Bootstrap inference in regressions with estimated factors and serial correlation

Antoine Djogbenou, Sílvia Gonçalves and Benoit Perron*
Université de Montréal

February 17, 2014

Abstract

This paper considers bootstrap inference in a factor-augmented regression context where the errors could potentially be serially correlated. This generalizes results in Gonçalves and Perron (2013) and makes the bootstrap applicable to forecasting contexts where the forecast horizon is greater than one. We propose and justify two residual-based approaches, a block wild bootstrap (BWB) and a dependent wild bootstrap (DWB). Our simulations document improvement in coverage rates of confidence intervals for the coefficients when using BWB or DWB relative to both asymptotic theory and the wild bootstrap when serial correlation is present in the regression errors.

Keywords: Factor model, bootstrap, serial correlation, forecast.

1 Introduction

Factor-augmented regressions have become quite popular in research in finance and economics since the seminal paper of Stock and Watson (2002). They are often used in a forecasting context as they allow to summarize a large number of predictors with a small number of indexes.

Because these indices are treated as latent factors in an approximate factor model, the estimated regression contains estimated regressors which poses challenges for inference. Under regularity conditions, Bai and Ng (2006) derived the asymptotic distribution of regression estimates. One of the key conditions used in their work is that $\sqrt{T}/N \to 0$. In that case, the error in estimating the factors can be neglected and inference can proceed as if they were observed.

Gonçalves and Perron (2013) (GP (2013) thereafter) showed that the finite sample properties of the asymptotic approach of Bai and Ng (2006) can be poor, especially if N is not sufficiently large relative to T. In particular, estimation of factors leads to an asymptotic bias term in the OLS estimator if $\sqrt{T}/N \to c$ and $c \neq 0$. They provided a set of high level conditions under which any residual-based bootstrap method is valid in this context and showed that a bootstrap algorithm based on the wild bootstrap removes this bias and outperforms the asymptotic approach of Bai and Ng (2006) in

*We are grateful for comments from seminar participants at the Toulouse School of Economics and Pompeu Fabra University as well as from participants at the Workshop on Bootstrap Methods for Time Series, Copenhagen, Denmark, September 2013, and the MAESG conference in Emory, Atlanta, November 2013. Gonçalves acknowledges financial support from the NSERC and MITACS whereas Perron acknowledges financial support from the SSHRC and MITACS.

simulation experiments. This wild bootstrap algorithm is only valid when the forecasting horizon is one because it does not reproduce serial correlation. In general, when the forecasting horizon is larger than one, the residuals in the factor-augmented regression will follow a moving average process.

In this paper, we extend the work of Bai and Ng (2006) and Gonçalves and Perron (2013) by considering errors that are serially correlated. Bai and Ng effectively ruled out possible serial correlation since their estimator of the asymptotic variance of the scaled average of the scores is only consistent with heteroskedasticity. We begin by providing an asymptotic theory under general assumptions on the serial correlation of the error term (of the strong mixing type) and proposing a consistent estimator of the covariance matrix in that case. As in Gonçalves and Perron (2013), we allow $\sqrt{T}/N \to c > 0$ so that a bias term appears in the asymptotic distribution. Secondly, we propose two residual-based bootstrap schemes and show that they provide valid inference in this context. The first scheme which we call the block wild bootstrap (BWB) was proposed by Yeh (1998) for a linear regression with fixed scalar regressor and strong mixing errors. It is implemented by separating the residuals into non-overlapping blocks of observations and multiplying the elements of each block by the same realization of an external variable. The fact that each element in a block is multiplied by the same external draw generates correlation among the elements within a block but enforces independence across blocks. The second scheme we consider is the dependent wild bootstrap (DWB) originally proposed by Shao (2010) in the context of the smooth function model with time series observations. The DWB differs from the BWB by smoothing the external draws across blocks. Our main contribution is to show that these two methods are valid in the context of a factor augmented regression model with estimated factors and serially correlated errors, characterized by a strong mixing assumption.

The remainder of the paper is organized as follows. Section 2 introduces our assumptions, provides the asymptotic distribution of the OLS estimator, and proposes a consistent estimator of the covariance matrix. Section 3 considers bootstrap inference using our two proposed algorithms. Section 4 presents our simulation experiments, and Section 5 concludes. Mathematical proofs appear in the Appendix.

2 Assumptions and asymptotic results

We consider the following standard factor-augmented regression model,

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

where y_{t+h} denotes the variable of interest, for example GDP growth or inflation, with h the forecast horizon. The vector W_t contains observed regressors (including for instance lags of y_t), while the $r \times 1$ vector F_t consists of *latent* factors which help forecast y_{t+h}. These are thought as common latent factors in a panel factor model given by

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

where λ_i, $i = 1, \ldots, N$, are the $r \times 1$ factor loadings and e_{it} is an idiosyncratic error term, $i = 1, \ldots, N$, $t = 1, \ldots, T$. We will denote the set of regressors as $z_t = \left(F_t', W_t'\right)'$, $t = 1, \ldots, T$.

We impose the following assumptions. Throughout, $\|M\| = \left(trace\left(M'M\right)\right)^{1/2}$ denotes the Euclidean norm, $M > 0$ denotes positive definiteness for a square matrix, and C represents a generic finite constant.

Assumption 1 (factor model)

a) $E\|F_t\|^4 \leq C$ and $\Sigma_F = \lim_{T\to\infty} E\left(\frac{1}{T}F'F\right) = \lim_{T\to\infty} E\left(\frac{1}{T}\sum_{t=1}^{T} F_t F_t'\right) > 0.$

b) $\|\lambda_i\| \leq C$ if λ_i are deterministic, or $E\|\lambda_i\| \leq C$ if not, and $\frac{1}{N}\Lambda'\Lambda = \frac{1}{N}\sum_{i=1}^{N}\lambda_i\lambda_i' \longrightarrow_p \Sigma_\Lambda > 0.$

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

Assumption 2 (Idiosyncratic errors)

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

b) $E\left(e_{it}e_{js}\right) = \sigma_{ij,ts}$, $|\sigma_{ij,ts}| \leq \overline{\sigma}_{ij}$ for all (t, s) and $|\sigma_{ij,ts}| \leq \tau_{st}$ for all (i, j) with $\frac{1}{N}\sum_{i,j=1}^{N}\overline{\sigma}_{ij} \leq C$, $\frac{1}{T}\sum_{t,s=1}^{T}\tau_{st} \leq C$ and $\frac{1}{NT}\sum_{i,j,t,s=1}|\sigma_{ij,ts}| \leq C.$

c) $E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(e_{it}e_{is} - E\left(e_{it}e_{is}\right)\right)\right|^4 \leq C$ for all (t, s).

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

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

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

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

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

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

$$\Gamma_t \equiv Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it}\right).$$

Assumption 4 (Weak dependence betwen ε_{t+h} and e_{it})

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

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

Assumption 5 (Moments and dependence of the score vector) For some $r > 2$,

a) $E\left(z_t\varepsilon_{t+h}\right) = 0$, $E\left\|z_t\right\|^{2r} < C$ and $E\left(\varepsilon_{t+h}^{2r}\right) < C$.

b) $\left\{\left(z_t', \varepsilon_{t+h}\right)\right\}$ is a fourth order stationary strong mixing sequence of size $-\frac{2r}{r-2}$.

c) $\Sigma_{zz} = \lim_{T\longrightarrow\infty} E\left(\frac{1}{T}\sum_{t=1}^{T} z_t z_t'\right) > 0$.

d) $\Omega = \lim_{T\longrightarrow\infty} Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}\right) > 0$.

Assumptions 1-4 are identical to those of Gonçalves and Perron (2013). Assumption 5 contains the fundamental difference with GP (2013). We replace the high level central limit theorem assumption of GP (2013, cf. Assumption 5(c)) by more primitive assumptions that allow us to show consistency of the bootstrap in this context. Specifically, we impose a strong mixing assumption on $\left(z_t', \varepsilon_{t+h}\right)$ and require the existence of slightly more than four finite moments for these random variables (which is a strengthening of the moment conditions used by GP (2013)). Under these assumptions, we can show that a central limit theorem holds for the regression scores (using the latent factors), thus verifying Assumption 5 of GP (2013). Our strong mixing assumption allows for quite general serial dependence, including the class of stationary ARMA processes.

To estimate the factor-augmented regression, it is necessary to use an estimator of the latent factors F_t. It is well known that factor models suffer from a lack of identification. As shown by Bai (2003), the principal component $\widetilde{F}_t$ is only consistent with a rotation of F_t, denoted by HF_t, where H denotes the associated rotation matrix. Bai showed that the rotation matrix H is given by

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

where $\tilde{V}$ is a $r \times r$ diagonal matrix with the r largest eigenvalues of XX'/NT, in decreasing order on the diagonal.

It is useful to rewrite the model as

$$y_{t+h} = \widehat{z}'\delta + \alpha' H^{-1}\left(H \cdot F_t - \widetilde{F}_t\right) + \varepsilon_{t+h},$$

where $\delta' = \left(\alpha' H^{-1}\ \beta'\right)$ and $\widehat{z}_t' = \left(\widetilde{F}_t', W_t'\right)$. The consequence of the lack of identification of the factor model is that the coefficients associated with the estimated factors are rotated versions of those associated with the true latent factors. Bai and Ng (2013) provide three sets of conditions under which $H_0 = p\lim H = diag\left(\pm 1\right)$. Under those conditions, α will be identified up to sign.

The OLS estimator from regressing y_{t+h} on $\tilde{F}_t$ and W_t is given by

$$\widehat{\delta} = \left(\widehat{\alpha}', \widehat{\beta}'\right)' = \left(\sum_{t=1}^{T-h} \widehat{z}_t \widehat{z}_t'\right)^{-1} \sum_{t=1}^{T-h} \widehat{z}_t y_{t+h}.$$

and it will be such that $\hat{\delta} \to^P \delta \equiv \left(\ \alpha' H^{-1}\quad \beta'\ \right)'$ under our assumptions. We denote $\Phi_0 = diag\left(H_0, I\right)$. The following theorem provides the asymptotic distribution of the OLS estimator. The proof is in the Appendix.

Theorem 2.1 *Under Assumptions 1-5, if* $\dfrac{\sqrt{T}}{N} \longrightarrow c < \infty$, *as* $N, T \to \infty$, *then*

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

with

$$\Sigma_\delta = \Phi_0'^{-1} \Sigma_{zz}^{-1} \Omega \Sigma_{zz}^{-1} \Phi_0^{-1},$$

and

$$\Delta_\delta = \begin{pmatrix} \Delta_\alpha \\ \Delta_\beta \end{pmatrix} = \left(\Phi_0 \Sigma_{zz} \Phi_0'\right)^{-1} \begin{pmatrix} \Sigma_{\widetilde{F}} + V\Sigma_{\widetilde{F}} V \\ \Sigma_{W\widetilde{F}} V \Sigma_{\widetilde{F}} V^{-1} \end{pmatrix} \left(H_0^{-1}\right)' \alpha$$

where $\Sigma_{W\widetilde{F}} = p\lim\left(\dfrac{W'\widetilde{F}}{T}\right)$, $\Sigma_{\widetilde{F}} = V^{-1} Q \Gamma Q' V^{-1}$, $Q = p\lim \dfrac{\widetilde{F}F}{T}$, *and* $V = p\lim \widetilde{V}$.

Theorem 2.1 follows from Theorem 2.1 of GP (2013), where the asymptotic normality of the OLS estimator was obtained under a high level CLT assumption on the regression scores. Instead, here we allow dependence of unknown form by assuming a mixing condition on the regressors and on the regression errors. This primitive condition will be useful to establish the consistency of the BWB and DWB in Section 3, as well as the consistency of a HAC estimator of Ω, as we prove next. Note that under this mixing condition, Ω is not necessarily of the form $\Omega = E\left(z_t z_t' \varepsilon_{t+h}^2\right)$ assumed by Bai and Ng (2006).

To carry out inference or construct prediction intervals, a consistent covariance estimator of Σ_δ is required. As we allow for serial correlation in the score, a HAC estimator of Σ_δ is appropriate,

$$\widehat{\Sigma}_\delta = \left(\frac{1}{T}\widehat{z}'\widehat{z}\right)^{-1} \widehat{\Omega} \left(\frac{1}{T}\widehat{z}'\widehat{z}\right)^{-1}$$

with

$$\widehat{\Omega} = \hat{\Xi}_0 + \sum_{j=1}^{T-h-1} k\left(\frac{j}{M_T}\right)\left[\hat{\Xi}_j + \hat{\Xi}_j'\right],$$

where $\hat{\Xi}_j = \frac{1}{T}\sum_{t=1}^{T-h-j} \widehat{z}_t \widehat{z}_{t+j}' \widehat{\varepsilon}_{t+h} \widehat{\varepsilon}_{t+h+j}$ is the autocovariance matrix of the scores, $k(\cdot)$ is a kernel function, and M_T is a bandwidth.

To prove consistency of this estimator, restrictions must be placed on the kernel function $k(\cdot)$ and bandwidth M_T. We will consider kernels in the family $\mathcal{K}_1$ as in Andrews and Monahan (1992):

$$\mathcal{K}_1 = \left\{ \begin{array}{l} k(\cdot) : \mathbb{R} \longrightarrow [-1,1],\ k(0) = 1,\ k(x) = k(-x) \text{ for } x \in \mathbb{R}, \\ \int_{-\infty}^{+\infty} |k(x)|\, dx < \infty, \\ k(\cdot) \text{ is continuous at 0 and at all but a finite number of points.} \end{array} \right\}.$$

In addition, we must strengthen Assumptions 3 and 5. Assumption 3.d) is replaced by:

Assumption 3'

d) $E\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_t\right\|^4\right) \leq C$ where $E(\lambda_i e_{it}) = 0$ for all $(i,\ t)$.

while Assumptions 5.a) and b) are replaced by:

Assumption 5' For some $r > 2$,

a) $E(z_t \varepsilon_{t+h}) = 0$, $E\|z_t\|^{4r} < C$ and $E\left(\varepsilon_{t+h}^{4r}\right) < C$.

b) $\left\{\left(z_t',\ \varepsilon_{t+h}\right)\right\}$ is a fourth order stationary strong mixing sequence of size $-\frac{3r}{r-2}$.

The other parts of these two assumptions remain as before. By strengthening Assumption 5.a) by Assumption 5'.a) we have that $E\|z_t \varepsilon_{t+h}\|^{2r} < C$, which is sufficient for the proof of our next result. Assumption 5' is analogous to the assumptions made in Andrews (1991) to prove consistency of the HAC estimator.

Lemma 2.1 *Suppose that Assumptions 1-5, with Assumptions 3 and 5 strengthened by Assumptions 3' and 5' respectively, hold. Suppose further that $k(\cdot)$ belongs to the set $\mathcal{K}_1$ and that $M_T \longrightarrow \infty$ as $T \to \infty$ such that $\frac{M_T{}^2}{T} \to 0$. If $\frac{\sqrt{T}}{N} \to c < \infty$ as $N, T \to \infty$, $\widehat{\Sigma}_\delta \to^p \Sigma_\delta$.*

This lemma demonstrates that a HAC covariance estimator is consistent for Σ_δ despite the presence of estimated regressors. This implies that, as in Bai and Ng (2006), asymptotic inference can be carried out as if the factors were observed if $\sqrt{T}/N \to 0$ since in that case, the asymptotic distribution of $\sqrt{T}\left(\hat{\delta} - \delta\right)$ is centered at 0. If $\sqrt{T}/N \to c > 0$, Lemma 2.1 shows that HAC estimation is still possible, but inference is complicated by the need to account for the bias term in the asymptotic distribution. As in GP (2013), we consider the bootstrap to accomplish this in the next section.

