\sqrt{T} / N \rightarrow 0$. By Condition $A^{*} 5, \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \hat{z}_{t} \varepsilon_{t+h}^{*}$ is normally distributed, and so it must be $O_{P^{*}}(1)$, which implies that

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \hat{z}_{t}^{*} \varepsilon_{t+h}^{*}=\Phi_{0}^{*} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \hat{z}_{t} \varepsilon_{t+h}^{*}+o_{P^{*}}(1)
$$

under our assumptions. Similarly, by adding and subtracting appropriately and using Condition $A^{*} 1$ and $A^{*} 2$, we can show that

$$
\frac{1}{T} \sum_{t=1}^{T} \hat{z}_{t}^{*} \hat{z}_{t}^{* \prime}=\Phi_{0}^{*}\left(\frac{1}{T} \sum_{t=1}^{T} \hat{z}_{t} \hat{z}_{t}^{\prime}\right) \Phi_{0}^{* \prime}+o_{P^{*}}(1)=\Phi_{0}^{*}\left(\Phi_{0} \Sigma_{z z} \Phi_{0}^{\prime}\right) \Phi_{0}^{* \prime}+o_{P^{*}}(1)
$$

where $p \lim \frac{1}{T} \sum_{t=1}^{T} \hat{z}_{t} \hat{z}_{t}^{\prime}=\Phi_{0} \Sigma_{z z} \Phi_{0}^{\prime}>0$ given Assumptions A-F. It follows that with probability approaching one,

$$
\begin{aligned}
\sqrt{T}\left(\hat{\delta}^{*}-\delta^{*}\right) & =\Phi_{0}^{* /-1}\left(\Phi_{0} \Sigma_{z z} \Phi_{0}^{\prime}\right)^{-1} \Phi_{0}^{*-1} \Phi_{0}^{*} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \hat{z}_{t} \varepsilon_{t+h}^{*}+o_{P^{*}}(1) \\
& =\underbrace{\Phi_{0}^{*-1}\left(\Phi_{0} \Sigma_{z z} \Phi_{0}^{\prime}\right)^{-1}\left(\Phi_{0} \Sigma_{z z, \varepsilon} \Phi_{0}^{\prime}\right)^{1 / 2} \underbrace{\left(\Phi_{0} \Sigma_{z z . \varepsilon} \Phi_{0}^{\prime}\right)^{-1 / 2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \hat{z}_{t} \varepsilon_{t+h}^{*}}_{\rightarrow d^{*} N(0, I) \text { by Condition } A^{*} 5 .}}_{\rightarrow d^{d^{*}} N\left(0, \Phi_{0}^{* \prime-1} \Sigma_{\delta} \Phi_{0}^{*-1}\right)}+o_{P^{*}}(1),
\end{aligned}
$$

where $\Sigma_{\delta} \equiv\left(\Phi_{0} \Sigma_{z z} \Phi_{0}^{\prime}\right)^{-1}\left(\Phi_{0} \Sigma_{z z . \varepsilon} \Phi_{0}^{\prime}\right)\left(\Phi_{0} \Sigma_{z z} \Phi_{0}^{\prime}\right)^{-1}=\Phi_{0}^{\prime-1} \Sigma_{z z}^{-1} \Sigma_{z z . \varepsilon} \Sigma_{z z}^{-1} \Phi_{0}^{-1}$.
Part b) follows by noting that the main diagonal elements of $\Sigma_{\delta}^{*}$ coincide with those of $\Sigma_{\delta}$. An application of Polya's theorem provides the uniform convergence result.
Proof of Corollary [3.1. We only prove part a) since part b) follows trivially. Since $\delta^{*}=\delta_{0}^{*}+$ $O_{P^{*}}\left(\delta_{N T}^{-2}\right)$, where $\delta_{0}^{*} \equiv \Phi_{0}^{* /-1} \hat{\delta}$, we have that

$$
\sqrt{T}\left(\hat{\delta}^{*}-\delta^{*}\right)=\sqrt{T}\left(\hat{\delta}^{*}-\delta_{0}^{*}\right)+\sqrt{T}\left(\delta^{*}-\delta_{0}^{*}\right),
$$

where the second term is of order $O_{P^{*}}(\sqrt{T} / N)=o_{P^{*}}(1)$ under our assumptions. The result then follows from Theorem 3.1 by pre-multiplying $\sqrt{T}\left(\hat{\delta}^{*}-\delta_{0}^{*}\right)$ by $\Phi_{0}^{* \prime}$ and noting that $\Phi_{0}^{* \prime} \delta_{0}^{*}=\hat{\delta}$.

## B Appendix B. Bootstrap factor estimation results

In this Appendix, we give a set of high level conditions on $\left\{e_{i t}^{*}\right\}$ and $\left\{\varepsilon_{t+h}^{*}\right\}$ that can be used to appropriately control the error incurred in estimating the bootstrap factors $\tilde{F}_{t}^{*}$. In particular, these conditions will imply the first four parts of Condition $A^{*}$ given in Section 3. In Appendix C we will verify these conditions for the wild bootstrap algorithm proposed in Section 4.

Let

$$
X_{t}^{*}=\tilde{\Lambda} \tilde{F}_{t}+e_{t}^{*}
$$

where $e_{t}^{*}=\left(e_{1 t}^{*}, \ldots, e_{N t}^{*}\right)^{\prime}$ satisfies the following conditions.
Condition B1. $\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T}\left|\gamma_{s t}^{*}\right|^{2}=O_{P}(1)$, where $\gamma_{s t}^{*}=E^{*}\left(\frac{1}{N} \sum_{i=1}^{N} e_{i t}^{*} e_{i s}^{*}\right)$.
Condition B2. $\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} E^{*}\left|\frac{1}{\sqrt{N}} \sum_{i=1}^{N}\left(e_{i t}^{*} e_{i s}^{*}-E^{*}\left(e_{i t}^{*} e_{i s}^{*}\right)\right)\right|^{2}=O_{P}(1)$.
Condition B3. $\frac{1}{T} \sum_{t=1}^{T} E^{*}\left\|\frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{\sqrt{N}}\right\|^{2}=O_{P}(1)$.
Under Conditions B1-B3, we can prove the following result.

Lemma B. 1 Suppose $X_{t}^{*}=\tilde{\Lambda} \tilde{F}_{t}+e_{t}^{*}$, where $\left\{e_{i t}^{*}\right\}$ satisfies Conditions B1, B2, B3. It follows that

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

in probability, where $\delta_{N T}=\min (\sqrt{N}, \sqrt{T})$.
Lemma B. 1 is the bootstrap analogue of Lemma A.1.(i) of Bai and Ng (2006).
To prove the next result, we need the following additional conditions.
Condition B4. $\frac{1}{T} \sum_{t=1}^{T} E^{*}\left\|\frac{1}{\sqrt{T N}} \sum_{s=1}^{T} \sum_{i=1}^{N} \tilde{F}_{s}\left(e_{i t}^{*} e_{i s}^{*}-E^{*}\left(e_{i t}^{*} e_{i s}^{*}\right)\right)\right\|^{2}=O_{P}(1)$.
Condition B4 is the bootstrap analogue of Assumption F1 in Bai (2003).
Condition B5. $\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s} \hat{z}_{t}^{\prime} \gamma_{s t}^{*}=O_{P}(1)$, where $\hat{z}_{t}=\left(\tilde{F}_{t}^{\prime}, W_{t}^{\prime}\right)^{\prime}$.
Condition B6. $E^{*}\left\|\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{\sqrt{N}} \hat{z}_{t}^{\prime}\right\|^{2}=O_{P}(1)$.
Condition B6 is the bootstrap analogue of Assumption F2 in Bai (2003).
Lemma B. 2 Suppose $X_{t}^{*}=\tilde{\Lambda} \tilde{F}_{t}+e_{t}^{*}$, where $\left\{e_{i t}^{*}\right\}$ satisfies Conditions B1-B6. It follows that
a) $\frac{1}{T} \sum_{t=1}^{T}\left(\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}\right) \hat{z}_{t}^{\prime}=O_{P^{*}}\left(\delta_{N T}^{-2}\right)$, in probability, where $\hat{z}_{t}=\left(\tilde{F}_{t}^{\prime}, W_{t}^{\prime}\right)^{\prime}$.
b) $\frac{1}{T} \sum_{t=1}^{T}\left(\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}\right) \hat{z}_{t}^{* \prime}=O_{P^{*}}\left(\delta_{N T}^{-2}\right)$, in probability, where $\hat{z}_{t}^{*}=\left(\tilde{F}_{t}^{* \prime}, W_{t}^{\prime}\right)^{\prime}$.

Lemma B. 2 is the bootstrap analogue of Lemma A.1.(ii) and (iii) of Bai and Ng (2006).
Our next result controls the order of magnitude of $\frac{1}{T} \sum_{t=1}^{T-h}\left(\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}\right) \varepsilon_{t+h}^{*}$, where $\left\{\varepsilon_{t+h}^{*}\right\}$ is a bootstrap sample from $\left\{\hat{\varepsilon}_{t+h}\right\}$.

The following three additional conditions suffice to prove the result.
Condition B7. $\frac{1}{T} \sum_{t=1}^{T-h} \sum_{s=1}^{T} \tilde{F}_{s} \varepsilon_{t+h}^{*} \gamma_{s t}^{*}=O_{P}^{*}$ (1), in probability.
Condition B8. $E^{*}\left\|\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{\sqrt{N}} \varepsilon_{t+h}^{*}\right\|^{2}=O_{P}(1)$.
Condition B9. $\frac{1}{T} \sum_{t=1}^{T-h} E^{*}\left|\varepsilon_{t+h}^{*}\right|^{2}=O_{P}(1)$.
Lemma B. 3 Suppose $X_{t}^{*}=\tilde{\Lambda} \tilde{F}_{t}+e_{t}^{*}$, where $\left\{e_{i t}^{*}\right\}$ satisfies Conditions B1-B9. Suppose in addition that $\left\{\varepsilon_{t+h}^{*}: t=1, \ldots, T-h\right\}$ is a bootstrap sample obtained independently of $\left\{e_{t}^{*}\right\}$ such that Conditions B7-B9 are verified. It follows that

$$
\frac{1}{T} \sum_{t=1}^{T-h}\left(\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}\right) \varepsilon_{t+h}^{*}=O_{P^{*}}\left(\delta_{N T}^{-2}\right)
$$

in probability.

