

Bootstrap prediction intervals for factor models

Sílvia Gonçalves and Benoit Perron*

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

October 12, 2012

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 (2011).

Keywords: factor model, bootstrap, forecast, conditional mean.

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. 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,

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

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

*Gonçalves acknowledges financial support from the NSERC and MITACS whereas Perron acknowledges financial support from the SSHRCC and MITACS.

model,

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

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 $\{(y_t, X_t, W_t) : t = 1, \dots, T\}$, 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 $\hat{y}_{T+h|T} = \hat{\alpha}' \tilde{F}_T + \hat{\beta}' W_T$.

The main goal of this paper is to propose bootstrap prediction intervals for y_{T+h} and $y_{T+h|T}$. Because we treat factors as latent, forecasts for y_{T+h} and its conditional mean depend both on estimated factors and regression coefficients. These two sources of parameter uncertainty must be accounted for when constructing prediction intervals, as recently 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 (2011) 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 $\sqrt{T}/N \rightarrow c$ and $c \neq 0$. Gonçalves and Perron (2011) 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.

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 (2011) are used in Section 5 to show the asymptotic validity of wild bootstrap prediction intervals. Section 6 presents our simulation experiments, and Section 7 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 $z_t = (F_t' \ W_t')'$, where z_t is $p \times 1$, with $p = r + q$. Following Bai and Ng (2006), we make the following assumptions.

Assumption 1

- (a) $E \|F_t\|^4 \leq M$ and $\frac{1}{T} \sum_{t=1}^T F_t F_t' \rightarrow^P \Sigma_F > 0$, where Σ_F is a non-random $r \times r$ matrix.
- (b) The loadings λ_i are either deterministic such that $\|\lambda_i\| \leq M$, or stochastic such that $E \|\lambda_i\|^4 \leq M$. In either case, $\Lambda' \Lambda / N \rightarrow^P \Sigma_\Lambda > 0$, 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) $E(e_{it}) = 0$, $E|e_{it}|^4 \leq M$.
- (b) $E(e_{it}e_{js}) = \sigma_{ij,ts}$, $|\sigma_{ij,ts}| \leq \bar{\sigma}_{ij}$ for all (t, s) , $|\sigma_{ij,ts}| \leq \tau_{ts}$ for all (i, j) . Furthermore, $\sum_{s=1}^T \tau_{ts} \leq M$, for each t , and $\frac{1}{NT} \sum_{t,s,i,j} |\sigma_{ij,ts}| \leq M$.
- (c) For every (t, s) , $E \left| N^{-1/2} \sum_{i=1}^N (e_{it}e_{is} - E(e_{it}e_{is})) \right|^4 \leq M$.
- (d) $\frac{1}{NT^2} \sum_{t,s,l,u} \sum_{i,j} |Cov(e_{it}e_{is}, e_{jl}e_{ju})| < \Delta < \infty$.
- (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) $E(\varepsilon_{t+h} | y_t, F_t, y_{t-1}, F_{t-1}, \dots) = 0$ for any $h > 0$, and F_t and ε_t are independent of the idiosyncratic errors e_{is} for all (i, s, t) .
- (b) $E \|z_t\|^4 \leq M$ and $\frac{1}{T} \sum_{t=1}^T z_t z_t' \rightarrow^P \Sigma_{zz} > 0$.
- (c) 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 is similar to Assumption E of Bai and Ng (2006). It assumes that ε_{t+h} is a

martingale difference sequence with respect to the information available at time t , thus ruling out by construction serial dependence in the regression error. For $h = 1$, the m.d.s. assumption is not very strict but it may become so for $h > 1$. For simplicity, we follow Bai and Ng (2006) and maintain this assumption throughout.

2.2 Normal-theory prediction 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 $\tilde{F} = (\tilde{F}_1 \dots \tilde{F}_T)'$ 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 $\tilde{\Lambda} = (\tilde{\lambda}_1, \dots, \tilde{\lambda}_N)' = X'\tilde{F}(\tilde{F}'\tilde{F})^{-1} = X'\tilde{F}/T$.