3 Bootstrap inference

3.1 General residual-based bootstrap: review

In this section, we consider bootstrap inference on the coefficients of the factor-augmented regression. The proposed bootstrap scheme resamples the idiosyncratic and regression residuals separately and is similar to the one in Gonçalves and Perron (2013) with the difference that in the second step, residuals $\{\widehat{\varepsilon}_{t+h}\}$ are resampled by either the block wild bootstrap or the dependent wild bootstrap. As usual, we will denote with asterisks quantities in the bootstrap world. We will also denote by E^* (and Var^*) the expectation (and variance) under the bootstrap measure P^*.

Bootstrap algorithm

1. For $t = 1, \ldots, T$, generate
$$X_t^* = \tilde{\Lambda}\tilde{F}_t + e_t^*,$$
where $\{e_{it}^*\}$ is a resampled version of $\left\{\tilde{e}_{it} = X_{it} - \tilde{\lambda}_i'\tilde{F}_t\right\}$. In this step, we use the wild bootstrap and set
$$e_{it}^* = \widetilde{e}_{it} \cdot \eta_{it}, i = 1, .., N, t = 1, ..., T$$
where η_{it} is a draw from an external random variable that is i.i.d. with mean 0 and variance 1 over i and t.

2. Estimate the bootstrap factors $\left\{\tilde{F}_t^* : t = 1, \ldots, T\right\}$ by principal components using X^*.

3. For $t = 1, \ldots, T-h$, generate
$$y_{t+h}^* = \hat{\alpha}'\tilde{F}_t + \hat{\beta}'W_t + \varepsilon_{t+h}^*,$$
where the error term ε_{t+h}^* is a resampled version of $\hat{\varepsilon}_{t+h}$. In this step, we will use either the block wild bootstrap or the dependent wild bootstrap as detailed below to accommodate serial correlation in ε_{t+h}.

4. Regress y_{t+h}^* generated in step 3 on the bootstrap estimated factors $\tilde{F}_t^*$ obtained in step 2 and on the observed regressors W_t and obtain the OLS estimator $\hat{\delta}^*$:
$$\widehat{\delta}^* = \left(\sum_{t=1}^{T-h} \widehat{z}_t^* \widehat{z}_t^{*\prime}\right)^{-1} \sum_{t=1}^{T-h} \widehat{z}_t^* y_{t+h}^*$$
where $\hat{z}_t^* = \left(\tilde{F}_t^{*\prime}, W_t'\right)'$.

5. Repeat steps 1-4 B times.

As in the sample, the principal component estimator in the bootstrap consistently estimates the space of factors only. The specific rotation that is estimated is given by the bootstrap analogue of the

H matrix:

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

where $\tilde{V}^*$ is the $r \times r$ diagonal matrix containing on the main diagonal the r largest eigenvalues of $X^* X^{*\prime}/NT$, in decreasing order. Note that contrary to H, which depends on unknown population parameters, H^* is fully observed. Using the results in Bai and Ng (2013) , H^* converges asymptotically to a diagonal matrix with $+1$ or -1 on the main diagonal, see Gonçalves and Perron (2013) for more details.

The consequence of this lack of identification is that the bootstrap OLS estimator estimates $\delta^* = \left(\widehat{\alpha}' H^{*-1} \ \widehat{\beta}' \right)' = \left(\Phi^{*-1} \right)' \widehat{\delta}$ which is different from $\hat{\delta}$. Gonçalves and Perron (2013) suggested using a rotated version of this estimator, $\tilde{\delta} = \Phi^{*\prime} \widehat{\delta}^*$ for bootstrap inference, and we will do the same here.

The next assumption is a modified version of Assumptions 6-8 in GP(2013) applied to our context.

Assumption 6

a) λ_i are either deterministic such that $\|\lambda_i\| \leq C < \infty$, or stochastic such that $E \|\lambda_i\|^{12} \leq C < \infty$ for all i, and $E \|F_t\|^{12} \leq C < \infty$.

b) $E |e_{it}|^{12} \leq C < \infty$, for all (i, t) and $E (e_{it} e_{js}) = 0$, if $i \neq j$.

c) z_t and ε_{t+h} are independent of e_{is} for all (i, t, s).

Assumption 6.b) excludes cross-sectional dependence among idiosyncratic errors as in Assumption 8 of Gonçalves and Perron (2013). This is required because we use the wild bootstrap in step 1 of the bootstrap algorithm which destroys such dependence. We could relax this assumption if we were willing to assume that $\sqrt{T}/N \to 0$ as in Bai and Ng (2006). In that case, the bias term of the OLS estimator is 0, and this is the only quantity that depends on the properties of the idiosyncratic errors asymptotically. In that situation, factor estimation error does not matter asymptotically, and the key condition for bootstrap validity is to replicate the properties of the regression errors ε_{t+h}, as we are doing here with our two proposed blocking methods.

We now consider the two bootstrap schemes to generate ε^*_{t+h} in step 3 of this algorithm.

3.2 Block wild bootstrap

The first scheme we consider is the block wild bootstrap (BWB) first proposed by Yeh (1998) and analyzed in other contexts by Shao (2011) and Urbain and Smeekes (2013).

First, we form non-overlapping blocks of size b of consecutive residuals. For simplicity, we assume that $(T - h)/b = k$ where k is an integer and denotes the number of blocks of size b. For $l = 1, \ldots, b$ and $j = 1, \ldots, k$, we let

$$y^*_{(j-1)b+l+h} = \widehat{\alpha}' \widetilde{F}_{(j-1)b+l} + \hat{\beta}' W_{(j-1)b+l} + \varepsilon^*_{(j-1)b+l+h}, \tag{2}$$

where

$$\varepsilon^*_{(j-1)b+l+h} = \widehat{\varepsilon}_{(j-1)b+l+h} \cdot \nu_j$$

and ν_j is an external random variable with mean 0, variance 1, and independent and identically distributed across blocks. In other words, the bootstrap data is obtained by multiplying each residual by an external variable that is the same for all observations within a block. The BWB can also be represented as:

$$\text{First block: } \begin{pmatrix} \varepsilon^*_{1+h} \\ \vdots \\ \varepsilon^*_{b+h} \end{pmatrix} = \begin{pmatrix} \hat{\varepsilon}_{1+h} \\ \vdots \\ \hat{\varepsilon}_{b+h} \end{pmatrix} \nu_1,$$

$$\text{Second block: } \begin{pmatrix} \varepsilon^*_{b+1+h} \\ \vdots \\ \varepsilon^*_{2b+h} \end{pmatrix} = \begin{pmatrix} \hat{\varepsilon}_{b+1+h} \\ \vdots \\ \hat{\varepsilon}_{2b+h} \end{pmatrix} \nu_2,$$

$$\vdots$$

$$\text{Last block: } \begin{pmatrix} \varepsilon^*_{T-b+1} \\ \vdots \\ \varepsilon^*_{T} \end{pmatrix} = \begin{pmatrix} \hat{\varepsilon}_{T-b+1} \\ \vdots \\ \hat{\varepsilon}_{T} \end{pmatrix} \nu_k.$$

Once we have drawn a bootstrap sample $\left\{\varepsilon^*_{t+h}\right\}_{t=1}^{T-h}$, we can obtain observations on y^*_{t+h} through (2) and compute the OLS estimator $\hat{\delta}^*$. The next theorem shows the consistency of the bootstrap based on the rotated version of this estimator, $\Phi^{*\prime}\widehat{\delta}^*$.

Theorem 3.1 *Under the same assumptions as in Lemma 2.1, assuming* $E^* \left|\eta_{it}\right|^4 \leq C < \infty$, $i = 1, \ldots, N$, $t = 1, \ldots, T$, $E^* \left|\nu_j\right|^{4q} \leq C < \infty$, $j = 1, \ldots, k$, *for some* $q > 1$, *if* $\frac{\sqrt{T}}{N} \longrightarrow c < \infty$ *and* $b_T \longrightarrow \infty$ *such that* $\frac{b_T^2}{T} \longrightarrow 0$, *as* $N, T \longrightarrow \infty$, *then*

$$\sup_{x \in \mathbb{R}^{\dim(\delta)}} \left| P^* \left(\sqrt{T} \left(\Phi^{*\prime}\widehat{\delta}^* - \hat{\delta} \right) \leq x \right) - P \left(\sqrt{T} \left(\hat{\delta} - \delta \right) \leq x \right) \right| \to^P 0.$$

3.3 Dependent wild bootstrap

In this section, we consider the dependent wild bootstrap as an alternative to the block wild bootstrap. The dependent wild bootstrap was proposed by Shao (2010) and differs from the BWB by the fact that the draws of the external variable are smoothed across observations.

The DWB is implemented by multiplying each residual by a variable which is a local weighted average of external draws. The local weighting makes neighboring observations dependent, and this explains why it is valid under serial correlation. In other words, the DWB observations are obtained as:

$$\varepsilon^*_{t+h} = \widehat{\varepsilon}_{t+h} \cdot w^*_{t+h},$$

where w_{t+h}^* is the typical element of a vector w^* of length $T-h$ of random draws with mean 0 and covariance matrix K, with typical element $K_{ij} = E^*\left(w_i^* \cdot w_j^*\right) = k_{dwb}\left(\frac{j-i}{l_T}\right)$, with $k_{dwb}(\cdot)$ a kernel function and l_T a bandwidth parameter. Following Shao (2010), we assume that w^* is l_T-dependent. In our simulations, we set $w^* = K^{1/2}w$, where $w \sim N(0, I_{T-h})$. Because the choices of kernel and bandwidth used to construct the DWB observations do not need to coincide with the choices of kernel and bandwidth used to construct the HAC estimator $\hat{\Omega}$, we use different notations here.

We make the same assumptions as for the block wild bootstrap with the addition of the following restriction on the class of kernels.

Assumption 7 $k_{dwb} : \mathbb{R} \to [0,1]$ is symmetric with compact support on $[-1,1]$, $k_{dwb}(0) = 1$, $\lim_{x\to 0}\{1 - k_{dwb}(x)\}/|x|^q \neq 0$ for some $q \in (0,\ 1]$ such that $\psi(\xi) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} k_{dwb}(x)e^{i\xi x}dx \geq 0$ for all $\xi \in \mathbb{R}$.

The condition $\psi(\xi) \geq 0$ ensures that the matrix K is positive definite (see Shao (2010)). These assumptions are satisfied by the Bartlett and Parzen kernels but not for the truncated, quadratic spectral and the Tukey-Hanning kernels (see Andrews (1991), Davidson and De Jong (2000) and Shao (2010)).

The following theorem justifies the dependent wild bootstrap for inference on δ.

Theorem 3.2 *Under the same assumptions as in Lemma 2.1 and Assumption 7, and assuming* $E^*|\eta_{it}|^4 \leq C < \infty$, $E^*|w_t^*|^{2r} \leq C < \infty$, $t = 1, \ldots, T-h$, *for some* $r > 2$, *if* $\frac{\sqrt{T}}{N} \longrightarrow c < \infty$ *and* $l_T \longrightarrow \infty$ *such that* $T^{-1}l_T^{2(r+1)/r} \longrightarrow 0$, *as* $N, T \longrightarrow \infty$, *then*

$$\sup_{x\in\mathbb{R}^{\dim(\delta)}}\left|P^*\left(\sqrt{T}\left(\Phi^{*\prime}\widehat{\delta}^* - \hat{\delta}\right) \leq x\right) - P\left(\sqrt{T}\left(\hat{\delta} - \delta\right) \leq x\right)\right| \to^P 0.$$

This result is the DWB analog of Theorem 3.1 for the BWB. Both theorems allow us to use these two methods for constructing percentile confidence intervals using the bootstrap. In order to construct percentile-t intervals (Hall, 1992), we need a consistent estimator of the variance of $\sqrt{T}\left(\Phi^{*\prime}\widehat{\delta}^* - \hat{\delta}\right)$ to define studentized statistics. This estimator is given by $\Phi^{*\prime}\widehat{\Sigma}_\delta^*\Phi^*$, where

$$\widehat{\Sigma}_\delta^* = \left(\frac{1}{T}\widehat{z}^{*\prime}\widehat{z}^*\right)^{-1}\widehat{\Omega}^*\left(\frac{1}{T}\widehat{z}^{*\prime}\widehat{z}^*\right)^{-1}$$

with $\widehat{\Omega}^*$ being a HAC estimator

$$\widehat{\Omega}^* = \hat{\Xi}_0^* + \sum_{j=1}^{T-h} k^*\left(\frac{j}{M_T^*}\right)\left[\hat{\Xi}_j^* + \hat{\Xi}_j^{*\prime}\right]$$

where $k^*(\cdot)$ and M_T^* denote the kernel function and the bandwidth parameter used in the bootstrap HAC estimator and $\hat{\Xi}_j^* = \frac{1}{T}\sum_{t=1}^{T-h-j}\widehat{z}_t^*\widehat{z}_{t+j}^{*\prime}\widehat{\varepsilon}_{t+h}^*\widehat{\varepsilon}_{t+h+j}^*$.

The consistency of $\widehat{\Sigma}_\delta^*$ is formalized in the next lemma.

Lemma 3.1 *Suppose the assumptions of Theorems 3.1 and 3.2 hold for the DWB and the BWB, respectively. Let* $k^*(\cdot)$ *belong to the set* $\mathcal{K}_1$ *and* $M_T^* \longrightarrow \infty$ *as* $T \to \infty$ *such that* $\frac{M_T^{*2}}{T} \to 0$. *If* $\frac{\sqrt{T}}{N} \to c < \infty$ *as* $N, T \to \infty$,

$$\widehat{\Sigma}_\delta^* = \left(\frac{1}{T}\widehat{z}^{*\prime}\widehat{z}^*\right)^{-1}\widehat{\Omega}^*\left(\frac{1}{T}\widehat{z}^{*\prime}\widehat{z}^*\right)^{-1} \to^{P^*} \Sigma_\delta^* = \left(\Phi_0^{*\prime}\right)^{-1}\Sigma_\delta\left(\Phi_0^*\right)^{-1},$$

in probability.

This result implies that the distribution of studentized statistic for the j^{th} coefficient $\frac{\sqrt{T}\left(\left(\Phi^{*\prime}\widehat{\delta}^*\right)_j - \hat{\delta}_j\right)}{\sqrt{\left(\Phi^{*\prime}\hat{\Sigma}_\delta^*\Phi^*\right)_{jj}}}$ approximates the distribution of $\frac{\sqrt{T}\left(\widehat{\delta}_j - \delta_j\right)}{\sqrt{\left(\hat{\Sigma}_\delta\right)_{jj}}}$. These can be used to form symmetric or equal-tailed percentile-t confidence intervals.