Lemma B. 3 is the bootstrap analogue of Lemma A.1.(iv) of Bai and Ng (2006). The proof of this result follows exactly the proof of Lemma B.2, using the additional conditions, and therefore its proof is omitted.
Proof of Lemma B.1. The proof is based on the following identity:

$$
\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}=\tilde{V}^{*-1}\left(\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_{s}^{*} \gamma_{s t}^{*}+\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_{s}^{*} \zeta_{s t}^{*}+\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_{s}^{*} \eta_{s t}^{*}+\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_{s}^{*} \xi_{s t}^{*}\right)
$$

where

$$
\begin{aligned}
& \gamma_{s t}^{*}=E^{*}\left(\frac{1}{N} \sum_{i=1}^{N} e_{i s}^{*} e_{i t}^{*}\right), \zeta_{s t}^{*}=\frac{1}{N} \sum_{i=1}^{N}\left(e_{i s}^{*} e_{i t}^{*}-E^{*}\left(e_{i s}^{*} e_{i t}^{*}\right)\right) \\
& \eta_{s t}^{*}=\frac{1}{N} \sum_{i=1}^{N} \tilde{\lambda}_{i}^{\prime} \tilde{F}_{s} e_{i t}^{*}=\tilde{F}_{s}^{\prime} \frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{N} \text { and } \xi_{s t}^{*}=\frac{1}{N} \sum_{i=1}^{N} \tilde{\lambda}_{i}^{\prime} \tilde{F}_{t} e_{i s}^{*}=\eta_{t s}^{*} .
\end{aligned}
$$

Ignoring $\tilde{V}^{*-1}$ (which is $O_{P^{*}}(1)$ ), it follows that

$$
\frac{1}{T} \sum_{t=1}^{T}\left\|\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}\right\|^{2} \leq \frac{1}{T} \sum_{t=1}^{T}\left(a_{t}+b_{t}+c_{t}+d_{t}\right)
$$

where

$$
a_{t}=T^{-2}\left\|\sum_{s=1}^{T} \tilde{F}_{s}^{*} \gamma_{s t}^{*}\right\|^{2}, \quad b_{t}=T^{-2}\left\|\sum_{s=1}^{T} \tilde{F}_{s}^{*} \zeta_{s t}^{*}\right\|^{2}, \quad c_{t}=T^{-2}\left\|\sum_{s=1}^{T} \tilde{F}_{s}^{*} \eta_{s t}^{*}\right\|^{2}, \quad d_{t}=T^{-2}\left\|\sum_{s=1}^{T} \tilde{F}_{s}^{*} \xi_{s t}^{*}\right\|^{2}
$$

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

$$
\begin{aligned}
\frac{1}{T} \sum_{t=1}^{T} a_{t} \leq \frac{1}{T} & \underbrace{\left(\frac{1}{T} \sum_{s=1}^{T}\left\|\tilde{F}_{s}^{*}\right\|^{2}\right)} \underbrace{\left(\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \gamma_{s t}^{* 2}\right)}_{=\tilde{F}_{P}^{*}(1) \text { by Condition B1 }}=r \text { because } \frac{\tilde{F}^{*} \mid \tilde{F}^{*}}{T}=I_{r}
\end{aligned}=O_{P}\left(\frac{1}{T}\right) .
$$

For the second term, we have that

$$
\frac{1}{T} \sum_{t=1}^{T} b_{t} \leq \underbrace{\left(\frac{1}{T} \sum_{s=1}^{T}\left\|\tilde{F}_{s}^{*}\right\|^{2}\right)}_{=r} \underbrace{\left(\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T}\left|\zeta_{s t}^{*}\right|^{2}\right)}_{O_{P^{*}}\left(\frac{1}{N}\right)}=O_{P^{*}}\left(\frac{1}{N}\right) \cdot
$$

In particular, by Condition B2, we can show that

$$
\begin{aligned}
\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} E^{*}\left|\zeta_{s t}^{*}\right|^{2} & =\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} E^{*}\left|\frac{1}{N} \sum_{i=1}^{N}\left(e_{i s}^{*} e_{i t}^{*}-E^{*}\left(e_{i s}^{*} e_{i t}^{*}\right)\right)\right|^{2} \\
& =\frac{1}{N} \frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} E^{*}\left|\frac{1}{\sqrt{N}} \sum_{i=1}^{N}\left(e_{i s}^{*} e_{i t}^{*}-E^{*}\left(e_{i s}^{*} e_{i t}^{*}\right)\right)\right|^{2}=O_{P}\left(\frac{1}{N}\right),
\end{aligned}
$$

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

For the third term,

$$
\frac{1}{T} \sum_{t=1}^{T} c_{t}=\frac{1}{T} \sum_{t=1}^{T} T^{-2}\left\|\sum_{s=1}^{T} \tilde{F}_{s}^{*} \tilde{F}_{s}^{\prime} \frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{N}\right\|^{2}=\frac{1}{T} \sum_{t=1}^{T}\left\|\frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{N}\right\|^{2}\left\|T^{-1} \sum_{s=1}^{T} \tilde{F}_{s}^{*} \tilde{F}_{s}^{\prime}\right\|^{2}
$$

We have that

$$
\left\|T^{-1} \sum_{s=1}^{T} \tilde{F}_{s}^{*} \tilde{F}_{s}^{\prime}\right\|^{2} \leq \underbrace{\left(T^{-1} \sum_{s=1}^{T}\left\|\tilde{F}_{s}^{*}\right\|^{2}\right)}_{=r} \underbrace{\left(T^{-1} \sum_{s=1}^{T}\left\|\tilde{F}_{s}\right\|^{2}\right)}_{=r}=r^{2}=O_{P}(1),
$$

whereas by Condition B3 and Markov's inequality,

$$
\frac{1}{T} \sum_{t=1}^{T}\left\|\frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{N}\right\|^{2}=O_{P^{*}}\left(\frac{1}{N}\right)
$$

The fourth term in $d_{t}$ follows by the same arguments, using Condition B3.
Proof of Lemma B.2. Proof of part a). Using the above identity, we have that

$$
\frac{1}{T} \sum_{t=1}^{T}\left(\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}\right) \hat{z}_{t}^{\prime}=\tilde{V}^{*-1}(I+I I+I I I+I V),
$$

where

$$
\begin{aligned}
I & =\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s}^{*} \gamma_{s t}^{*} \hat{z}_{t}^{\prime}, \quad I I=\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s}^{*} \zeta_{s t}^{*} \hat{z}_{t}^{\prime}, \\
I I I & =\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s}^{*} \eta_{s t}^{*} \hat{z}_{t}^{\prime}, \quad \text { and } I V=\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s}^{*} \xi_{s t}^{*} \hat{z}_{t}^{\prime} .
\end{aligned}
$$

Start with $I$. We can write

$$
I=\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T}\left(\tilde{F}_{s}^{*}-H^{*} \tilde{F}_{s}\right) \gamma_{s t}^{*} \hat{z}_{t}^{\prime}+H^{*} \frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s} \hat{z}_{t}^{\prime} \gamma_{s t}^{*} \equiv I_{1}+I_{2} .
$$

$I_{2}=O_{P}\left(\frac{1}{T}\right)$ by Condition B3. For $I_{1}$, repeated application of the Cauchy-Schwartz inequality implies that

$$
\begin{aligned}
I_{1} & =\frac{1}{T} \sum_{s=1}^{T}\left(\tilde{F}_{s}^{*}-H^{*} \tilde{F}_{s}\right)\left(\frac{1}{T} \sum_{t=1}^{T} \gamma_{s t}^{*} \hat{z}_{t}^{\prime}\right) \leq\left(\frac{1}{T} \sum_{s=1}^{T}\left\|\tilde{F}_{s}^{*}-H^{*} \tilde{F}_{s}\right\|^{2}\right)^{1 / 2}\left(\frac{1}{T} \sum_{s=1}^{T}\left\|\frac{1}{T} \sum_{t=1}^{T} \gamma_{s t}^{*} \hat{z}_{t}^{\prime}\right\|^{2}\right)^{1 / 2} \\
& \leq \frac{1}{\sqrt{T}}(\underbrace{\frac{1}{T} \sum_{s=1}^{T}\left\|\tilde{F}_{s}^{*}-H^{*} \tilde{F}_{s}\right\|^{2}}_{O_{P *}\left(\delta_{N T}^{-2}\right)})^{1 / 2}(\underbrace{\frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T}\left|\gamma_{s t}^{*}\right|^{2}}_{O_{P}(1) \text { by Condition B1 }}(\underbrace{\frac{1}{T} \sum_{t=1}^{T}\left\|\hat{z}_{t}\right\|^{2}}_{O_{P}(1)}))^{1 / 2}=O_{P^{*}}\left(\frac{1}{\sqrt{T} \delta_{N T}}\right),
\end{aligned}
$$

provided $\frac{1}{T} \sum_{t=1}^{T}\left\|\hat{z}_{t}\right\|^{2}=O_{P}(1)$, which follows under Bai and Ng's (2006) assumptions. In particular,
given that $\hat{z}_{t}=\left(\tilde{F}_{t}^{\prime}, W_{t}^{\prime}\right)^{\prime}$,

$$
\frac{1}{T} \sum_{t=1}^{T}\left\|\hat{z}_{t}\right\|^{2}=\frac{1}{T} \sum_{t=1}^{T}\left\|\tilde{F}_{t}\right\|^{2}+\frac{1}{T} \sum_{t=1}^{T}\left\|W_{t}\right\|^{2}=O_{P}(1)
$$

if $E\left\|W_{t}\right\|^{2}=O(1)$, since $\frac{1}{T} \sum_{t=1}^{T}\left\|\tilde{F}_{t}\right\|^{2}=r$. Thus, $I=O_{P^{*}}\left(\frac{1}{\sqrt{T} \delta_{N T}}\right)+O_{P}\left(\frac{1}{T}\right)=O_{P^{*}}\left(\delta_{N T}^{-2}\right)$, in probability. Next, consider $I I$. We have that

$$
I I=\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T}\left(\tilde{F}_{s}^{*}-H^{*} \tilde{F}_{s}\right) \zeta_{s t}^{*} \hat{z}_{t}^{\prime}+H^{*} \frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s} \zeta_{s t}^{*} \hat{z}_{t}^{\prime} \equiv I I_{1}+I I_{2}
$$

We can show that

$$
I I_{1} \leq \underbrace{\left(\frac{1}{T} \sum_{s=1}^{T}\left\|\tilde{F}_{s}^{*}-H^{*} \tilde{F}_{s}\right\|^{2}\right)^{1 / 2}}_{O_{P^{*}}\left(\delta_{N T}^{-1}\right)} \underbrace{\left(\frac{1}{T} \sum_{s=1}^{T}\left\|\frac{1}{T} \sum_{t=1}^{T} \zeta_{s t}^{*} \hat{z}_{t}^{\prime}\right\|^{2}\right)^{1 / 2}}_{O_{P^{*}}\left(\frac{1}{\sqrt{N T}}\right)}=O_{P^{*}}\left(\frac{1}{\sqrt{N T} \delta_{N T}}\right)
$$

Indeed, by Cauchy-Schwartz inequality,

$$
\begin{aligned}
\frac{1}{T} \sum_{s=1}^{T} E^{*}\left\|\frac{1}{T} \sum_{t=1}^{T} \zeta_{s t}^{*} \hat{z}_{t}^{\prime}\right\|^{2} & \leq \frac{1}{T} \sum_{s=1}^{T} E^{*}\left\{\left(\frac{1}{T} \sum_{t=1}^{T}\left|\zeta_{s t}^{*}\right|^{2}\right)\left(\frac{1}{T} \sum_{t=1}^{T}\left\|\hat{z}_{t}\right\|^{2}\right)\right\} \\
& =\frac{1}{T} \underbrace{\left(\frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} E^{*}\left|\zeta_{s t}^{*}\right|^{2}\right)}_{O_{P}(1) \text { by Condition B2 }} \underbrace{\left(\frac{1}{T} \sum_{t=1}^{T}\left\|\hat{z}_{t}\right\|^{2}\right)}_{O_{P}(1)}=O_{P}\left(\frac{1}{T}\right)
\end{aligned}
$$