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

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

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 $y_{T+h} = \alpha'F_T + \beta'W_T$ 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

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

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

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

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 $H^{-1}\alpha$. 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

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

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

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

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

$$\hat{\Sigma}_\alpha = \left(\sum_{t=1}^{T-h} \hat{z}_t \hat{z}'_t \right)^{-1} \left(\sum_{t=1}^{T-h} \hat{z}_t \hat{z}'_t \hat{\varepsilon}_{t+h}^2 \right) \left(\sum_{t=1}^{T-h} \hat{z}_t \hat{z}'_t \right)^{-1}, \quad (8)$$

and

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

where $\tilde{\Gamma}_T$ is an estimator of $\Gamma_T = \lim_{N \rightarrow \infty} Var\left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i e_{iT}\right)$ and 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 prediction interval for $y_{T+h|T}$ given by

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

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

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

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. $N(0, \sigma_\varepsilon^2)$.

An important condition that justifies (10) and (11) is that $\sqrt{T}/N \rightarrow 0$. This condition ensures that the term reflecting the parameter estimation uncertainty in the forecast error decomposition (4), $\sqrt{T}(\hat{\delta} - \delta)$, 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 (2011), when $\sqrt{T}/N \rightarrow c \neq 0$,

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

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

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

Thus, the forecast error is asymptotically $N(0, \alpha' H^{-1} \Sigma_{\tilde{F}_T} H^{-1} \alpha)$. Since $N \hat{B}_T = (N/T) \hat{z}'_T \hat{\Sigma}_\delta \hat{z}_T + \hat{\alpha}' \hat{\Sigma}_{\tilde{F}_T} \hat{\alpha} = \alpha' H^{-1} \Sigma_{\tilde{F}_T} H^{-1} \alpha + o_P(1)$, the studentized forecast error given in (6) is still $N(0, 1)$ as $N, T \rightarrow \infty$. For the studentized forecast error associated with forecasting y_{T+h} , the variance of \hat{y}_{T+h} is asymptotically (as $N, T \rightarrow \infty$) 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 prediction intervals

Following Gonçalves and Perron (2011), we consider the following bootstrap DGP:

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

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

where $\{e_t^* = (e_{1t}^*, \dots, e_{Nt}^*)'\}$ denotes a bootstrap sample from $\{\tilde{e}_t = X_t - \tilde{\Lambda} \tilde{F}_t\}$ and $\{\varepsilon_{t+h}^*\}$ is a resampled version of $\{\hat{\varepsilon}_{t+h} = y_{t+h} - \hat{\alpha}' \tilde{F}_t - \hat{\beta}' W_t\}$. A specific method to generate e_t^* and ε_{t+h}^* is discussed in Section 5.

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

$$\hat{\delta}^* = \left(\sum_{t=1}^{T-h} \hat{z}_t^* \hat{z}_t^{*'} \right)^{-1} \sum_{t=1}^{T-h} \hat{z}_t^* y_{t+h}^*,$$

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

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

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

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

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

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

2. Estimate the bootstrap factors $\{\tilde{F}_t^* : t = 1, \dots, T\}$ using X^* .
3. For $t = 1, \dots, T - h$, 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

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

and bootstrap standard errors

$$\hat{B}_T^* = \frac{1}{T} \hat{z}_T^* \hat{\Sigma}_\delta^* \hat{z}_T^{*'} + \frac{1}{N} \hat{\alpha}^* \hat{\Sigma}_{\tilde{F}_T^*}^* \hat{\alpha}^{*'}, \quad (14)$$

where the choice of $\hat{\Sigma}_\delta^*$ and $\hat{\Sigma}_{\tilde{F}_T^*}^*$ depends on the assumptions on 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 $|s_{T+h}^*|$.

7. Repeat this process B times, resulting in statistics $\{s_{T+h,1}^*, \dots, s_{T+h,B}^*\}$ and $\{|s_{T+h,1}^*|, \dots, |s_{T+h,B}^*|\}$.

8. Compute the corresponding empirical quantiles:

- (a) For equal-tailed percentile- t bootstrap intervals, $q_{1-\alpha}^*$ is the empirical $1 - \alpha$ quantile of $\{s_{T+h,1}^*, \dots, s_{T+h,B}^*\}$.
- (b) For symmetric percentile- t bootstrap intervals, $q_{|\cdot|, 1-\alpha}^*$ is the empirical $1 - \alpha$ quantile of $\{|s_{T+h,1}^*|, \dots, |s_{T+h,B}^*|\}$.

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

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

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

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

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 $\{y_{1+h}^*, \dots, y_T^*, y_{T+1}^*, \dots, y_{T+h}^*\}$ using

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

where $\{\varepsilon_{1+h}^*, \dots, \varepsilon_T^*, \varepsilon_{T+1}^*, \dots, \varepsilon_{T+h}^*\}$ is a bootstrap sample obtained from $\{\hat{\varepsilon}_{1+h}, \dots, \hat{\varepsilon}_T\}$.

