2016s-19

Bootstrap prediction intervals for factor models

Sílvia Gonçalves, Benoit Perron, Antoine Djogbenou

Série Scientifique/Scientific Series

2016s-19

Bootstrap prediction intervals for factor models

Sílvia Gonçalves, Benoit Perron, Antoine Djogbenou

Série Scientifique
Scientific Series

Montréal
Avril/April 2016

CIRANO
Allier savoir et décision

Centre interuniversitaire de recherche en analyse des organisations

ISSN 2292-0838 (en ligne)

Bootstrap prediction intervals for factor models*

Sílvia Gonçalves†*, Benoit Perron*‡*, Antoine Djogbenou*§

Résumé/abstract

We propose bootstrap prediction intervals for an observation *h* periods into the future and its conditional mean. We assume that these forecasts are made using a set of factors extracted from a large panel of variables. Because we treat these factors as latent, our forecasts depend both on estimated factors and estimated regression coefficients. Under regularity conditions, Bai and Ng (2006) proposed the construction of asymptotic intervals under Gaussianity of the innovations. The bootstrap allows us to relax this assumption and to construct valid prediction intervals under more general conditions. Moreover, even under Gaussianity, the bootstrap leads to more accurate intervals in cases where the cross-sectional dimension is relatively small as it reduces the bias of the OLS estimator as shown in a recent paper by Gonçalves and Perron (2014).

Mots clés/keywords : factor model, bootstrap, forecast, conditional mean

---

* We are grateful for comments from Lutz Kilian and seminar participants at the University of California, San Diego, the University of Southern California, Queen's, Pompeu Fabra, the Toulouse School of Economics, Duke, Western, Sherbrooke, and Michigan as well as from participants at the 23rd meeting of the (EC)^2, the 2013 North American Winter Meeting of the Econometric Society, the 2013 meeting of the Canadian Economics Association, the 2013 Joint Statistical Meetings, the conference "Bootstrap Methods for Time Series" in Copenhagen, the MAESG conference in Atlanta, and the 2014 Canadian Econometric Study Group. We also thank Tatevik Sekhposyan for providing the data for the empirical application. Gonçalves acknowledges financial support from the NSERC and MITACS whereas Perron acknowledges financial support from the SSHRC and MITACS.
† Département de sciences économiques, CIREQ and CIRANO, Université de Montréal.
‡ Département de sciences économiques, CIREQ and CIRANO, Université de Montréal.
§ Département de sciences économiques, CIREQ and CIRANO, Université de Montréal.

1 Introduction

Forecasting using factor-augmented regression models has become increasingly popular since the seminal paper of Stock and Watson (2002). The main idea underlying the so-called diffusion index forecasts is that when forecasting a given variable of interest, a large number of predictors can be summarized by a small number of indexes when the data follows an approximate factor model. The indexes are the latent factors driving the panel factor model and can be estimated by principal components. Point forecasts can be obtained by running a standard OLS regression augmented with the estimated factors.

In this paper, we consider the construction of prediction intervals in factor-augmented regression models using the bootstrap. In particular, our main contribution is to show the consistency of bootstrap intervals for a future target variable and its conditional mean. Our results allow for the construction of bootstrap prediction intervals without assuming Gaussianity and with better finite-sample properties than those based on asymptotic theory.

To be more specific, suppose that y_{t+h} denotes the variable to be forecast (where h is the forecast horizon) and let X_t be a N-dimensional vector of candidate predictors. We assume that y_{t+h} follows a factor-augmented regression model,

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

where W_t is a vector of observed regressors (including for instance lags of y_t) which jointly with F_t help forecast y_{t+h}. The r-dimensional vector F_t describes the common latent factors in the panel factor model,

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

where the $r \times 1$ vector λ_i contains the factor loadings and e_{it} is an idiosyncratic error term.

The goal is to forecast y_{T+h} or its conditional mean $y_{T+h|T} = \alpha' F_T + \beta' W_T$ using $\{(y_t, X_t, W_t) : t = 1, \ldots, T\}$, the available data at time T. Since factors are not observed, the diffusion index forecast approach typically involves a two-step procedure: in the first step we estimate F_t by principal components (yielding $\tilde{F}_t$) and in the second step we regress y_{t+h} on W_t and $\tilde{F}_t$ to obtain the regression coefficients. The point forecast is then constructed as $\hat{y}_{T+h|T} = \hat{\alpha}' \tilde{F}_T + \hat{\beta}' W_T$. Because we treat factors as latent, point forecasts depend both on estimated factors and regression coefficients. These two sources of parameter uncertainty must be accounted for when constructing prediction intervals and confidence intervals, as shown by Bai and Ng (2006).

Under regularity conditions, Bai and Ng (2006) derived the asymptotic distribution of regression estimates and the corresponding forecast errors and proposed the construction of asymptotic intervals. Our motivation for using the bootstrap as an alternative method of inference is twofold. First, 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. This was recently shown by Gonçalves and Perron (2014) in the context of confidence intervals for the regression coefficients, and as we will show below, the same is true in the context of prediction intervals. In particular, estimation of factors leads to an asymptotic bias term in the OLS estimator if $\sqrt{T}/N \to c$ and $c \neq 0$. Gonçalves and Perron (2014) proposed a bootstrap method that removes this bias and outperforms the asymptotic approach of Bai and Ng (2006). Second, the bootstrap allows for the construction of prediction intervals for y_{T+h} that are consistent under more general assumptions than the asymptotic approach of Bai and Ng (2006). In particular, the bootstrap does not require the Gaussianity assumption on the regression errors that justifies the asymptotic prediction intervals of Bai and Ng (2006). As our simulations show, prediction intervals based on the Gaussianity assumption perform poorly when the regression error is asymmetrically distributed whereas the bootstrap prediction intervals do not suffer significant size distortions.

We apply our procedure to forecasting inflation changes using quarterly observations on the US GDP deflator for the period 1973-2014. The resulting bootstrap intervals differ in interesting ways from the asymptotic ones in specific periods. In particular, the 95% equal-tailed percentile-t bootstrap intervals are shifted downwards and lie entirely below 0 following the financial crisis of 2008 and during the last quarter of 2011. These periods were marked by a significant concern of deflation. Our intervals are more consistent with such concerns than the asymptotic ones which include some probability of increasing inflation.

The remainder of the paper is organized as follows. Section 2 introduces our forecasting model and considers asymptotic prediction intervals. Section 3 describes two bootstrap prediction algorithms. Section 4 presents a set of high level assumptions on the bootstrap idiosyncratic errors under which the bootstrap distribution of the estimated factors at a given time period is consistent for the distribution of the sample estimated factors. These results together with the results of Gonçalves and Perron (2014) and Djogbenou, Gonçalves, and Perron (2014) regarding inference on the coefficients are used in Section 5 to show the asymptotic validity of wild bootstrap prediction intervals. Section 6 presents our simulation experiments, while Section 7 presents an empirical illustration of our methods. Finally, Section 8 concludes. Mathematical proofs appear in the Appendix.

2 Prediction intervals based on asymptotic theory

This section introduces our assumptions and reviews the asymptotic theory-based prediction intervals proposed by Bai and Ng (2006).

2.1 Assumptions

Let $z_t = \begin{pmatrix} F_t' & W_t' \end{pmatrix}'$, where z_t is $p \times 1$, with $p = r + q$. Following Bai and Ng (2006), we make the following assumptions.

Assumption 1

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

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

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

Assumption 2

(a) $E\left(e_{it}\right) = 0$, $E\left|e_{it}\right|^4 \leq M$.

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

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

(d) $\frac{1}{NT^2}\sum_{t,s,l,u}\sum_{i,j}\left|Cov\left(e_{it}e_{is}, e_{jl}e_{ju}\right)\right| < M < \infty$.

(e) For each t, $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it} \rightarrow^d N(0,\Gamma_t)$, where $\Gamma_t \equiv \lim_{N\rightarrow\infty} Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it}\right) > 0$.

Assumption 3 The variables $\{\lambda_i\}$, $\{F_t\}$ and $\{e_{it}\}$ are three mutually independent groups. Dependence within each group is allowed.

Assumption 4

(a) $E(\varepsilon_{t+h}) = 0$ and $E|\varepsilon_{t+h}|^4 < M$.

(b) $E(\varepsilon_{t+h}|y_t, z_t, y_{t-1}, z_{t-1}, \ldots) = 0$ for any $h > 0$, and (z_t', ε_t) are independent of the idiosyncratic errors e_{is} for all (i,s,t).

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

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

Assumptions 1 and 2 are standard in the approximate factors literature, allowing in particular for weak cross sectional and serial dependence in e_{it} of unknown form. Assumption 3 assumes independence among the factors, the factor loadings and the idiosyncratic error terms. We could allow for weak dependence among these three groups of variables at the cost of introducing restrictions on this dependence. Assumption 4 imposes moment conditions on $\{\varepsilon_{t+h}\}$, on $\{z_t\}$ and on the score vector $\{z_t\varepsilon_{t+h}\}$. Part c) requires $\{z_t z_t'\}$ to satisfy a law of large numbers. Part d) requires the score to satisfy a central limit theorem, where Ω denotes the limiting variance of the scaled average of the scores. We generalize the form of the covariance matrix assumed in Bai and Ng (2006) to allow for serial correlation as this will generally be the case when the forecast horizon is greater than 1.

2.2 Normal-theory intervals

As described in Section 1, the diffusion index forecasts are based on a two step estimation procedure. The first step consists of extracting the common factors $\tilde{F}_t$ from the N-dimensional panel X_t. In particular, given X, we estimate F and Λ with the method of principal components. F is estimated with the $T \times r$ matrix $\tilde{F} = \begin{pmatrix} \tilde{F}_1 & \ldots & \tilde{F}_T \end{pmatrix}'$ composed of $\sqrt{T}$ times the eigenvectors corresponding to the r largest eigenvalues of XX'/TN (arranged in decreasing order), where the normalization $\frac{\tilde{F}'\tilde{F}}{T} = I_r$ is used. The matrix containing the estimated loadings is then $\tilde{\Lambda} = \left(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_N\right)' = X'\tilde{F}\left(\tilde{F}'\tilde{F}\right)^{-1} = X'\tilde{F}/T$.

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

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

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

Suppose the object of interest is $y_{T+h|T}$, the conditional mean of $y_{T+h} = \alpha' F_T + \beta' W_T + \varepsilon_{T+h}$ at time T. The point forecast is $\hat{y}_{T+h|T} = \hat{\alpha}'\tilde{F}_T + \hat{\beta}' W_T$ and the forecast error is given by

$$\hat{y}_{T+h|T} - y_{T+h|T} = \frac{1}{\sqrt{T}} \hat{z}_T' \sqrt{T}\left(\hat{\delta} - \delta\right) + \frac{1}{\sqrt{N}} \alpha' H^{-1} \sqrt{N}\left(\tilde{F}_T - HF_T\right), \tag{4}$$

where $\delta \equiv \begin{pmatrix} \alpha' H^{-1} & \beta' \end{pmatrix}'$ is the probability limit of $\hat{\delta}$. The matrix H is defined as

$$H = \tilde{V}^{-1} \frac{\tilde{F}'F}{T} \frac{\Lambda'\Lambda}{N}, \tag{5}$$

where $\tilde{V}$ is the $r \times r$ diagonal matrix containing on the main diagonal the r largest eigenvalues of XX'/NT, in decreasing order (cf. Bai (2003)). It arises because factor models are only identified up to rotation, implying that the principal component estimator $\tilde{F}_t$ converges to HF_t, and the OLS estimator $\hat{\alpha}$ converges to $H^{-1\prime}\alpha$. It must be noted that forecasts do not depend on this rotation since the product is uniquely identified.

The above decomposition shows that the asymptotic distribution of the forecast error depends on two sources of uncertainty: the first is the usual parameter estimation uncertainty associated with estimation of α and β, and the second is the factors estimation uncertainty. Under Assumptions 1-4, and assuming that $\sqrt{T}/N \rightarrow 0$ and $\sqrt{N}/T \rightarrow 0$ as $N, T \rightarrow \infty$, Bai and Ng (2006) show that the studentized forecast error

$$\frac{\hat{y}_{T+h|T} - y_{T+h|T}}{\sqrt{\hat{B}_T}} \rightarrow^d N(0,1), \tag{6}$$

where $\hat{B}_T$ is a consistent estimator of the asymptotic variance of $\hat{y}_{T+h|T}$ given by

$$\hat{B}_T = \widehat{Var}\left(\hat{y}_{T+h|T}\right) = \frac{1}{T}\hat{z}_T'\hat{\Sigma}_\delta \hat{z}_T + \frac{1}{N}\hat{\alpha}'\hat{\Sigma}_{\tilde{F}_T}\hat{\alpha}. \tag{7}$$

Here, $\hat{\Sigma}_\delta$ consistently estimates $\Sigma_\delta = Var\left(\sqrt{T}\left(\hat{\delta} - \delta\right)\right)$ and $\hat{\Sigma}_{\tilde{F}_T}$ consistently estimates $\Sigma_{\tilde{F}_T} = Var\left(\sqrt{N}\left(\tilde{F}_T - HF_T\right)\right)$. In particular, under Assumptions 1-4,

$$\hat{\Sigma}_\delta = \left(T^{-1}\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t'\right)^{-1}\hat{\Omega}_T\left(T^{-1}\sum_{t=1}^{T-h}\hat{z}_t\hat{z}_t'\right)^{-1}, \tag{8}$$

where $\hat{\Omega}_T$ is a heteroskedasticity and autocorrelation consistent (HAC) estimator of

$\Omega = \lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h} z_t\varepsilon_{t+h}\right)$, and

$$\hat{\Sigma}_{\tilde{F}_T} = \tilde{V}^{-1}\tilde{\Gamma}_T\tilde{V}^{-1}, \tag{9}$$

where $\tilde{\Gamma}_T$ is an estimator of $\Gamma_T = \lim_{N\to\infty} Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{iT}\right)$ which depends on the cross sectional dependence and heterogeneity properties of e_{iT}. Bai and Ng (2006) provide three different estimators of Γ_T. Section 5 below considers such an estimator.

The central limit theorem result in (6) justifies the construction of an asymptotic $100(1-\alpha)\%$ level confidence interval for $y_{T+h|T}$ given by

$$\left(\hat{y}_{T+h|T} - z_{1-\alpha/2}\sqrt{\hat{B}_T}, \hat{y}_{T+h|T} + z_{1-\alpha/2}\sqrt{\hat{B}_T}\right), \tag{10}$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of a standard normal distribution.