4 Simulation results

In this section, we report results of a simulation experiment to document the properties of the bootstrap inference procedures above. Our design follows Gonçalves, Perron, and Djogbenou (2013) closely. We consider a single factor model,

$$y_{t+h} = \alpha F_t + \varepsilon_{t+h},$$

where F_t is an AR(1) process:

$$F_t = .8F_{t-1} + u_t,$$

with u_t drawn for a normal distribution with mean 0 and variance $1 - .8^2$ independently over time.

We consider three possibilities for the error term ε_{t+h}. In the first two designs, we set $h = 1$ or 12 and let the error term follow an MA$(h-1)$ as is appropriate if the forecasting model is correctly specified. In each case, the MA process is (as in Cheng and Hansen, 2013):

$$\varepsilon_{t+h} = \sum_{j=0}^{h-1} .8^j v_{t+h-j},$$

and v_t is drawn $N\left(0, \left(\frac{1}{\sum_{j=0}^{h-1} .8^{2j}}\right)^2\right)$ so that ε_{t+h} has variance 1.

Finally, in the last design, we set $h = 1$ and generate ε_{t+h} from an AR(1) process:

$$\varepsilon_{t+h} = .8\varepsilon_{t+h-1} + v_{t+h},$$

with v_{t+h} drawn for a normal with expectation 0 and variance $\left(1 - .8^2\right)$. This design is plausible for cases where the forecasting model is dynamically misspecified.

As in Gonçalves, Perron, and Djogbenou (2013), the $(T \times N)$ matrix of panel variables is generated as:

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

where λ_i is drawn from a $U[0,1]$ distribution (independent across i) and e_{it} is heteroskedastic but independent over i and t. The variance of e_{it} is drawn from $U[.5, 1.5]$ for each i.

We consider asymptotic and bootstrap confidence intervals at a nominal level of 95% for the regression coefficient. Asymptotic inference is conducted using a HAC estimator with a quadratic spectral kernel and with bandwidth selected by the data-based rule from Andrews (1991), both in the original sample and in the bootstrap samples. We consider three bootstrap schemes for generating ε^*_{t+h} in step 3 of our algorithm: the wild bootstrap, the block wild bootstrap with block size equal to the integer part of the bandwidth choice in the sample, and the dependent wild bootstrap with Bartlett kernel and bandwidth equal to the one selected in the sample.

We consider two values for each of N and T, 50 and 100, so that we have a total of four sample sizes. For all our bootstrap schemes, we let $\eta_{it} \sim N(0,1)$. Moreover, for the BWB, we let $\nu_j \sim N(0,1)$ whereas we let $w^* = K^{1/2} w$, with $w \sim N(0, I_{T-h})$ for the DWB. We set the number of replications to 5,000 and the number of bootstrap to 399.

Table 1 reports our simulation results. We report coverage rates of confidence intervals, the bias of the estimators, the length of the confidence intervals, and the bandwidth choices made in the sample and in the bootstrap.

The first set of results are coverage rates of the confidence intervals. We report results for the OLS estimator, the OLS estimator if we did not have to estimate the factors, and six bootstrap intervals. We report coverage rates of symmetric-t and equal-tailed-t intervals for the wild bootstrap (WB), the block wild bootstrap (BWB) and dependent wild bootstrap (DWB). Remember that the wild bootstrap is not valid with serial correlation.

The results for the first DGP are similar to those of Gonçalves and Perron (2013). The OLS estimator suffers from severe undercoverage. These distortions come from the presence of a bias associated with the estimation of the factor. This is illustrated in two ways: first, the OLS estimator with the true factor has coverage much closer to the nominal level, and second, the bias results show that the OLS estimator is biased (downward) when the factor must be estimated (and this bias goes down with N and T), while the estimator is essentially unbiased when we use the true factor.

The bootstrap is successful in removing this bias and providing more reliable inference. Whereas coverage is only 57% with $N = T = 50$ for asymptotic theory, symmetric bootstrap intervals have a coverage rate of about 87% and equal-tailed intervals about 89%. As N and T increase, coverage rates approach their nominal levels. With this design, all three bootstrap methods are asymptotically valid, and we see only small differences among them.

It is interesting to note that the equal-tailed intervals are much shorter than the symmetric intervals. This is because the sampling distribution of the OLS estimator is shifted to the left, and imposing symmetry around 0 is inappropriate in this case and entails a cost We also see that the equal-tailed intervals provide slightly better coverage than the symmetric ones.

Many of the same features are reproduced in the other two designs. The OLS estimator is still biased due to the estimation of the factor, but the effect on coverage is not as dramatic as the bias of the estimator is unaffected by the change in design but its variance increases. Thus, the t-statistic is less shifted to the left than in the first design, and the overall effect is that coverage improves. We do see the effect of serial correlation on the deterioration of inference for the OLS estimator with true factor.

In the last two designs, we see differences among bootstrap methods. The wild bootstrap does not reproduce serial correlation and leads to intervals with lower coverage rates with equal-tailed intervals. On the other hand, we see little difference with the symmetric-t intervals. The fact that the wild bootstrap does not reproduce serial correlation is highlighted by the selected bandwidths. The selected bandwidth in the wild bootstrap is similar to the selected bandwidth when the data was i.i.d in the first design. The selected bandwidth in the BWB and DWB are lower than in the sample but large enough to capture some of the serial correlation in the bootstrap errors.

Moreover, the dependent wild bootstrap provides slightly better coverage than the BWB. However, contrary to the first design, the symmetric intervals provide much better coverage than the equal-tailed intervals. This is due to the fact that the bias is less important in these designs than in the first one relative to the variance. Nevertheless, the equal-tailed intervals are much shorter than the symmetric ones.

Conclusion

In this paper, we theoretically justify two bootstrap methods for inference on the coefficients in factor-augmented regressions with serial correlation. Serial correlation naturally arises in a multi-step forecasting context or in a forecasting model that is dynamically misspecified. Our proposed bootstrap algorithm resamples the idiosyncratic errors with the wild bootstrap and the regression errors with either the block wild bootstrap or dependent wild bootstrap. Both methods are proved to provide valid inference under strong mixing dependence despite factor estimation error.

The results in this paper can be used to construct valid prediction intervals for the conditional mean or the realization of the variable of interest h periods into the future. This extension of the current results is explored in a recent paper by Gonçalves, Perron, and Djogbenou (2013).

A Appendix 1: Proofs of results in Sections 2 and 3

Throughout this appendix, we let $\delta_{NT} = \min\left[\sqrt{N}, \sqrt{T}\right]$. We first state an auxiliary result which strengthens the results in Lemma A.1 in Bai (2003) and Theorem 1 in Bai and Ng (2002), followed by its proof. We then prove Theorems 2.1, 3.1 and 3.2 and Lemmas 2.1 and 3.1.

Lemma A.1 *Under Assumptions 1, 2, 3 and 4.a) strengthened by Assumption 3'.d), as* $\frac{\sqrt{T}}{N} \longrightarrow c < \infty$, *when* $N, T \longrightarrow \infty$, $\sum_{t=1}^{T} \left\| \widetilde{F}_t - HF_t \right\|^4 = O_p(1)$.

Proof of Lemma A.1. We have the following identity

$$\widetilde{F}_t - HF_t = \widetilde{V}^{-1}\left(\underbrace{\frac{1}{T}\sum_{s=1}^{T}\widetilde{F}_s\gamma_{st}}_{a_t} + \underbrace{\frac{1}{T}\sum_{s=1}^{T}\widetilde{F}_s\zeta_{st}}_{b_t} + \underbrace{\frac{1}{T}\sum_{s=1}^{T}\widetilde{F}_s\eta_{st}}_{c_t} + \underbrace{\frac{1}{T}\sum_{s=1}^{T}\widetilde{F}_s\xi_{st}}_{d_t} \right),$$

where $\gamma_{st} = E\left(\frac{1}{N}\sum_{i=1}^{N} e_{is}e_{it}\right)$, $\zeta_{st} = \frac{1}{N}\sum_{i=1}^{N}\left(e_{is}e_{it} - E\left(\frac{1}{N}\sum_{i=1}^{N} e_{is}e_{it}\right)\right)$, $\eta_{st} = \frac{1}{N}\sum_{i=1}^{N}\lambda_i' F_s e_{it}$, and $\xi_{st} = \frac{1}{N}\sum_{i=1}^{N}\lambda_i' F_t e_{is}$. By the c-r inequality, it follows that

$$\sum_{t=1}^{T}\left\|\widetilde{F}_t - HF_t\right\|^4 \leq 4^3 \left\|\widetilde{V}^{-1}\right\|^4 \left(\sum_{t=1}^{T}\|a_t\|^4 + \sum_{t=1}^{T}\|b_t\|^4 + \sum_{t=1}^{T}\|c_t\|^4 + \sum_{t=1}^{T}\|d_t\|^4\right).$$

Note that $\frac{1}{T}\sum_{t=1}^{T}\|a_t\|^4 \leq T\left(\frac{1}{T}\sum_{t=1}^{T}\|a_t\|^2\right)^2$ and $\frac{1}{T}\sum_{t=1}^{T}\|a_t\|^2 = O_p(1/T)$ (Bai and Ng (2002)), implying that $\frac{1}{T}\sum_{t=1}^{T}\|a_t\|^4 = O_p(1/T)$ and $\sum_{t=1}^{T}\|a_t\|^4 = O_p(1)$. Similarly, by repeated application of Cauchy-Schwarz inequality,

$$\begin{aligned}
\sum_{t=1}^{T}\|b_t\|^4 &= \frac{1}{T^4}\sum_{t=1}^{T}\left\|\sum_{s=1}^{T}\widetilde{F}_s\zeta_{st}\right\|^4 \\
&= \frac{1}{T^4}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\sum_{u=1}^{T}\sum_{s_1=1}^{T}\sum_{u_1=1}^{T}\widetilde{F}_s'\widetilde{F}_u\zeta_{st}\zeta_{ut}\widetilde{F}_{s_1}'\widetilde{F}_{u_1}\zeta_{s_1t}\zeta_{u_1t}\right) \\
&\leq \frac{1}{T^4}\sum_{s=1}^{T}\sum_{u=1}^{T}\sum_{s_1=1}^{T}\sum_{u_1=1}^{T}\left[\left|\widetilde{F}_s'\widetilde{F}_u\widetilde{F}_{s_1}'\widetilde{F}_{u_1}\right|\left(\left(\sum_{t=1}^{T}\zeta_{st}^2\zeta_{ut}^2\right)^{\frac{1}{2}}\left(\sum_{t=1}^{T}\zeta_{s_1t}^2\zeta_{u_1t}^2\right)^{\frac{1}{2}}\right)\right] \\
&\leq \left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{u=1}^{T}\left|\widetilde{F}_s'\widetilde{F}_u\right|\left(\sum_{t=1}^{T}\zeta_{st}^2\zeta_{ut}^2\right)^{\frac{1}{2}}\right]^2 \\
&\leq \left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{u=1}^{T}\left(\widetilde{F}_s'\widetilde{F}_u\right)^2\right]\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\zeta_{st}^2\zeta_{ut}^2\right)\right].
\end{aligned}$$

Hence,

$$\sum_{t=1}^{T}\|b_t\|^4 \leq \left[\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{F}_s\right\|^2\right]^2\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\zeta_{st}^2\zeta_{ut}^2\right)\right],$$

where $\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{F}_s\right\|^2 = O_p(1)$ (see GP(2013)) and $E\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\zeta_{st}^2\zeta_{ut}^2\right)\right] = O\left(\left(\frac{\sqrt{T}}{N}\right)^2\right)$. Indeed,

$$\begin{aligned}E\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\zeta_{st}^2\zeta_{ut}^2\right)\right] &\leq \frac{1}{T^2}\sum_{s=1}^{T}\sum_{u=1}^{T}\sum_{t=1}^{T}\left[E\left(\zeta_{st}^4\right)\right]^{\frac{1}{2}}\left[E\left(\zeta_{ut}^4\right)\right]^{\frac{1}{2}}\\ &\leq \frac{1}{T^2}\sum_{s=1}^{T}\sum_{u=1}^{T}\sum_{t=1}^{T}\left[\max_{s,t}E\left(\zeta_{st}^4\right)\right]^{\frac{1}{2}}\left[\max_{u,t}E\left(\zeta_{ut}^4\right)\right]^{\frac{1}{2}}\\ &\leq T\left[\max_{s,t}E\left(\zeta_{st}^4\right)\right] = O\left(\frac{T}{N^2}\right),\end{aligned}$$

since $\max_{s,t}E\left(\zeta_{st}^4\right) = O\left(\frac{1}{N^2}\right)$ by Assumption 2.c. Thus, $\sum_{t=1}^{T}\|b_t\|^4 = O_p\left(\left(\frac{\sqrt{T}}{N}\right)^2\right)$. Thirdly,

$$\sum_{t=1}^{T}\|c_t\|^4 = \sum_{t=1}^{T}\left\|\frac{1}{T}\frac{1}{N}\sum_{s=1}^{T}\widetilde{F}_sF_s'\Lambda e_t\right\|^4 \leq \sum_{t=1}^{T}\left\|\frac{1}{N}\Lambda e_t\right\|^4\left\|\frac{1}{T}\sum_{s=1}^{T}\widetilde{F}_sF_s'\right\|^4,$$

implying that

$$\sum_{t=1}^{T}\|c_t\|^4 \leq \frac{T}{N^2}\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_t\right\|^4\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{F}_s\right\|^2\right)^2\left(\frac{1}{T}\sum_{s=1}^{T}\|F_s\|^2\right)^2 = O_p\left(\left(\frac{\sqrt{T}}{N}\right)^2\right),$$

since $\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{F}_s\right\|^2 = O_p(1)$, $\frac{1}{T}\sum_{s=1}^{T}\|F_s\|^2 = O_p(1)$. The proof that $\sum_{t=1}^{T}\|d_t\|^4 = O_p\left(\frac{T}{N^2}\right)$ is similar and therefore omitted. Thus, $\sum_{t=1}^{T}\left\|\widetilde{F}_t - HF_t\right\|^4 = O_p(1) + O_p\left(\frac{T}{N^2}\right) = O_p(1)$ as $\frac{\sqrt{T}}{N} \longrightarrow c < \infty$. ■

Proof of Theorem 2.1. This result follows from an application of Theorem 2.1 of GP (2013). In particular, under Assumptions 1-5, using Theorem 5.3 of Gallant and White (1988), we have that $\Omega^{-\frac{1}{2}}\frac{1}{\sqrt{T}}z'\varepsilon \to^d N(0, I)$, which verifies Assumption 5.b) of GP (2013). Our moment conditions on z_t and ε_{t+h} imply those of GP (2013) and Assumptions 1-4 are identical to the remaining assumptions of GP (2013). ■

Proof of Lemma 2.1. One can write

$$\widehat{\Omega} = \frac{1}{T}\left[\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\widehat{z}_t\widehat{\varepsilon}_{t+h}\widehat{z}_s'\widehat{\varepsilon}_{s+h}k\left(\frac{s-t}{M_T}\right)\right] \equiv A_{1T} + A_{2T} + A_{3T} + A_{2T}' + A_{4T} + A_{5T} + A_{3T}' + A_{5T}' + A_{6T},$$

with

$$
\begin{aligned}
A_{1T} &= \frac{1}{T}\left[\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\widehat{z}_t\varepsilon_{t+h}\widehat{z}_s'\varepsilon_{s+h}k\left(\frac{s-t}{M_T}\right)\right],\\
A_{2T} &= \frac{1}{T}\left[\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\widehat{z}_t\varepsilon_{t+h}\left(\delta-\widehat{\delta}\right)'\widehat{z}_s\widehat{z}_s'k\left(\frac{s-t}{M_T}\right)\right],\\
A_{3T} &= \frac{1}{T}\left[\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\widehat{z}_t\varepsilon_{t+h}\alpha'H^{-1}\left(HF_s-\widetilde{F}_s\right)\widehat{z}_s'k\left(\frac{s-t}{M_T}\right)\right],\\
A_{4T} &= \frac{1}{T}\left[\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\widehat{z}_t\widehat{z}_t'\left(\delta-\widehat{\delta}\right)\left(\delta-\widehat{\delta}\right)'\widehat{z}_s\widehat{z}_s'k\left(\frac{s-t}{M_T}\right)\right],\\
A_{5T} &= \frac{1}{T}\left[\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\widehat{z}_t\widehat{z}_t'\left(\delta-\widehat{\delta}\right)\alpha'H^{-1}\left(HF_s-\widetilde{F}_s\right)\widehat{z}_s'k\left(\frac{s-t}{M_T}\right)\right],
\end{aligned}
$$

and

$$
A_{6T}=\frac{1}{T}\left[\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\widehat{z}_t\left(HF_t-\widetilde{F}_t\right)'\left(H^{-1}\right)'\alpha\alpha'H^{-1}\left(HF_s-\widetilde{F}_s\right)\widehat{z}_s'k\left(\frac{s-t}{M_T}\right)\right].
$$

Next we show that $A_{iT}=o_P(1)$, for $i=2,\ldots,6$, and that A_{1T} converges to $\Phi_0\Omega\Phi_0'$.