4. Not making use of the stretch $\{y_{T+1}^*, \dots, y_{T+h}^*\}$, 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 standard error as

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

where $\hat{\sigma}_\varepsilon^{*2}$ is a consistent estimator of $\sigma_\varepsilon^2 = \text{Var}(\varepsilon_{T+h})$ 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 $|s_{T+h}^*|$.

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

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

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

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

The main differences between the two algorithms is that in step 3 of Algorithm 2 we generate observations for y_{t+h}^* for $t = 1, \dots, T$ 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 $t = 1, \dots, T - h$, but the bootstrap variance \hat{C}_T^* contains an extra term $\hat{\sigma}_\varepsilon^{*2}$ that reflects the uncertainty associated with the error ε_{T+h} associated with the new observation y_{T+h} .

4 Bootstrap distribution of estimated factors

The asymptotic validity of the bootstrap prediction 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 prediction error for the conditional mean is given by

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

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

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

where \tilde{V}^* is the $r \times r$ diagonal matrix containing on the main diagonal the r largest eigenvalues of $X^* X^{*'} / NT$, in decreasing order. Note that contrary to H , which depends on unknown population parameters, H^* is fully observed.

Adding and subtracting appropriately, we can write

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

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 $T_{NT}^* \xrightarrow{P^*} 0$, in probability, when for any $\delta > 0$, $P^*(|T_{NT}^*| > \delta) = o_P(1)$. We write $T_{NT}^* = O_{P^*}(1)$, in probability, when for all $\delta > 0$ there exists $M_\delta < \infty$ such that $\lim_{N, T \rightarrow \infty} P[P^*(|T_{NT}^*| > M_\delta) > \delta] = 0$. Finally, we write $T_{NT}^* \xrightarrow{d^*} D$, in probability, if conditional on a sample with probability that converges to one,

T_{NT}^* weakly converges to the distribution D under P^* , i.e. $E^*(f(T_{NT}^*)) \rightarrow^P E(f(D))$ 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 prediction error captures the two forms of estimation uncertainty in (4) provided: (1) the bootstrap distribution of $\sqrt{T}(\Phi^{*'}\hat{\delta}^* - \hat{\delta})$ is a consistent estimator of the distribution of $\sqrt{T}(\hat{\delta} - \delta)$, and (2) the bootstrap distribution of $\sqrt{N}(H^{*-1}\tilde{F}_T^* - \tilde{F}_T)$ is a consistent estimator of the distribution of $\sqrt{N}(\tilde{F}_T - HF_T)$. Gonçalves and Perron (2011) discussed conditions for the consistency of the bootstrap distribution of $\sqrt{T}(\hat{\delta} - \delta)$. Here we propose a set of conditions that justifies using the bootstrap to consistently estimate the distribution of the estimated factors $\sqrt{N}(\tilde{F}_t - HF_t)$ at each point t .

Condition \mathcal{A} .

- A.1. (a) $T^{-1} \sum_{t=1}^T \sum_{s=1}^T |\gamma_{st}^*|^2 = O_P(1)$, and (b) for each t , $\sum_{s=1}^T |\gamma_{st}^*|^2 = O_P(1)$, where $\gamma_{st}^* = E^*\left(\frac{1}{N} \sum_{i=1}^N e_{it}^* e_{is}^*\right)$.
- A.2. (a) $\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T E^* \left| \frac{1}{\sqrt{N}} \sum_{i=1}^N (e_{it}^* e_{is}^* - E^*(e_{it}^* e_{is}^*)) \right|^2 = O_P(1)$, and (b) for each t , $\frac{1}{T} \sum_{s=1}^T E^* \left| \frac{1}{\sqrt{N}} \sum_{i=1}^N (e_{it}^* e_{is}^* - E^*(e_{it}^* e_{is}^*)) \right|^2 = O_P(1)$.
- A.3. For each t , $E^* \left\| \frac{1}{\sqrt{TN}} \sum_{s=1}^T \sum_{i=1}^N \tilde{F}_s (e_{it}^* e_{is}^* - E^*(e_{it}^* e_{is}^*)) \right\|^2 = O_P(1)$.
- A.4. $E^* \left\| \frac{1}{\sqrt{TN}} \sum_{t=1}^T \sum_{i=1}^N \tilde{F}_t \tilde{\lambda}_i' e_{it}^* \right\|^2 = O_P(1)$.
- A.5. $\frac{1}{T} \sum_{t=1}^T E^* \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{\lambda}_i e_{it}^* \right\|^2 = O_P(1)$.
- A.6. For each t , $\Gamma_t^{*-1/2} \frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{\lambda}_i e_{it}^* \rightarrow^{d^*} N(0, I_r)$, in probability, where $\Gamma_t^* = Var^*\left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{\lambda}_i e_{it}^*\right)$ is uniformly definite positive.