When the object of interest is a prediction interval for y_{T+h}, Bai and Ng (2006) propose

$$\left(\hat{y}_{T+h|T} - z_{1-\alpha/2}\sqrt{\hat{C}_T}, \hat{y}_{T+h|T} + z_{1-\alpha/2}\sqrt{\hat{C}_T}\right), \tag{11}$$

where

$$\hat{C}_T = \hat{B}_T + \hat{\sigma}_\varepsilon^2,$$

with $\hat{B}_T$ as above and $\hat{\sigma}_\varepsilon^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2$. The validity of (11) depends on the additional assumption that ε_t is i.i.d. $N\left(0,\sigma_\varepsilon^2\right)$.

An important condition that justifies (10) and (11) is that $\sqrt{T}/N \to 0$. This condition ensures that the term reflecting the parameter estimation uncertainty in the forecast error decomposition (4), $\sqrt{T}\left(\hat{\delta}-\delta\right)$, is asymptotically normal with a mean of zero and a variance-covariance matrix that does not depend on the factors estimation uncertainty. As was recently shown by Gonçalves and Perron (2014), when $\sqrt{T}/N \to c \neq 0$,

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

where Δ_δ is a bias term that reflects the contribution of the factors estimation error to the asymptotic distribution of the regression estimates $\hat{\delta}$. In this case, the two terms in (4) will depend on the factors estimation uncertainty and a natural question is whether this will have an effect on the prediction intervals (10) and (11) derived by Bai and Ng (2006) under the assumption that $c = 0$. As we argue next, these intervals remain valid even when $c \neq 0$. The main reason is that when $\sqrt{T}/N \to c \neq 0$, the ratio $N/T \to 0$, which implies that the parameter estimation uncertainty associated with δ is dominated asymptotically by the uncertainty from having to estimate F_T.

More formally, when $\sqrt{T}/N \to c \neq 0$, $N/T \to 0$ and the convergence rate of $\hat{y}_{T+h|T}$ is $\sqrt{N}$, implying that

$$\begin{aligned}\sqrt{N}\left(\hat{y}_{T+h|T}-y_{T+h|T}\right) &= \sqrt{N/T}\sqrt{T}\left(\hat{\delta}-\delta\right)'\hat{z}_T+\alpha' H^{-1}\sqrt{N}\left(\tilde{F}_T-HF_T\right) \\ &= \alpha' H^{-1}\sqrt{N}\left(\tilde{F}_T-HF_T\right)+o_P(1).\end{aligned}$$

Thus, the forecast error is asymptotically $N\left(0, \alpha' H^{-1}\Sigma_{\tilde{F}_T}H^{-1\prime}\alpha\right)$. Since $N\hat{B}_T = (N/T)\,\hat{z}_T'\hat{\Sigma}_\delta\hat{z}_T + \hat{\alpha}'\hat{\Sigma}_{\tilde{F}_T}\hat{\alpha} = \alpha' H^{-1}\Sigma_{\tilde{F}_T}H^{-1\prime}\alpha + o_P(1)$, the studentized forecast error given in (6) is still $N(0,1)$ as $N, T \to \infty$. For the studentized forecast error associated with forecasting y_{T+h}, the variance of $\hat{y}_{T+h}$ is asymptotically (as $N, T \to \infty$) dominated by the variance of the error term σ_ε^2, implying that neither the parameter estimation uncertainty nor the factors estimation uncertainty contribute to the asymptotic variance.

3 Description of bootstrap intervals

Following Gonçalves and Perron (2014), we consider the following bootstrap data-generating process:

$$X_t^* = \tilde{\Lambda}\tilde{F}_t + e_t^*, \tag{12}$$

$$y_{t+h}^* = \hat{\alpha}'\tilde{F}_t + \hat{\beta}'W_t + \varepsilon_{t+h}^*, \tag{13}$$

where $\left\{e_t^* = (e_{1t}^*, \ldots, e_{Nt}^*)'\right\}$ denotes a bootstrap sample from $\left\{\tilde{e}_t = X_t - \tilde{\Lambda}\tilde{F}_t\right\}$ and $\left\{\varepsilon_{t+h}^*\right\}$ is a resampled version of $\left\{\hat{\varepsilon}_{t+h} = y_{t+h} - \hat{\alpha}'\tilde{F}_t - \hat{\beta}'W_t\right\}$.

Our goal in this section is to describe two general bootstrap algorithms that can be used to compute intervals for $y_{T+h|T}$ and y_{T+h} for *any* choice of $\{e_t^*\}$ and $\left\{\varepsilon_{t+h}^*\right\}$. The specific method of generating $\{e_t^*\}$ and $\left\{\varepsilon_{t+h}^*\right\}$ will depend on the assumptions we make on $\{e_{it}\}$ and $\{\varepsilon_{t+h}\}$, respectively. In Section 5 we describe several methods. For example, we rely on the wild bootstrap to generate both $\{e_t^*\}$ and $\left\{\varepsilon_{t+1}^*\right\}$ when constructing confidence intervals for $y_{T+1|T}$. The wild bootstrap is justified in this setting since we assume away cross sectional dependence in e_{it} and we assume that ε_{t+1} is a m.d.s. when $h = 1$. For one-step ahead prediction intervals we strengthen the m.d.s. assumption to an i.i.d. assumption on ε_{t+1}, and therefore we generate ε_{t+1}^* using the i.i.d. bootstrap. For multi-step prediction intervals, we generate ε_{t+h}^* with either the block wild bootstrap or the dependent wild bootstrap of Djogbenou et al. (2014) to account for possible serial correlation.

We estimate the factors by the method of principal components using the bootstrap panel data set $\{X_t^* : t = 1, \ldots, T\}$. We let $\tilde{F}^* = \left(\tilde{F}_1^*, \ldots, \tilde{F}_T^*\right)'$ denote the $T \times r$ matrix of bootstrap estimated factors which equal the r eigenvectors of $X^* X^{*\prime}/NT$ (multiplied by $\sqrt{T}$) corresponding to the r largest eigenvalues. The $N \times r$ matrix of estimated bootstrap loadings is given by $\tilde{\Lambda}^* = \left(\tilde{\lambda}_1^*, \ldots, \tilde{\lambda}_N^*\right)' = X^{*\prime} \tilde{F}^*/T$. We then run a regression of y_{t+h}^* on $\tilde{F}_t^*$ and W_t using observations $t = 1, \ldots, T - h$. We let $\hat{\delta}^*$ denote the corresponding OLS estimator

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

where $\hat{z}_t^* = \left(\tilde{F}_t^{*\prime}, W_t'\right)'$.

The steps for obtaining a bootstrap confidence interval for $y_{T+h|T}$ are as follows.

Algorithm 1 (Bootstrap confidence interval for $y_{T+h|T}$)

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

$$X_t^* = \tilde{\Lambda} \tilde{F}_t + e_t^*,$$

   where $\{e_{it}^*\}$ is a resampled version of $\left\{\tilde{e}_{it} = X_{it} - \tilde{\lambda}_i' \tilde{F}_t\right\}$.

2. Estimate the bootstrap factors $\left\{\tilde{F}_t^* : t = 1, \ldots, T\right\}$ 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 $\hat{\varepsilon}_{t+h}$.

4. Regress y^*_{t+h} generated in step 3 on the bootstrap estimated factors $\tilde{F}^*_t$ obtained in step 2 and on the fixed regressors W_t and obtain the OLS estimator $\hat{\delta}^*$.

5. Obtain bootstrap forecasts

$$\hat{y}^*_{T+h|T} = \hat{\alpha}^{*\prime}\tilde{F}^*_T + \hat{\beta}^{*\prime}W_T \equiv \hat{\delta}^{*\prime}\hat{z}^*_T,$$

and bootstrap variance

$$\hat{B}^*_T = \frac{1}{T}\hat{z}^{*\prime}_T\hat{\Sigma}^*_\delta\hat{z}^*_T + \frac{1}{N}\hat{\alpha}^{*\prime}\hat{\Sigma}^*_{\tilde{F}_T}\hat{\alpha}^*, \tag{14}$$

where the choice of $\hat{\Sigma}^*_\delta$ and $\hat{\Sigma}^*_{\tilde{F}_T}$ depends on the properties of ε^*_{t+h} and e^*_{it}.

6. Let $y^*_{T+h|T} = \hat{\alpha}'\tilde{F}_T + \hat{\beta}'W_T$ and compute bootstrap prediction errors:

   (a) For equal-tailed percentile-t bootstrap intervals, compute studentized bootstrap prediction errors as

   $$s^*_{T+h} = \frac{\hat{y}^*_{T+h|T} - y^*_{T+h|T}}{\sqrt{\hat{B}^*_T}}.$$

   (b) For symmetric percentile-t bootstrap intervals, compute $\left|s^*_{T+h}\right|$.

7. Repeat this process B times, resulting in statistics $\left\{s^*_{T+h,1}, \ldots, s^*_{T+h,B}\right\}$ and $\left\{\left|s^*_{T+h,1}\right|, \ldots, \left|s^*_{T+h,B}\right|\right\}$.

8. Compute the corresponding empirical quantiles:

   (a) For equal-tailed percentile-t bootstrap intervals, $q^*_{1-\alpha}$ is the empirical $1-\alpha$ quantile of $\left\{s^*_{T+h,1}, \ldots, s^*_{T+h,B}\right\}$.

   (b) For symmetric percentile-t bootstrap intervals, $q^*_{|\cdot|,1-\alpha}$ is the empirical $1-\alpha$ quantile of $\left\{\left|s^*_{T+h,1}\right|, \ldots, \left|s^*_{T+h,B}\right|\right\}$.

A $100(1-\alpha)\%$ equal-tailed percentile-t bootstrap interval for $y_{T+h|T}$ is given by

$$EQ^{1-\alpha}_{y_{T+h|T}} \equiv \left(\hat{y}_{T+h|T} - q^*_{1-\alpha/2}\sqrt{\hat{B}_T}, \hat{y}_{T+h|T} - q^*_{\alpha/2}\sqrt{\hat{B}_T}\right), \qquad (15)$$

whereas a $100(1-\alpha)\%$ symmetric percentile-t bootstrap interval for $y_{T+h|T}$ is given by

$$SY^{1-\alpha}_{y_{T+h|T}} \equiv \left(\hat{y}_{T+h|T} - q^*_{|\cdot|,1-\alpha}\sqrt{\hat{B}_T}, \hat{y}_{T+h|T} + q^*_{|\cdot|,1-\alpha}\sqrt{\hat{B}_T}\right), \qquad (16)$$

When prediction intervals for a new observation y_{T+h} are the object of interest, the algorithm reads as follows.

Algorithm 2 (Bootstrap prediction interval for y_{T+h})

1. Identical to Algorithm 1.

2. Identical to Algorithm 1.

3. Generate $\left\{y^*_{1+h}, \ldots, y^*_T, y^*_{T+1}, \ldots, y^*_{T+h}\right\}$ using

$$y^*_{t+h} = \hat{\alpha}'\tilde{F}_t + \hat{\beta}'W_t + \varepsilon^*_{t+h},$$

   where $\left\{\varepsilon^*_{1+h}, \ldots, \varepsilon^*_T, \varepsilon^*_{T+1}, \ldots, \varepsilon^*_{T+h}\right\}$ is a bootstrap sample obtained from $\{\hat{\varepsilon}_{1+h}, \ldots, \hat{\varepsilon}_T\}$.

4. Not making use of the stretch $\left\{y^*_{T+1}, \ldots, y^*_{T+h}\right\}$, compute $\hat{\delta}^*$ as in Algorithm 1.

5. Obtain the bootstrap point forecast $\hat{y}^*_{T+h|T}$ as in Algorithm 1 but compute its variance as

$$\hat{C}^*_T = \hat{B}^*_T + \hat{\sigma}^{*2}_{\varepsilon},$$

where $\hat{\sigma}_{\varepsilon}^{*2}$ is a consistent estimator of $\sigma_{\varepsilon}^{2} = Var\left(\varepsilon_{T+h}\right)$ and $\hat{B}_T^*$ is as in Algorithm 1.

6. Let $y_{T+h}^* = \hat{\alpha}'\tilde{F}_T + \hat{\beta}'W_T + \varepsilon_{T+h}^*$ and compute bootstrap prediction errors:

   (a) For equal-tailed percentile-t bootstrap intervals, compute studentized bootstrap prediction errors as

$$s_{T+h}^* = \frac{\hat{y}_{T+h|T}^* - y_{T+h}^*}{\sqrt{\hat{C}_T^*}}.$$

   (b) For symmetric percentile-t bootstrap intervals, compute $\left|s_{T+h}^*\right|$.

7. Identical to Algorithm 1.

8. Identical to Algorithm 1.

A $100(1-\alpha)\,\%$ equal-tailed percentile-t bootstrap interval for y_{T+h} is given by

$$EQ_{y_{T+h}}^{1-\alpha} \equiv \left(\hat{y}_{T+h|T} - q_{1-\alpha/2}^*\sqrt{\hat{C}_T}, \hat{y}_{T+h|T} - q_{\alpha/2}^*\sqrt{\hat{C}_T}\right), \tag{17}$$

whereas a $100(1-\alpha)\,\%$ symmetric percentile-t bootstrap interval for y_{T+h} is given by

$$SY_{y_{T+h}}^{1-\alpha} \equiv \left(\hat{y}_{T+h|T} - q_{|\cdot|,1-\alpha}^*\sqrt{\hat{C}_T}, \hat{y}_{T+h|T} + q_{|\cdot|,1-\alpha}^*\sqrt{\hat{C}_T}\right). \tag{18}$$

The main differences between the two algorithms is that in step 3 of Algorithm 2 we generate observations for y_{t+h}^* for $t = 1, \ldots, T$ instead of stopping at $t = T-h$. This allows us to obtain a bootstrap observation for y_{T+h}^*, the bootstrap analogue of y_{T+h}, which we will use in constructing the studentized statistic s_{T+h}^* in step 6 of Algorithm 2. The point forecast is identical to Algorithm 1 and relies only on observations for $t = 1, \ldots, T-h$, but the bootstrap variance $\hat{C}_T^*$ contains an extra term $\hat{\sigma}_\varepsilon^{*2}$ that reflects the uncertainty associated with the error of the new observation ε_{T+h}.

Note that Algorithm 2 generates bootstrap point forecasts $\hat{y}^*_{T+h|T}$ and bootstrap future observations y^*_{T+h} that are conditional on W_T. This is important because the point forecast $\hat{y}_{T+h|T}$ depends on W_T. When W_t contains lagged dependent variables (e.g. $W_t = y_t$ and $h = 1$), steps 5 and 6 of Algorithm 2 set $W_T = y_T$ when computing $\hat{y}^*_{T+1|T}$ and y^*_{T+1}. This is effectively equivalent to setting $y_T^* = y_T$ for the purposes of computing these quantities. However, Step 3 of Algorithm 2 generates observations on $\left\{y^*_{t+1} : t = 1, \ldots, T\right\}$ that do not necessarily satisfy the requirement that $y_T^* = y_T$. As recently discussed by Pan and Politis (2014), we can account for parameter estimation uncertainty in predictions generated by autoregressive models by relying on a forward bootstrap method that contains two steps: one step generates the bootstrap data by relying on the forward representation of the model. This step accounts for parameter estimation uncertainty even if $y_T^* \neq y_T$. In a second step, we evaluate the bootstrap prediction and future observation conditional on the last value(s) of the observed variable. Our Algorithm 2 can be viewed as a version of the forward bootstrap method of Pan and Politis (2014) when some of the regressors are latent factors that need to be estimated.