(1) Starting from A_{2T}, we can write

$$
\begin{aligned}
A_{2T} &= \frac{1}{T}\left[\sum_{t=1}^{T-h}\widehat{z}_t\varepsilon_{t+h}\left(\delta-\widehat{\delta}\right)'\widehat{z}_t\widehat{z}_t'\right]\\
&\quad+\frac{1}{T}\left[\sum_{\tau=1}^{T-h-1}\sum_{t=1}^{T-h-\tau}\widehat{z}_t\varepsilon_{t+h}\left(\delta-\widehat{\delta}\right)'\widehat{z}_{t+\tau}\widehat{z}_{t+\tau}'k\left(\frac{\tau}{M_T}\right)\right]\\
&\quad+\frac{1}{T}\left[\sum_{\tau=1}^{T-h-1}\sum_{t=1}^{T-h-\tau}\widehat{z}_{t+\tau}\varepsilon_{t+h+\tau}\left(\delta-\widehat{\delta}\right)'\widehat{z}_t\widehat{z}_t'k\left(\frac{\tau}{M_T}\right)\right]\\
&\equiv A_{2T,1}+A_{2T,2}+A_{2T,3}.
\end{aligned}
$$

By repeated application of Cauchy-Schwarz inequality,

$$
\begin{aligned}
\|A_{2T,1}\| &\leq \frac{1}{T}\left\|\delta-\widehat{\delta}\right\|\left(\sum_{t=1}^{T-h}\|\widehat{z}_t\|^4\right)^{\frac{1}{4}}\left(\sum_{t=1}^{T-h}\|\varepsilon_{t+h}\|^4\right)^{\frac{1}{4}}\left(\sum_{t=1}^{T-h}\|\widehat{z}_t\|^4\right)^{\frac{1}{2}}\\
&\leq \left\|\delta-\widehat{\delta}\right\|\left(\frac{1}{T}\sum_{t=1}^{T-h}\|\widehat{z}_t\|^4\right)^{\frac{3}{4}}\left(\frac{1}{T}\sum_{t=1}^{T-h}\|\varepsilon_{t+h}\|^4\right)^{\frac{1}{4}}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\|A_{2T,2}\| &\leq \left\|\delta-\widehat{\delta}\right\|\sum_{\tau=1}^{T-h-1}\left|k\left(\frac{\tau}{M_T}\right)\right|\left(\frac{1}{T}\sum_{t=1}^{T-h-\tau}\|\widehat{z}_t\|^4\right)^{\frac{1}{4}}\left(\frac{1}{T}\sum_{t=1}^{T-h-\tau}\|\varepsilon_{t+h}\|^4\right)^{\frac{1}{4}}\left(\frac{1}{T}\sum_{t=1}^{T-h-\tau}\|\widehat{z}_{t+\tau}\|^4\right)^{\frac{1}{2}}\\
&\leq \left\|\delta-\widehat{\delta}\right\|\sum_{\tau=1}^{T-h}\left|k\left(\frac{\tau}{M_T}\right)\right|\left(\frac{1}{T}\sum_{t=1}^{T-h}\|\widehat{z}_t\|^4\right)^{\frac{3}{4}}\left(\frac{1}{T}\sum_{t=1}^{T-h}\|\varepsilon_{t+h}\|^4\right)^{\frac{1}{4}},
\end{aligned}
$$

and

$$\|A_{2T,3}\| \leq \left\|\delta-\widehat{\delta}\right\| \sum_{\tau=1}^{T-h} \left|k\left(\frac{\tau}{M_T}\right)\right| \left(\frac{1}{T}\sum_{t=1}^{T-h} \|\widehat{z}_t\|^4\right)^{\frac{3}{4}} \left(\frac{1}{T}\sum_{t=1}^{T-h} \|\varepsilon_{t+h}\|^4\right)^{\frac{1}{4}}.$$

It follows that,

$$\|A_{2T}\| \leq 2\left\|\delta-\widehat{\delta}\right\| \sum_{\tau=0}^{T-h} \left|k\left(\frac{\tau}{M_T}\right)\right| \left(\frac{1}{T}\sum_{t=1}^{T-h} \|\widehat{z}_t\|^4\right)^{\frac{3}{4}} \left(\frac{1}{T}\sum_{t=1}^{T-h} \|\varepsilon_{t+h}\|^4\right)^{\frac{1}{4}}.$$

By Theorem 2.1, $\left\|\delta-\widehat{\delta}\right\| = O_p\left(T^{-\frac{1}{2}}\right)$. Since $\frac{1}{M_T}\sum_{\tau=0}^{T-h}\left|k\left(\frac{\tau}{M_T}\right)\right| \to \int_0^{+\infty} |k(x)|\, dx < \infty$, we have that $\sum_{\tau=1}^{T-h}\left|k\left(\frac{\tau}{M_T}\right)\right| = O(M_T)$. Similarly, we can show that $\sum_{t=1}^{T-h}\|\widehat{z}_t\|^4 \leq 8\Phi\sum_{t=1}^{T-h}\|z_t\|^4 + 8\sum_{t=1}^{T-h}\|\widehat{z}_t-\Phi z_t\|^4$ by using the decomposition $\widehat{z}_t = \Phi z_t + (\widehat{z}_t - \Phi z_t)$ and the c-r inequality. It follows that $\sum_{t=1}^{T-h}\|\widehat{z}_t\|^4 = O_p(T)$ given that $\sum_{t=1}^{T-h}\|z_t\|^4 = O_p(T)$ by Markov's inequality and the moment conditions on z_t and given that $\sum_{t=1}^{T-h}\|\widehat{z}_t-\Phi z_t\|^4 = O_p(1)$ (see Lemma A.1). Hence, $A_{2T} = O_p\left(\frac{M_T}{\sqrt{T}}\right) = o_p(1)$ since we assume that $M_T/T^{1/2} \to 0$.

(2) By the decomposition $\widehat{z}_t = \Phi z_t + (\widehat{z}_t - \Phi z_t)$, we can write

$$\begin{aligned}
A_{3T} &= A_{3T,1} + A_{3T,2}, \text{ where} \\
A_{3T,1} &= \frac{1}{T}\left[\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\left(\widetilde{F}_t - HF_t\right)\varepsilon_{t+h}\alpha' H^{-1}\left(HF_s-\widetilde{F}_s\right)\widehat{z}_s' k\left(\frac{s-t}{M_T}\right)\right] \text{ and} \\
A_{3T,2} &= \frac{1}{T}\Phi\left[\sum_{t=1}^{T-h}\sum_{s=1}^{T-h} z_t\varepsilon_{t+h}\alpha' H^{-1}\left(HF_s-\widetilde{F}_s\right)\widehat{z}_s' k\left(\frac{s-t}{M_T}\right)\right].
\end{aligned}$$

Using the same arguments as for A_{2T}, we can bound

$$\begin{aligned}
\|A_{3T,1}\| &\leq \frac{2}{T}\left\|\alpha' H^{-1}\right\|\left(\sum_{\tau=0}^{T-h}\left|k\left(\frac{\tau}{M_T}\right)\right|\right)\left[\left(\sum_{t=1}^{T-h}\left\|\left(\widetilde{F}_t-HF_t\right)\varepsilon_{t+h}\right\|^2\right)^{\frac{1}{2}}\left(\sum_{t=1}^{T-h}\left\|\left(\widetilde{F}_t-HF_t\right)\varepsilon_{t+h}\right\|^2\right)^{\frac{1}{2}}\right] \\
&\leq \frac{2}{T}\left\|\alpha' H^{-1}\right\|\left(\sum_{\tau=0}^{T-h}\left|k\left(\frac{\tau}{M_T}\right)\right|\right)\left[\left(\sum_{t=1}^{T-h}\left\|\widetilde{F}_t-HF_t\right\|^4\right)^{\frac{1}{2}}\left(\sum_{t=1}^{T-h}\varepsilon_{t+h}^4\right)^{\frac{1}{2}}\right],
\end{aligned}$$

where $\sum_{\tau=0}^{T-h}\left|k\left(\frac{\tau}{M_T}\right)\right| = O\left(M_T\right)$, $\sum_{t=1}^{T-h}\left\|\widetilde{F}_t - HF_t\right\|^4 = O_p\left(1\right)$ and $\sum_{t=1}^{T-h}\varepsilon_{t+h}^4 = O_p\left(T\right)$. Hence, $A_{3T,1} = O_p\left(\frac{M_T}{\sqrt{T}}\right)$. Similarly,

$$\begin{aligned}
\left\|A_{3T,2}\right\| &\leq \frac{\|\Phi\|}{T}\left[\sum_{s=1}^{T-h}\left\|\alpha' H^{-1}\left(HF_s - \widetilde{F}_s\right)\widehat{z}_s'\right\|^2\right]^{\frac{1}{2}}\left[\sum_{s=1}^{T-h}\left\|\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}k\left(\frac{s-t}{M_T}\right)\right\|^2\right]^{\frac{1}{2}} \\
&\leq \frac{\|\Phi\|}{T}\left\|\alpha' H^{-1}\right\|\left[\sum_{s=1}^{T-h}\left\|HF_s - \widetilde{F}_s\right\|^4\right]^{\frac{1}{4}}\left[\sum_{s=1}^{T-h}\left\|\widehat{z}_s\right\|^4\right]^{\frac{1}{4}}\left[\sum_{s=1}^{T-h}\left\|\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}k\left(\frac{s-t}{M_T}\right)\right\|^2\right]^{\frac{1}{2}} \\
&= O_p\left(\left(\frac{M_T}{\sqrt{T}}\right)^{\frac{1}{2}}\right),
\end{aligned}$$

where in particular $\sum_{s=1}^{T-h}\left\|\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}k\left(\frac{s-t}{M_T}\right)\right\|^2 = O_p\left(M_T\cdot T\right)$. To show this, note that

$$\begin{aligned}
E\left[\sum_{s=1}^{T-h}\left\|\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}k\left(\frac{s-t}{M_T}\right)\right\|^2\right] &= \sum_{s=1}^{T-h}\left[\sum_{t=1}^{T-h}\sum_{t_1=1}^{T-h} E\left(z_t'\varepsilon_{t+h}z_{t_1}\varepsilon_{t_1+h}\right)k\left(\frac{s-t}{M_T}\right)k\left(\frac{s-t_1}{M_T}\right)\right] \\
&\leq \sum_{t=1}^{T-h}\sum_{t_1=1}^{T-h}\left|E\left(z_t'\varepsilon_{t+h}z_{t_1}\varepsilon_{t_1+h}\right)\right|\sum_{s=1}^{T-h}\left|k\left(\frac{s-t}{M_T}\right)\right|\left|k\left(\frac{s-t_1}{M_T}\right)\right| \\
&\leq \frac{1}{2}\sum_{t=1}^{T-h}\sum_{t_1=1}^{T-h}\left|E\left(z_t'\varepsilon_{t+h}z_{t_1}\varepsilon_{t_1+h}\right)\right|\left[\sum_{s=1}^{T-h}\left|k\left(\frac{s-t}{M_T}\right)\right|^2 + \sum_{s=1}^{T-h}\left|k\left(\frac{s-t_1}{M_T}\right)\right|^2\right] \\
&\leq \underbrace{\left[\sum_{\tau=-\infty}^{+\infty}\left|k\left(\frac{\tau}{M_T}\right)\right|^2\right]}_{O(M_T)}\underbrace{\sum_{t=1}^{T-h}\sum_{t_1=1}^{T-h}\left|E\left(z_t'\varepsilon_{t+h}z_{t_1}\varepsilon_{t_1+h}\right)\right|}_{O_p(T)} = O_P\left(M_T\cdot T\right).
\end{aligned}$$

Indeed, $\frac{1}{M_T}\sum_{\tau=-\infty}^{+\infty}\left|k\left(\frac{\tau}{M_T}\right)\right|^2 \to \int_{-\infty}^{+\infty} k\left(x\right)^2 dx \leq \int_{-\infty}^{+\infty}\left|k\left(x\right)\right| dx < \infty$ (as $\left|k\left(x\right)\right| \in [0,\ 1]$) and $\sum_{\tau=-\infty}^{+\infty}\left|k\left(\frac{\tau}{M_T}\right)\right|^2 = O\left(M_T\right)$. Moreover,

$$\begin{aligned}
&\sum_{t=1}^{T-h}\sum_{t_1=1}^{T-h}\left|E\left(z_t'\varepsilon_{t+h}z_{t_1}\varepsilon_{t_1+h}\right)\right| \\
&\leq \sum_{t=1}^{T-h} E\left\|z_t\varepsilon_{t+h}\right\|^2 + 2\sum_{\tau=1}^{T-h-1}\sum_{j=1}^{T-h-\tau}\left|E\left(z_t'\varepsilon_{t+h}z_{t+\tau}\varepsilon_{t+\tau+h}\right)\right| \\
&\leq \sum_{t=1}^{T-h} E\left\|z_t\epsilon_{t+h}\right\|^2 + 2\cdot 2\left(2^{1-\frac{1}{r}}+1\right)\sum_{\tau=1}^{T-h-1}\sum_{j=1}^{T-h-\tau}\left[E\left(\left\|z_t\varepsilon_{t+h}\right\|^r\right)\right]^{\frac{1}{r}}\left[E\left(\left\|z_{t+\tau}\varepsilon_{t+\tau+h}\right\|^r\right)\right]^{\frac{1}{r}}\left(\alpha(\tau)\right)^{\frac{r-2}{r}},
\end{aligned}$$

using Corollary 14.3 of Davidson (1994) with $p = r > 2$ and $r > \dfrac{p}{p-1}$ and the fact that $E\left\|z_t\varepsilon_{t+h}\right\|^r \leq C < \infty$. Thus,

$$\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\left|E\left(z_s'\varepsilon_{s+h}z_t\varepsilon_{t+h}\right)\right| = O\left(T\right).$$