Condition \mathcal{A} is the bootstrap analogue of Bai's (2003) assumptions used to derive the limiting distribution of $\sqrt{N}(\tilde{F}_t - HF_t)$. Gonçalves and Perron (2011) also relied on similar high level assumptions to study the bootstrap distribution of $\sqrt{T}(\hat{\delta}^* - \delta^*)$. 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 $\mathcal{A}.1.(b)$ and $\mathcal{A}.2.(b)$ expand Conditions A*(b) and A*(c) in Gonçalves and Perron (2011) 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 (2011)

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 $N, T \rightarrow \infty$ such that $\sqrt{N}/T^{3/4} \rightarrow 0$, 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

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

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 $\sqrt{N}/T \rightarrow 0$, $\sqrt{N} \left(\tilde{F}_t - H F_t \right) \rightarrow^d N(0, \Pi_t)$, 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 ($\sqrt{N}/T^{3/4} \rightarrow 0$ instead of $\sqrt{N}/T \rightarrow 0$) 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 $E \|F_t\|^q \leq M$ for all t , then the required rate restriction is $\sqrt{N}/T^{1-1/q} \rightarrow 0$. See Remarks 1 and 2 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 $N, T \rightarrow \infty$ such that $\sqrt{N}/T^{3/4} \rightarrow 0$, $\sqrt{N} \left(H^{*-1} \tilde{F}_t^* - \tilde{F}_t \right) \rightarrow^{d^*} N(0, \Pi_t)$, in probability, where $\Pi_t = V^{-1} Q \Gamma_t Q' V^{-1}$ is the asymptotic covariance matrix of $\sqrt{N} \left(\tilde{F}_t - H F_t \right)$.*

Corollary 4.1 justifies using the bootstrap to construct confidence intervals for the rotated factors $H F_t$ 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 prediction 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 $\sqrt{N} \left(\tilde{F}_t - H F_t \right)$. A similar rotation was discussed in Gonçalves and Perron (2011) in the context of bootstrapping the regression coefficients $\hat{\delta}$.

5 Validity of bootstrap prediction intervals

5.1 Intervals for $y_{T+h|T}$

When prediction intervals for the conditional mean $y_{T+h|T}$ are the object of interest, we use a two-step wild bootstrap scheme, as in Gonçalves and Perron (2011). Specifically, we rely on Algorithm 1 and we let

$$\varepsilon_{t+h}^* = \hat{\varepsilon}_{t+h} \cdot v_{t+h}, \quad t = 1, \dots, T-h, \quad (20)$$

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

$$e_{it}^* = \tilde{e}_{it} \cdot \eta_{it}, \quad t = 1, \dots, T, \quad i = 1, \dots, N, \quad (21)$$

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

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 $q > 1$, $E |\varepsilon_{t+h}|^{4q} \leq M < \infty$, for all t, h .

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

The martingale difference sequence assumption on ε_{t+h} (Assumption 4(b)) motivates the use of the wild bootstrap in (20). This assumption rules out serial correlation in ε_{t+h} but allows for conditional heteroskedasticity and was already used by Bai and Ng (2006).

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 $\sqrt{T}/N \rightarrow c = 0$). Since factors estimation uncertainty depends on the cross sectional correlation of the idiosyncratic errors e_{it} (via $\Gamma_T = \lim_{N \rightarrow \infty} \text{Var} \left(1/\sqrt{N} \sum_{i=1}^N \lambda_i e_{iT} \right)$), 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 (2011) 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 standard errors, \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) $\Gamma_T = \lim_{N \rightarrow \infty} \text{Var} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i e_{iT} \right)$ under Assumption 6. We compute \hat{B}_T^* using (14) and relying on the 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 $\varepsilon_{t+h}^* = \hat{\varepsilon}_{t+h} \cdot v_{t+h}$ and $e_{it}^* = \tilde{e}_{it} \cdot \eta_{it}$, where $v_{t+h} \sim i.i.d.(0, 1)$ for all $t = 1, \dots, T-h$ and $\eta_{it} \sim i.i.d.(0, 1)$ for all $i = 1, \dots, N$; $t = 1, \dots, T$, and v_{t+h} and η_{it} are mutually independent. Moreover, assume that $E^* |\eta_{it}|^4 < C$ for all (i, t) and $E^* |v_{t+h}|^4 < C$ for all t . If $\sqrt{T}/N \rightarrow c$, where $0 \leq c < \infty$, and $\sqrt{N}/T^{11/12} \rightarrow 0$, then conditional on $\{y_t, X_t, W_t : t = 1, \dots, T\}$,*

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

in probability.