4 Bootstrap distribution of estimated factors

The asymptotic validity of the bootstrap intervals for y_{T+h} and $y_{T+h|T}$ described in the previous section depends on the ability of the bootstrap to capture two sources of estimation error: the parameter estimation error and the factors estimation error. In particular, the bootstrap estimation error for the conditional mean is given by

$$\hat{y}^*_{T+h|T} - y^*_{T+h|T} = \frac{1}{\sqrt{T}} \hat{z}^{*\prime}_T \sqrt{T} \left(\hat{\delta}^* - \delta^* \right) + \frac{1}{\sqrt{N}} \hat{\alpha}' H^{*-1} \sqrt{N} \left(\tilde{F}^*_T - H^* F_T \right),$$

where $\delta^* = \Phi^{*\prime -1} \hat{\delta}$ and $\Phi^* = diag\left(H^*, I_q\right)$. Here, H^* is the bootstrap analogue of the rotation matrix H defined in (5), i.e.

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

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 (2014) for more details.

Adding and subtracting appropriately, we can write

$$\hat{y}^*_{T+h|T} - y^*_{T+h|T} = \frac{1}{\sqrt{T}} \hat{z}'_T \sqrt{T} \left(\Phi^{*\prime} \hat{\delta}^* - \hat{\delta} \right) + \frac{1}{\sqrt{N}} \hat{\alpha}' \sqrt{N} \left(H^{*-1} \tilde{F}^*_T - \tilde{F}_T \right) + o_{P^*}(1). \tag{19}$$

As usual in the bootstrap literature, we use P^* to denote the bootstrap probability measure, conditional on a given sample; E^* and Var^* denote the corresponding bootstrap expected value and variance operators. For any bootstrap statistic T^*_{NT}, we write $T^*_{NT} = o_{P^*}(1)$, in probability, or $T^*_{NT} \to^{P^*} 0$, in probability, when for any $\delta > 0$, $P^*\left(|T^*_{NT}| > \delta\right) = o_P(1)$. We write $T^*_{NT} = O_{P^*}(1)$, in probability, when for all $\delta > 0$ there exists $M_\delta < \infty$ such that $\lim_{N,T \to \infty} P\left[P^*\left(|T^*_{NT}| > M_\delta\right) > \delta\right] = 0$. Finally, we write $T^*_{NT} \to^{d^*} D$, in probability, if conditional on a sample with probability that converges to one, T^*_{NT} weakly converges to the distribution D under P^*, i.e. $E^*\left(f\left(T^*_{NT}\right)\right) \to^P E\left(f\left(D\right)\right)$ for all bounded and uniformly continuous functions f. See Chang and Park (2003) for similar notation and for several useful bootstrap asymptotic properties.

The stochastic expansion (19) shows that the bootstrap estimation error captures the two forms of estimation uncertainty in (4) provided: (1) the bootstrap distribution of $\sqrt{T}\left(\Phi^{*\prime}\hat{\delta}^* - \hat{\delta}\right)$ is a consistent estimator of the distribution of $\sqrt{T}\left(\hat{\delta} - \delta\right)$, and (2) the bootstrap distribution of $\sqrt{N}\left(H^{*-1}\tilde{F}^*_T - \tilde{F}_T\right)$ is a consistent estimator of the distribution of $\sqrt{N}\left(\tilde{F}_T - HF_T\right)$. Gonçalves and Perron (2014) discussed conditions for the consistency of the bootstrap distribution of $\sqrt{T}\left(\hat{\delta} - \delta\right)$. Here we propose a set of conditions that justifies using the bootstrap to consistently estimate the distribution of the estimated factors $\sqrt{N}\left(\tilde{F}_t - HF_t\right)$ at each point t.

Condition $\mathcal{A}$.

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

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

$\mathcal{A}$.3. For each t, $E^*\left\|\frac{1}{\sqrt{TN}}\sum_{s=1}^T\sum_{i=1}^N \tilde{F}_s\left(e^*_{it}e^*_{is} - E^*\left(e^*_{it}e^*_{is}\right)\right)\right\|^2 = O_P(1)$.

$\mathcal{A}$.4. $E^*\left\|\frac{1}{\sqrt{TN}}\sum_{t=1}^T\sum_{i=1}^N \tilde{F}_t\tilde{\lambda}'_i e^*_{it}\right\|^2 = O_P(1)$.

$\mathcal{A}$.5. $\frac{1}{T}\sum_{t=1}^T E^*\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N \tilde{\lambda}_i e^*_{it}\right\|^2 = O_P(1)$.

$\mathcal{A}$.6. For each t, $\Gamma^{*-1/2}_t \frac{1}{\sqrt{N}}\sum_{i=1}^N \tilde{\lambda}_i e^*_{it} \to^{d^*} N(0, I_r)$, in probability, where $\Gamma^*_t = Var^*\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \tilde{\lambda}_i e^*_{it}\right)$ is uniformly positive definite.

Condition $\mathcal{A}$ is the bootstrap analogue of Bai's (2003) assumptions used to derive the limiting distribution of $\sqrt{N}\left(\tilde{F}_t - HF_t\right)$. Gonçalves and Perron (2014) also relied on similar high level assumptions to study the bootstrap distribution of $\sqrt{T}\left(\hat{\delta}^{*}-\delta^{*}\right)$. In particular, Conditions $\mathcal{A}.4$ and $\mathcal{A}.5$ correspond to their Conditions B*(c) and B*(d), respectively. Since our goal here is to characterize the limiting distribution of the bootstrap estimated factors at each point t, we need to complement some of their other conditions by requiring boundedness in probability of some bootstrap moments at each point in time t (in addition to boundedness in probability of the time average of these bootstrap moments; e.g. Conditions $\mathcal{A}.1$ and $\mathcal{A}.2$ expand Conditions A*(b) and A*(c) in Gonçalves and Perron (2014) in this manner). We also require that a central limit theorem applies to the scaled cross sectional average of $\tilde{\lambda}_i e_{it}^{*}$, at each time t (Condition $\mathcal{A}.6$). This high level condition ensures asymptotic normality for the bootstrap estimated factors. It was not required by Gonçalves and Perron (2014) because their goal was only to consistently estimate the distribution of the regression estimates, not of the estimated factors.

Theorem 4.1 *Suppose Assumptions 1 and 2 hold. Under Condition $\mathcal{A}$, as $N, T \to \infty$ such that $\sqrt{N}/T^{3/4} \to 0$, we have that for each t,*

$$\sqrt{N}\left(\tilde{F}_t^{*}-H^{*}\tilde{F}_t\right)=H^{*}\tilde{V}^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\lambda}_i e_{it}^{*}+o_{P^{*}}(1),$$

in probability, which implies that

$$\Pi_t^{*-1/2}\sqrt{N}\left(H^{*-1}\tilde{F}_t^{*}-\tilde{F}_t\right)\to^{d^{*}} N\left(0, I_r\right),$$

in probability, where $\Pi_t^{*}=\tilde{V}^{-1}\Gamma_t^{*}\tilde{V}^{-1}$.

Theorem 1.(i) of Bai (2003) shows that under regularity conditions weaker than Assumptions 1 and 2 and provided $\sqrt{N}/T \to 0$, $\sqrt{N}\left(\tilde{F}_t-HF_t\right)\to^{d} N\left(0, \Pi_t\right)$, where $\Pi_t=V^{-1}Q\Gamma_t Q'V^{-1}$,

$Q = p\lim\left(\frac{\tilde{F}'F}{T}\right)$. Theorem 4.1 is its bootstrap analogue. A stronger rate condition ($\sqrt{N}/T^{3/4} \to 0$ instead of $\sqrt{N}/T \to 0$) is used to show that the remainder terms in the stochastic expansion of $\sqrt{N}\left(\tilde{F}_t^* - H^*\tilde{F}_t\right)$ are asymptotically negligible. This rate condition is a function of the number of finite moments for F_s we assume. In particular, if we replace Assumption 1(a) with $E\left\|F_t\right\|^q \leq M$ for all t, then the required rate restriction is $\sqrt{N}/T^{1-1/q} \to 0$. See Remarks 1 and 3 below.

To prove the consistency of Π_t^* for Π_t we impose the following additional condition.

Condition $\mathcal{B}$. For each t, $p\lim\Gamma_t^* = Q\Gamma_t Q'$.

Condition $\mathcal{B}$ requires that Γ_t^*, the bootstrap variance of the scaled cross sectional average of the scores $\tilde{\lambda}_i e_{it}^*$, be consistent for $Q\Gamma_t Q'$. This in turn requires that we resample $\tilde{e}_{it}$ in a way that preserves the cross sectional dependence and heterogeneity properties of e_{it}.

Corollary 4.1 *Under Assumptions 1 and 2 and Conditions $\mathcal{A}$ and $\mathcal{B}$, we have that for each t, as $N, T \to \infty$ such that $\sqrt{N}/T^{3/4} \to 0$, $\sqrt{N}\left(H^{*-1}\tilde{F}_t^* - \tilde{F}_t\right) \to^{d^*} N\left(0, \Pi_t\right)$, in probability, where $\Pi_t = V^{-1}Q\Gamma_t Q'V^{-1}$ is the asymptotic covariance matrix of $\sqrt{N}\left(\tilde{F}_t - HF_t\right)$.*

Corollary 4.1 justifies using the bootstrap to construct confidence intervals for the rotated factors HF_t provided Conditions $\mathcal{A}$ and $\mathcal{B}$ hold. These conditions are high level conditions that can be checked for any particular bootstrap scheme used to generate e_{it}^*. We verify them for a wild bootstrap in Section 5 when proving the consistency of bootstrap confidence intervals for the conditional mean.

The fact that factors and factor loadings are not separately identified implies the need to rotate the bootstrap estimated factors in order to consistently estimate the distribution of the sample factor estimates, i.e. we use $\sqrt{N}\left(H^{*-1}\tilde{F}_t^* - \tilde{F}_t\right)$ to approximate the distribution of $\sqrt{N}\left(\tilde{F}_t - HF_t\right)$. A similar rotation was discussed in Gonçalves and Perron (2014) in the context of bootstrapping the regression coefficients $\hat{\delta}$.

5 Validity of bootstrap intervals

5.1 Confidence intervals for $y_{T+1|T}$

We begin by considering intervals for next period's conditional mean. For this purpose, we use a two-step wild bootstrap scheme, as in Gonçalves and Perron (2014). Specifically, we rely on Algorithm 1 and we let

$$\varepsilon^*_{t+1} = \hat{\varepsilon}_{t+1} \cdot v_{t+1}, \quad t = 1, \ldots, T-1, \tag{20}$$

with v_{t+1} i.i.d.$(0,1)$, and

$$e^*_{it} = \tilde{e}_{it} \cdot \eta_{it}, \quad t = 1, \ldots, T, \ i = 1, \ldots, N, \tag{21}$$

where η_{it} is i.i.d.$(0,1)$ across (i,t), independently of v_{t+1}.

To prove the asymptotic validity of this method we strengthen Assumptions 1-4 as follows.

Assumption 5. λ_i are either deterministic such that $\|\lambda_i\| \leq M < \infty$, or stochastic such that $E\|\lambda_i\|^{12} \leq M < \infty$ for all i; $E\|F_t\|^{12} \leq M < \infty$; $E|e_{it}|^{12} \leq M < \infty$, for all (i,t); and for some $q > 1$, $E|\varepsilon_{t+1}|^{4q} \leq M < \infty$, for all t.

Assumption 6. $E(e_{it}e_{js}) = 0$ if $i \neq j$.

With $h = 1$, our Assumption 4(b) on ε_{t+h} becomes a martingale difference sequence assumption, and the wild bootstrap in (20) is natural. This assumption rules out serial correlation in

ε_{t+1} but allows for conditional heteroskedasticity. Below, we consider the case where $h > 1$.

Assumption 6 assumes the absence of cross sectional correlation in the idiosyncratic errors and motivates the use of the wild bootstrap in (21). As the results in the previous sections show, prediction intervals for y_{T+h} or $y_{T+h|T}$ are a function of the factors estimation uncertainty even when this source of uncertainty is asymptotically negligible for the estimation of the distribution of the regression coefficients (i.e. even when $\sqrt{T}/N \to c = 0$). Since factors estimation uncertainty depends on the cross sectional correlation of the idiosyncratic errors e_{it} (via $\Gamma_T = \lim_{N\to\infty} Var\left(1/\sqrt{N}\sum_{i=1}^N \lambda_i e_{iT}\right)$), bootstrap prediction intervals need to mimic this form of correlation to be asymptotically valid. Contrary to the pure time series context, a natural ordering does not exist in the cross sectional dimension, which implies that proposing a nonparametric bootstrap method (e.g. a block bootstrap) that replicates the cross sectional dependence is challenging if a parametric model is not assumed. Therefore, we follow Gonçalves and Perron (2014) and use a wild bootstrap to generate e_{it}^* under Assumption 6.

The bootstrap percentile-t method, as described in Algorithm 1 and equations (15) and (16), requires the choice of two variances, $\hat{B}_T$ and its bootstrap analogue $\hat{B}_T^*$. To compute $\hat{B}_T$ we use (7), where $\hat{\Sigma}_\delta$ is given in (8). $\hat{\Sigma}_{\tilde{F}_T}$ is given in (9), where

$$\tilde{\Gamma}_T = \frac{1}{N}\sum_{i=1}^N \tilde{\lambda}_i \tilde{\lambda}_i' \tilde{e}_{iT}^2$$

is estimator 5(a) in Bai and Ng (2006), and it is a consistent estimator of (a rotated version of) $\Gamma_T = \lim_{N\to\infty} Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \lambda_i e_{iT}\right)$ under Assumption 6. We compute $\hat{B}_T^*$ using (14) and relying on the heteroskedasticity-robust bootstrap analogues of $\hat{\Sigma}_\delta$ and $\hat{\Sigma}_{\tilde{F}_T}$.