For $I I_{2}$, ignoring $H^{*}$ (which is $O_{P^{*}}(1)$ since $H^{*}=H_{0}^{*}+O_{P^{*}}\left(\delta_{N T}^{-2}\right)$, where $H_{0}^{*}=\operatorname{diag}( \pm 1)$ ), we have that

$$
I I_{2}=\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s} \zeta_{s t}^{*} \hat{z}_{t}^{\prime}=\frac{1}{\sqrt{T N}} \frac{1}{T} \sum_{t=1}^{T} \underbrace{\left(\frac{1}{\sqrt{T N}} \sum_{s=1}^{T} \sum_{i=1}^{N} \tilde{F}_{s}\left(e_{i t}^{*} e_{i s}^{*}-E^{*}\left(e_{i t}^{*} e_{i s}^{*}\right)\right)\right)}_{\equiv m_{t}^{*}} \hat{z}_{t}^{\prime} \equiv \frac{1}{\sqrt{T N}} \frac{1}{T} \sum_{t=1}^{T} m_{t}^{*} \hat{z}_{t}^{\prime}
$$

We can show that $\frac{1}{T} \sum_{t=1}^{T} m_{t}^{*} \hat{z}_{t}^{\prime}=O_{P^{*}}(1)$, implying that $I I_{2}=O_{P^{*}}\left(\frac{1}{\sqrt{N T}}\right)$ in probability. By Cauchy-Schwartz inequality, we have that

$$
\frac{1}{T} \sum_{t=1}^{T} m_{t}^{*} \hat{z}_{t}^{\prime} \leq\left(\frac{1}{T} \sum_{t=1}^{T}\left\|m_{t}^{*}\right\|^{2}\right)^{1 / 2}\left(\frac{1}{T} \sum_{t=1}^{T}\left\|\hat{z}_{t}\right\|^{2}\right)^{1 / 2}=O_{P^{*}}(1),
$$

provided $\frac{1}{T} \sum_{t=1}^{T}\left\|m_{t}^{*}\right\|^{2}=O_{P^{*}}(1)$, or $\frac{1}{T} \sum_{t=1}^{T} E^{*}\left\|m_{t}^{*}\right\|^{2}=O_{P}(1)$, by Markov's inequality. But

$$
\frac{1}{T} \sum_{t=1}^{T} E^{*}\left\|m_{t}^{*}\right\|^{2}=\frac{1}{T} \sum_{t=1}^{T} E^{*}\left\|\frac{1}{\sqrt{T N}} \sum_{s=1}^{T} \sum_{i=1}^{N} \tilde{F}_{s}\left(e_{i t}^{*} e_{i s}^{*}-E^{*}\left(e_{i t}^{*} e_{i s}^{*}\right)\right)\right\|^{2}=O_{P}(1)
$$

by Condition B4.

Next, consider

$$
I I I=\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s}^{*} \eta_{s t}^{*} \hat{z}_{t}^{\prime}=\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T}\left(\tilde{F}_{s}^{*}-H^{*} \tilde{F}_{s}\right) \eta_{s t}^{*} \hat{z}_{t}^{\prime}+H^{*} \frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s} \eta_{s t}^{*} \hat{z}_{t}^{\prime} \equiv I I I_{1}+I I I_{2} .
$$

Starting with $I I I_{1}$, we have that

$$
I I I_{1} \leq \underbrace{\left(\frac{1}{T} \sum_{s=1}^{T}\left\|\tilde{F}_{s}^{*}-H^{*} \tilde{F}_{s}\right\|^{2}\right)^{1 / 2}}_{O_{P^{*}}\left(\delta_{N T}^{-1}\right)} \underbrace{\left(\frac{1}{T} \sum_{s=1}^{T}\left\|\frac{1}{T} \sum_{t=1}^{T} \eta_{s t}^{*} \hat{z}_{t}^{\prime}\right\|^{2}\right)^{1 / 2}}_{O_{P^{*}}\left(\frac{1}{\sqrt{N}}\right)}
$$

since

$$
\frac{1}{T} \sum_{s=1}^{T} E^{*}\left\|\frac{1}{T} \sum_{t=1}^{T} \eta_{s t}^{*} \hat{z}_{t}^{\prime}\right\|^{2} \leq \frac{1}{T} \sum_{s=1}^{T} E^{*}\left(\frac{1}{T} \sum_{t=1}^{T} \eta_{s t}^{* 2}\right)\left(\frac{1}{T} \sum_{t=1}^{T}\left\|\hat{z}_{t}\right\|^{2}\right)=\frac{1}{T^{2}} \sum_{s=1}^{T} \sum_{t=1}^{T} E^{*}\left(\eta_{s t}^{* 2}\right) \times O_{P}(1),
$$

and we can show that under Condition B3 $\frac{1}{T^{2}} \sum_{s=1}^{T} \sum_{t=1}^{T} E^{*}\left(\eta_{s t}^{* 2}\right)=O_{P}\left(\frac{1}{N}\right)$. Indeed,

$$
\begin{aligned}
\frac{1}{T^{2}} \sum_{s=1}^{T} \sum_{t=1}^{T} \eta_{s t}^{* 2} & =\frac{1}{T^{2}} \sum_{s=1}^{T} \sum_{t=1}^{T}\left\|\tilde{F}_{s}^{\prime} \frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{N}\right\|^{2}=\frac{1}{T^{2}} \sum_{s=1}^{T} \sum_{t=1}^{T}\left\|\tilde{F}_{s}\right\|^{2}\left\|\frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{N}\right\|^{2} \\
& =\left(\frac{1}{T} \sum_{s=1}^{T}\left\|\tilde{F}_{s}\right\|^{2}\right)\left(\frac{1}{T} \sum_{t=1}^{T}\left\|\frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{\sqrt{N}}\right\|^{2}\right) \frac{1}{N}=O_{P^{*}}\left(\frac{1}{N}\right),
\end{aligned}
$$

given Condition B3.
Ignoring again $H^{*}$ and replacing $\eta_{s t}^{*}$ with its definition, we have that

$$
I I I_{2}=\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s} \eta_{s t}^{*} \hat{z}_{t}^{\prime}=\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s}\left(\tilde{F}_{s}^{\prime} \frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{N}\right) \hat{z}_{t}^{\prime}=\underbrace{\left(\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_{s} \tilde{F}_{s}^{\prime}\right)}_{O_{P}(1)}\left(\frac{1}{T} \sum_{t=1}^{T} \frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{N} \hat{z}_{t}^{\prime}\right) .
$$

For the second factor, we can write

$$
\frac{1}{T} \sum_{t=1}^{T} \frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{N} \hat{z}_{t}^{\prime}=\frac{1}{\sqrt{T N}}\left(\frac{1}{\sqrt{T N}} \sum_{t=1}^{N} \tilde{\Lambda}^{\prime} e_{t}^{*} \hat{z}_{t}^{\prime}\right)
$$

and use Condition B6 to show that $\frac{1}{\sqrt{T N}} \sum_{t=1}^{N} \tilde{\Lambda}^{\prime} e_{t}^{*} \hat{z}_{t}^{\prime}=O_{P^{*}}(1)$. This implies that $I I I_{2}=O_{P^{*}}\left(\frac{1}{\sqrt{N T}}\right)$.
For $I V$, write

$$
I V=\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T}\left(\tilde{F}_{s}^{*}-H^{*} \tilde{F}_{s}\right) \xi_{s t}^{*} \hat{z}_{t}^{\prime}+H^{*} \frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s} \xi_{s t}^{*} \hat{z}_{t}^{\prime} \equiv I V_{1}+I V_{2}
$$

and note that $\xi_{s t}^{*}=\eta_{t s}^{*}$. It follows that $I V_{1}=O_{P^{*}}\left(\frac{1}{\delta_{N T} \sqrt{N}}\right)$ by the exact same arguments as we used
for $I I I_{1}$. For $I V_{2}$, we have that

$$
\begin{aligned}
I V_{2} & =\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s} \xi_{s t}^{*} \hat{z}_{t}^{\prime}=\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s} \eta_{t s}^{*} \hat{z}_{t}^{\prime}=\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s}\left(\tilde{F}_{t}^{\prime} \frac{\tilde{\Lambda}^{\prime} e_{s}^{*}}{N}\right) \hat{z}_{t}^{\prime} \\
& =\frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s}\left(\frac{e_{s}^{*} \tilde{\Lambda}}{N} \tilde{F}_{t}\right) \hat{z}_{t}^{\prime}=\left(\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_{s} \frac{e_{s}^{*} \tilde{\Lambda}}{N}\right)\left(\frac{1}{T} \sum_{t=1}^{T} \tilde{F}_{t} \hat{z}_{t}^{\prime}\right)=O_{P^{*}}\left(\frac{1}{\sqrt{N T}}\right) \times O_{P}(1),
\end{aligned}
$$

given that Condition B 6 can be used to show that $\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_{s} \frac{e_{s}^{* *} \tilde{\Lambda}}{N}=O_{P^{*}}\left(\frac{1}{\sqrt{N T}}\right)$.
Proof of part b). Let $\hat{z}_{t}^{*}=\left(\tilde{F}_{t}^{* \prime}, W_{t}^{\prime}\right)^{\prime}$, and write $\hat{z}_{t}^{*}=\hat{z}_{t}^{*}-\bar{z}_{t}^{*}+\bar{z}_{t}^{*}$, where

$$
\bar{z}_{t}^{*}=\binom{H^{*} \tilde{F}_{t}}{W_{t}}=\left(\begin{array}{cc}
H^{*} & 0 \\
0 & I
\end{array}\right)\binom{\tilde{F}_{t}}{W_{t}} \equiv \Phi^{*} \hat{z}_{t}
$$

It follows that

$$
\frac{1}{T} \sum_{t=1}^{T}\left(\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}\right) \hat{z}_{t}^{* \prime}=\frac{1}{T} \sum_{t=1}^{T}\left(\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}\right)\left(\hat{z}_{t}^{*}-\bar{z}_{t}^{*}\right)^{\prime}+\frac{1}{T} \sum_{t=1}^{T}\left(\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}\right) \bar{z}_{t}^{* \prime} \equiv A^{*}+B^{*}
$$

Note that

$$
\hat{z}_{t}^{*}-\bar{z}_{t}^{*}=\binom{\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}}{W_{t}-W_{t}}=\binom{\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}}{0}
$$

which implies that

$$
A^{*}=\left(\frac{1}{T} \sum_{t=1}^{T}\left(\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}\right)\left(\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}\right)^{\prime} \quad 0\right)=\left(\begin{array}{ll}
O_{P^{*}}\left(\delta_{N T}^{-2}\right) & 0
\end{array}\right)=O_{P^{*}}\left(\delta_{N T}^{-2}\right),
$$

given that $\frac{1}{T} \sum_{t=1}^{T}\left\|\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}\right\|^{2}=O_{P^{*}}\left(\delta_{N T}^{-2}\right)$.
For $B^{*}$, we have that

$$
B^{*}=\frac{1}{T} \sum_{t=1}^{T}\left(\tilde{F}_{t}^{*}-H^{*} \tilde{F}_{t}\right) \hat{z}_{t}^{\prime} \Phi^{* \prime}=O_{P^{*}}\left(\delta_{N T}^{-2}\right) \times O_{P^{*}}(1)=O_{P^{*}}\left(\delta_{N T}^{-2}\right)
$$

given part a) and the fact that $H^{*}=O_{P^{*}}(1)$.