This implies that $A_{3T} = O_p\left(\frac{M_T}{\sqrt{T}}\right) + O_p\left(\left(\frac{M_T}{\sqrt{T}}\right)^{\frac{1}{2}}\right) = O_p\left(\frac{M_T}{\sqrt{T}}\right) = o_p\left(1\right).$

(3) As previously, we can bound

$$\begin{aligned}
\left\|A_{4T}\right\| &\leq 2\frac{1}{T}\left\|\delta-\widehat{\delta}\right\|^2\left(\sum_{\tau=0}^{T-h}\left|k\left(\frac{\tau}{M_T}\right)\right|\left(\sum_{t=1}^{T-h}\left\|\widehat{z}_t\right\|^4\right)^{\frac{1}{2}}\left(\sum_{t=1}^{T-h}\left\|\widehat{z}_t\right\|^4\right)^{\frac{1}{2}}\right)\\
&\leq 2\frac{1}{T}\underbrace{\left\|\delta-\widehat{\delta}\right\|^2}_{O_p(T^{-1})}\left(\underbrace{\sum_{\tau=0}^{T-h}\left|k\left(\frac{\tau}{M_T}\right)\right|}_{O(M_T)}\right)\left(\underbrace{\sum_{t=1}^{T-h}\left\|\widehat{z}_t\right\|^4}_{O_p(T)}\right) = O_P\left(\frac{M_T}{T}\right) = o_p\left(1\right).
\end{aligned}$$

(4) For A_{5T}, we have that

$$\begin{aligned}
\left\|A_{5T}\right\| &\leq 2\underbrace{\left\|\delta-\widehat{\delta}\right\|}_{O_p\left(T^{-\frac{1}{2}}\right)}\left\|\alpha'H^{-1}\right\|\frac{1}{T}\underbrace{\sum_{\tau=0}^{T-h}\left|k\left(\frac{\tau}{M_T}\right)\right|}_{O(M_T)}\left(\underbrace{\sum_{t=1}^{T-h}\left\|\widehat{z}_t\widehat{z}_t'\right\|^2}_{T}\right)^{\frac{1}{2}}\\
&\quad\times\left(\underbrace{\sum_{s=1}^{T-h}\left\|HF_s-\widetilde{F}_s\right\|^4}_{O_p(1)}\right)^{\frac{1}{4}}\left(\underbrace{\sum_{t=1}^{T-h}\left\|\widehat{z}_s\right\|^4}_{O_p(T)}\right)^{\frac{1}{4}},
\end{aligned}$$

which implies that $A_{5T} = O_p\left(\frac{M_T}{T^{\frac{3}{4}}}\right) = o_p\left(1\right)$ given that $M_T/T^{1/2} = o\left(1\right)$.

(5) For A_{6T}, we can show that $A_{6T} = O_p\left(M_T/\sqrt{T}\right) = o_p\left(1\right)$ by using the same arguments as above.

(6) Finally, to show that $A_{1T} \rightarrow_p \Phi_0\Omega\Phi_0'$, by replacing $\widehat{z}_t$ by $\Phi z_t + \left(\widehat{z}_t - \Phi z_t\right)$, we have that

$$\begin{aligned}
A_{1T} &= A_{1T,1} + A_{1T,2} + A_{1T,3} + A_{1T,4}, \text{ where}\\
A_{1T,1} &= \Phi\frac{1}{T}\left[\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}z_t\varepsilon_{t+h}z_s'\varepsilon_{s+h}k\left(\frac{s-t}{M_T}\right)\right]\Phi',\\
A_{1T,2} &= \frac{1}{T}\left[\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\left(\widehat{z}_t-\Phi z_t\right)\varepsilon_{t+h}\left(\widehat{z}_s-\Phi z_s\right)'\varepsilon_{s+h}k\left(\frac{s-t}{M_T}\right)\right],\\
A_{1T,3} &= \Phi\frac{1}{T}\left[\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}z_t\varepsilon_{t+h}\left(\widehat{z}_s-\Phi z_s\right)'\varepsilon_{s+h}k\left(\frac{s-t}{M_T}\right)\right] \text{ and}\\
A_{1T,4} &= A_{1T,3}'.
\end{aligned}$$

By arguments similar to those already used, we can show that the last three terms are $O_p\left(M_T/\sqrt{T}\right) = o_p(1)$. To show that $A_{1T,1} \to^p \Phi_0 \Omega \Phi_0'$, note that under our assumptions, $E\|z_t\varepsilon_{t+h}\|^{2r} < C$ and $\{z_t\varepsilon_{t+h}\}$ is a strong mixing sequence of size $-\frac{3r}{r-2}$. The result then follows by Proposition 1.b of Andrews (1991) and the fact that $\Phi = \Phi_0 + o_p(1)$. Since $\frac{1}{T}\widehat{z}'\widehat{z} = \Phi_0 \Sigma_{zz} \Phi_0' + o_p(1)$ with $\Sigma_{zz} > 0$, we conclude that

$$\widehat{\Sigma}_\delta = \left(\frac{1}{T}\widehat{z}'\widehat{z}\right)^{-1} \widehat{\Omega} \left(\frac{1}{T}\widehat{z}'\widehat{z}\right)^{-1} \to^P \Sigma_\delta = \left(\Phi_0'\right)^{-1} \Sigma_{zz}^{-1} \Omega \Sigma_{zz}^{-1} \Phi_0^{-1}.$$

■

Proof of Theorem 3.1. We verify Conditions A*-F* of GP (2013). Because our bootstrap scheme relies on the wild bootstrap to generate e_{it}^*, as in GP (2013), conditions that only involve this random variable were already verified by GP (2013). In particular, Conditions A*, B* and F* are satisfied under our assumptions (see proof of Theorem 4.1 of GP (2013)). Hence, we only need to verify Conditions C*, D* and E*.

- Starting with Condition C*(a), by the independence between e_{it}^* and ε_{s+h}^*, and the fact that e_{it}^* is independent across (i,t), it follows that

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T} E^* \left| \frac{1}{\sqrt{TN}} \sum_{s=1}^{T-h} \sum_{i=1}^{N} \varepsilon_{s+h}^* \left(e_{it}^* e_{is}^* - E\left(e_{it}^* e_{is}^*\right)\right)\right|^2 \\
= \; & \frac{1}{T}\sum_{t=1}^{T} \frac{1}{T} \sum_{s=1}^{T-h}\sum_{l=1}^{T-h} E^*\left(\varepsilon_{s+h}^*\varepsilon_{l+h}^*\right) E^* \left[\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(e_{it}^* e_{is}^* - E\left(e_{it}^* e_{is}^*\right)\right)\right)\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\left(e_{jt}^* e_{jl}^* - E\left(e_{jt}^* e_{jl}^*\right)\right)\right)\right] \\
= \; & \frac{1}{T}\sum_{t=1}^{T} \frac{1}{T} \sum_{s=1}^{T-h}\sum_{l=1}^{T-h} E^*\left(\varepsilon_{s+h}^*\varepsilon_{l+h}^*\right) \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} Cov^*\left(e_{it}^* e_{is}^*, e_{jt}^* e_{jl}^*\right) \\
= \; & \frac{1}{T}\sum_{t=1}^{T} \frac{1}{T} \sum_{s=1}^{T-h} E^*\left(\varepsilon_{s+h}^{*2}\right) \frac{1}{N}\sum_{i=1}^{N} \widetilde{e}_{it}^2 \widetilde{e}_{is}^2 Var^*\left(\eta_{it}\eta_{is}\right) \\
= \; & \frac{1}{T}\sum_{t=1}^{T} \frac{1}{T} \sum_{s=1}^{T-h} \widehat{\varepsilon}_{s+h}^2 \frac{1}{N}\sum_{i=1}^{N} \widetilde{e}_{it}^2 \widetilde{e}_{is}^2 Var^*\left(\eta_{it}\eta_{is}\right) \\
\leq \; & M \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{e}_{it}^2\right)\left(\frac{1}{T}\sum_{s=1}^{T-h}\widehat{\varepsilon}_{s+h}^2 \widetilde{e}_{is}^2\right),
\end{aligned}$$

  where the third equality uses the fact that $Cov^*\left(e_{it}^* e_{is}^*, e_{jt}^* e_{jl}^*\right) = 0$ for $i \neq j$ or $s \neq l$, and the fourth equality uses the fact that $E^*\left(\varepsilon_{s+h}^{*2}\right) = \widehat{\varepsilon}_{s+h}^2$, given that $E^*\left(\eta_j^2\right) = 1$ for all $j = 1, \ldots, k$. The inequality relies on a bound for $Var^*\left(\eta_{it}\eta_{is}\right)$ under our assumptions. The result follows by an application of Cauchy-Schwarz inequality given in particular the fact that $\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{e}_{it}^4/NT = O_P(1)$ and $\sum_{s=1}^{T-h}\widehat{\varepsilon}_{s+h}^4/T = O_P(1)$ under our assumptions.

- For Condition C*(b),

$$\begin{aligned}
E^*\left\|\frac{1}{\sqrt{TN}}\sum_{t=1}^{T-h}\sum_{i=1}^{N}\widetilde{\lambda}_i e_{it}^*\varepsilon_{t+h}^*\right\|^2 &= \frac{1}{TN}E^*\left(\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\varepsilon_{t+h}^*\varepsilon_{s+h}^*\left(\sum_{i=1}^{N}\widetilde{\lambda}_i' e_{it}^*\right)\left(\sum_{j=1}^{N}\widetilde{\lambda}_j e_{js}^*\right)\right)\\
&= \frac{1}{TN}\left[\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}E^*\left(\varepsilon_{t+h}^*\varepsilon_{s+h}^*\right)E^*\left[\left(\sum_{i=1}^{N}\widetilde{\lambda}_i' e_{it}^*\right)\left(\sum_{j=1}^{N}\widetilde{\lambda}_j e_{js}^*\right)\right]\right]\\
&= \frac{1}{TN}\left[\sum_{t=1}^{T-h}E^*\left(\varepsilon_{t+h}^{*2}\right)\left(\sum_{i=1}^{N}\widetilde{\lambda}_i'\widetilde{\lambda}_i E^*\left(e_{it}^{*2}\right)\right)\right]\\
&= \frac{1}{TN}\left[\sum_{t=1}^{T-h}\widehat{\varepsilon}_{t+h}^2\left(\sum_{i=1}^{N}\left\|\widetilde{\lambda}_i\right\|^2\widetilde{e}_{it}^2\right)\right],
\end{aligned}$$

where the third equality uses the fact that $E^*\left(e_{it}^* e_{js}^*\right)=0$ whenever $i\neq j$ or $t\neq s$, and the fourth equality the fact that $E^*\left(\varepsilon_{t+h}^{*2}\right)=\widehat{\varepsilon}_{t+h}^2$. The rest of the proof follows exactly the proof of GP (2013) (cf. Proof of Theorem 4.1).

- The proof of Condition C*(c) follows the proof in GP (2013) closely with the only difference that we show that $\frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_{t+h}^{*4}=O_{p^*}(1)$ in probability. Indeed,

$$E^*\left(\frac{1}{T}\sum_{t=1}^{T-h}\varepsilon_{t+h}^{*4}\right)=\frac{1}{T}\sum_{j=1}^{k}\sum_{l=1}^{b}\widehat{\varepsilon}_{(j-1)b+l+h}^4 E^*\left(\nu_j^4\right)\leq M\frac{1}{T}\sum_{j=1}^{k}\sum_{l=1}^{b}\widehat{\varepsilon}_{(j-1)b+l+h}^4=M\frac{1}{T}\sum_{t=1}^{T-h}\widehat{\varepsilon}_{t+h}^4,$$

since $E^*\left(\nu_j^4\right)\leq M<\infty$. Because $\frac{1}{T}\sum_{t=1}^{T-h}\widehat{\varepsilon}_{t+h}^4=O_P(1)$ under our assumptions, this proves the desired result.

- For Condition D*(a), we have that for any $i=1,\ldots,b$ and $j=1,\ldots,k$,

$$\begin{aligned}
E^*\left(\varepsilon_{i+(j-1)b+h}^*\right) &= \widehat{\varepsilon}_{(j-1)b+i+h}E^*\left(\nu_j\right)=0\\
\text{and } \frac{1}{T}\sum_{t=1}^{T-h}E^*\left|\varepsilon_{t+h}^*\right|^2 &= \frac{1}{T}\sum_{t=1}^{T-h}\widehat{\varepsilon}_{t+h}^2\leq\left(\frac{1}{T}\sum_{t=1}^{T-h}\widehat{\varepsilon}_{t+h}^4\right)^{\frac{1}{2}}=O_p(1).
\end{aligned}$$

- For Condition D*(b), let

$$\xi_j^*\equiv\Omega^{*-\frac{1}{2}}\frac{1}{\sqrt{b}}\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\varepsilon_{(j-1)b+l+h}^*=\Omega^{*-\frac{1}{2}}\frac{1}{\sqrt{b}}\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\widehat{\varepsilon}_{(j-1)b+l+h}\cdot\nu_j,$$

where ν_j are i.i.d. $(0,1)$ across j. We can write

$$\Omega^{*-\frac{1}{2}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\widehat{z}_t\varepsilon_{t+h}^*=\frac{1}{\sqrt{k}}\sum_{j=1}^{k}\xi_j^*,$$

where ξ_j^* are conditionally independent for $j = 1, \ldots, k$, with

$$E^*\left(\xi_j^*\right) = 0 \text{ and } Var^*\left(\frac{1}{\sqrt{k}}\sum_{j=1}^{k}\xi_j^*\right) = I.$$

It suffices to show that for some $d > 1$, $Z_T \equiv \frac{1}{k^d}\sum_{j=1}^{k} E^*\left\|\xi_j^*\right\|^{2d} = o_p(1)$. Replacing ξ_j^* by its definition and using the fact that $k = (T-h)/b$, we have that

$$\begin{aligned} Z_T &= \left(\frac{T}{T-h}\right)^d \frac{1}{T^d}\sum_{j=1}^{k} E^*\left\|\Omega^{*-\frac{1}{2}}\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\varepsilon^*_{(j-1)b+l+h}\right\|^{2d} \\ &\leq C\left\|\Omega^{*-\frac{1}{2}}\right\|^{2d}\frac{1}{T^d}\sum_{p=1}^{k}E^*\left(\left\|\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\varepsilon^*_{(j-1)b+l+h}\right\|\right)^{2d}. \end{aligned}$$