Theorem 5.1 *Suppose Assumptions 1-6 hold and we use Algorithm 1 with* $\varepsilon_{t+1}^* = \hat{\varepsilon}_{t+1} \cdot v_{t+1}$

and $e^*_{it} = \tilde{e}_{it} \cdot \eta_{it}$, *where* $v_{t+1} \sim i.i.d.(0,1)$ *for all* $t = 1, \ldots, T-1$ *and* $\eta_{it} \sim i.i.d.(0,1)$ *for all* $i = 1, \ldots, N$; $t = 1, \ldots, T$, *and* v_{t+1} *and* η_{it} *are mutually independent. Moreover, assume that* $E^* |\eta_{it}|^4 < C$ *for all* (i,t) *and* $E^* |v_{t+1}|^4 < C$ *for all* t. *If* $\sqrt{T}/N \to c$, *where* $0 \leq c < \infty$, *and* $\sqrt{N}/T^{11/12} \to 0$, *then conditional on* $\{y_t, X_t, W_t : t = 1, \ldots, T\}$,

$$\frac{\hat{y}^*_{T+1|T} - y^*_{T+1|T}}{\sqrt{\hat{B}^*_T}} \to^{d^*} N(0,1),$$

in probability.

Remark 1 *The rate restriction* $\sqrt{N}/T^{11/12} \to 0$ *is slightly stronger than the rate used by Bai (2003) (cf.* $\sqrt{N}/T \to 0$*). It is weaker than the restriction* $\sqrt{N}/T^{3/4} \to 0$ *used in Theorem 4.1 and Corollary 4.1 because we have strengthened the number of factor moments that exist from 4 to 12 (compare Assumption 5 with Assumption 1(a)). See Remark 3 in the Appendix.*

Remark 2 *Since* $\frac{\hat{y}_{T+1|T} - y_{T+1|T}}{\sqrt{\hat{B}_T}} \to^d N(0,1)$, *as shown by Bai and Ng (2006), Theorem 5.1 implies that bootstrap confidence intervals for* $y_{T+1|T}$ *obtained with Algorithm 1 have the correct coverage probability asymptotically.*

5.2 Prediction intervals for y_{T+1}

In this section we provide a theoretical justification for bootstrap prediction intervals for y_{T+1} as described in Algorithm 2. In particular, our goal is to prove that a bootstrap prediction interval contains the future observation y_{T+1} with unconditional probability that converges to the nominal level as $N, T \to \infty$.

We add the following assumption.

Assumption 7. ε_{t+1} is i.i.d.$(0, \sigma^2_\varepsilon)$ with a continuous distribution function $F_\varepsilon(x) = P(\varepsilon_{t+1} \leq x)$.

Assumption 7 strengthens the m.d.s. Assumption 4.(b) by requiring the regression errors to be i.i.d. However, and contrary to Bai and Ng (2006), F_ε does not need to be Gaussian. The continuity assumption on F_ε is used below to prove that the Kolmogorov distance between the bootstrap distribution of the studentized forecast error and the distribution of its sample analogue converges in probability to zero.

Let the studentized forecast error be defined as

$$s_{T+1} \equiv \frac{\hat{y}_{T+1|T} - y_{T+1}}{\sqrt{\hat{B}_T + \hat{\sigma}_\varepsilon^2}},$$

where $\hat{\sigma}_\varepsilon^2$ is a consistent estimate of $\sigma_\varepsilon^2 = Var\left(\varepsilon_{T+1}\right)$ and $\hat{B}_T = \widehat{Var}\left(\hat{y}_{T+1|T}\right) = \frac{1}{T}\hat{z}_T'\hat{\Sigma}_\delta \hat{z}_T + \frac{1}{N}\hat{\alpha}'\hat{\Sigma}_{\tilde{F}_T}\hat{\alpha}$. Given Assumption 7, we can use

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{T}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1}^2 \quad \text{and} \quad \hat{\Sigma}_\delta = \hat{\sigma}_\varepsilon^2\left(\frac{1}{T}\sum_{t=1}^{T-1}\hat{z}_t\hat{z}_t'\right)^{-1}. \tag{22}$$

Our goal is to show that the bootstrap can be used to estimate consistently $F_{T,s}\left(x\right) = P\left(s_{T+1} \leq x\right)$, the distribution function of s_{T+1}. Note that we can write

$$\begin{aligned} \hat{y}_{T+1|T} - y_{T+1} &= \left(\hat{y}_{T+1|T} - y_{T+1|T}\right) + \left(y_{T+1|T} - y_{T+1}\right) \\ &= -\varepsilon_{T+1} + O_P\left(1/\delta_{NT}\right), \end{aligned}$$

given that $\hat{y}_{T+1|T} - y_{T+1|T} = O_P\left(\frac{1}{\delta_{NT}}\right)$, where $\delta_{NT} = \min\left(\sqrt{N}, \sqrt{T}\right)$ (this follows under the assumptions of Theorem 5.1). Since $\hat{\sigma}_\varepsilon^2 \to^P \sigma_\varepsilon^2$ and $\hat{B}_T = O_P\left(1/\delta_{NT}^2\right) = o_P\left(1\right)$, it follows that

$$s_{T+1} = -\frac{\varepsilon_{T+1}}{\sigma_\varepsilon} + o_P\left(1\right). \tag{23}$$

Thus, as $N, T \to \infty$, s_{T+1} converges in distribution to the random variable $-\frac{\varepsilon_{T+1}}{\sigma_\varepsilon}$, i.e.

$$F_{T,s}(x) \equiv P(s_{T+1} \le x) \to P\left(-\frac{\varepsilon_{T+1}}{\sigma_\varepsilon} \le x\right) = 1 - F_\varepsilon(-x\sigma_\varepsilon) \equiv F_{\infty,s}(x),$$

for all $x \in \mathbb{R}$. If we assume that ε_{t+1} is i.i.d. $N(0, \sigma_\varepsilon^2)$, as in Bai and Ng (2006), then $F_\varepsilon(-x\sigma_\varepsilon) = \Phi(-x) = 1 - \Phi(x)$, implying that $F_{T,s}(x) \to \Phi(x)$, i.e. $s_{T+1} \to^d N(0,1)$. Nevertheless, this is not generally true unless we make the Gaussianity assumption. We note that although asymptotically the variance of the prediction error $\hat{y}_{T+1|T} - y_{T+1}$ does not depend on any parameter nor factors estimation uncertainty (as it is dominated by σ_ε^2 for large N and T), we still suggest using $\hat{C}_T = \hat{B}_T + \hat{\sigma}_\varepsilon^2$ to studentize $\hat{y}_{T+1|T} - y_{T+1}$ since $\hat{\sigma}_\varepsilon^2$ will underestimate the true forecast variance for finite T and N. Politis (2013) and Pan and Politis (2014) discuss notions of asymptotic validity that require taking into account the estimation of the condition mean. More specifically, in addition to requiring that the interval contains the true observation with the desired nominal coverage probability asymptotically, they require the bootstrap to capture parameter estimation uncertainty. To the extent that their definitions can be extended to the case of generated regressors, we expect our bootstrap intervals to satisfy these stricter notions of validity.

Next we show that the bootstrap yields a consistent estimate of the distribution of s_{T+1} without assuming that ε_{t+1} is Gaussian. Our proposal is based on a two-step residual based bootstrap scheme, as described in Algorithm 2 and equations (17) and (18), where in step 3 we generate $\left\{\varepsilon_2^*, \ldots, \varepsilon_T^*, \varepsilon_{T+1}^*\right\}$ as a random sample obtained from the centered residuals $\left\{\hat{\varepsilon}_2 - \bar{\hat{\varepsilon}}, \ldots, \hat{\varepsilon}_T - \bar{\hat{\varepsilon}}\right\}$. Resampling in an i.i.d. fashion is justified under Assumption 7. We recenter the residuals because $\bar{\hat{\varepsilon}}$ is not necessarily zero unless W_t contains a constant regressor.

Nevertheless, since $\bar{\hat{\varepsilon}} = o_P(1)$, resampling the uncentered residuals is also asymptotically valid in our context. We compute $\hat{B}_T^*$ and $\hat{\sigma}_\varepsilon^{*2}$ using the bootstrap analogues of $\hat{\Sigma}_\delta$ and $\hat{\sigma}_\varepsilon^2$ introduced in (22). Note that $\hat{\sigma}_\varepsilon^{*2}$ is a consistent estimator of σ_ε^2 and $\hat{B}_T^* = o_{P^*}(1)$, in probability.

As above, we can write

$$\begin{aligned} \hat{y}_{T+1|T}^* - y_{T+1}^* &= \left(\hat{y}_{T+1|T}^* - y_{T+1|T}^*\right) + \left(y_{T+1|T}^* - y_{T+1}^*\right) \\ &= -\varepsilon_{T+1}^* + O_{P^*}\left(1/\delta_{NT}\right), \end{aligned}$$

in probability, which in turn implies

$$s_{T+1}^* \equiv \frac{\hat{y}_{T+1|T}^* - y_{T+1}^*}{\sqrt{\hat{B}_T^* + \hat{\sigma}_\varepsilon^{*2}}} = -\frac{\varepsilon_{T+1}^*}{\sigma_\varepsilon} + o_{P^*}(1). \tag{24}$$

Thus, $F_{T,s}^*(x) = P^*\left(s_{T+1}^* \leq x\right)$, the bootstrap distribution of s_{T+1}^* (conditional on the sample) is asymptotically the same as the bootstrap distribution of $-\frac{\varepsilon_{T+1}^*}{\sigma_\varepsilon}$.

Let $F_{T,\varepsilon}^*$ denote the bootstrap distribution function of ε_t^*. It is clear from the stochastic expansions (23) and (24) that the crucial step is to show that ε_{T+1}^* converges weakly in probability to ε_{T+1}, i.e. $d\left(F_{T,\varepsilon}^*, F_\varepsilon\right) \to^P 0$ for any metric that metrizes weak convergence. In the following we use Mallows metric which is defined as $d_2(F_X, F_Y) = \left(\inf\left(E|X-Y|^2\right)\right)^{1/2}$ over all joint distributions for the random variables X and Y having marginal distributions F_X and F_Y, respectively.

Lemma 5.1 *Under Assumptions 1-7, and as $T, N \to \infty$ such that $\sqrt{T}/N \to c$, $0 \leq c < \infty$, $d_2\left(F_{T,\varepsilon}^*, F_\varepsilon\right) \to^P 0$.*

Corollary 5.1 *Under the same assumptions as Theorem 5.1 strengthened by Assumption 7,* *we have that*

$$\sup_{x\in\mathbb{R}} \left|F^*_{T,s}(x) - F_{\infty,s}(x)\right| \to 0,$$

in probability.

Corollary 5.1 implies the asymptotic validity of the bootstrap prediction intervals given in (17) and (18), where asymptotic validity means that the interval contains y_{T+1} with unconditional probability converging to the nominal level asymptotically. Specifically, we can show that $P\left(y_{T+1} \in EQ^{1-\alpha}_{y_{T+1}}\right) \to 1-\alpha$ and $P\left(y_{T+1} \in SY^{1-\alpha}_{y_{T+1}}\right) \to 1-\alpha$ as $N, T \to \infty$. See e.g. Beran (1987) and Wolf and Wunderli (2015, Proposition 1). For instance,

$$\begin{aligned} P\left(y_{T+1} \in EQ^{1-\alpha}_{y_{T+1}}\right) &= P\left(s_{T+1} \leq q^*_{1-\alpha/2}\right) - P\left(s_{T+1} \leq q^*_{\alpha/2}\right) \\ &= P\left(F^*_{T,s}(s_{T+1}) \leq 1-\alpha/2\right) - P\left(F^*_{T,s}(s_{T+1}) \leq \alpha/2\right). \end{aligned}$$

Given Corollary 5.1, we have that $F^*_{T,s}(s_{T+1}) = F_{\infty,s}(s_{T+1}) + o_P(1)$, and we can show that $F_{\infty,s}(s_{T+1}) \to^d U[0,1]$. Indeed, for any x,

$$P\left(F_{\infty,s}(s_{T+1}) \leq x\right) = P\left(s_{T+1} \leq F^{-1}_{\infty,s}(x)\right) \equiv F_{T,s}\left(F^{-1}_{\infty,s}(x)\right) \to F_{\infty,s}\left(F^{-1}_{\infty,s}(x)\right) = x.$$

A stronger result than that implied by Corollary 5.1 would be to prove that $P\left(y_{T+1} \in EQ^{1-\alpha}_{y_{T+1}} | z_T\right) \to 1-\alpha$, where $z_T = (F'_T, W'_T)'$. Nevertheless, to claim asymptotic validity of the bootstrap prediction intervals conditional on the regressors would require stronger assumptions, namely the assumption that ε_{T+1} is independent of z_T. Such a strong exogeneity assumption is unlikely to be satisfied in economics.

5.3 Multi-horizon forecasting, $h > 1$

Finally, we consider the case where the forecasting horizon, h, is larger than 1. The main complication in this case is the fact that the regression errors ε_{t+h} in the factor-augmented regression will generally be serially correlated to order $h-1$. This serial correlation affects the distribution of $\sqrt{T}\left(\hat{\delta}-\delta\right)$ since the form of Ω is different in this case, as it includes autocovariances of the score process.

We modify our two algorithms above by drawing ε^*_{t+h} using the block wild bootstrap (BWB) algorithm proposed in Djogbenou et al. (2014). The idea is to separate the sample residuals $\hat{\varepsilon}_{t+h}$ into non-overlapping blocks of b consecutive observations. For simplicity, we assume that $\frac{T-h}{b}$, the number of such blocks, is an integer. Then, we generate our bootstrap errors by multiplying each residual within a block by the *same* draw of an external variable, i.e.

$$\varepsilon^*_{i+(j-1)b} = \hat{\varepsilon}_{i+(j-1)b}\eta_j$$

for $j = 1, \ldots, \frac{T-h}{b}$, $i = 1+h, \ldots, h+b$, and $\eta_j \sim i.i.d.\,(0,1)$. The fact that each residual within a block is multiplied by the same external draw preserves the time series dependence. We let $b = h$ because we use the fact that $\varepsilon_{t+h} \sim MA\,(h-1)$ under Assumption 4(b). For $h = 1$, this algorithm is the same as the wild bootstrap. Djogbenou et al. (2014) show that this algorithm allows for valid bootstrap inference in a regression model with estimated factors and general mixing conditions on the error term. The moving average structure obtained in a forecasting context (assuming correct specification) obviously satisfies these mixing conditions, and this ensures that this block wild bootstrap algorithm replicates the distribution of $\sqrt{T}\left(\hat{\delta}-\delta\right)$ after rotating the estimated parameter in the bootstrap world. Thus, the result of Theorem 5.1 holds in this more general context since $h > 1$ does not affect factor estimation.