## C Appendix C. Proofs of results in Section 4

First, we state an auxiliary result and its proof. Then we prove Lemma 4.1 and Theorem 4.1.
Lemma C. 1 Suppose Assumptions A-F hold. If, in addition, for some $p \geq 2, E\left|e_{i t}\right|^{2 p} \leq M<\infty$, $E\left\|\lambda_{i}\right\|^{p} \leq M<\infty$ and $E\left\|F_{t}\right\|^{p} \leq M<\infty$, it follows that
(i) $\frac{1}{T} \sum_{t=1}^{T}\left\|\tilde{F}_{t}-H F_{t}\right\|^{p}=O_{P}(1)$;
(ii) $\frac{1}{N} \sum_{i=1}^{N}\left\|\tilde{\lambda}_{i}-H^{-1 \prime} \lambda_{i}\right\|^{p}=O_{P}(1)$;
(iii) $\frac{1}{T N} \sum_{t=1}^{T} \sum_{i=1}^{N} \tilde{e}_{i t}^{p}=O_{P}(1)$.

## Proof of Lemma C. 1 ,

Proof of (i). We rely on the following identity (see Bai and Ng (2002), proof of Theorem 1):

$$
\tilde{F}_{t}-H F_{t}=\tilde{V}^{-1}\left(\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_{s} \psi_{s t}+\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_{s} \eta_{s t}+\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_{s} \xi_{s t}\right)
$$

where

$$
\psi_{s t}=\frac{1}{N} \sum_{i=1}^{N} e_{i s} e_{i t} ; \quad \eta_{s t}=\frac{1}{N} \sum_{i=1}^{N} \lambda_{i}^{\prime} F_{s} e_{i t} ; \quad \text { and } \xi_{s t}=\eta_{t s} .
$$

It follows that by the $c-r$ inequality,

$$
\frac{1}{T} \sum_{t=1}^{T}\left\|\tilde{F}_{t}-H F_{t}\right\|^{p} \leq 3^{p-1}\left\|\tilde{V}^{-1}\right\|^{p}\left(\frac{1}{T} \sum_{t=1}^{T} a_{t}+\frac{1}{T} \sum_{t=1}^{T} b_{t}+\frac{1}{T} \sum_{t=1}^{T} c_{t}\right)
$$

where

$$
a_{t}=\frac{1}{T^{p}}\left\|\sum_{s=1}^{T} \tilde{F}_{s} \psi_{s t}\right\|^{p} ; \quad b_{t}=\frac{1}{T^{p}}\left\|\sum_{s=1}^{T} \tilde{F}_{s} \eta_{s t}\right\|^{p} ; \text { and } c_{t}=\frac{1}{T^{p}}\left\|\sum_{s=1}^{T} \tilde{F}_{s} \xi_{s t}\right\|^{p} .
$$

Let $\chi_{s t}$ denote either $\psi_{s t}, \eta_{s t}$ or $\xi_{s t}$. We can write

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

where the inequality follows by Cauchy-Schwartz. It follows that

$$
\begin{aligned}
\frac{1}{T} \sum_{t=1}^{T} \frac{1}{T^{p}}\left\|\sum_{s=1}^{T} \tilde{F}_{s} \chi_{s t}\right\|^{p} & \leq \frac{1}{T} \sum_{t=1}^{T}(\underbrace{\frac{1}{T} \sum_{s=1}^{T}\left\|\tilde{F}_{s}\right\|^{2}}_{=r})^{p / 2}\left(\frac{1}{T} \sum_{s=1}^{T}\left|\chi_{s t}\right|^{2}\right)^{p / 2} \leq r^{p / 2} \frac{1}{T} \sum_{t=1}^{T}\left(\frac{1}{T} \sum_{s=1}^{T}\left|\chi_{s t}\right|^{2}\right)^{p / 2} \\
& \leq r^{p / 2} \frac{1}{T^{2}} \sum_{t=1}^{T} \sum_{s=1}^{T}\left|\chi_{s t}\right|^{p}
\end{aligned}
$$

where the last inequality follows again by the $c-r$ inequality. Thus it suffices to show that $E\left|\chi_{s t}\right|^{p} \leq$ $M<\infty$ to prove that the above term is $O_{P}(1)$. Starting with $\chi_{s t}=\psi_{s t}$,

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

given that we assume $E\left|e_{i t}\right|^{2 p} \leq M<\infty$. When $\chi_{s t}=\eta_{s t}$, we have that

$$
E\left|\eta_{s t}\right|^{p}=E\left|\frac{1}{N} \sum_{i=1}^{N} \lambda_{i}^{\prime} F_{s} e_{i t}\right|^{p} \leq \frac{1}{N} \sum_{i=1}^{N} E\left|\lambda_{i}^{\prime} F_{s} e_{i t}\right|^{p}=\frac{1}{N} \sum_{i=1}^{N} E\left\|\lambda_{i}\right\|^{p} E\left\|F_{s}\right\|^{p} E\left|e_{i t}\right|^{p} \leq M^{3},
$$

where we have used the independence between the three groups of random variables $\left\{\lambda_{i}\right\},\left\{F_{s}\right\}$ and $\left\{e_{i t}\right\}$ in obtaining the first equality, and the assumptions that $E\left\|\lambda_{i}\right\|^{p} \leq M, E\left\|F_{s}\right\|^{p} \leq M$, and $E\left|e_{i t}\right|^{2 p} \leq M$ in obtaining the last inequality. The term that depends on $\chi_{s t}=\xi_{s t}$ can be dealt with similarly.

Proof of (ii). Note that $\tilde{\Lambda}=\frac{X^{\prime} \tilde{F}}{T}$, which implies that $\tilde{\Lambda}^{\prime}=\frac{\tilde{F}^{\prime} X}{T}$. Since $X=F \Lambda^{\prime}+e$, it follows that

$$
\tilde{\Lambda}^{\prime}=\frac{\tilde{F}^{\prime} F}{T} \Lambda^{\prime}+\frac{\tilde{F}^{\prime} e}{T},
$$

thus implying that $\tilde{\lambda}_{i}=\frac{\tilde{F}^{\prime} F}{T} \lambda_{i}+\frac{\tilde{F}^{\prime} e_{i}}{T}$, where $e_{i}=\left(e_{i 1}, \ldots, e_{i T}\right)^{\prime}$. We can write

$$
\frac{\tilde{F}^{\prime} F H^{\prime}}{T}=\frac{\tilde{F}^{\prime} \tilde{F}}{T}-\frac{\tilde{F}^{\prime}\left(\tilde{F}-F H^{\prime}\right)}{T}=I_{r}-\frac{\tilde{F}^{\prime}\left(\tilde{F}-F H^{\prime}\right)}{T},
$$

from which it follows that

$$
\begin{aligned}
\tilde{\lambda}_{i} & =\frac{\tilde{F}^{\prime} F H^{\prime}}{T} H^{\prime-1} \lambda_{i}+\frac{\tilde{F}^{\prime} e_{i}}{T} \\
& =H^{\prime-1} \lambda_{i}-T^{-1} \tilde{F}^{\prime}\left(\tilde{F}-F H^{\prime}\right) H^{\prime-1} \lambda_{i}+T^{-1}\left(\tilde{F}-F H^{\prime}\right)^{\prime} e_{i}+T^{-1}\left(F H^{\prime}\right)^{\prime} e_{i} .
\end{aligned}
$$

Thus,

$$
\frac{1}{N} \sum_{i=1}^{N}\left\|\tilde{\lambda}_{i}-H^{-1 \prime} \lambda_{i}\right\|^{p} \leq 3^{p-1}\binom{\frac{1}{N} \sum_{i=1}^{N}\left\|T^{-1} \tilde{F}^{\prime}\left(\tilde{F}-F H^{\prime}\right) H^{\prime-1} \lambda_{i}\right\|^{p}}{+\frac{1}{N} \sum_{i=1}^{N}\left\|T^{-1}\left(\tilde{F}-F H^{\prime}\right)^{\prime} e_{i}\right\|^{p}+\frac{1}{N} \sum_{i=1}^{N}\left\|T^{-1}\left(F H^{\prime}\right)^{\prime} e_{i}\right\|^{p}}
$$

For the first term, we have that

$$
\frac{1}{N} \sum_{i=1}^{N}\left\|T^{-1} \tilde{F}^{\prime}\left(\tilde{F}-F H^{\prime}\right) H^{\prime-1} \lambda_{i}\right\|^{p} \leq\left\|T^{-1 / 2} \tilde{F}\right\|^{p}\left\|T^{-1 / 2}\left(\tilde{F}-F H^{\prime}\right)\right\|^{p}\left\|H^{\prime-1}\right\|^{p} \frac{1}{N} \sum_{i=1}^{N}\left\|\lambda_{i}\right\|^{p} .
$$

Now,

$$
\left\|T^{-1 / 2} \tilde{F}\right\|^{p}=\left(T^{-1}\|\tilde{F}\|^{2}\right)^{p / 2}=\left(T^{-1} \sum_{t=1}^{T}\left\|\tilde{F}_{t}\right\|^{2}\right)^{p / 2}=r^{p / 2}
$$

given that $\tilde{F}^{\prime} \tilde{F} / T=I_{r}$. Similarly, under Assumptions A-F,

$$
\left\|T^{-1 / 2}\left(\tilde{F}-F H^{\prime}\right)\right\|^{p}=\left(T^{-1} \sum_{t=1}^{T}\left\|\tilde{F}_{t}-H F_{t}\right\|^{2}\right)^{p / 2}=O_{P}\left(\delta_{N T}^{-p}\right)=O_{P}(1) .
$$

Since $\left\|H^{\prime-1}\right\|^{p}=O_{P}(1)$, it follows that the first term is $O_{P}(1)$ provided $E\left\|\lambda_{i}\right\|^{p} \leq M<\infty$, which holds by assumption.

For the second term,

$$
\frac{1}{N} \sum_{t=1}^{N}\left\|T^{-1}\left(\tilde{F}-F H^{\prime}\right)^{\prime} e_{i}\right\|^{p}=\left\|T^{-1 / 2}\left(\tilde{F}-F H^{\prime}\right)\right\|^{p} \frac{1}{N} \sum_{t=1}^{N}\left\|T^{-1 / 2} e_{i}\right\|^{p},
$$

where the first factor is $O_{P}(1)$ and the second factor is dominated by

$$
\frac{1}{N} \sum_{i=1}^{N}\left(\left\|T^{-1 / 2} e_{i}\right\|^{2}\right)^{p / 2}=\frac{1}{N} \sum_{t=1}^{N}\left(T^{-1} e_{i}^{\prime} e_{i}\right)^{p / 2}=\frac{1}{N} \sum_{t=1}^{N}\left(T^{-1} \sum_{t=1}^{T} e_{i t}^{2}\right)^{p / 2} \leq \frac{1}{N T} \sum_{t=1}^{N} \sum_{t=1}^{T} e_{i t}^{p},
$$

which is $O_{P}(1)$ given the assumption that $E\left|e_{i t}\right|^{p} \leq M$. The third term can be bounded similarly using in particular the fact that $E\left\|F_{t}\right\|^{2} \leq M<\infty$.