Since $\left\|\Omega^{*-\frac{1}{2}}\right\|^{2d} = O_P(1)$ as $\Omega^* \to_p \Phi_0\Omega\Phi_0'$ (see Condition E*) and $\Phi_0\Omega\Phi_0' > 0$, it suffices to show that the second factor is $o_P(1)$. Noting that

$$\begin{aligned} \varepsilon^*_{(j-1)b+l+h} &= \hat{\varepsilon}_{(j-1)b+l+h}\cdot\nu_j = \varepsilon_{(j-1)b+l+h}\cdot\nu_j \\ &\quad -\widehat{z}'_{(j-1)b+l}\left(\widehat{\delta}-\delta\right)\cdot\nu_j + \left(HF_{(j-1)b+l} - \widetilde{F}_{(j-1)b+l}\right)'\left(H^{-1}\right)'\alpha\cdot\nu_j, \end{aligned}$$

we have that

$$\begin{aligned} &\frac{1}{T^d}\sum_{j=1}^{k}E^*\left\|\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\varepsilon^*_{(j-1)b+l+h}\right\|^{2d} \\ \leq\ & 3^{2d-1}\frac{1}{T^d}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\varepsilon_{(j-1)b+l+h}\right\|^{2d} E^*\left|\nu_j\right|^{2d} \\ &+3^{2d-1}\frac{1}{T^d}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\widehat{z}'_{(j-1)b+l}\left(\widehat{\delta}-\delta\right)\right\|^{2d}E^*\left|\nu_j\right|^{2d} \\ &+3^{2d-1}\frac{1}{T^d}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\left(HF_{(j-1)b+l}-\widetilde{F}_{(j-1)b+l}\right)'\left(H^{-1}\right)'\alpha\right\|^{2d}E^*\left|\nu_j\right|^{2d} \\ \equiv\ & (a)+(b)+(c). \end{aligned}$$

Since $E^*\left|\nu_j\right|^{2d} \leq M$, it suffices to show that each of (a), (b) and (c) is $o_p(1)$. Starting with (a), replacing $\widehat{z}_{(j-1)b+l}$ by $\Phi z_{(j-1)b+l} + \left(\widehat{z}_{(j-1)b+l} - \Phi z_{(j-1)b+l}\right)$ and using the c-r inequality, it follows that

$$\begin{aligned} \frac{1}{T^d}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\varepsilon_{(j-1)b+l+h}\right\|^{2d} &\leq \left\|\Phi\right\|^{2d}\frac{2^{2d-1}}{T^d}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}z_{(j-1)b+l}\varepsilon_{(j-1)b+l+h}\right\|^{2d} \\ &\quad +\frac{2^{2d-1}}{T^d}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}\left(\widehat{z}_{(j-1)b+l}-\Phi z_{(j-1)b+l}\right)\varepsilon_{(j-1)b+l+h}\right\|^{2d}. \end{aligned}$$

Noting that for any $d>1$, $\sum_{j=1}^{k}|a_j|^{2d}\leq\left(\sum_{j=1}^{k}|a_j|^2\right)^d$ for any a_j, we have that

$$\begin{aligned}
&\frac{1}{T^d}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}\left(\widehat{z}_{(j-1)b+l}-\Phi z_{(j-1)b+l}\right)\varepsilon_{(j-1)b+l+h}\right\|^{2d}\\
\leq\;&\frac{1}{T^d}\left(\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}\left(\widehat{z}_{(j-1)b+l}-\Phi z_{(j-1)b+l}\right)\varepsilon_{(j-1)b+l+h}\right\|^{2}\right)^d=O_p\left(b/\sqrt{T}\right)^d=o_p(1),
\end{aligned}$$

where by the c-r inequality,

$$\begin{aligned}
\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}\left(\widehat{z}_{(j-1)b+l}-\Phi z_{(j-1)b+l}\right)\varepsilon_{(j-1)b+l+h}\right\|^{2}&\leq b\sum_{j=1}^{k}\sum_{l=1}^{b}\left\|\left(\widehat{z}_{(j-1)b+l}-\Phi z_{(j-1)b+l}\right)\varepsilon_{(j-1)b+l+h}\right\|^2\\
&= b\sum_{t=1}^{T-h}\left\|\left(\widehat{z}_t-\Phi z_t\right)\varepsilon_{t+h}\right\|^2\\
&\leq b\left(\sum_{t=1}^{T-h}\left\|\widehat{z}_t-\Phi z_t\right\|^4\right)^{\frac{1}{2}}\left(\sum_{t=1}^{T-h}\varepsilon_t^4\right)^{\frac{1}{2}}=O_p\left(b\sqrt{T}\right),
\end{aligned}$$

given that $\sum_{t=1}^{T-h}\|\widehat{z}_t-\Phi z_t\|^4=O_p(1)$ and $\sum_{t=1}^{T-h}\varepsilon_t^4=O_p(T)$. It now remains to show that for some $d>1$,

$$\frac{1}{T^d}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}z_{(j-1)b+l}\varepsilon_{(j-1)b+l+h}\right\|^{2d}=o_p(1).$$

For d such that $1<d<2$,

$$\begin{aligned}
&E\left(\frac{1}{T^d}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}z_{(j-1)b+l}\varepsilon_{(j-1)b+l+h}\right\|^{2d}\right)\leq\frac{b^d}{T^d}\sum_{j=1}^{k}\left[E\left\|\frac{1}{\sqrt{b}}\sum_{l=1}^{b}z_{(j-1)b+l}\varepsilon_{(j-1)b+l+h}\right\|^2\right]^d\\
\leq\;&\frac{b^d}{T^d}\sum_{j=1}^{k}\left[C+2\cdot 2\left(2^{1-\frac{1}{r}}+1\right)C^{\frac{2}{r}}\sum_{\tau=1}^{b-1}(\alpha(\tau))^{\frac{r-2}{r}}\right]^d\\
\leq\;&\frac{b^d}{T^d}k\cdot M'=O\left[\left(\frac{b}{T}\right)^{d-1}\right]=o(1),
\end{aligned}$$

where the first inequality follows by Jensen's inequality, the second one by Corollary 14.3 of Davidson (1994) and the last one uses the fact that $M'=\left[C+2\cdot 2\left(2^{1-\frac{1}{r}}+1\right)C^{\frac{2}{r}}\sum_{\tau=1}^{\infty}(\alpha(\tau))^{\frac{r-2}{r}}\right]^d<\infty$ under our assumptions. For (b), by repeated application of the Cauchy-Schwarz and the c-r inequalities, we have that

$$\frac{1}{T^d}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\widehat{z}'_{(j-1)b+l}\left(\widehat{\delta}-\delta\right)\right\|^{2d}\leq\frac{b^d}{T^d}\left\|\widehat{\delta}-\delta\right\|^{2d}\sum_{j=1}^{k}\left(\sum_{l=1}^{b}\left\|\widehat{z}_{(j-1)b+l}\right\|^4\right)^d$$

$$= O\left(b^d/T^d\right)\cdot O_p\left(1/T^d\right)\cdot O_p\left(k\cdot b^d\right)=O_p\left(b^{2d-1}/T^{2d-1}\right)=o_p\left(1\right).$$

For (c), using similar arguments as above, we have that

$$\begin{aligned}
&\frac{1}{T^d}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\left(HF_{(j-1)b+l}-\widetilde{F}_{(j-1)b+l}\right)'\left(H^{-1}\right)'\alpha\right\|^{2d}\\
\leq\;&\frac{1}{T^d}\left(\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\left(HF_{(j-1)b+l}-\widetilde{F}_{(j-1)b+l}\right)'\left(H^{-1}\right)'\alpha\right\|^{2}\right)^d\\
\leq\;&\left\|\left(H^{-1}\right)'\alpha\right\|^{2d}\frac{1}{T^d}\left(b\sum_{j=1}^{k}\sum_{l=1}^{b}\left\|\widehat{z}_{(j-1)b+l}\left(HF_{(j-1)b+l}-\widetilde{F}_{(j-1)b+l}\right)'\right\|^{2}\right)^d\\
\leq\;&\left\|\left(H^{-1}\right)'\alpha\right\|^{2d}\frac{b^d}{T^d}\left(\sum_{t=1}^{T-h}\left\|\widehat{z}_t\right\|^4\right)^{\frac{d}{2}}\left(\sum_{t=1}^{T-h}\left\|HF_t-\widetilde{F}'_t\right\|^4\right)^{\frac{d}{2}}=O_p\left(\left(b/T^{\frac{1}{2}}\right)^d\right),
\end{aligned}$$

which is $o_p\left(1\right)$ since $b/\sqrt{T}=o\left(1\right).$

- For Condition E*, since ν_j is independent over $j=1,\ldots,k$, with $Var^*\left(\nu_j\right)=1$, we have that

$$\begin{aligned}
\Omega^* &= \frac{1}{T}Var^*\left(\sum_{j=1}^{k}\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\widehat{\varepsilon}_{(j-1)b+l+h}\cdot\nu_j\right)\\
&= \frac{1}{T}\sum_{j=1}^{k}\left(\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\widehat{\varepsilon}_{(j-1)b+l+h}\right)\left(\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\widehat{\varepsilon}_{(j-1)b+l+h}\right)'\\
&\equiv I+II+II'+III,
\end{aligned}$$

where

$$\begin{aligned}
I &= \frac{1}{T}\sum_{j=1}^{k}\left(\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\varepsilon_{(j-1)b+l+h}\right)\left(\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\varepsilon_{(j-1)b+l+h}\right)' \\
II &= \frac{1}{T}\sum_{j=1}^{k}\left(\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\varepsilon_{(j-1)b+l+h}\right) \\
&\quad\times\left(\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\left[-\widehat{z}'_{(j-1)b+l}\left(\widehat{\delta}-\delta\right)+\alpha' H^{-1}\left(HF_{(j-1)b+l}-\widetilde{F}_{(j-1)b+l}\right)\right]\right)' \\
III &= \frac{1}{T}\sum_{j=1}^{k}\left(\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\left[-\widehat{z}'_{(j-1)b+l}\left(\widehat{\delta}-\delta\right)+\alpha' H^{-1}\left(HF_{(j-1)b+l}-\widetilde{F}_{(j-1)b+l}\right)\right]\right) \\
&\quad\times\left(\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\left[-\widehat{z}'_{(j-1)b+l}\left(\widehat{\delta}-\delta\right)+\alpha' H^{-1}\left(HF_{(j-1)b+l}-\widetilde{F}_{(j-1)b+l}\right)\right]\right)'.
\end{aligned}$$

Starting with I, and using the fact that $\widehat{z}_{(j-1)b+l}=\Phi z_{(j-1)b+l}+\left(\widehat{z}_{(j-1)b+l}-\Phi z_{(j-1)b+l}\right)$, we can write

$$I = I_1 + I_2 + I_2' + I_3$$

where

$$\begin{aligned}
I_1 &= \frac{1}{T}\Phi\sum_{j=1}^{k}\left(\sum_{l=1}^{b}z_{(j-1)b+l}\varepsilon_{(j-1)b+l+h}\right)\left(\sum_{l=1}^{b}z_{(j-1)b+l}\varepsilon_{(j-1)b+l+h}\right)'\Phi', \\
I_2 &= \frac{1}{T}\Phi\sum_{j=1}^{k}\left(\sum_{l=1}^{b}z_{(j-1)b+l}\varepsilon_{(j-1)b+l+h}\right)\left(\sum_{l=1}^{b}\left(\widehat{z}_{(j-1)b+l}-\Phi z_{(j-1)b+l}\right)\varepsilon_{(j-1)b+l+h}\right)'
\end{aligned}$$

and

$$I_3 = \frac{1}{T}\sum_{j=1}^{k}\left(\sum_{l=1}^{b}\left(\widehat{z}_{(j-1)b+l}-\Phi z_{(j-1)b+l}\right)\varepsilon_{(j-1)b+l+h}\right)\left(\sum_{l=1}^{b}\left(\widehat{z}_{(j-1)b+l}-\Phi z_{(j-1)b+l}\right)\varepsilon_{(j-1)b+l+h}\right)'.$$

For I_1, as $\Phi=\Phi_0+o_p(1)$, it is sufficient to show that

$$\frac{1}{T}\sum_{j=1}^{k}\left(\sum_{l=1}^{b}z_{(j-1)b+l}\varepsilon_{(j-1)b+l}\right)\left(\sum_{l=1}^{b}z_{(j-1)b+l}\varepsilon_{(j-1)b+l}\right)'\longrightarrow^{P}\Omega=\lim_{T\longrightarrow\infty}Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}z_t\varepsilon_{t+h}\right).$$

This follows by an application of Theorem 3.1 of Lahiri (2003). Indeed, under our assumptions for $r>2$, $\sum_{j=1}^{\infty}\alpha(j)^{\frac{r-2}{r}}<\infty$ and $b^{-1}+(T-h)^{-1}b\longrightarrow 0$. For I_2, by Cauchy-Schwarz inequality, we have that

$$\begin{aligned}
I_2 &\leq \|\Phi\|\left(\frac{1}{T}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}z_{(j-1)b+l}\varepsilon_{(j-1)b+l+h}\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}\left(\widehat{z}_{(j-1)b+l}-\Phi z_{(j-1)b+l}\right)\varepsilon_{(j-1)b+l+h}\right\|^2\right)^{\frac{1}{2}} \\
&= O_p\left(b^2/T\right)^{\frac{1}{4}}=o_p\left(1\right),
\end{aligned}$$

where the first factor is $O_p(1)$ by a mixingale inequality. For the second factor,

$$\frac{1}{T}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}\left(\widehat{z}_{(j-1)b+l}-\Phi z_{(j-1)b+l}\right)\varepsilon_{(j-1)b+l+h}\right\|^2 \leq \frac{b}{T}\sum_{j=1}^{k}\sum_{l=1}^{b}\left\|\left(\widehat{z}_{(j-1)b+l}-\Phi z_{(-j1)b+l}\right)\varepsilon_{(j-1)b+l+h}\right\|^2$$

$$\leq \frac{b}{T}\sum_{t=1}^{T-h}\left\|\left(\widehat{z}_t-\Phi z_t\right)\varepsilon_{t+h}\right\|^2 \leq \frac{b}{T}\left(\sum_{t=1}^{T-h}\left\|\widehat{z}_t-\Phi z_t\right\|^4\right)^{\frac{1}{2}}\left(\sum_{t=1}^{T-h}\varepsilon_{t+h}^4\right)^{\frac{1}{2}} = O_p\left(b/\sqrt{T}\right)$$

since $\sum_{t=1}^{T-h}\left\|\widehat{z}_t-\Phi z_t\right\|^4 = O_p(1)$ and $\sum_{t=1}^{T-h}\varepsilon_{t+h}^4 = O_p(T)$. The same argument can be used to show that $\|I_3\| = O_p\left(b/\sqrt{T}\right) = o_p(1)$. This shows that $I = o_p(1)$. Next, we show that $II = o_p(1)$. Letting $\widehat{X}_j \equiv \sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\varepsilon_{(j-1)b+l+h}$, we can write

$$\|II\| \leq \left(\frac{1}{T}\sum_{j=1}^{k}\left\|\widehat{X}_j\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\left(-\widehat{z}'_{(j-1)b+l}\left(\widehat{\delta}-\delta\right)+\alpha' H^{-1}\left(HF_{(j-1)b+l}-\widetilde{F}_{(j-1)b+l}\right)\right)\right\|^2\right)^{\frac{1}{2}}$$

where

$$\frac{1}{T}\sum_{j=1}^{k}\left\|\widehat{X}_j\right\|^2 = trace\left[\frac{1}{T}\sum_{j=1}^{k}\left(\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\varepsilon_{(j-1)b+l+h}\right)\left(\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\varepsilon_{(j-1)b+l+h}\right)'\right] = trace(I)$$