For the forecast of the new observation, y_{T+h}, the crucial condition for asymptotic validity of the bootstrap prediction intervals is to capture the marginal distribution of $\varepsilon_{T+h.}$. This means that the i.i.d. bootstrap can still be used in step 2 of algorithm 2 to generate ε_{t+h}^* despite the serial correlation in ε_{t+h}. Alternatively, we can also amend the block wild bootstrap by generating $\hat{\varepsilon}_{t+h}^*$ as above for $t = 1, \ldots, T-h$ and generating ε_{T+h}^* as a draw from the empirical distribution function of $\hat{\varepsilon}_t$, $t = 1, \ldots, T-h$. We will compare these two approaches in the simulation experiment below.

6 Simulations

In this section, we report results from a simulation experiment to analyze the properties of the normal asymptotic intervals as well as their bootstrap counterparts analyzed above. The data-generating process is similar to the one used in Gonçalves and Perron (2014). We consider the single factor model:

$$y_{t+h} = .5F_t + \varepsilon_{t+h} \tag{25}$$

where F_t is an autoregressive process:

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

with u_t drawn from a normal distribution independently over time with a variance of $(1 - .8^2)$. We use the backward representation of this autoregressive process to make sure that all sample paths have $F_T = 1$. We will consider two forecasting horizons, $h = 1$ and $h = 4$.

The regression error ε_{t+h} will be homoskedastic with expectation 0, variance 1 and will have a moving average structure to accommodate multi-horizon forecasting:

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

and to analyze the effects of deviations from normality, we report results for two distributions for v_t :

$$\begin{aligned} \text{Normal:} \quad & v_t \sim \left(\frac{1}{\sum_{j=0}^{h-1} .8^{2j}} \right) N(0,1) \\ \text{Mixture} \quad : \quad & v_t \sim \left(\frac{1}{\sum_{j=0}^{h-1} .8^{2j}} \right) \frac{1}{\sqrt{10}} \left[pN(-1,1) + (1-p) N(9,1) \right], \end{aligned}$$

where p is distributed as $Bernoulli\,(.9)$. The particular mixture distribution we are using is similar to the one proposed by Pascual, Romo and Ruiz (2004). Most of the data is drawn from a $N(-1,1)$ but about 10% will come from a second normal with a much larger mean of 9. The scaling term in parentheses ensures that the variance of ε_{t+h} is 1 regardless of h. We have also considered other distributions such as the uniform, exponential, and χ^2 but do not report these results for brevity.

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%. Asymptotic inference is conducted by using a HAC estimator (quadratic spectral kernel with bandwidth set to h) to account for possible serial correlation.

We use Algorithms 1 and 2 described above to generate the bootstrap data with $B = 999$ bootstrap replications. The idiosyncratic errors are always drawn using the wild bootstrap in step 1. In step 3, three bootstrap schemes are analyzed to draw ε_t^*: the first one draws the residuals with replacement in an i.i.d. fashion, the second one uses the wild bootstrap, while the last one redraws the residuals using the block wild bootstrap with a block size equal to h. The first two methods are only valid when $h = 1$, while the last one is valid for both values of h. In all applications of the wild bootstrap and block wild bootstrap, the external variable has a standard normal distribution. With the wild bootstrap, we use the heteroskedasticity-robust variance estimator, while we use the HAC one with block size equal to h for the block wild bootstrap.

We consider two types of bootstrap intervals: symmetric percentile-t and equal-tailed percentile-t. We report experiments based on 5,000 replications and with three values for T (50, 100, and 200) and 4 values for N (50, 100, 150, and 200).

We report results graphically for the conditional mean $y_{T+h|T}$ and for the new observation y_{T+h}. We report the frequency of times the 95% confidence interval is to the left or right of the true parameter. Each figure has three rows corresponding to $T = 50$, $T = 100$, and $T = 200$ with N on the horizontal axis, and in the last column, we show the average length of the corresponding confidence intervals relative to the length of the "ideal"confidence intervals obtained with the 2.5% and 97.5% quantiles from the empirical distribution simulated for each N and T 1,000,000 times as endpoints. To keep the figures readable, we report results for two bootstrap methods in each figure. For the conditional mean, we report results for the wild bootstrap and block wild bootstrap (with differences thus only coming from the block size since the two methods are the same for a block size equal to 1). For the observation, we report results using the iid and block wild bootstrap since we require an i.i.d. assumption for the construction of intervals for this quantity.

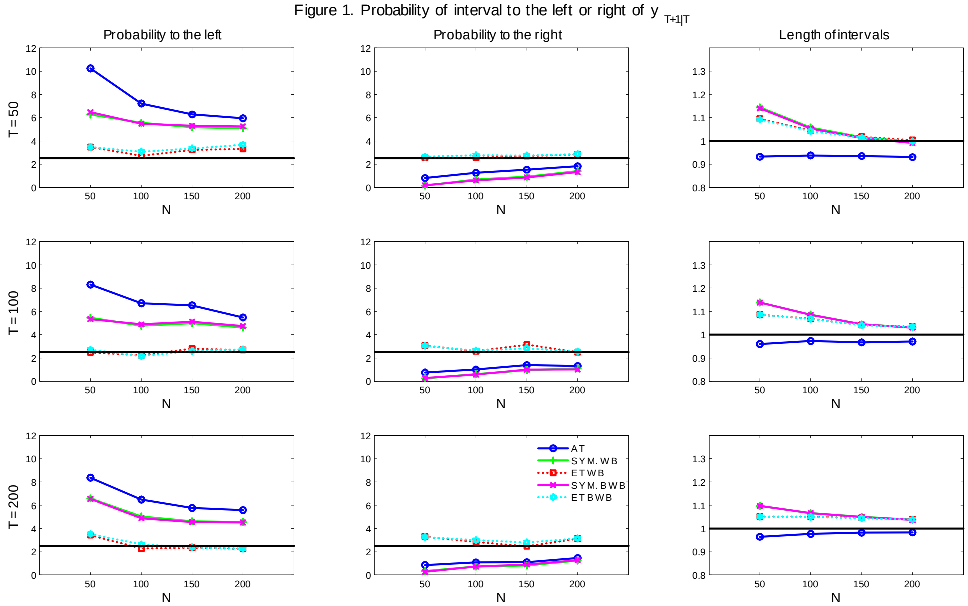
It turns out that the distribution of ε_{T+h} noticeably affects the results for y_{T+h} only. As a consequence, we only report results with Gaussian ε_{t+h} for the conditional mean. On the other hand, the results of y_{T+h} are dominated by the behavior of ε_{t+h}. Thus, the contribution of the conditional mean from the contribution of ε_{T+h} in the forecasts of y_{T+h} are clearly separated.

6.1 Forecasting horizon $h = 1$

We start by presenting results when we are interested in making a prediction for next period's value. For this horizon, because ε_{t+1} does not have serial correlation, the wild bootstrap and block wild bootstrap methods are identical with reported differences due to simulation error.

Conditional mean, $y_{T+1|T}$ The results for the conditional mean are presented in Figure 1. Asymptotic theory (blue line) shows large distortions that decrease with an increasing N. For example, for $N = T = 50$, the 95% confidence interval does not include the true mean in 11% of the replications instead of the nominal 5%. This number is reduced to 7.8% when $N = 200$ and $T = 50$. Moreover, we see that most of these instances are in one direction, when the confidence interval is to the left of the true value. This can be explained by a bias in the estimation of the parameter δ as documented by Gonçalves and Perron (2014) due to the estimation uncertainty in the factors. This bias is negative, thus shifting the distribution of the conditional mean to the left, leading to more rejections on the left side and fewer on the right side than predicted by the asymptotic normal distribution.

Figure 1. Probability of interval to the left or right of $y_{T+1|T}$



Note: The figures in the first two columns report the fraction of confidence intervals that lie to the left or to the right of the conditional mean for each method as a function of the cross-sectional dimension N. Each row corresponds to a different time series dimension. The last column reports the length of the confidence intervals relative to the length of the "ideal" intervals obtained as the 2.5% and 97.5% quantiles of the empirical distribution.

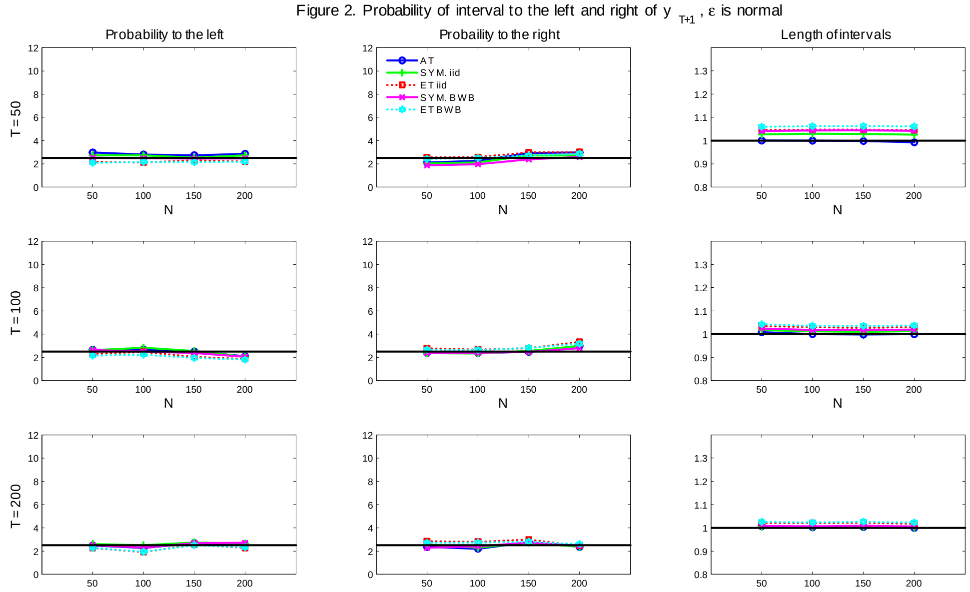
The presence of bias is reflected in the bootstrap distribution of $\hat{y}^*_{T+1|T}$ which is also shifted to the left. This is illustrated by a large difference between the bootstrap symmetric and equal-tailed intervals. The symmetric intervals reproduce the pattern of more coverage to the left than to the right, while equal-tailed intervals distribute coverage more or less equally in both tails. In both cases, the total rejection rates are closer to their nominal level than with asymptotic theory, for example with $N = T = 50$, the wild bootstrap does not include the true value in 6.7% of the replications with the symmetric intervals and 6.1% for the equal-tailed intervals.

This phenomenon is also reflected in the length of intervals. The asymptotic intervals are shortest (and least accurate). The equal-tailed intervals are typically slightly shorter than the corresponding symmetric intervals.

Forecast of y_{T+1} We next consider the prediction of y_{T+1} in Figures 2 and 3. As mentioned before, given our parameter configuration, the uncertainty is dominated by the underlying error term ε_{T+1} and not estimation uncertainty. This is the reason asymptotic intervals rely on the normality assumption. This provides a motivation for the bootstrap, and the effect of non-normality is highlighted in our figures.

Figure 2 shows that under normality, inference for y_{T+1} is quite accurate for all methods, and it is essentially unaffected by the values of N and T as predicted since it is dominated by the behavior of ε_{t+h}. All methods perform similarly, though we see that the asymptotic intervals that make the correct Gaussianity assumption are shorter than those based on the bootstrap. The iid bootstrap also produces slightly narrower intervals than the block wild bootstrap.

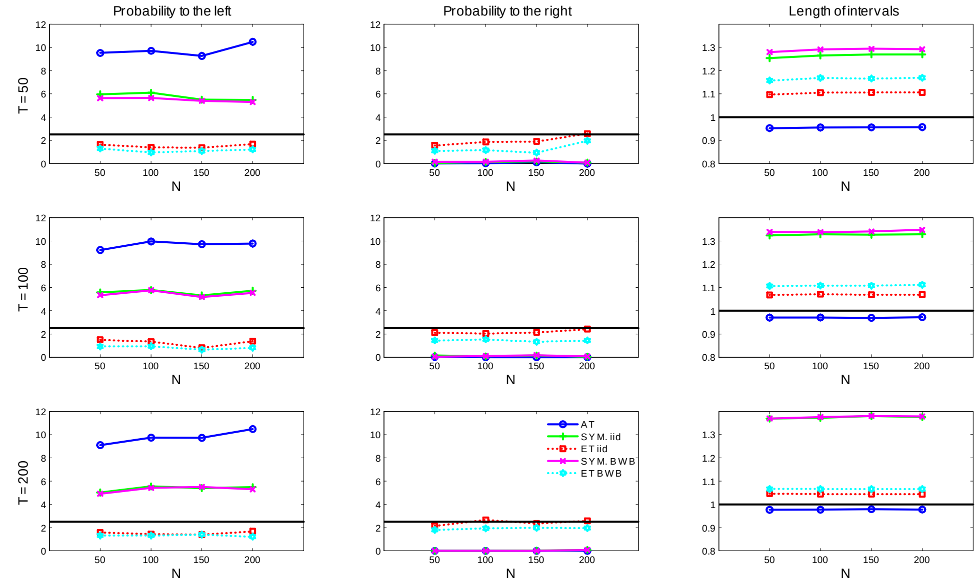
Figure 2. Probability of interval to the left and right of y_{T+1}, ε is normal



Note: The figures in the first two columns report the fraction of confidence intervals that lie to the left or to the right of the observation for each method as a function of the cross-sectional dimension N. Each row corresponds to a different time series dimension. The last column reports the length of the confidence intervals relative to the length of the "ideal" intervals obtained as the 2.5% and 97.5% quantiles of the empirical distribution.

Figure 3 provides the same information when the errors are drawn from a mixture of normals. We see problems with asymptotic theory, and these come almost exclusively in the form a confidence interval to the left of the true value. This is due to the fact that we have falsely imposed that errors are Gaussian, whereas the true distribution is bimodal. On the other hand, the bootstrap corrects these difficulties. The symmetric intervals do so by reducing coverage on the left side to between 5 and 6% and having almost no coverage to the right. The equal-tailed intervals distribute coverage more evenly by reducing undercoverage on the right side and pretty much eliminating the over-coverage on the left side. Because they allow for asymmetry, the equal-tailed intervals are shorter than the symmetric ones. Similarly, the i.i.d. bootstrap that makes the correct assumption that ε_{T+1} is i.i.d. produces slightly more accurate coverage and shorter intervals.