Proof of (iii). We can write

$$
\tilde{e}_{i t}=e_{i t}-\lambda_{i}^{\prime} H^{-1}\left(\tilde{F}_{t}-H F_{t}\right)-\left(\tilde{\lambda}_{i}-H^{-1 \prime} \lambda_{i}\right)^{\prime} \tilde{F}_{t},
$$

which implies that

$$
\begin{aligned}
\frac{1}{N T} \sum_{t=1}^{T} \sum_{i=1}^{N}\left|\tilde{e}_{i t}\right|^{p} \leq & 3^{p-1} \frac{1}{N T} \sum_{t=1}^{T} \sum_{i=1}^{N}\left|e_{i t}\right|^{p}+3^{p-1} \frac{1}{N} \sum_{i=1}^{N}\left\|\lambda_{i}\right\|^{p}\left\|H^{-1}\right\|^{p} \frac{1}{T} \sum_{t=1}^{T}\left\|\tilde{F}_{t}-H F_{t}\right\|^{p} \\
& +3^{p-1} \frac{1}{N} \sum_{i=1}^{N}\left\|\tilde{\lambda}_{i}-H^{-1 \prime} \lambda_{i}\right\|^{p} \frac{1}{T} \sum_{t=1}^{T}\left\|\tilde{F}_{t}\right\|^{p}
\end{aligned}
$$

The first term is $O_{P}(1)$ given that $E\left|e_{i t}\right|^{p}=O(1)$; the second term is $O_{P}(1)$ since $E\left\|\lambda_{i}\right\|^{p}=O$ (1) and given part (i); and the third term is $O_{P}(1)$ given parts (ii) and (iii), since in particular

$$
\frac{1}{T} \sum_{t=1}^{T}\left\|\tilde{F}_{t}\right\|^{p} \leq \frac{1}{T} \sum_{t=1}^{T}\left\|H F_{t}+\left(\tilde{F}_{t}-H F_{t}\right)\right\|^{p} \leq 2^{p-1}\left(\|H\|^{p} \frac{1}{T} \sum_{t=1}^{T}\left\|F_{t}\right\|^{p}+\frac{1}{T} \sum_{t=1}^{T}\left\|\tilde{F}_{t}-H F_{t}\right\|^{p}\right)=O_{P}(1)
$$

Proof of Lemma 4.1. To verify the first four parts of Condition $A^{*}$, we apply Lemmas B.1, B. 2 and B. 3 in Appendix B. This entails checking that Conditions B1-B9 in Appendix B hold for the factor-augmented wild bootstrap. We start with Condition B1. For the wild bootstrap, we have that

$$
\gamma_{s t}^{*}=E^{*}\left(\frac{1}{N} \sum_{i=1}^{N} e_{i t}^{*} t_{i s}^{*}\right)=\frac{1}{N} \sum_{i=1}^{N} E^{*}\left(e_{i t}^{*} e_{i s}^{*}\right)=\frac{1}{N} \sum_{i=1}^{N} \tilde{e}_{i t} \tilde{e}_{i s} 1(t=s) .
$$

Thus, condition B1 becomes

$$
\frac{1}{T} \sum_{t=1}^{T}\left(\frac{1}{N} \sum_{i=1}^{N} \tilde{e}_{i t}^{2}\right)^{2}=O_{P}(1)
$$

This expression is bounded by $\frac{1}{T} \sum_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} \tilde{e}_{i t}^{4}$, which is $O_{P}(1)$ under our assumptions by an application of Lemma C. 1 (iii) with $p=4$.

For Condition B2, note that for any $(t, s)$,

$$
\begin{aligned}
E^{*} \left\lvert\, \frac{1}{\sqrt{N}} \sum_{i=1}^{N}\left(e_{i t}^{*} e_{i s}^{*}-\left.E^{*}\left(e_{i t}^{*} e_{i s}^{*}\right)\right|^{2}\right.\right. & =\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} E^{*}\left(\left(e_{i t}^{*} e_{i s}^{*}-E^{*}\left(e_{i t}^{*} e_{i s}^{*}\right)\right)\left(e_{j t}^{*} e_{j s}^{*}-E^{*}\left(e_{j t}^{*} e_{j s}^{*}\right)\right)\right) \\
& =\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \operatorname{Cov}^{*}\left(e_{i t}^{*} e_{i s}^{*}, e_{j t}^{*} e_{j s}^{*}\right) .
\end{aligned}
$$

For the wild bootstrap where $e_{i t}^{*}=\tilde{e}_{i t} \eta_{i t}$, with $\eta_{i t}$ i.i.d. across $(i, t)$,

$$
\operatorname{Cov}^{*}\left(e_{i t}^{*} e_{i s}^{*}, e_{j t}^{*} e_{j s}^{*}\right)=\tilde{e}_{i t} \tilde{e}_{i s} \tilde{e}_{j t} \tilde{e}_{j s} \operatorname{Cov}\left(\eta_{i t} \eta_{i s}, \eta_{j t} \eta_{j s}\right)=\left\{\begin{array}{c}
\tilde{e}_{i t}^{2} \tilde{e}_{i s}^{2} \operatorname{Var}\left(\eta_{i t} \eta_{i s}\right) \quad \text { if } i=j \\
0 \quad \text { if } i \neq j
\end{array},\right.
$$

which implies that

$$
E^{*}\left|\frac{1}{\sqrt{N}} \sum_{i=1}^{N}\left(e_{i t}^{*} e_{i s}^{*}-E^{*}\left(e_{i t}^{*} e_{i s}^{*}\right)\right)\right|^{2}=\frac{1}{N} \sum_{i=1}^{N} \tilde{e}_{i t}^{2} \tilde{e}_{i s}^{2} \operatorname{Var}\left(\eta_{i t} \eta_{i s}\right) .
$$

Thus, condition B2 becomes

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

for some constants $\bar{\eta}$ and $C$.
For Condition B3, we have that

$$
E^{*}\left\|\frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{\sqrt{N}}\right\|^{2}=E^{*}\left\|\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{\lambda}_{i} e_{i t}^{*}\right\|^{2}=\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{\lambda}_{i}^{\prime} \tilde{\lambda}_{j} \underbrace{E^{*}\left(e_{i t}^{*} e_{j t}^{*}\right)}_{=0 \text { for } i \neq j}=\frac{1}{N} \sum_{i=1}^{N} \tilde{\lambda}_{i}^{\prime} \tilde{\lambda}_{i} \tilde{e}_{i t}^{2},
$$

since $\operatorname{Cov}\left(\eta_{i t}, \eta_{j t}\right)=0$ and $\operatorname{Var}\left(\eta_{i t}\right)=1$. Thus, Condition B3 becomes

$$
\begin{aligned}
& \frac{1}{T} \sum_{t=1}^{T} E^{*}\left\|\frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{\sqrt{N}}\right\|^{2}=\frac{1}{T} \sum_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} \tilde{\lambda}_{i}^{\prime} \tilde{\lambda}_{i} \tilde{e}_{i t}^{2}=\frac{1}{N} \sum_{i=1}^{N}\left\|\tilde{\lambda}_{i}\right\|^{2}\left(\frac{1}{T} \sum_{t=1}^{T} \tilde{e}_{i t}^{2}\right) \\
\leq & \left(\frac{1}{N} \sum_{i=1}^{N}\left\|\tilde{\lambda}_{i}\right\|^{4}\right)^{1 / 2}\left(\frac{1}{N} \sum_{i=1}^{N}\left(\frac{1}{T} \sum_{t=1}^{T} \tilde{e}_{i t}^{2}\right)^{2}\right)^{1 / 2} \leq\left(\frac{1}{N} \sum_{i=1}^{N}\left\|\tilde{\lambda}_{i}\right\|^{4}\right)^{1 / 2}\left(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \tilde{e}_{i t}^{4}\right)^{1 / 2}=O_{P}(1),
\end{aligned}
$$

under our assumptions, by an application of Lemma C.1. In particular, note that

$$
\frac{1}{N} \sum_{i=1}^{N}\left\|\tilde{\lambda}_{i}\right\|^{4} \leq 2^{3}\left(\frac{1}{N} \sum_{i=1}^{N}\left\|H^{-1^{\prime}} \lambda_{i}\right\|^{4}+\frac{1}{N} \sum_{i=1}^{N}\left\|\tilde{\lambda}_{i}-H^{-1 \prime} \lambda_{i}\right\|^{4}\right)
$$

and use Lemma C. 1 (ii) with $p=4$ to bound the second term. The first term is bounded by the assumptions on $\left\{\lambda_{i}\right\}$.

For Condition B4, we have that

$$
\begin{aligned}
& \frac{1}{T} \sum_{t=1}^{T} E^{*}\left\|\frac{1}{\sqrt{T N}} \sum_{s=1}^{T} \sum_{i=1}^{N} \tilde{F}_{s}\left(e_{i t}^{*} e_{i s}^{*}-E^{*}\left(e_{i t}^{*} e_{i s}^{*}\right)\right)\right\|^{2}=\frac{1}{T} \sum_{t=1}^{T} \frac{1}{T} E^{*}\left\|\sum_{s=1}^{T} \tilde{F}_{s}\left[\frac{1}{\sqrt{N}} \sum_{i=1}^{N}\left(e_{i t}^{*} e_{i s}^{*}-E^{*}\left(e_{i t}^{*} e_{i s}^{*}\right)\right)\right]\right\|^{2} \\
= & \frac{1}{T} \sum_{t=1}^{T} \frac{1}{T} \sum_{s=1}^{T} \sum_{l=1}^{T} \tilde{F}_{s}^{\prime} \tilde{F}_{l} E^{*}\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N}\left(e_{i t}^{*} e_{i s}^{*}-E^{*}\left(e_{i t}^{*} e_{i s}^{*}\right)\right) \frac{1}{\sqrt{N}} \sum_{j=1}^{N}\left(e_{j t}^{*} e_{j l}^{*}-E^{*}\left(e_{j t}^{*} e_{j l}^{*}\right)\right)\right)
\end{aligned}
$$

where

$$
\begin{aligned}
& E^{*}\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N}\left(e_{i t}^{*} e_{i s}^{*}-E^{*}\left(e_{i t}^{*} e_{i s}^{*}\right)\right) \frac{1}{\sqrt{N}} \sum_{j=1}^{N}\left(e_{j t}^{*} e_{j l}^{*}-E^{*}\left(e_{j t}^{*} e_{j l}^{*}\right)\right)\right) \\
= & \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} E^{*}\left[\left(e_{i t}^{*} e_{i s}^{*}-E^{*}\left(e_{i t}^{*} e_{i s}^{*}\right)\right)\left(e_{j t}^{*} e_{j l}^{*}-E^{*}\left(e_{j t}^{*} e_{j l}^{*}\right)\right)\right]=\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \operatorname{Cov}^{*}\left(e_{i t}^{*} e_{i s}^{*}, e_{j t}^{*} e_{j l}^{*}\right) .
\end{aligned}
$$

Using the properties of the wild bootstrap, in particular the bootstrap cross sectional independence, we can show that
$\operatorname{Cov}^{*}\left(e_{i t}^{*} e_{i s}^{*}, e_{j t}^{*} e_{j l}^{*}\right)=0$ when $i \neq j$ for any $t, s, l$.