is $O_p(1)$ by the fact that $I = O_p(1)$. Using twice the c-r inequality, one can write

$$\begin{aligned}
&\frac{1}{T}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\left[-\widehat{z}'_{(j-1)b+l}\left(\widehat{\delta}-\delta\right)+\alpha' H^{-1}\left(HF_{(j-1)b+l}-\widetilde{F}_{(j-1)b+l}\right)\right]\right\|^2 \\
\leq\ & \frac{2}{T}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\widehat{z}'_{(j-1)b+l}\left(\widehat{\delta}-\delta\right)\right\|^2 + \frac{2}{T}\sum_{j=1}^{k}\left\|\sum_{l=1}^{b}\widehat{z}_{(j-1)b+l}\alpha' H^{-1}\left(HF_{(j-1)b+l}-\widetilde{F}_{(j-1)b+l}\right)\right\|^2 \\
\leq\ & II_1 + II_2,
\end{aligned}$$

where

$$\begin{aligned}
II_1 &= \frac{2b}{T}\sum_{j=1}^{k}\sum_{l=1}^{b}\left\|\widehat{z}_{(j-1)b+l}\widehat{z}'_{(j-1)b+l}\left(\widehat{\delta}-\delta\right)\right\|^2 \\
&= \frac{2b}{T}\sum_{t=1}^{T-h}\left\|\widehat{z}_t\widehat{z}'_t\left(\widehat{\delta}-\delta\right)\right\|^2 \leq \frac{2b}{T}\left\|\widehat{\delta}-\delta\right\|^2\sum_{t=1}^{T-h}\left\|\widehat{z}_t\right\|^4 \\
&= O\left(\frac{b}{T}\right)O_p\left(\frac{1}{T}\right)O_p(T) = O_p\left(\frac{b}{T}\right) = o_p(1),
\end{aligned}$$

and

$$
\begin{aligned}
II_2 &= \frac{2b}{T}\sum_{j=1}^{k}\sum_{l=1}^{b}\left\|\widehat{z}_{(j-1)b+l}\alpha' H^{-1}\left(HF_{(j-1)b+l}-\widetilde{F}_{(j-1)b+l}\right)\right\|^2 \\
&= \frac{2b}{T}\sum_{t=1}^{T-h}\left\|\widehat{z}_t\alpha' H^{-1}\left(HF_t-\widetilde{F}_t\right)\right\|^2 \\
&\leq \frac{2b}{T}\left\|\alpha' H^{-1}\right\|^2\left(\sum_{t=1}^{T-h}\left\|\widehat{z}_t\right\|^4\right)^{\frac{1}{2}}\left(\sum_{t=1}^{T-h}\left\|HF_t-\widetilde{F}_t\right\|^4\right)^{\frac{1}{2}} = O_p\left(b/\sqrt{T}\right),
\end{aligned}
$$

since in particular $\sum_{t=1}^{T-h}\left\|HF_t-\widetilde{F}_t\right\|^4 = O_p(1)$. Hence, $II = o_p(1)$.

Finally, note that $\|III\| \leq II_1 + II_2 = o_p(1)$, which completes the proof. ■

Proof of Theorem 3.2. The proof follows closely that of Theorem 3.1, so we only highlight the main differences. As in that proof, only Conditions C*, D* and E* of GP (2013) need to be verified, now with

$$
\varepsilon^*_{s+h} = \hat{\varepsilon}_{s+h}\cdot w^*_{s+h},
$$

where w^* is l_T-dependent with mean zero and covariance matrix K, a $(T-h)\times(T-h)$ matrix with typical element given by $K_{ij} = k_{dwb}\left(\frac{j-i}{l_T}\right)$, where $k_{dwb}(\cdot)$ is a kernel function and l_T is a bandwidth parameter. Conditions C*(a) and (b) follow immediately by noting that $E^*\left(\varepsilon^{*2}_{s+h}\right) = \hat{\varepsilon}_{s+h}$ since $Var^*(w^*) = K$ with diagonal elements equal to one. For Condition C*(c), we have that

$$
E^*\left(\frac{1}{T}\sum_{t=1}^{T-h}\varepsilon^{*4}_{t+h}\right) = \frac{1}{T}\sum_{t=1}^{T-h}\widehat{\varepsilon}^4_{t+h}E^*\left(w^{*4}_{t+h}\right) \leq M\frac{1}{T}\sum_{t=1}^{T-h}\widehat{\varepsilon}^4_{t+h} = M\frac{1}{T}\sum_{t=1}^{T-h}\widehat{\varepsilon}^4_{t+h} = O_P(1),
$$

since $E^*\left(w^{*4}_{t+h}\right) \leq M < \infty$. Condition D*(a) follows exactly as in the proof of Theorem 3.1, with w^*_t replacing v_j. To prove D*(b), noting that

$$
\varepsilon^*_{t+h} = \hat{\varepsilon}_{t+h}\cdot w^*_{t+h} = \varepsilon_{t+h}\cdot w^*_{t+h} - \widehat{z}'_t\left(\widehat{\delta}-\delta\right)\cdot w^*_{t+h} + \left(HF_t-\widetilde{F}_t\right)'\left(H^{-1}\right)'\alpha\cdot w^*_{t+h},
$$

and $\widehat{z}_t = \Phi z_t + \left(\widehat{z}_t - \Phi z_t\right)$, we have that

$$
\Omega^{*-\frac{1}{2}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\widehat{z}_t\varepsilon^*_{t+h} = J_1 + J_2 + J_3 + J_4,
$$

where

$$
\begin{aligned}
J_1 &= \Omega^{*-\frac{1}{2}}\frac{1}{\sqrt{T}}\Phi\sum_{t=1}^{T-h}z_t\varepsilon_{t+h}\cdot w^*_{t+h}, \\
J_2 &= \Omega^{*-\frac{1}{2}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\left(\widehat{z}_t-\Phi z_t\right)\varepsilon_{t+h}\cdot w^*_{t+h}, \\
J_3 &= -\Omega^{*-\frac{1}{2}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\widehat{z}_t\widehat{z}'_t\left(\widehat{\delta}-\delta\right)\cdot w^*_{t+h}
\end{aligned}
$$

and

$$J_4 = \Omega^{*-\frac{1}{2}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T-h} \widehat{z}_t \left(HF_t - \widetilde{F}_t \right)' \left(H^{-1} \right)' \alpha \cdot w_{t+h}^*.$$

We first show that $J_i = o_{p^*}(1)$ for $i = 2$, 3 and 4. Starting with J_2, we have that

$$\|J_2\| \leq \left\| \Omega^{*-\frac{1}{2}} \right\| \frac{1}{\sqrt{T}} \left\| \sum_{t=1}^{T-h} (\widehat{z}_t - \Phi z_t) \varepsilon_{t+h} \cdot w_{t+h}^* \right\| = O_{p^*}\left(\left(l_T^2 / T \right)^{1/4} \right) = o_{p^*}(1),$$

since $E^* \left\| \sum_{t=1}^{T-h} (\widehat{z}_t - \Phi z_t) \varepsilon_{t+h} \cdot w_{t+h}^* \right\|^2 = O_p\left(l_T T^{1/2} \right)$ (note that this term coincides with $B''_{2T,1}$ above) and $\Omega^* \to^p \Phi_0 \Omega \Phi'_0 > 0$, implying that $\Omega^* > 0$ with probability converging to one, as we will show next. For J_3, we have the following bound,

$$\|J_3\| \leq \left\| \Omega^{*-\frac{1}{2}} \right\| \frac{1}{\sqrt{T}} \left\| \sum_{t=1}^{T-h} \widehat{z}_t \widehat{z}'_t \left(\widehat{\delta} - \delta \right) \cdot w_{t+h}^* \right\| = O_{p^*}\left(\left(l_T / T \right)^{1/2} \right) = o_{p^*}(1),$$

where $E^* \left\| \sum_{t=1}^{T-h} \widehat{z}_t \widehat{z}'_t \left(\widehat{\delta} - \delta \right) \cdot w_{t+h}^* \right\|^2 = O_p(l_T)$ as we will show in the proof of Condition E*. Similarly, $J_4 = O_{p^*}\left(l_T^{\frac{1}{2}} / T^{\frac{1}{4}} \right) = o_{p^*}(1)$. It remains to show that $J_1 \to^{d^*} N(0, I)$ in probability. For this purpose, we use Theorem 3.1 of Shao (2010). Indeed, under our assumptions, Shao's (2010) assumptions are satisfied. In particular, as $\{z_t \varepsilon_{t+h}\}$ are strong mixing of size $-\frac{3r}{r-2}$ for some $r > 2$ with $E \|z_t \varepsilon_{t+h}\|^{2r} < C < \infty$, we have that $\sum_{j=1}^{\infty} (\alpha(j))^{\frac{r}{r+2}} < \infty$ wich implies his Assumption 3.1. We also have that $\sum_{j=1}^{\infty} j^2 (\alpha(j))^{\frac{r-2}{r}} < \infty$ and $E \|z_t \varepsilon_{t+h}\|^{2r} < C < \infty$, thus verifying Shao's Assumption 3.2 (by Lemma 1 of Andrews (1991)).

For Condition E*, following Lemma 2.1, we can write

$$\Omega^* = B_{1T} + B_{2T} + B_{3T} + B'_{2T} + B_{4T} + B_{5T} + B'_{3T} + B'_{5T} + B_{6T}, \text{ with}$$

$$
\begin{aligned}
B_{1T} &= \frac{1}{T}E^{*}\left[\left(\sum_{t=1}^{T-h}\widehat{z}_t\varepsilon_{t+h}w_{t+h}^{*}\right)\left(\sum_{t=1}^{T-h}\widehat{z}_t\varepsilon_{t+h}w_{t+h}^{*}\right)'\right]\\
B_{2T} &= \frac{1}{T}E^{*}\left[\left(\sum_{t=1}^{T-h}\widehat{z}_t\varepsilon_{t+h}w_{t+h}^{*}\right)\left(\sum_{t=1}^{T-h}\widehat{z}_t\widehat{z}_t'\left(\delta-\widehat{\delta}\right)w_{t+h}^{*}\right)'\right]\\
B_{3T} &= \frac{1}{T}E^{*}\left[\left(\sum_{t=1}^{T-h}\widehat{z}_t\varepsilon_{t+h}w_{t+h}^{*}\right)\left(\sum_{t=1}^{T-h}\widehat{z}_t\left(HF_t-\widetilde{F}_t\right)'\left(H^{-1}\right)'\alpha w_{t+h}^{*}\right)'\right]\\
B_{4T} &= \frac{1}{T}E^{*}\left[\left(\sum_{t=1}^{T-h}\widehat{z}_t\widehat{z}_t'\left(\delta-\widehat{\delta}\right)w_{t+h}^{*}\right)\left(\sum_{t=1}^{T-h}\widehat{z}_t\widehat{z}_t'\left(\delta-\widehat{\delta}\right)w_{t+h}^{*}\right)'\right]\\
B_{5T} &= \frac{1}{T}E^{*}\left[\left(\sum_{t=1}^{T-h}\widehat{z}_t\widehat{z}_t'\left(\delta-\widehat{\delta}\right)w_{t+h}^{*}\right)\left(\sum_{t=1}^{T-h}\widehat{z}_t\left(HF_t-\widetilde{F}_t\right)'\left(H^{-1}\right)'\alpha w_{t+h}^{*}\right)'\right]
\end{aligned}
$$

and

$$
B_{6T}=\frac{1}{T}E^{*}\left[\left(\sum_{t=1}^{T-h}\widehat{z}_t\left(HF_t-\widetilde{F}_t\right)'\left(H^{-1}\right)'\alpha w_{t+h}^{*}\right)\left(\sum_{t=1}^{T-h}\widehat{z}_t\left(HF_t-\widetilde{F}_t\right)'\left(H^{-1}\right)'\alpha w_{t+h}^{*}\right)'\right].
$$

We show that each of B_{iT}, $i=2,3,...,6$ are $o_p(1)$ and that B_{1T} converges in probability to $\Phi_0\Omega\Phi_0'$. Starting with B_{2T},

$$
\|B_{2T}\|\leq\frac{1}{T}\Bigg(\underbrace{E^{*}\left\|\sum_{t=1}^{T-h}\widehat{z}_t\varepsilon_{t+h}w_{t+h}^{*}\right\|^{2}}_{\equiv B_{2T,1}}\Bigg)^{\frac{1}{2}}\Bigg(\underbrace{E^{*}\left\|\sum_{t=1}^{T-h}\widehat{z}_t\widehat{z}_t'\left(\delta-\widehat{\delta}\right)w_{t+h}^{*}\right\|^{2}}_{\equiv B_{2T,2}}\Bigg)^{\frac{1}{2}},
$$

where

$$
\begin{aligned}
B_{2T,1} &= E^{*}\left\|\sum_{t=1}^{T-h}\Phi z_t\varepsilon_{t+h}w_{t+h}^{*}+\sum_{t=1}^{T-h}\left(\widehat{z}_t-\Phi z_t\right)\varepsilon_{t+h}w_{t+h}^{*}\right\|^{2}\\
&\leq 2E^{*}\left\|\sum_{t=1}^{T-h}\Phi z_t\varepsilon_{t+h}w_{t+h}^{*}\right\|^{2}+2E^{*}\left\|\sum_{t=1}^{T-h}\left(\widehat{z}_t-\Phi z_t\right)\varepsilon_{t+h}w_{t+h}^{*}\right\|^{2}\\
&\leq 2\underbrace{\|\Phi\|}_{=O_p(1)}{}^{2}\underbrace{E^{*}\left\|\sum_{t=1}^{T-h}z_t\varepsilon_{t+h}w_{t+h}^{*}\right\|^{2}}_{\equiv B_{2T,1}'}+\underbrace{2E^{*}\left\|\sum_{t=1}^{T-h}\left(\widehat{z}_t-\Phi z_t\right)\varepsilon_{t+h}w_{t+h}^{*}\right\|^{2}}_{\equiv B_{2T,1}''},
\end{aligned}
$$

with

$$\begin{aligned}
B'_{2T,1} &= T \cdot trace\left\{\frac{1}{T}E^*\left[\left(\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}w^*_{t+h}\right)\left(\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}w^*_{t+h}\right)'\right]\right\} \\
&= T \cdot trace\left(\frac{1}{T}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\left(z_t\varepsilon_{t+h}\right)\left(z'_s\varepsilon_{s+h}\right)E^*\left(w^*_{t+h}w^*_{s+h}\right)\right) \\
&= T \cdot trace\left(\frac{1}{T}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\left(z_t\varepsilon_{t+h}\right)\left(z'_s\varepsilon_{s+h}\right)k_{dwb}\left(\frac{t-s}{l_T}\right)\right) \\
&= O_P\left(T\right),
\end{aligned}$$

since the matrix inside the trace operator is a HAC estimator which converges to Ω under our assumptions. Also,

$$\begin{aligned}
B''_{2T,1} &= E^*\left\|\sum_{t=1}^{T-h}\left(\widetilde{F}_t - HF_t\right)\varepsilon_{t+h}w^*_{t+h}\right\|^2 \\
&= \sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\left(\widetilde{F}_t - HF_t\right)'\left(\widetilde{F}_s - HF_s\right)\varepsilon_{t+h}\varepsilon_{s+h}\underbrace{E^*\left(w^*_{t+h}w^*_{s+h}\right)}_{=k_{dwb}\left(\frac{t-s}{l_T}\right)} \\
&\leq 2\sum_{\tau=0}^{T-h}\left|k\left(\frac{\tau}{l_T}\right)\right|\left[\left(\sum_{t=1}^{T-h}\left\|\widetilde{F}_t - HF_t\right\|^2\varepsilon^2_{t+h}\right)^{\frac{1}{2}}\left(\sum_{t=1}^{T-h}\left\|\widetilde{F}_t - HF_t\right\|^2\varepsilon^2_{t+h}\right)^{\frac{1}{2}}\right] \\
&\leq 2l_T\left\{\underbrace{\left(\sum_{t=1}^{T-h}\left\|\widetilde{F}_t - HF_t\right\|^4\right)}_{=O_p(1)}\underbrace{\left(\sum_{t=1}^{T-h}\varepsilon^4_{t+h}\right)}_{=O_p(T)}\right\}^{\frac{1}{2}}\underbrace{\left\{\frac{1}{l_T}\sum_{\tau=0}^{T-h}\left|k_{dwb}\left(\frac{\tau}{l_T}\right)\right|\right\}}_{=O(1)} \\
&= O_p\left(l_T T^{\frac{1}{2}}\right).
\end{aligned}$$