Figure 3. Probability of interval to the left or right of y_{T+1}, ε is mixture



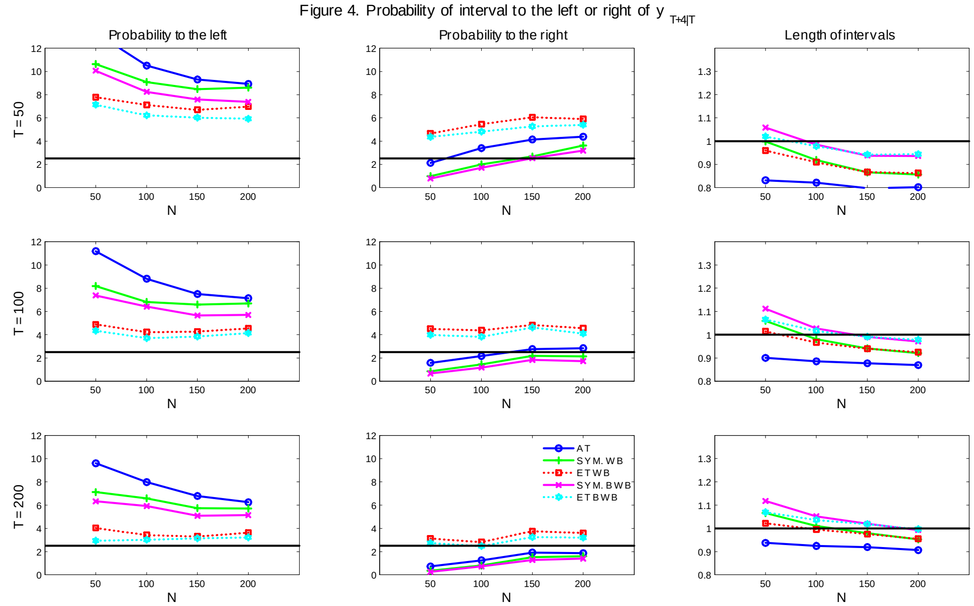
Note: see Figure 2.

6.2 Multi-horizon forecasting

In Figures 4-6, we report the same results as before but for $h = 4$ instead of $h = 1$. Because the error term is now a moving average of order 3, the wild bootstrap and block wild bootstrap (a block size equal to 4 is used) are no longer identical.

Figure 4 reports the results for the conditional mean, $y_{T+4|T}$. The main difference with Figure 1 is that there is a gap between the accuracy of the intervals based on the wild bootstrap and on the block wild bootstrap. As before, the equal-tailed intervals provide more accurate intervals and smaller length because they capture the bias in the distribution.

Figure 4. Probability of interval to the left or right of $y_{T+4|T}$

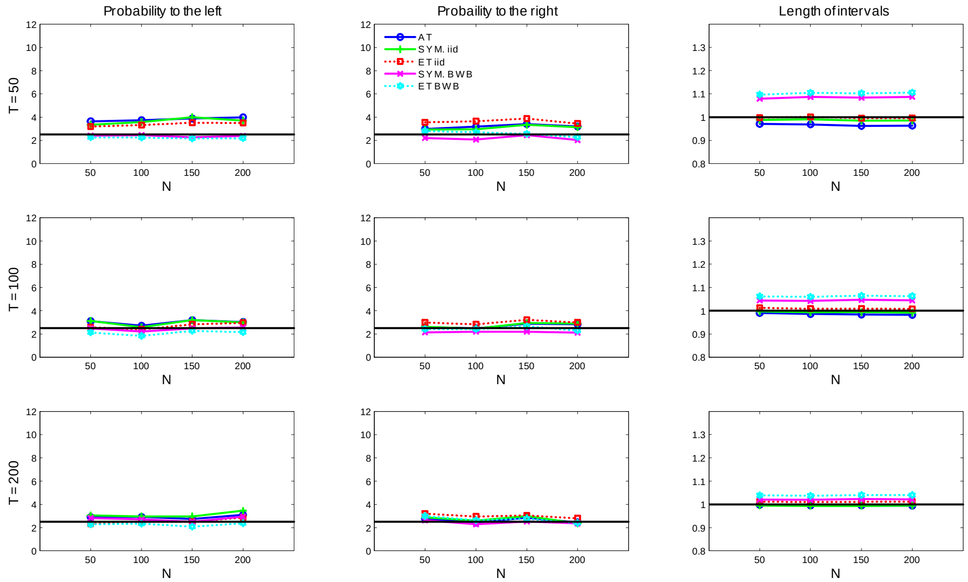


Note: see Figure 1.

While there is a difference in coverage between the wild bootstrap and the block wild bootstrap, it is not very large. This feature can be explained by the fact that factors are estimated. The forecast error variance has two parts, one due to the estimation of the parameters and one due to the estimation of the factors (see equation (4)). Serial correlation only affects the first term in that expression, and thus its effect is dampened by the presence of the second term which is usually not present in a typical forecasting context where predictors are observed.

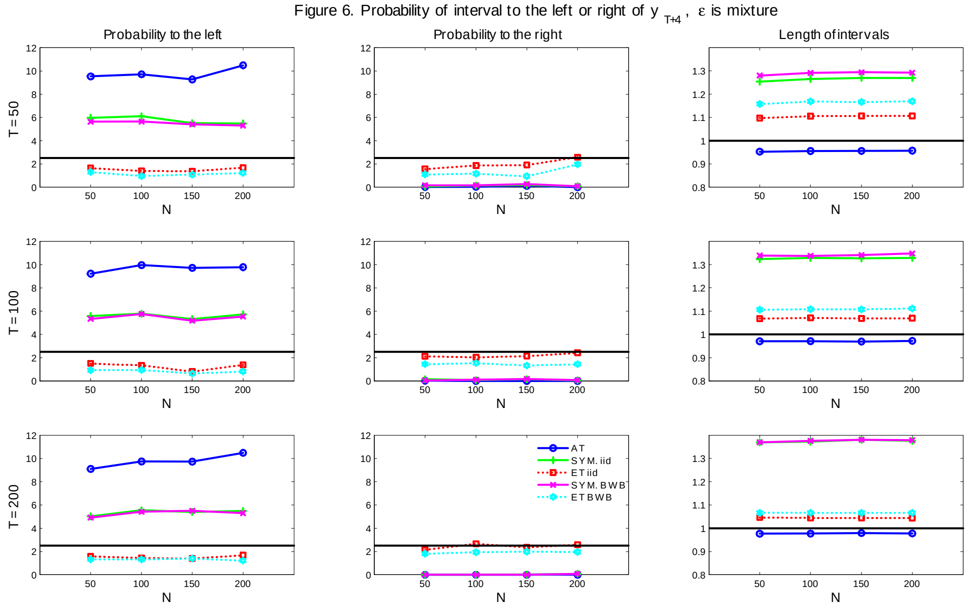
Figures 5 and 6 give the results for the new observation, y_{T+4}. Overall, we see that serial correlation does not seem to affect inference on y_{T+h} much. There are some effects when $T = 50$, but this seems related to difficulties in estimating the distribution of ε_{T+4} with serial correlation. Otherwise, the figures and conclusions are similar to those in Figures 2 and 3 with the exception of the fact that the block wild bootstrap leads to much wider intervals than the i.i.d. bootstrap with some improvement in coverage for $T = 50$.

Figure 5. Probability of interval to the left and right of y_{T+4}, ε is normal



Note: see Figure 2.

Figure 6. Probability of interval to the left or right of y_{T+4}, ε is mixture



Note: see Figure 2.

7 Empirical illustration

In this section, we use the dataset of Stock and Watson (2003) and Rossi and Sekhposyan (2014), updated to the first quarter of 2014, to illustrate the properties of asymptotic and bootstrap intervals.[1]

We consider forecast intervals for changes in the inflation rate measured by the quarterly growth rate of the GDP deflator ($PGDP$) at annual rate:

$$\Delta\pi_t = \left[\ln\left(\frac{PDGP_t}{PGDP_{t-1}}\right) - \ln\left(\frac{PDGP_{t-1}}{PGDP_{t-2}}\right)\right] \times 400.$$

There is a total of $N = 29$ series on asset prices, measures of economic activity, wages and prices, and money used to construct forecasts, see Rossi and Sekhposyan (2014) for details. The inflation rate is not included in the data used for extracting the factors. In order to have a balanced panel, our sample covers the period 1973q1-2014q1.

We construct forecasts from the factor-augmented autoregressive model:

$$\Delta\hat{\pi}_{t+h} = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\phi}_j \Delta\pi_{t-j+1} + \sum_{j=1}^{r} \hat{\alpha}_j \tilde{F}_{j,t}.$$

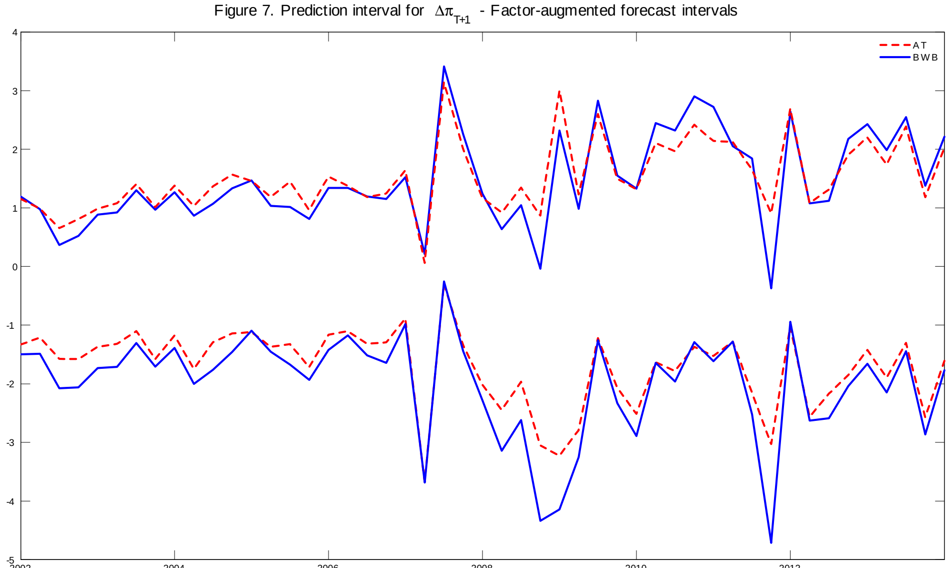
We compute forecast intervals for $h = 1$ for the last 50 observations in the sample. This means that the forecasts are made each period from the third quarter of 2001 until the end of 2013. We use a rolling window of 40 observations to estimate factors and parameters as in Rossi and Sekhposyan. We also follow Rossi and Sekhposyan and first choose the AR order p for each time period using BIC and then augment with the estimated factors. In each period,

[1] We thank Tatevik Sekhposyan for providing us with the data.

we select the number of factors such that the factors explain a minimum of 60% of the total variance of the panel after centering and rescaling. Three factors are selected by this approach in 40 out of the 50 periods, and 4 for the remaining 10 periods.

The factor-based forecasts reduce the root mean squared error of the forecasts by about 13% relative to autoregressive forecasts. In Figure 7, we report prediction intervals for the factor-augmented forecasts. The dashed red lines represent the bounds of the (pointwise) 95% prediction interval based on the asymptotic theory of Bai and Ng (2006) for each date. This interval is symmetric around the point forecast by construction since it is based on the normal distribution. We also report bootstrap intervals based on the block wild bootstrap (BWB) for ε^*_{t+h} with block size equal to the bandwidth selected by the Andrews (1991) rule and the wild bootstrap for e^*_t. Other methods for drawing ε^*_{t+h} lead to very similar intervals, and we do not report them to ease exposition (they are available from the authors upon request). The reported intervals were constructed as equal-tailed percentile-t intervals and are based on $B = 9999$ bootstrap replications.

Figure 7. Prediction interval for $\Delta\pi_{T+1}$ - Factor-augmented forecast intervals

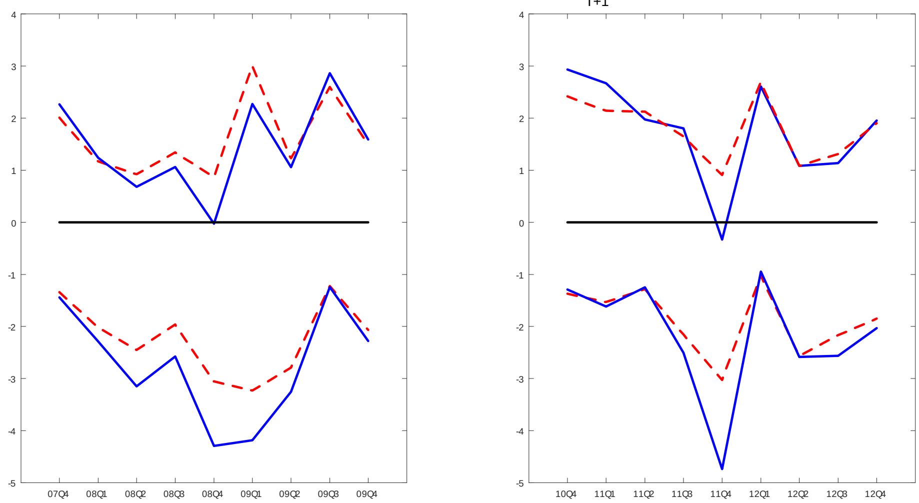


Note: The dashed red lines represent the bounds of the (pointwise) 95% prediction interval based on the asymptotic theory of Bai and Ng (2006) for each date. The solid blue line are bounds of the 95% equal-tailed percentile-t bootstrap intervals based on the block wild bootstrap (BWB) with block size equal to the bandwidth selected by the Andrews (1991) based on B=9999 bootstrap replications.

While both sets of intervals in Figure 7 are similar, there are noticeable differences that can be attributed either to bias in the estimation of the parameters or to non-normality in the distribution of the error term. Rossi and Sekhposyan (2014) find fairly strong evidence of non-normality of the forecast errors for this series, and this is likely an important source of the differences between the asymptotic and equal-tailed intervals.

The behavior of the bootstrap intervals during specific periods is quite interesting. For example, early in the sample, the bootstrap intervals are consistently shifted down relative to the asymptotic intervals. Figure 8 highlights two periods where the bootstrap interval lies completely below 0. The left panel presents the same intervals as Figure 7 around the fourth quarter of 2008. We see that the bootstrap intervals are shifted down for most of the reported period, and the upper limit of the bootstrap interval drops just below 0 (it is -.07%) in the fourth quarter of 2008. On the other hand, the asymptotic interval contains 0 with an upper limit of about 1%. This means that policy makers concerned about a sudden reduction in inflation following the collapse of Lehman Brothers would have underestimated the probability of a reduction in inflation had they based their decision on the asymptotic intervals.

Figure 8. Prediction interval for $\Delta\pi_{T+1}$ - 2008 and 2011



Note: see Figure 7.

Similarly, the right panel of Figure 8 focuses on the intervals around the fourth quarter of 2011. As in 2008, the bootstrap interval for the fourth quarter of 2011 is shifted down and includes only negative values, whereas the corresponding asymptotic interval includes positive inflation changes. At that time, many central banks were concerned about deflation risk, and relying on asymptotic intervals would have given them the impression that large reductions in inflation were much less likely than suggested by the bootstrap interval (the change in inflation turned out to be -2 percentage points).