When $i=j$,

$$
\operatorname{Cov}^{*}\left(e_{i t}^{*} e_{i s}^{*}, e_{j t}^{*} e_{j l}^{*}\right)=\operatorname{Cov}^{*}\left(e_{i t}^{*} e_{i s}^{*}, e_{i t}^{*} e_{i l}^{*}\right)=\left\{\begin{array}{cc}
0 & \text { if } s \neq l \\
\operatorname{Var}^{*}\left(e_{i t}^{*} e_{i s}^{*}\right)=\tilde{e}_{i t}^{2} \tilde{e}_{i s}^{2} \operatorname{Var}^{*}\left(\eta_{i t} \eta_{i s}\right) & \text { if } s=l .
\end{array}\right.
$$

It follows that

$$
\begin{aligned}
& \frac{1}{T} \sum_{t=1}^{T} E^{*}\left\|\frac{1}{\sqrt{T N}} \sum_{s=1}^{T} \sum_{i=1}^{N} \tilde{F}_{s}\left(e_{i t}^{*} e_{i s}^{*}-E^{*}\left(e_{i t}^{*} e_{i s}^{*}\right)\right)\right\|^{2} \\
= & \frac{1}{T} \sum_{t=1}^{T} \frac{1}{T} \sum_{s=1}^{T} \tilde{F}_{s}^{\prime} \tilde{F}_{s} \frac{1}{N} \sum_{i=1}^{N} \tilde{e}_{i t}^{2} \tilde{e}_{i s}^{2} V a r^{*}\left(\eta_{i t} \eta_{i s}\right) \leq \bar{\eta} \frac{1}{N} \sum_{i=1}^{N}\left(\frac{1}{T} \sum_{t=1}^{T} \tilde{e}_{i t}^{2}\right)\left(\frac{1}{T} \sum_{s=1}^{T} \tilde{F}_{s}^{\prime} \tilde{F}_{s} \tilde{e}_{i s}^{2}\right) \\
= & \bar{\eta} \frac{1}{N} \sum_{i=1}^{N}\left(\frac{1}{T} \sum_{t=1}^{T} \tilde{e}_{i t}^{2}\right)\left(\frac{1}{T} \sum_{s=1}^{T}\left\|\tilde{F}_{s}\right\|^{2} \tilde{e}_{i s}^{2}\right) \leq \bar{\eta}\left[\frac{1}{N} \sum_{i=1}^{N}\left(\frac{1}{T} \sum_{t=1}^{T} \tilde{e}_{i t}^{2}\right)^{2}\right]^{1 / 2}\left[\frac{1}{N} \sum_{i=1}^{N}\left(\frac{1}{T} \sum_{s=1}^{T}\left\|\tilde{F}_{s}\right\|^{2} \tilde{e}_{i s}^{2}\right)^{2}\right]^{1 / 2} \\
\leq & \bar{\eta}\left[\frac{1}{N} \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{e}_{i t}^{4}\right]^{1 / 2}\left[\frac{1}{T} \sum_{s=1}^{T}\left\|\tilde{F}_{s}\right\|^{4} \frac{1}{N} \frac{1}{T} \sum_{i=1}^{N} \sum_{s=1}^{T} \tilde{e}_{i s}^{4}\right]^{1 / 2} .
\end{aligned}
$$

This expression is $O_{P}(1)$ under our assumptions. For Condition B5, under the wild bootstrap, in particular the bootstrap time series independence, we have that

$$
\begin{aligned}
& \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{F}_{s} \hat{z}_{t}^{\prime} \gamma_{s t}^{*}=\frac{1}{T} \sum_{t=1}^{T} \tilde{F}_{t} \hat{z}_{t}^{\prime} \gamma_{t t}^{*}=\frac{1}{T} \sum_{t=1}^{T} \tilde{F}_{t} \hat{z}_{t}^{\prime}\left(\frac{1}{N} \sum_{i=1}^{N} \tilde{e}_{i t}^{2}\right) \\
\leq & \left(\frac{1}{T} \sum_{t=1}^{T}\left\|\tilde{F}_{t} \hat{z}_{t}^{\prime}\right\|^{2}\right)^{1 / 2}\left[\frac{1}{T} \sum_{t=1}^{T}\left(\frac{1}{N} \sum_{i=1}^{N} \tilde{e}_{i t}^{2}\right)^{2}\right]^{1 / 2} \leq\left(\frac{1}{T} \sum_{t=1}^{T}\left\|\tilde{F}_{t}\right\|^{4} \frac{1}{T} \sum_{t=1}^{T}\left\|\hat{z}_{t}\right\|^{4}\right)^{1 / 4}\left[\frac{1}{T} \frac{1}{N} \sum_{t=1}^{T} \sum_{i=1}^{N} \tilde{e}_{i t}^{4}\right]^{1 / 2},
\end{aligned}
$$

which is $O_{P}(1)$ under our assumptions. Next consider Condition B6. We have that

$$
\begin{aligned}
& E^{*}\left\|\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{\sqrt{N}} \hat{z}_{t}^{\prime}\right\|^{2}=\frac{1}{T N} E^{*}\left\|\sum_{t=1}^{T}\left(\sum_{i=1}^{N} \tilde{\lambda}_{i} e_{i t}^{*}\right) \hat{z}_{t}^{\prime}\right\|^{2} \\
&= \frac{1}{T N} E^{*}\left\{\operatorname{tr}\left(\sum_{t=1}^{T} \sum_{s=1}^{T} \hat{z}_{t}\left(\sum_{i=1}^{N} \tilde{\lambda}_{i}^{\prime} e_{i t}^{*}\right)\left(\sum_{j=1}^{N} \tilde{\lambda}_{j} e_{j s}^{*}\right) \hat{z}_{s}^{\prime}\right)\right\} \\
&=\frac{1}{T N} \operatorname{tr}\left(\sum_{t=1}^{T} \sum_{s=1}^{T} \hat{z}_{t}\left(\sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{\lambda}_{i}^{\prime} \tilde{\lambda}_{j} E^{*}\left(e_{i t}^{*} e_{j s}^{*}\right)\right) \hat{z}_{s}^{\prime}\right) \\
&=\frac{1}{T N} \operatorname{tr}\left(\sum_{t=1}^{T} \hat{z}_{t}\left(\sum_{i=1}^{N} \tilde{\lambda}_{i}^{\prime} \tilde{\lambda}_{i} E^{*}\left(e_{i t}^{* 2}\right)\right) \hat{z}_{t}^{\prime}\right)=\frac{1}{T N} \sum_{t=1}^{T} \hat{z}_{t}^{\prime} \hat{z}_{t}\left(\sum_{i=1}^{N} \tilde{\lambda}_{i}^{\prime} \tilde{\lambda}_{i} \tilde{e}_{i t}^{2}\right) \\
&=\frac{1}{T} \sum_{t=1}^{T}\left\|\hat{z}_{t}\right\|^{2} \frac{1}{N} \sum_{i=1}^{N}\left\|\tilde{\lambda}_{i}\right\|^{2} \tilde{e}_{i t}^{2} \leq \underbrace{\left(\frac{1}{T} \sum_{t=1}^{T}\left\|\hat{z}_{t}\right\|^{4}\right)}_{=O_{P}(1) \text { under our assumptions }}{ }^{1 / 2}\left(\frac{1}{T} \sum_{t=1}^{T}\left(\frac{1}{N} \sum_{i=1}^{N}\left\|\tilde{\lambda}_{i}\right\|^{2} \tilde{e}_{i t}^{2}\right)^{2}\right)^{1 / 2} .
\end{aligned}
$$

But by Cauchy-Schwartz,

$$
\frac{1}{T} \sum_{t=1}^{T}\left(\frac{1}{N} \sum_{i=1}^{N}\left\|\tilde{\lambda}_{i}\right\|^{2} \tilde{e}_{i t}^{2}\right)^{2} \leq \underbrace{\frac{1}{N} \sum_{i=1}^{N}\left\|\tilde{\lambda}_{i}\right\|^{4} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} \tilde{e}_{i t}^{4}}_{O_{P}(1)}=O_{P}(1)
$$

under our conditions. For Condition B7, 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. $(0,1)$.
Because $\gamma_{s t}^{*}=0$ for $t \neq s$ and by repeated application of Cauchy-Schwartz inequality, we have that

$$
\begin{aligned}
& \frac{1}{T} \sum_{t=1}^{T-h} \sum_{s=1}^{T} \tilde{F}_{s} \varepsilon_{t+h}^{*} \gamma_{s t}^{*}=\frac{1}{T} \sum_{t=1}^{T} \tilde{F}_{t} \varepsilon_{t+h}^{*} \gamma_{t t}^{*}=\frac{1}{T} \sum_{t=1}^{T-h} \tilde{F}_{t} \varepsilon_{t+h}^{*}\left(\frac{1}{N} \sum_{i=1}^{N} \tilde{e}_{i t}^{2}\right) \\
\leq & \left(\frac{1}{T} \sum_{t=1}^{T-h}\left\|\tilde{F}_{t} \varepsilon_{t+h}^{*}\right\|^{2}\right)^{1 / 2}\left[\frac{1}{T} \sum_{t=1}^{T-h}\left(\frac{1}{N} \sum_{i=1}^{N} \tilde{e}_{i t}^{2}\right)^{2}\right]^{1 / 2} \leq(\underbrace{\frac{1}{T} \sum_{t=1}^{T}\left\|\tilde{F}_{t}\right\|^{4} \frac{1}{T} \sum_{t=1}^{T-h}\left\|\varepsilon_{t+h}^{*}\right\|^{4}}_{O_{P}(1)})_{O_{P^{*}(1)}}^{1 / 4} \underbrace{\left[\frac{1}{T} \frac{1}{N} \sum_{t=1}^{T} \sum_{i=1}^{N} \tilde{e}_{i t}^{4}\right]^{1 / 2}}_{=O_{P}(1)},
\end{aligned}
$$

under our assumptions, provided in particular that $\frac{1}{T} \sum_{t=1}^{T-h} \varepsilon_{t+h}^{* 4}=O_{P^{*}}(1)$ 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 $\varepsilon_{t+h}^{*}$, we have that

$$
\frac{1}{T} \sum_{t=1}^{T-h} E^{*}\left(\varepsilon_{t+h}^{* 4}\right)=\frac{1}{T} \sum_{t=1}^{T-h} \hat{\varepsilon}_{t+h}^{4} \underbrace{E^{*}\left(v_{t+h}^{4}\right)}_{\leq C<\infty \text { by assumption }} \leq \frac{1}{T} \sum_{t=1}^{T-h} \hat{\varepsilon}_{t+h}^{4},
$$

which is verified under our conditions. Next, we verify Condition B8. We have that

$$
\begin{aligned}
& E^{*}\left\|\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{\sqrt{N}} \varepsilon_{t+h}^{*}\right\|^{2} \\
& E^{*}\left\|\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \frac{\tilde{\Lambda}^{\prime} e_{t}^{*}}{\sqrt{N}} \varepsilon_{t+h}^{*}\right\|^{2}=\frac{1}{T N} E^{*}\left\|\sum_{t=1}^{T-h}\left(\sum_{i=1}^{N} \tilde{\lambda}_{i} e_{i t}^{*}\right) \varepsilon_{t+h}^{*}\right\|^{2} \\
&=\frac{1}{T N} E^{*}\left\{\left(\sum_{t=1}^{T-h} \sum_{s=1}^{T-h} \varepsilon_{t+h}^{*} \varepsilon_{s+h}^{*}\left(\sum_{i=1}^{N} \tilde{\lambda}_{i}^{\prime} e_{i t}^{*}\right)\left(\sum_{j=1}^{N} \tilde{\lambda}_{j} e_{j s}^{*}\right)\right)\right\} .
\end{aligned}
$$