Hence, $B_{2T,1} = O_p\left(T\right)$ given that under our assumptions $l_T = o\left(T^{1/2}\right)$. Similarly, we also have that

$$\begin{aligned}
B_{2T,2} &= E^*\left[\left(\sum_{t=1}^{T-h}\widehat{z}_t\widehat{z}'_t\left(\delta - \widehat{\delta}\right)w^*_{t+h}\right)'\left(\sum_{t=1}^{T-h}\widehat{z}_t\widehat{z}'_t\left(\delta - \widehat{\delta}\right)w^*_{t+h}\right)\right] \\
&= \left[\left(\delta - \widehat{\delta}\right)'\left(\sum_{s=1}^{T-h}\sum_{t=1}^{T-h}\widehat{z}_t\widehat{z}'_t\widehat{z}_s\widehat{z}'_sE^*\left(w^*_{t+h}w^*_{s+h}\right)\right)\left(\delta - \widehat{\delta}\right)\right]. \\
&\leq 2\left\|\delta - \widehat{\delta}\right\|^2\left(\sum_{\tau=0}^{T-h}\left|k_{dwb}\left(\frac{\tau}{l_T}\right)\right|\sum_{t=1}^{T-h}\left\|\widehat{z}_t\right\|^4\right) \\
&= O_p\left(T^{-1}\right)\cdot O_p\left(T\right)\cdot O\left(l_T\right) = O\left(l_T\right).
\end{aligned}$$

Hence,

$$B_{2T} = O_p\left(\frac{1}{T}T^{\frac{1}{2}} \cdot l_T^{\frac{1}{2}}\right) = O_p\left(\left(\frac{l_T}{T}\right)^{\frac{1}{2}}\right) = o_p(1),$$

since $\frac{l_T}{T} = o(1)$. For B_{3T}, by Cauchy-Schwarz inequality,

$$\|B_{3T}\| \leq \frac{1}{T}\left(E^*\left\|\sum_{t=1}^{T-h}\widehat{z}_t\varepsilon_{t+h}w_{t+h}^*\right\|^2\right)^{\frac{1}{2}}\left(E^*\left\|\sum_{t=1}^{T-h}\widehat{z}_t\left(HF_t - \widetilde{F}_t\right)'\left(H^{-1}\right)'\alpha w_{t+h}^*\right\|^2\right)^{\frac{1}{2}} = O_p\left(\left(l_T^2/T\right)^{\frac{1}{4}}\right).$$

Indeed, $E^*\left\|\sum_{t=1}^{T-h}\widehat{z}_t\varepsilon_{t+h}w_{t+h}^*\right\|^2 = O_p(T)$, whereas we have that

$$\begin{aligned}
& E^*\left\|\sum_{t=1}^{T-h}\widehat{z}_t\left(HF_t - \widetilde{F}_t\right)'\left(H^{-1}\right)'\alpha w_{t+h}^*\right\|^2 \leq \left\|H^{-1}\right\|^2\|\alpha\|^2 E^*\left\|\sum_{t=1}^{T-h}\widehat{z}_t\left(HF_t - \widetilde{F}_t\right)' w_{t+h}^*\right\|^2 \\
\leq\ & \left\|H^{-1}\right\|^2\|\alpha\|^2 \, trace\left[\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\left(HF_t - \widetilde{F}_t\right)\widehat{z}_t'\widehat{z}_s\left(HF_s - \widetilde{F}_s\right)' E^*\left(w_{t+h}^* w_{s+h}^*\right)\right] \\
\leq\ & \left\|H^{-1}\right\|^2\|\alpha\|^2\left[\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\widehat{z}_t'\widehat{z}_s\left(HF_s - \widetilde{F}_s\right)'\left(HF_t - \widetilde{F}_t\right)k_{dwb}\left(\frac{t-s}{l_T}\right)\right] \\
\leq\ & 2\left\|H^{-1}\right\|^2\|\alpha\|^2\left[\left(\sum_{t=1}^{T-h}\|\widehat{z}_t\|^4\right)^{\frac{1}{2}}\left(\sum_{t=1}^{T-h}\left\|HF_t - \widetilde{F}_t\right\|^4\right)^{\frac{1}{2}}\sum_{\tau=0}^{T-h}\left|k_{dwb}\left(\frac{\tau}{l_T}\right)\right|\right] = O_p\left(l_T \cdot T^{\frac{1}{2}}\right).
\end{aligned}$$

For B_{4T}, note that $\|B_{4T}\| \leq \frac{1}{T}E^*\left[\left\|\left(\sum_{t=1}^{T-h}\widehat{z}_t\widehat{z}_t'\left(\delta - \widehat{\delta}\right)w_{t+h}^*\right)\right\|^2\right] = O_p\left(\frac{l_T}{T}\right) = o_p(1)$. For B_{5T}, by Cauchy-Schwarz inequality, we have that

$$\begin{aligned}
\|B_{5T}\| &\leq \frac{1}{T}\left(E^*\left\|\left(\sum_{t=1}^{T-h}\widehat{z}_t\widehat{z}_t'\left(\delta - \widehat{\delta}\right)w_{t+h}^*\right)\right\|^2\right)^{1/2}\left(E^*\left\|\sum_{t=1}^{T-h}\widehat{z}_t\left(HF_t - \widetilde{F}_t\right)'\left(H^{-1}\right)'\alpha w_{t+h}^*\right\|^2\right)^{1/2} \\
&= O\left(T^{-1}\right)\cdot O_p\left(l_T^{\frac{1}{2}}\right)\cdot O_p\left(l_T^{\frac{1}{2}}T^{\frac{1}{4}}\right) = O_p\left(\frac{l_T}{T^{1/2}}\frac{1}{T^{1/4}}\right) = o_p(1),
\end{aligned}$$

given that $l_T = o\left(T^{1/2}\right)$. For B_{6T}, noting that

$$\|B_{6T}\| \leq \frac{1}{T}E^*\left\|\sum_{t=1}^{T-h}\widehat{z}_t\left(HF_t - \widetilde{F}_t\right)'\left(H^{-1}\right)'\alpha w_{t+h}^*\right\|^2,$$

where $E^*\left\|\sum_{t=1}^{T-h}\widehat{z}_t\left(HF_t-\widetilde{F}_t\right)'\left(H^{-1}\right)'\alpha w_{t+h}^*\right\|^2=O_p\left(l_T\cdot T^{\frac{1}{2}}\right)$, we have $B_{6T}=O_p\left(l_T/\sqrt{T}\right)=o_p(1)$. Finally, we have that

$$\begin{aligned}B_{1T} &= \frac{1}{T}E^*\left(\sum_{t=1}^{T-h}\widehat{z}_t\varepsilon_{t+h}w_{t+h}^*\right)\left(\sum_{t=1}^{T-h}\widehat{z}_t\varepsilon_{t+h}w_{t+h}^*\right)' \\ &= \frac{1}{T}\Phi E^*\left(\sum_{t=1}^{T-h}z_t\varepsilon_{t+h}w_{t+h}^*\right)\left(\sum_{t=1}^{T-h}z_t\varepsilon_{t+h}w_{t+h}^*\right)'\Phi'+R_T^*,\end{aligned}$$

where $\|R^*\|\leq\frac{1}{T}B_{2T,1}''=O_p\left(\frac{l_T}{T^{1/2}}\right)=o_p(1)$. For the first term, we have that

$$\frac{1}{T}E^*\left(\sum_{t=1}^{T-h}z_t\varepsilon_{t+h}w_{t+h}^*\right)\left(\sum_{t=1}^{T-h}z_t\varepsilon_{t+h}w_{t+h}^*\right)'=\frac{1}{T}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}z_tz_s'\varepsilon_{t+h}\varepsilon_{s+h}k_{dwb}\left(\frac{s-t}{l_T}\right),$$

which converges in probability towards Ω under our assumptions. This implies that $B_{1T}\longrightarrow_p\Phi_0\Omega\Phi_0'$. ■

Proof of Lemma 3.1. The proof follows closely that of Lemma 2.1 and therefore we omit the details. ■

References

[1] Andrews, D. W K, 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.

[2] Andrews, D. W K and C. J. Monahan, 1992. "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953-66.

[3] Bai, J., 2003. "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135-172.

[4] Bai, J. and S. Ng, 2002. "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191-221.

[5] Bai, J. and S. Ng, 2006. "Confidence Intervals for Diffusion Index Forecasts and Inference with Factor-augmented Regressions," *Econometrica*, 74, 1133-1150.

[6] Bai, J. and S. Ng, 2013. "Principal Components Estimation and Identification of the Factors," *Journal of Econometrics*, 176, 18-29.

[7] Cheng, X. and B. E. Hansen, 2013. "Forecasting with Factor-Augmented Regression: A Frequentist Model Averaging Approach," *Journal of Econometrics*, forthcoming.

[8] Davidson, J., 1994. *Stochastic Limit Theory: An Introduction for Econometricians*, Oxford University Press.

[9] De Jong, R. and J. Davidson, 2000. "Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices," *Econometrica*, 68, 407-424.

[10] Gallant, R. and H. White, 1987. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Blackwell.

[11] Gonçalves, S., B. Perron, and A. Djogbenou (2013). "Bootstrap prediction intervals for factor models," mimeo, Université de Montréal.

[12] Gonçalves, S. and B. Perron, 2013. "Bootstrapping Factor-Augmented Regressions," *Journal of Econometrics*, forthcoming.

[13] Hall, P., 1992. *The Bootstrap and Edgeworth Expansion*, Springer.

[14] Lahiri, S.N., 2003. *Resampling Methods for Dependent Data,* Springer-Verlag.

[15] Shao, X., 2010. "The Dependent Wild Bootstrap," *Journal of the American Statistical Association,* 05, 218-235.

[16] Shao, X., 2011. "A Bootstrap-assisted Spectral Test of White Noise under Unknown Dependence," *Journal of Econometrics,* 162, 213-224.

[17] 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.

[18] Smeekes, S. and J.-P. Urbain, 2013. "Unit Root Testing Using Modified Wild Bootstrap Methods," mimeo*,* Maastricht University.

[19] Yeh, A. B., 1998. "A Bootstrap Procedure in Linear Regression with Nonstationary Errors," *The Canadian Journal of Statistics Association*, 26(1), 149-160.

Table 1. Simulation results

			MA(h-1) errors								AR(1) errors			
			h = 1				h = 12				h = 1			
		N =	50		100		50		100		50		100	
		T =	50	100	50	100	50	100	50	100	50	100	50	100
Coverage rates for coefficient		OLS	56.9	54.3	75.8	77.8	68.7	71.2	75.3	81.0	61.5	68.5	72.1	79.5
		True Ft	92.0	93.8	91.7	93.5	80.5	86.0	80.6	85.7	77.7	84.8	78.7	86.0
	Symmetric t	WB	87.0	89.3	91.1	92.5	82.9	88.8	84.7	90.1	84.3	90.3	86.3	91.0
		BWB	86.9	89.4	90.9	92.4	84.3	88.9	86.0	90.5	85.1	90.0	87.3	91.2
		DWB	86.9	89.5	90.8	92.7	84.5	89.2	86.3	90.5	85.1	89.8	87.3	91.5
	Equal-tailed t	WB	89.1	91.0	90.7	92.7	74.1	80.6	74.5	81.2	75.5	80.6	77.0	82.7
		BWB	89.0	91.3	91.1	92.5	77.2	84.5	78.4	85.4	78.7	85.3	80.9	86.7
		DWB	89.0	90.9	90.5	92.4	77.9	85.0	79.0	85.9	78.9	85.7	81.3	87.1
Length of intervals		OLS	0.55	0.40	0.54	0.39	0.92	0.69	0.92	0.70	0.75	0.62	0.76	0.63
		True Ft	0.57	0.40	0.57	0.40	0.98	0.72	0.97	0.73	0.81	0.66	0.81	0.66
	Symmetric t	WB	0.99	0.72	0.82	0.57	1.45	1.16	1.23	0.96	1.38	1.15	1.17	0.93
		BWB	1.00	0.72	0.83	0.57	1.56	1.17	1.36	0.99	1.45	1.14	1.25	0.95
		DWB	1.00	0.73	0.83	0.57	1.56	1.17	1.35	1.00	1.45	1.14	1.25	0.95
	Equal-tailed t	WB	0.70	0.48	0.65	0.44	1.05	0.78	0.98	0.75	0.96	0.75	0.90	0.71
		BWB	0.71	0.48	0.65	0.44	1.20	0.87	1.15	0.85	1.09	0.84	1.04	0.80
		DWB	0.70	0.48	0.65	0.44	1.20	0.88	1.15	0.85	1.09	0.84	1.04	0.80
Bias		OLS	-0.21	-0.16	-0.14	-0.10	-0.20	-0.17	-0.13	-0.10	-0.21	-0.16	-0.14	-0.10
		True Ft	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		WB	-0.13	-0.12	-0.10	-0.08	-0.14	-0.12	-0.10	-0.08	-0.14	-0.12	-0.10	-0.08
		BWB	-0.13	-0.12	-0.10	-0.08	-0.14	-0.12	-0.10	-0.08	-0.14	-0.12	-0.10	-0.08
		DWB	-0.13	-0.12	-0.10	-0.08	-0.14	-0.12	-0.10	-0.08	-0.14	-0.12	-0.10	-0.08
Bandwidth choices		OLS	1.59	1.64	1.59	1.64	4.09	5.70	4.51	6.23	4.56	6.01	4.95	6.46
		WB	1.50	1.55	1.51	1.56	1.56	1.65	1.57	1.68	1.59	1.67	1.60	1.70
		BWB	1.56	1.62	1.57	1.63	2.64	3.75	2.99	4.29	2.85	3.90	3.32	4.47
		DWB	1.56	1.62	1.56	1.64	2.47	3.67	2.75	4.16	2.74	3.85	3.10	4.37