8 Conclusion

In this paper, we have proposed the bootstrap to construct valid prediction intervals for models involving estimated factors. We considered two objects of interest: the conditional mean $y_{T+h|T}$ and the realization y_{T+h}. The bootstrap improves considerably on asymptotic theory for the conditional mean when the factors are relevant because of the bias in the estimation of the regression coefficients. However, our simulation results suggest that allowing for serial correlation, as is relevant when the forecasting horizon is greater than 1, is not very important in practice. For the observation, the bootstrap allows the construction of valid intervals without having to make strong distributional assumptions such as normality as was done in previous work by Bai and Ng (2006) .

One key assumption that we had to make to establish our results is that the idiosyncratic errors in the factor models are cross-sectionally independent. This is certainly restrictive, but it allows for the use of the wild bootstrap on the idiosyncratic errors. Non-parametric bootstrapping under more general conditions remains a challenge. The results in this paper could be used to prove the validity of a scheme in that context by showing the conditions $\mathcal{A}$ and $\mathcal{B}$ are satisfied.

A Appendix

The proof of Theorem 4.1 requires the following auxiliary result, which is the bootstrap analogue of Lemma A.2 of Bai (2003). It is based on the following identity that holds for each t:

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

where

$$\begin{aligned} \gamma_{st}^* &= E^* \left(\frac{1}{N} \sum_{i=1}^{N} e_{is}^* e_{it}^* \right), \ \zeta_{st}^* = \frac{1}{N} \sum_{i=1}^{N} \left(e_{is}^* e_{it}^* - E^* \left(e_{is}^* e_{it}^* \right) \right), \\ \eta_{st}^* &= \frac{1}{N} \sum_{i=1}^{N} \tilde{\lambda}_i' \tilde{F}_s e_{it}^* = \tilde{F}_s' \frac{\tilde{\Lambda}' e_t^*}{N} \text{ and } \xi_{st}^* = \frac{1}{N} \sum_{i=1}^{N} \tilde{\lambda}_i' \tilde{F}_t e_{is}^* = \eta_{ts}^*. \end{aligned}$$

Lemma A.1 *Assume Assumptions 1 and 2 hold. Under Condition $\mathcal{A}$, we have that for each t, in probability, as $N, T \to \infty$,*

(a) $T^{-1} \sum_{s=1}^{T} \tilde{F}_s^* \gamma_{st}^* = O_{P^*} \left(\frac{1}{\sqrt{T} \delta_{NT}} \right) + O_{P^*} \left(\frac{1}{T^{3/4}} \right);$

(b) $T^{-1} \sum_{s=1}^{T} \tilde{F}_s^* \zeta_{st}^* = O_{P^*} \left(\frac{1}{\sqrt{N} \delta_{NT}} \right);$

(c) $T^{-1} \sum_{s=1}^{T} \tilde{F}_s^* \eta_{st}^* = O_{P^*} \left(\frac{1}{\sqrt{N}} \right);$

(d) $T^{-1} \sum_{s=1}^{T} \tilde{F}_s^* \xi_{st}^* = O_{P^*} \left(\frac{1}{\sqrt{N} \delta_{NT}} \right).$

Remark 3 *The term $O_{P^*} \left(1/T^{3/4} \right)$ that appears in (a) is of a larger order of magnitude than the corresponding term in Bai (2003, Lemma A.2(i)), which is $O_P(1/T)$. The reason why we obtain this larger term is that we rely on Bonferroni's inequality and Chebyshev's inequality* *to bound* $\max_{1\le s\le T}\|F_s\| = O_P\left(T^{1/4}\right)$ *using the fourth order moment assumption on* F_s *(cf. Assumption 1(a)). In general, if* $E\|F_s\|^q \le M$ *for all* s*, then* $\max_{1\le s\le T}\|F_s\| = O_P\left(T^{1/q}\right)$ *and we will obtain a term of order* $O_{P^*}\left(1/T^{1-1/q}\right)$.

Proof of Lemma A.1. The proof follows closely that of Lemma A.2 of Bai (2003). The only exception is (a), where an additional $O\left(\frac{1}{T^{3/4}}\right)$ term appears. In particular, we write

$$T^{-1}\sum_{s=1}^{T}\tilde{F}_s^*\gamma_{st}^* = T^{-1}\sum_{s=1}^{T}\left(\tilde{F}_s^* - H^*\tilde{F}_s\right)\gamma_{st}^* + H^*T^{-1}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}^* = a_t^* + b_t^*.$$

We use Cauchy-Schwartz and Condition $\mathcal{A}.1$ to bound a_t^* as follows

$$\begin{aligned}\|a_t^*\| &\le \left(T^{-1}\sum_{s=1}^{T}\left\|\tilde{F}_s^* - H^*\tilde{F}_s\right\|^2\right)^{1/2}\left(T^{-1}\sum_{s=1}^{T}|\gamma_{st}^*|^2\right)^{1/2} \\ &= O_{P^*}\left(\frac{1}{\delta_{NT}}\right)O_P\left(\frac{1}{\sqrt{T}}\right) = O_{P^*}\left(\frac{1}{\delta_{NT}\sqrt{T}}\right),\end{aligned}$$

where $T^{-1}\sum_{s=1}^{T}\left\|\tilde{F}_s^* - H^*\tilde{F}_s\right\|^2 = O_{P^*}\left(\delta_{NT}^{-2}\right)$ by Lemma 3.1 of Gonçalves and Perron (2014) (note that this lemma only requires Conditions A*(b), A*(c), and B*(d), which correspond to our Condition $\mathcal{A}.1$, $\mathcal{A}.2$ and $\mathcal{A}.5$). For b_t^*, we have that (ignoring H^*, which is $O_{P^*}(1)$),

$$b_t^* = T^{-1}\sum_{s=1}^{T}\tilde{F}_s\gamma_{st}^* = T^{-1}\sum_{s=1}^{T}\left(\tilde{F}_s - HF_s\right)\gamma_{st}^* + HT^{-1}\sum_{s=1}^{T}F_s\gamma_{st}^* = b_{1t}^* + b_{2t}^*,$$

where $b_{1t}^* = O_P\left(1/\delta_{NT}\sqrt{T}\right)$ using the fact that $T^{-1}\sum_{s=1}^{T}\left\|\tilde{F}_s - HF_s\right\|^2 = O_P\left(\delta_{NT}^{-2}\right)$ under Assumptions 1 and 2 and the fact that $T^{-1}\sum_{s=1}^{T}|\gamma_{st}^*|^2 = O_P(1/T)$ for each t by Condition

$\mathcal{A}.1$. For b_{2t}^*, note that (ignoring $H = O_P(1)$),

$$\|b_{2t}^*\| \le \underbrace{\left(\max_s \|F_s\|\right)}_{O_P\left(T^{1/4}\right)} \underbrace{T^{-1}\sum_{s=1}^{T} |\gamma_{st}^*|}_{O_P\left(\frac{1}{T}\right)} = O_P\left(\frac{1}{T^{3/4}}\right),$$

where we have used the fact that $E\|F_s\|^4 \le M$ for all s (Assumption 1) to bound $\max_s \|F_s\|$. Indeed, by Bonferroni's inequality and Chebyshev's inequality, we have that

$$P\left(T^{-1/4}\max_s \|F_s\| > M\right) \le \sum_{s=1}^{T} P\left(\|F_s\| > T^{1/4}M\right) \le \sum_{s=1}^{T} \frac{E\|F_s\|^4}{M^4 T} \le \frac{1}{M^3} \to 0$$

for M sufficiently large. For (b), we follow exactly the proof of Bai (2003) and use Condition $\mathcal{A}.2$ to bound $T^{-1}\sum_{s=1}^{T} \zeta_{st}^{*2} = O_{P^*}\left(\frac{1}{N}\right)$ for each t; similarly, we use Condition $\mathcal{A}.3$ to bound $\frac{1}{T}\sum_{s=1}^{T} \tilde{F}_s \zeta_{st}^*$ for each t. For (c), we bound $T^{-1}\sum_{s=1}^{T} \tilde{F}_s \eta_{st}^* = N^{-1}H^*\sum_{i=1}^{N} \tilde{\lambda}_i e_{it}^* = O_{P^*}\left(1/\sqrt{N}\right)$ by using Condition $\mathcal{A}.6$. This same condition is used to bound $T^{-1}\sum_{s=1}^{T} \eta_{st}^{*2} = O_{P^*}(1/N)$ for each t. Finally, for part (d), we use Condition $\mathcal{A}.4$ to bound $T^{-1}\sum_{s=1}^{T} \tilde{F}_s \xi_{st}^* = O_{P^*}\left(\frac{1}{\sqrt{NT}}\right)$ for each t and we use Condition $\mathcal{A}.5$ to bound $T^{-1}\sum_{s=1}^{T} \xi_{st}^{*2} = O_{P^*}(1/N)$ for each t.

Proof of Theorem 4.1. By Lemma A.1, it follows that the third term in $\sqrt{N}\left(\tilde{F}_t^* - H^*\tilde{F}_t\right)$ is the dominant one (it is $O_{P^*}(1)$); the first term is $O_{P^*}\left(\frac{\sqrt{N}}{\sqrt{T}\delta_{NT}}\right) + O_P\left(\frac{\sqrt{N}}{T^{3/4}}\right) = O_{P^*}\left(\frac{\sqrt{N}}{T^{3/4}}\right) = o_{P^*}(1)$ if $\sqrt{N}/T^{3/4} \to 0$ whereas the second and the fourth terms are $O_{P^*}(1/\delta_{NT}) = o_{P^*}(1)$ as

$N, T \to \infty$. Thus, we have that

$$\begin{aligned}
\sqrt{N}\left(\tilde{F}_t^* - H^*\tilde{F}_t\right) &= \tilde{V}^{*-1}\frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s^*\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\lambda}_i'\tilde{F}_s e_{it}^* + o_{P^*}(1) \\
&= \left[\tilde{V}^{*-1}\left(\frac{\tilde{F}^{*\prime}\tilde{F}}{T}\right)\left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)\right]\left(\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\lambda}_i e_{it}^* + o_{P^*}(1) \\
&= H^*\tilde{V}^{-1}\Gamma_t^{*1/2}\underbrace{\Gamma_t^{*-1/2}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\lambda}_i e_{it}^*}_{\to^{d^*} N(0,I_r) \text{ by Condition } \mathcal{A}.6} + o_{P^*}(1), \qquad (26)
\end{aligned}$$

given the definition of H^* and the fact that $\tilde{V} = \frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}$. Since $\det\left(\Gamma_t^*\right) > \epsilon > 0$ for all N and some ϵ, Γ_t^{*-1} exists and we can define $\Gamma_t^{*-1/2} = \left(\Gamma_t^{*1/2}\right)^{-1}$ where $\Gamma_t^{*1/2}\Gamma_t^{*1/2} = \Gamma_t^*$. Let $\Pi_t^{*-1/2} = \Gamma_t^{*-1/2}\tilde{V}$ and note that $\Pi_t^{*-1/2}$ is symmetric and it is such that $\left(\Pi_t^{*-1/2}\right)\left(\Pi_t^{*-1/2}\right) = \tilde{V}\Gamma_t^{*-1}\tilde{V} = \Pi_t^{*-1}$. The result follows by multiplying (26) by $\Pi_t^{*-1/2}H^{*-1}$ and using Condition $\mathcal{A}.6$.

Proof of Corollary 4.1. Condition $\mathcal{B}$ and the fact that $\tilde{V} \to^P V$ under our assumptions imply that $\Pi_t^* \to^P \Pi_t \equiv V^{-1}Q\Gamma_t Q'V^{-1}$. This suffices to show the result.

Proof of Theorem 5.1. Using the decomposition (19) and the fact that

$$\hat{z}_T^* = \Phi^*\hat{z}_T + \begin{pmatrix} \tilde{F}_T^* - H^*\tilde{F}_T \\ 0 \end{pmatrix},$$

where $\Phi^* = diag\left(H^*, I_q\right)$, it follows that

$$\hat{y}_{T+1|T}^* - y_{T+1|T}^* = \frac{1}{\sqrt{T}}\hat{z}_T'\sqrt{T}\left(\Phi^{*\prime}\hat{\delta}^* - \hat{\delta}\right) + \frac{1}{\sqrt{N}}\hat{\alpha}'\sqrt{N}\left(H^{*-1}\tilde{F}_T^* - \tilde{F}_T\right) + r_T^*,$$

where the remainder is

$$r_T^* = \frac{1}{\sqrt{T}} \left(\tilde{F}_T^* - H^* \tilde{F}_T \right)' \sqrt{T} \left(\hat{\alpha}^* - H^{*-1\prime} \hat{\alpha} \right) = O_{P^*} \left(\frac{1}{\sqrt{TN}} \right).$$

First, we argue that

$$\frac{\hat{y}_{T+1|T}^* - y_{T+1|T}^*}{\sqrt{B_T^*}} \to^{d^*} N(0,1), \tag{27}$$

where B_T^* is the asymptotic variance of $\hat{y}_{T+1|T}^* - y_{T+1|T}^*$, i.e. $B_T^* = aVar^* \left(\hat{y}_{T+1|T}^* - y_{T+1|T}^* \right) = \frac{1}{T} \hat{z}_T' \Sigma_\delta \hat{z}_T + \frac{1}{N} \hat{\alpha}' \Pi_T \hat{\alpha}$. To show (27), we follow the arguments of Bai and Ng (2006, proof of their Theorem 3) and show that (1) $Z_{1T}^* = \sqrt{T} \left(\Phi^{*\prime} \hat{\delta}^* - \hat{\delta} \right) \to^{d^*} N(-c\Delta_\delta, \Sigma_\delta)$; (2) $Z_{2T}^* = \sqrt{N} \left(H^{*-1} \tilde{F}_T^* - \tilde{F}_T \right) \to^{d^*} N(0, \Pi_T)$; (3) Z_{1T}^* and Z_{2T}^* are asymptotically independent (conditional on the original sample). Condition (1) follows from Gonçalves and Perron (2014) under Assumptions 1-6; (2) follows from Corollary 4.1 provided $\sqrt{N}/T^{11/12} \to 0$ and conditions $\mathcal{A}$ and $\mathcal{B}$ hold for the wild bootstrap (which we verify next); (3) holds because we generate e_t^* independently of ε_{t+1}^*.