By the independence between $\left\{e_{i t}^{*}\right\}$ and $\left\{\varepsilon_{s}^{*}\right\}$, we have that

$$
\begin{aligned}
& =\frac{1}{T N}(\sum_{t=1}^{T-h} \sum_{s=1}^{T-h} \underbrace{E^{*}\left(\varepsilon_{t+h}^{*} \varepsilon_{s+s}^{*}+h\right.}_{=0} \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{\lambda}_{i}^{\prime} \tilde{\lambda}_{j} \underbrace{E^{*}\left(e_{i t}^{*} e_{j s}^{*}\right)}_{=0 \text { if } i \neq j \text { or } t \neq s}) \\
& =\frac{1}{T N}\left(\sum_{t=1}^{T-h} E^{*}\left(\varepsilon_{t+h}^{* 2}\right) \sum_{i=1}^{N} \tilde{\lambda}_{i}^{\prime} \tilde{\lambda}_{i} E^{*}\left(e_{i t}^{*^{2}}\right)\right)=\frac{1}{T} \sum_{t=1}^{T-h} \hat{\varepsilon}_{t+h}^{2} \frac{1}{N} \sum_{i=1}^{N}\left\|\tilde{\lambda}_{i}\right\|^{2} \tilde{e}_{i t}^{2} \\
& \leq\left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{\varepsilon}_{t+h}^{4}\right)^{1 / 2}\left(\frac{1}{T} \sum_{t=1}^{T-h}\left(\frac{1}{N} \sum_{i=1}^{N}\left\|\tilde{\lambda}_{i}\right\|^{2} \tilde{e}_{i t}^{2}\right)^{2}\right)^{1 / 2}=O_{P}(1) \times O_{P}(1)
\end{aligned}
$$

under our assumptions and using the same arguments used above to show condition B6. Finally. we verify Condition B9. We have that

$$
\frac{1}{T} \sum_{t=1}^{T-h} E^{*}\left|\varepsilon_{t+h}^{*}\right|^{2}=\frac{1}{T} \sum_{t=1}^{T-h} \hat{\varepsilon}_{t+h}^{2} \leq\left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{\varepsilon}_{t+h}^{4}\right)^{1 / 2}=O_{P}(1)
$$

under our assumptions.
To conclude the proof, we only need to verify that the bootstrap CLT result (part 5 of Condition $A^{*}$ ) holds for the wild bootstrap. By the Cramer-Wold device, it suffices to show that

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \underbrace{\ell^{\prime}\left(\Phi_{0} \Sigma_{z z . \varepsilon} \Phi_{0}^{\prime}\right)^{-1 / 2} \hat{z}_{t} \varepsilon_{t+h}^{*}}_{\equiv w_{t}^{*}} \rightarrow d^{d^{*}} N(0,1),
$$

in probability for any $\ell$ such that $\ell^{\prime} \ell=1$. Note that $w_{t}^{*}$ is an heterogeneous array of independent random variables (given that $\varepsilon_{t+h}^{*}$ is conditionally independent but heteroskedastic). Thus, we apply a CLT for heterogeneous independent arrays. Note that $E^{*}\left(w_{t}^{*}\right)=0$ and

$$
\begin{aligned}
\operatorname{Var}^{*}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} w_{t}^{*}\right) & =\ell^{\prime}\left(\Phi_{0} \Sigma_{z z . \varepsilon} \Phi_{0}^{\prime}\right)^{-1 / 2} \operatorname{Var}^{*}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \hat{z}_{t} \varepsilon_{t+h}^{*}\right)\left(\Phi_{0} \Sigma_{z z . \varepsilon} \Phi_{0}^{\prime}\right)^{-1 / 2} \ell \\
& =\ell^{\prime}\left(\Phi_{0} \Sigma_{z z . \varepsilon} \Phi_{0}^{\prime}\right)^{-1 / 2}\left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{z}_{t} \hat{z}_{t}^{\prime} \hat{\varepsilon}_{t+h}^{2}\right)\left(\Phi_{0} \Sigma_{z z . \varepsilon} \Phi_{0}^{\prime}\right)^{-1 / 2} \ell .
\end{aligned}
$$

Under our assumptions, Bai and Ng (2006) showed that

$$
\frac{1}{T} \sum_{t=1}^{T-h} \hat{z}_{t} \hat{z}_{t}^{\prime} \hat{\varepsilon}_{t+h}^{2} \rightarrow^{P} \Phi_{0} \Sigma_{z z . \varepsilon} \Phi_{0}^{\prime}
$$

which implies that

$$
\operatorname{Var}^{*}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} w_{t}^{*}\right) \rightarrow^{P} \ell^{\prime}\left(\Phi_{0} \Sigma_{z z . \varepsilon} \Phi_{0}^{\prime}\right)^{-1 / 2}\left(\Phi_{0} \Sigma_{z z . \varepsilon} \Phi_{0}^{\prime}\right)\left(\Phi_{0} \Sigma_{z z . \varepsilon} \Phi_{0}^{\prime}\right)^{-1 / 2} \ell=1
$$

Thus, it suffices to verify Lyapunov's condition, i.e. for some $r>1$,

$$
\frac{1}{T^{r}} \sum_{t=1}^{T-h} E^{*}\left|w_{t}^{*}\right|^{2 r} \rightarrow^{P} 0
$$

Noting that $w_{t}^{*}=\ell^{\prime}\left(\Phi_{0} \Sigma_{z z . \varepsilon} \Phi_{0}^{\prime}\right)^{-1 / 2} \hat{z}_{t} \varepsilon_{t+h}^{*}$, we have that

$$
\begin{aligned}
\frac{1}{T^{r}} \sum_{t=1}^{T-h} E^{*}\left|w_{t}^{*}\right|^{2 r} & =\frac{1}{T^{r-1}} \frac{1}{T} \sum_{t=1}^{T-h}\left|\ell^{\prime}\left(\Phi_{0} \Sigma_{z z . \varepsilon} \Phi_{0}^{\prime}\right)^{-1 / 2} \hat{z}_{t}\right|^{2 r}\left|\hat{\varepsilon}_{t+h}\right|^{2 r} \\
& \leq \frac{1}{T^{r-1}}\|\ell\|^{2 r}\left\|\left(\Phi_{0} \Sigma_{z z . \varepsilon} \Phi_{0}^{\prime}\right)^{-1 / 2}\right\|^{2 r} \frac{1}{T} \sum_{t=1}^{T-h}\left|\hat{z}_{t}\right|^{2 r}\left|\hat{\varepsilon}_{t+h}\right|^{2 r} \underbrace{E^{*}\left|v_{t+r}\right|^{2 r}}_{\leq \Delta<\infty} \\
& \leq C \frac{1}{T^{r-1}}\left(\frac{1}{T} \sum_{t=1}^{T-h}\left|\hat{z}_{t}\right|^{4 r}\right)^{1 / 2}\left(\frac{1}{T} \sum_{t=1}^{T-h}\left|\hat{\varepsilon}_{t+h}\right|^{4 r}\right)^{1 / 2}=O_{P}\left(\frac{1}{T^{r-1}}\right)=o_{P}(1),
\end{aligned}
$$

by an application of Lemma C. 1 .
Proof of Theorem 4.1, The proof follows from Lemma 4.1, given Theorem 3.1 and Corollary 3.1,

## References

[1] Bai, J., 2003. "Inferential theory for factor models of large dimensions," Econometrica, 71, 135172.
[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, 2008. "Large dimensional factor analysis," Foundations and Trends in Econometrics, 3, 89-163.
[6] Chamberlain, G. and M. Rothschild, 1983. "Arbitrage, factor structure and mean-variance analysis in large asset markets," Econometrica, 51, 1305-1324.
[7] Connor, G. and R. A. Korajczyk, 1986. "Performance measurement with the arbitrage pricing theory," Journal of Financial Economics, 15, 373-394.
[8] 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.
[9] Eichengreen, B. A. Mody, M. Nedeljkovic, and L. Sarno, 2009. "How the Subprime Crisis Went Global: Evidence from Bank Credit Default Swap Spreads", NBER working paper 14904.
[10] Gospodinov, N. and S. Ng, 2010. "Commodity prices, convenience yields and inflation," manuscript, Concordia University.
[11] Groen, J. and G. Kapetanios, 2009. "Model Selection Criteria for Factor-Augmented Regressions", Federal Reserve Bank of New York, Staff report 363.
[12] Ludvigson, S. and S. Ng, 2007. "The empirical risk return relation: a factor analysis approach," Journal of Financial Economics, 83, 171-222.
[13] Ludvigson, S. and S. Ng, 2009a. "Macro factors in bond risk premia," Review of Financial Studies, 22, 5027-5067.
[14] Ludvigson, S. and S. Ng, 2009b. "A factor analysis of bond risk premia," Handbook of Applied Econometrics, forthcoming.
[15] Onatski, A., 2009a. "Testing Hypotheses About the Number of Factors in Large Factor Models," Econometrica, 77: 1447-1479.
[16] Onatski, A., 2009b. "Asymptotics of the principal components estimator of large factor models with weak factors," manuscript, Columbia University.
[17] Pagan, A., 1984. "Econometric Issues in the Analysis of Regressions with Generated Regressors," International Economic Review, 25, 221-247.
[18] 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.
[19] Yamamoto, Y., 2009. "Bootstrap inference for impulse response functions in factor-augmented vector autoregressions," manuscript.