Proof of Condition $\mathcal{A}$ for the wild bootstrap. We verify for $t = T$. We have that $\sum_{s=1}^T |\gamma_{sT}^*|^2 = \left(\frac{1}{N} \sum_{i=1}^N \tilde{e}_{iT}^2 \right)^2$. Thus, it suffices to show that $\frac{1}{N} \sum_{i=1}^N \tilde{e}_{iT}^2 = O_P(1)$. This follows by using the decomposition

$$\tilde{e}_{it} = e_{it} - \lambda_i' H^{-1} \left(\tilde{F}_t - HF_t \right) - \left(\tilde{\lambda}_i - H^{-1\prime} \lambda_i \right)' \tilde{F}_t,$$

which implies that

$$\frac{1}{N}\sum_{i=1}^{N}|\tilde{e}_{it}|^2 \leq 3\frac{1}{N}\sum_{i=1}^{N}|e_{it}|^2+3\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|^2\left\|H^{-1}\right\|^2\left\|\tilde{F}_T-HF_T\right\|^2$$
$$+3\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_i-H^{-1\prime}\lambda_i\right\|^2\left\|\tilde{F}_T\right\|^2.$$

The first term is $O_P(1)$ given that $E|e_{it}|^2=O(1)$; the second term is $O_P(1)$ since $E\|\lambda_i\|^2 = O(1)$ and given that $\left\|\tilde{F}_T-HF_T\right\|^2=O_P(1/N)=o_P(1)$; and the third term is $O_P(1)$ given Lemma C.1.(ii) of Gonçalves and Perron (2014) and the fact that $\left\|\tilde{F}_T\right\|^2=O_P(1)$. Next, we verify $\mathcal{A}.2$. For $t=T$, following the proof of Theorem 4.1 in Gonçalves and Perron (2014) (condition A*(c)), we have that

$$\begin{aligned}
&\frac{1}{T}\sum_{s=1}^{T}E^*\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(e^*_{iT}e^*_{is}-E^*\left(e^*_{iT}e^*_{is}\right)\right)\right|^2\\
=\;&\frac{1}{T}\sum_{s=1}^{T}\frac{1}{N}\sum_{i=1}^{N}\tilde{e}^2_{iT}\tilde{e}^2_{is}\underbrace{Var\left(\eta_{iT}\eta_{is}\right)}_{\leq\bar{\eta}}\leq\bar{\eta}\frac{1}{N}\sum_{i=1}^{N}\tilde{e}^2_{iT}\left(\frac{1}{T}\sum_{s=1}^{T}\tilde{e}^2_{is}\right)\\
\leq\;&\bar{\eta}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{e}^4_{iT}\right)^{1/2}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\tilde{e}^4_{is}\right)^{1/2}=O_P(1),
\end{aligned}\tag{28}$$

where the first factor in (28) can be bounded by an argument similar to that used above to bound $\frac{1}{N}\sum_{i=1}^{N}\tilde{e}^2_{iT}$, and the second factor can be bounded by Lemma C.1 (iii) of Gonçalves and Perron (2014). $\mathcal{A}.3$ follows by an argument similar to that used by Gonçalves and Perron

(2014) to verify Condition B*(b). In particular,

$$
\begin{aligned}
& E^{*}\left\|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T}\sum_{i=1}^{N}\tilde{F}_{s}\left(e_{is}^{*}e_{iT}^{*}-E^{*}\left(e_{is}^{*}e_{iT}^{*}\right)\right)\right\|^{2} \\
= & \frac{1}{T}\sum_{s=1}^{T}\tilde{F}_{s}'\tilde{F}_{s}\frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{iT}^{2}\tilde{e}_{is}^{2}Var^{*}\left(\eta_{iT}\eta_{is}\right)\leq\bar{\eta}\frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{iT}^{2}\left(\frac{1}{T}\sum_{s=1}^{T}\tilde{F}_{s}'\tilde{F}_{s}\tilde{e}_{is}^{2}\right) \\
\leq & \bar{\eta}\left[\frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{iT}^{4}\right]^{1/2}\left[\frac{1}{T}\sum_{s=1}^{T}\left\|\tilde{F}_{s}\right\|^{4}\frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{s=1}^{T}\tilde{e}_{is}^{4}\right]^{1/2}=O_{P}\left(1\right),
\end{aligned}
$$

under our assumptions. Conditions $\mathcal{A}.4$ and $\mathcal{A}.5$ correspond to Gonçalves and Perron's (2014) Conditions B*(c) and B*(d), respectively. Finally, we prove Condition $\mathcal{A}.6$ for $t=T$. Using the fact that $e_{iT}^{*}=\tilde{e}_{iT}\eta_{iT}$, where $\eta_{iT}\sim$ i.i.d. $(0,1)$ across i, note that

$$
\Gamma_{T}^{*}=Var^{*}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\lambda}_{i}e_{iT}^{*}\right)=\frac{1}{N}\sum_{i=1}^{N}\tilde{\lambda}_{i}\tilde{\lambda}_{i}'\tilde{e}_{iT}^{2}\rightarrow^{P}Q\Gamma_{T}Q',
$$

by Theorem 6 of Bai (2003), where $\Gamma_{T}\equiv\lim_{N\rightarrow\infty}Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_{i}e_{iT}\right)>0$ by assumption. Thus, Γ_{T}^{*} is uniformly positive definite. We now need to verify that

$$
\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\ell'\Gamma_{T}^{*-1/2}\tilde{\lambda}_{i}e_{iT}^{*}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\underbrace{\ell'\Gamma_{T}^{*-1/2}\tilde{\lambda}_{i}\tilde{e}_{iT}\eta_{iT}}_{=\omega_{iT}^{*}}\rightarrow^{d^{*}}N\left(0,1\right),
$$

in probability, for any ℓ such that $\ell'\ell=1$. Since ω_{iT}^{*} is an heterogeneous array of independent random variables (given that η_{it} is i.i.d.), we apply a CLT for heterogeneous independent arrays.

Note that $E^*(\omega^*_{iT}) = 0$ and

$$\begin{aligned} Var^*\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \omega^*_{iT}\right) &= \ell'(\Gamma^*_T)^{-1/2} Var^*\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \tilde{\lambda}_i \tilde{e}_{iT}\eta_{iT}\right)(\Gamma^*_T)^{-1/2}\ell \\ &= \ell'(\Gamma^*_T)^{-1/2}\left(\frac{1}{N}\sum_{i=1}^N \tilde{\lambda}_i\tilde{\lambda}_i'\tilde{e}^2_{iT}\right)(\Gamma^*_T)^{-1/2}\ell = \ell'\ell = 1. \end{aligned}$$

Thus, it suffices to verify Lyapunov's condition, i.e. for some $r > 1$, $\frac{1}{N^r}\sum_{i=1}^N E^*|\omega^*_{iT}|^{2r} \to^P 0$. We have that

$$\begin{aligned} \frac{1}{N^r}\sum_{i=1}^N E^*|\omega^*_{iT}|^{2r} &\leq \frac{1}{N^{r-1}}\|\ell\|^{2r}\left\|(\Gamma^*_T)^{-1/2}\right\|^{2r}\frac{1}{N}\sum_{i=1}^N\left\|\tilde{\lambda}_i\right\|^{2r}|\tilde{e}_{iT}|^{2r}\underbrace{E^*|\eta_{iT}|^{2r}}_{\leq M<\infty} \\ &\leq C\frac{1}{N^{r-1}}\left\|(\Gamma^*_T)^{-1/2}\right\|^{2r}\left(\frac{1}{N}\sum_{i=1}^N\left\|\tilde{\lambda}_i\right\|^{4r}\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^N|\tilde{e}_{iT}|^{4r}\right)^{1/2} = O_P\left(\frac{1}{N^{r-1}}\right) = o_P(1) \end{aligned}$$

Proof of Condition $\mathcal{B}$ for the wild bootstrap. $\Gamma^*_T = \frac{1}{N}\sum_{i=1}^N \tilde{\lambda}_i\tilde{\lambda}_i'\tilde{e}^2_{iT} \to^P Q\Gamma_T Q'$, by Theorem 6 of Bai (2003).

The result for the studentized statistic (where we replace B^*_T with an estimate $\hat{B}^*_T$) then follows by showing that $\hat{z}^{*\prime}_T\hat{\Sigma}^*_\delta\hat{z}^*_T - \hat{z}'_T\Sigma_\delta\hat{z}_T \to^{P^*} 0$, and $\hat{\alpha}^{*\prime}\hat{\Sigma}^*_{\tilde{F}_T}\hat{\alpha}^* - \hat{\alpha}'\hat{\Sigma}_{\tilde{F}_T}\hat{\alpha} \to^{P^*} 0$, in probability. This can be shown using the arguments in Bai and Ng (2006, Theorems 3.1) and Bai (2003, Theorem 6).

Proof of Lemma 5.1. Recall that $F_\varepsilon(x) = P(\varepsilon_t \leq x)$ and define the following empirical distribution functions,

$$F_{T,\hat{\varepsilon}-\bar{\hat{\varepsilon}}}(x) = \frac{1}{T-1}\sum_{t=1}^{T-1}1\left\{\hat{\varepsilon}_{t+1} - \bar{\hat{\varepsilon}} \leq x\right\} \text{ and } F_{T,\varepsilon}(x) = \frac{1}{T-1}\sum_{t=1}^{T-1}1\left\{\varepsilon_{t+1} \leq x\right\},$$

where $\bar{\hat{\varepsilon}} = \frac{1}{T-1}\sum_{t=1}^{T-1} \hat{\varepsilon}_{t+1}$. Note that $F_{T,\varepsilon^*}(x) = F_{T,\hat{\varepsilon}-\bar{\hat{\varepsilon}}}(x)$. It follows that

$$d_2\left(F_{T,\hat{\varepsilon}-\bar{\hat{\varepsilon}}}, F_\varepsilon\right) \leq d_2\left(F_{T,\hat{\varepsilon}-\bar{\hat{\varepsilon}}}, F_{T,\varepsilon}\right) + d_2\left(F_{T,\varepsilon}, F_\varepsilon\right),$$

where $d_2\left(F_{T,\varepsilon}, F_\varepsilon\right) = o_{a.s.}(1)$ by Lemma 8.4 of Bickel and Freedman (1981). Thus, it suffices to show that $d_2\left(F_{T,\hat{\varepsilon}-\bar{\hat{\varepsilon}}}, F_{T,\varepsilon}\right) = o_P(1)$. Let I be distributed uniformly on $\{1, \ldots, T-1\}$ and define $X_1 = \hat{\varepsilon}_{I+1} - \bar{\hat{\varepsilon}}$ and $Y_1 = \varepsilon_{I+1}$. We have that

$$\begin{aligned}\left(d_2\left(F_{T,\hat{\varepsilon}-\bar{\hat{\varepsilon}}}, F_{T,\varepsilon}\right)\right)^2 &\leq E\left(X_1 - Y_1\right)^2 = E_I\left(\hat{\varepsilon}_{I+1} - \bar{\hat{\varepsilon}} - \varepsilon_{I+1}\right)^2 = \frac{1}{T-1}\sum_{t=1}^{T-1}\left(\hat{\varepsilon}_{t+1} - \bar{\hat{\varepsilon}} - \varepsilon_{t+1}\right)^2 \\ &= \frac{1}{T-1}\sum_{t=1}^{T-1}\left(\hat{\varepsilon}_{t+1} - \varepsilon_{t+1}\right)^2 - 2\frac{1}{T-1}\sum_{t=1}^{T-1}\left(\hat{\varepsilon}_{t+1} - \varepsilon_{t+1}\right)\bar{\hat{\varepsilon}} + \left(\bar{\hat{\varepsilon}}\right)^2 \equiv A_1 + A_2 + A_3.\end{aligned}$$

We can write

$$\hat{\varepsilon}_{t+1} - \varepsilon_{t+1} = -\left(\tilde{F}_t - HF_t\right)'\hat{\alpha} - \left(\Phi z_t\right)'\left(\hat{\delta} - \delta\right),$$

where $\Phi = diag\left(H, I_q\right)$. This implies that

$$A_1 \leq 2\frac{1}{T-1}\sum_{t=1}^{T-1}\left\|\tilde{F}_t - HF_t\right\|^2 \left\|\hat{\alpha}\right\|^2 + 2\frac{1}{T-1}\sum_{t=1}^{T-1}\left\|\Phi z_t\right\|^2\left\|\hat{\delta} - \delta\right\|^2 = O_P\left(\frac{1}{\delta_{NT}^2}\right) + O_P\left(\frac{1}{T}\right) = o_P(1).$$

Similarly,

$$\bar{\hat{\varepsilon}} = \frac{1}{T-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1} = \frac{1}{T-1}\sum_{t=1}^{T-1}\left(\hat{\varepsilon}_{t+1} - \varepsilon_{t+1}\right) + \frac{1}{T-1}\sum_{t=1}^{T-1}\varepsilon_{t+1} = O_P\left(\frac{1}{\delta_{NT}}\right) + o_P(1),$$

where the first term is bounded by an argument similar to that used to bound A_1 (via the Cauchy-Schwartz inequality). This implies that A_2 and A_3 are also $o_P(1)$.

Proof of Corollary 5.1. Lemma 5.1 implies that $s^*_{T+1} \to^{d^*} 1 - F_\varepsilon(-x\sigma_\varepsilon)$, in probability. Since $s_{T+1} \to^d 1 - F_\varepsilon(-x\sigma_\varepsilon)$ and F_ε is everywhere continuous under Assumption 7, Polya's Theorem implies the result.

References

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

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

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

[4] Beran, R. 1987. "Prepivoting to Reduce Level Error of Confidence Sets", *Biometrika*, 74, 3, 457-68.

[5] Bickel, P. and D. Freedman, 1981. "Asymptotic Theory for the Bootstrap". *Annals of Statistics*, Vol. 9, No.6, 1196-1217.

[6] Chang, Y. and J. Park, 2003. "A Sieve Bootstrap for the Test of a Unit Root", *Journal of Time Series Analysis*, 24, 379-400.

[7] Djogbenou, A., S. Gonçalves, and B. Perron, 2014. "Bootstrap Inference for Regressions with Estimated Factors and Serial Correlation", *Journal of Time Series Analysis*, forthcoming.

[8] Gonçalves, S. and B. Perron, 2014. "Bootstrapping Factor-Augmented Regressions", *Journal of Econometrics*, 182, 156-173.

[9] Pan, L. and D. Politis, 2014. "Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions", *Journal of Statistical Planning and Inference*, forthcoming.

[10] Pascual, L., J. Romo, and E. Ruiz, 2001. "Effects of Parameter Estimation on Prediction Densities: a Bootstrap Approach", *International Journal of Forecasting*, 17 (1), 83-103.

[11] Politis, D. 2013. "Model-free Model-fitting and Predictive Distributions" (with discussion and rejoinder), *Test*, 22, 183-250.

[12] 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.

[13] Wolf, M. and D. Wunderli, 2015. "Bootstrap Joint Prediction Regions", *Journal of Time Series Analysis*, 36, 352–376.

1130, rue Sherbrooke Ouest, bureau 1400, Montréal (Québec) H3A 2M8
Tél. : 514-985-4000 • Téléc. : 514-985-4039
www.cirano.qc.ca • info@cirano.qc.ca

Centre interuniversitaire de recherche en analyse des organisations
Center for Interuniversity Research and Analysis on Organizations