Table 1. Coverage rate of confidence intervals, DGP 1 (homo - homo)

| alpha $=(0,0)^{\prime}$ | $N=50$ |  |  |  |  |  | $N=100$ |  |  |  |  |  | $N=200$ |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\mathrm{T}=50$ |  |  | $\mathrm{T}=100$ |  |  | $\mathrm{T}=50$ |  |  | $\mathrm{T}=100$ |  |  | $\mathrm{T}=50$ |  |  | $\mathrm{T}=100$ |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Asymptotic | 93.2 | 91.2 | 91.6 | 94.5 | 94.5 | 93.9 | 94.5 | 91.7 | 92.2 | 94.7 | 92.8 | 94.7 | 93.3 | 90.2 | 91.8 | 93.6 | 92.7 | 92.8 |
| wild, wild | 95.9 | 95.0 | 99.8 | 96.7 | 97.1 | 99.6 | 96.6 | 95.5 | 98.2 | 97.3 | 95.9 | 99.0 | 96.5 | 93.3 | 97.2 | 96.2 | 96.3 | 96.3 |
|  | 96.9 | 96.0 | 99.9 | 97.1 | 97.9 | 99.9 | 97.5 | 97.2 | 99.5 | 97.3 | 96.7 | 99.3 | 97.3 | 95.1 | 99.5 | 96.4 | 97.3 | 98.9 |
| iid, wild | 96.1 | 94.7 | 99.7 | 96.2 | 97.8 | 99.8 | 96.6 | 95.8 | 98.3 | 97.4 | 95.4 | 98.6 | 96.1 | 93.6 | 96.8 | 96.3 | 96.1 | 96.5 |
|  | 96.9 | 95.8 | 100.0 | 97.1 | 98.1 | 99.9 | 97.6 | 97.5 | 99.7 | 97.7 | 96.2 | 99.3 | 97.0 | 95.4 | 99.4 | 96.4 | 96.8 | 98.9 |
| iid, iid | 96.6 | 96.0 | 99.8 | 96.7 | 97.7 | 99.8 | 96.7 | 95.8 | 98.4 | 96.9 | 96.7 | 99.0 | 96.7 | 95.0 | 97.7 | 96.2 | 96.0 | 97.3 |
|  | 97.5 | 96.9 | 100.0 | 97.1 | 98.2 | 99.9 | 98.2 | 97.9 | 99.6 | 98.0 | 97.2 | 99.3 | 97.6 | 96.4 | 99.5 | 96.6 | 97.9 | 98.5 |
| alpha $=(1,1)^{\prime}$ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Asymptotic | 92.3 | 77.2 | 86.3 | 94.0 | 73.7 | 91.6 | 93.2 | 82.1 | 88.4 | 93.5 | 86.7 | 93.6 | 92.7 | 87.9 | 91.3 | 94.4 | 89.9 | 93.7 |
| wild, wild | 95.0 | 93.1 | 97.4 | 97.3 | 94.3 | 98.8 | 96.1 | 93.7 | 92.6 | 95.9 | 95.3 | 98.8 | 95.2 | 94.9 | 96.7 | 97.5 | 96.5 | 97.6 |
|  | 95.5 | 94.8 | 97.5 | 97.9 | 96.1 | 98.9 | 96.8 | 96.2 | 97.4 | 96.0 | 95.7 | 99.0 | 95.6 | 96.4 | 98.8 | 97.2 | 97.3 | 98.7 |
| iid, wild | 95.6 | 93.5 | 97.5 | 96.8 | 94.1 | 98.8 | 96.3 | 93.8 | 92.9 | 96.2 | 95.6 | 98.5 | 95.4 | 95.2 | 96.4 | 97.3 | 96.7 | 97.7 |
|  | 95.5 | 95.6 | 97.5 | 97.7 | 96.5 | 99.1 | 97.0 | 96.3 | 97.3 | 96.2 | 96.3 | 98.8 | 95.7 | 96.3 | 99.2 | 97.6 | 97.3 | 98.8 |
| iid, iid | 95.6 | 93.8 | 97.5 | 97.3 | 94.1 | 98.6 | 96.7 | 94.4 | 94.2 | 96.1 | 95.4 | 99.2 | 95.8 | 95.6 | 97.3 | 97.4 | 96.9 | 97.8 |
|  | 96.2 | 95.5 | 97.8 | 97.6 | 96.0 | 99.0 | 97.7 | 96.5 | 98.0 | 96.2 | 96.2 | 99.1 | 96.8 | 97.3 | 99.1 | 97.8 | 97.7 | 99.1 |

Note: The table reports the coverage rate of $95 \%$ confidence intervals for the constant and the coefficients on the two factors in the factoraugmented regression (12) with the factors replaced by their sample estimates. The first boostrap method is for the idiosyncratic errors, while the second one is for the regression errors. The two rows for each method are for a symmetric percentile interval and a symmetric percentile-t interval respectively.

Table 2. Coverage rate of confidence intervals, DGP 2 (hetero - hetero)

| alpha $=(0,0)^{\prime}$ | $N=50$ |  |  |  |  |  | $N=100$ |  |  |  |  |  | $N=200$ |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\mathrm{T}=50$ |  |  | $\mathrm{T}=100$ |  |  | $T=50$ |  |  | $\mathrm{T}=100$ |  |  | $T=50$ |  |  | $\mathrm{T}=100$ |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Asymptotic | 93.1 | 91.3 | 91.2 | 94.5 | 94.5 | 91.7 | 95.1 | 91.0 | 89.8 | 95.6 | 92.9 | 93.2 | 92.4 | 90.7 | 90.8 | 93.6 | 92.9 | 91.3 |
| wild, wild | 96.4 | 96.7 | 99.9 | 96.5 | 97.6 | 99.7 | 97.6 | 94.3 | 98.4 | 97.1 | 96.5 | 99.5 | 96.0 | 93.8 | 96.1 | 96.2 | 96.0 | 96.3 |
|  | 96.4 | 97.6 | 100.0 | 96.6 | 98.7 | 99.9 | 97.8 | 96.9 | 99.9 | 97.5 | 97.6 | 99.9 | 96.7 | 96.8 | 99.5 | 96.5 | 96.5 | 99.5 |
| iid, wild | 96.8 | 95.8 | 99.6 | 96.6 | 97.8 | 99.6 | 98.1 | 94.7 | 98.2 | 97.5 | 96.1 | 99.3 | 96.2 | 93.3 | 95.4 | 96.7 | 95.2 | 96.5 |
|  | 96.7 | 97.5 | 100.0 | 97.0 | 98.6 | 100.0 | 98.4 | 97.3 | 99.9 | 97.5 | 97.5 | 99.8 | 96.1 | 96.3 | 99.6 | 96.9 | 96.9 | 98.9 |
| alpha $=(1,1)^{\prime}$ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Asymptotic | 93.8 | 73.8 | 83.1 | 94.1 | 70.8 | 88.0 | 91.3 | 84.2 | 89.9 | 93.3 | 83.1 | 90.5 | 94.0 | 87.4 | 89.9 | 94.3 | 91.3 | 92.6 |
| wild, wild | 96.6 | 91.8 | 94.9 | 97.7 | 91.7 | 97.5 | 95.6 | 92.7 | 96.5 | 96.7 | 96.1 | 95.8 | 96.5 | 94.1 | 95.9 | 97.2 | 96.7 | 96.6 |
|  | 97.5 | 94.9 | 96.2 | 97.4 | 95.3 | 98.5 | 95.8 | 96.8 | 99.0 | 96.3 | 97.1 | 97.4 | 96.6 | 96.9 | 99.4 | 97.1 | 97.6 | 98.5 |
| iid, wild | 97.0 | 91.5 | 93.6 | 97.8 | 92.0 | 96.5 | 95.8 | 93.4 | 95.7 | 96.6 | 95.7 | 95.4 | 97.1 | 93.7 | 95.4 | 97.6 | 95.7 | 96.2 |
|  | 96.9 | 94.7 | 95.1 | 97.1 | 95.0 | 98.2 | 95.8 | 96.4 | 98.6 | 96.5 | 96.7 | 96.8 | 97.2 | 97.1 | 99.2 | 97.3 | 97.3 | 98.9 |

Note: $\quad$ See table 1.

Table 3. Coverage rate of confidence intervals, DGP 3 (hetero-AR)

|  | $N=50$ |  |  |  |  |  | $N=100$ |  |  |  |  |  | $N=200$ |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\mathrm{T}=50$ |  |  | $\mathrm{T}=100$ |  |  | $\mathrm{T}=50$ |  |  | $\mathrm{T}=100$ |  |  | $\mathrm{T}=50$ |  |  | $\mathrm{T}=100$ |  |  |
| alpha $=(0,0)^{\prime}$ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Asymptotic | 93.5 | 91.4 | 92.4 | 93.5 | 93.0 | 93.2 | 92.6 | 92.0 | 91.0 | 94.1 | 91.8 | 92.0 | 92.8 | 91.3 | 91.7 | 93.1 | 92.8 | 92.9 |
| wild, wild | 96.4 | 95.7 | 99.8 | 96.7 | 96.8 | 100.0 | 96.2 | 94.5 | 97.7 | 96.6 | 95.1 | 98.7 | 97.0 | 95.1 | 96.3 | 96.2 | 95.2 | 96.7 |
|  | 97.0 | 96.8 | 100.0 | 97.2 | 97.7 | 100.0 | 97.0 | 97.3 | 99.7 | 96.4 | 96.1 | 99.4 | 97.0 | 97.7 | 99.6 | 96.1 | 96.5 | 99.3 |
| iid, wild | 96.6 | 95.7 | 99.9 | 97.0 | 97.3 | 100.0 | 96.3 | 95.2 | 98.1 | 96.5 | 95.1 | 99.3 | 96.6 | 94.7 | 95.9 | 96.4 | 94.7 | 96.5 |
|  | 97.2 | 97.2 | 100.0 | 96.9 | 97.8 | 100.0 | 97.4 | 97.6 | 99.9 | 96.8 | 96.3 | 99.7 | 97.5 | 97.4 | 99.4 | 95.9 | 96.4 | 99.1 |
| alpha $=(1,1)^{\prime}$ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Asymptotic | 89.3 | 47 | 49.9 | 89.5 | 51.8 | 81.4 | 91.5 | 63.4 | 59.4 | 93.6 | 69.3 | 77.6 | 91.7 | 69.6 | 84.2 | 92.7 | 79.3 | 89.3 |
| wild, wild | 95.5 | 67.6 | 59.0 | 96.0 | 80.5 | 91.0 | 96.1 | 81.1 | 63.5 | 96.9 | 86.7 | 84.2 | 95.7 | 83.3 | 88.1 | 96.4 | 88.4 | 92.6 |
|  | 95.3 | 75.5 | 60.7 | 95.1 | 85.9 | 91.8 | 96.2 | 87.5 | 70.2 | 96.6 | 90.5 | 86.1 | 95.8 | 89.6 | 92.3 | 95.9 | 91.6 | 97.0 |
| iid, wild | 95.1 | 69.5 | 60.1 | 95.6 | 81.9 | 91.8 | 95.5 | 81.7 | 63.8 | 96.7 | 87 | 85.2 | 96.0 | 83.9 | 88.4 | 96.4 | 88.9 | 93 |
|  | 94.3 | 76.2 | 61.8 | 94.4 | 87.1 | 92.5 | 95.1 | 88 | 71.3 | 95.9 | 91.5 | 87.2 | 96.8 | 89.2 | 92.7 | 96.1 | 92 | 96.8 |

Note:
See table 1.


[^0]:    *We would like to thank Nikolay Gospodinov and Serena Ng for very useful discussions and comments. Gonçalves acknowledges financial support from the NSERC and MITACS whereas Perron acknowledges financial support from the SSHRCC and MITACS.