Remark 1 *The rate restriction $\sqrt{N}/T^{11/12} \rightarrow 0$ is slightly stronger than the rate used by Bai (2003) (cf. $\sqrt{N}/T \rightarrow 0$). It is weaker than the restriction $\sqrt{N}/T^{3/4} \rightarrow 0$ 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 2 in the Appendix.*

5.2 Intervals for y_{T+h}

In this section we provide a theoretical justification for bootstrap prediction intervals for y_{T+h} as described in Algorithm 2. We add the following assumption.

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

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

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

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

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

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

Note that we can write

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

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

$$s_{T+h} = -\frac{\varepsilon_{T+h}}{\sigma_\varepsilon} + o_P(1). \quad (23)$$

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

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

for all $x \in \mathbb{R}$. If we assume that ε_{t+h} 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 $F_{T,s}(x) \rightarrow \Phi(x)$, i.e. $s_{T+h} \xrightarrow{d} N(0, 1)$. Nevertheless, this is not generally true unless we make the Gaussianity assumption. We note that although asymptotically the variance of the prediction error $\hat{y}_{T+h|T} - y_{T+h}$ 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 $\hat{y}_{T+h|T} - y_{T+h}$ since $\hat{\sigma}_\varepsilon^2$ will underestimate the true forecast variance for finite T and N .

Next we show that the bootstrap yields a consistent estimate of the distribution of s_{T+h} without assuming that ε_{t+h} 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 $\{\varepsilon_{1+h}^*, \dots, \varepsilon_T^*, \varepsilon_{T+1}^*, \dots, \varepsilon_{T+h}^*\}$ as a random sample obtained from the centered residuals $\{\hat{\varepsilon}_{1+h} - \bar{\varepsilon}, \dots, \hat{\varepsilon}_T - \bar{\varepsilon}\}$. Resampling in an i.i.d. fashion is justified under Assumption 7. We recenter the residuals because $\bar{\varepsilon}$ is not necessarily zero unless W_t contains a constant regressor. Nevertheless, since $\bar{\varepsilon} = o_P(1)$, 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

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

in probability, which in turn implies

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

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

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+h}^* converges weakly in probability to ε_{T+h} , i.e. $d(F_{T,\varepsilon}^*, F_\varepsilon) \xrightarrow{P} 0$ for any metric that metrizes weak convergence. In the following we use Mallows metric which is defined as $d_2(F_X, F_Y) = \left(\inf(E|X - Y|^2)\right)^{1/2}$ 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 $T, N \rightarrow \infty$ such that $\sqrt{T}/N \rightarrow c$, $0 \leq c < \infty$, $d_2(F_{T,\varepsilon}^*, F_\varepsilon) \xrightarrow{P} 0$.*

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

$$\sup_{x \in \mathbb{R}} |F_{T,s}^*(x) - F_{\infty,s}(x)| \rightarrow 0,$$

in probability.

Corollary 5.1 implies the asymptotic validity of the bootstrap prediction intervals given in (17) and (18). Specifically, we can show that $P(y_{T+h} \in EQ_{y_{T+h}}^{1-\alpha}) \rightarrow 1 - \alpha$ and $P(y_{T+h} \in SY_{y_{T+h}}^{1-\alpha}) \rightarrow 1 - \alpha$ as $N, T \rightarrow \infty$. See e.g. Beran (1987) and Wolf and Wunderli (2012, Proposition 3.1). For instance,

$$\begin{aligned} P(y_{T+h} \in EQ_{y_{T+h}}^{1-\alpha}) &= P(s_{T+h} \leq q_{1-\alpha/2}^*) - P(s_{T+h} \leq q_{\alpha/2}^*) \\ &= P(F_{T,s}^*(s_{T+h}) \leq 1 - \alpha/2) - P(F_{T,s}^*(s_{T+h}) \leq \alpha/2). \end{aligned}$$

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

$$P(F_{\infty,s}(s_{T+h}) \leq x) = P(s_{T+h} \leq F_{\infty,s}^{-1}(x)) \equiv F_{T,s}(F_{\infty,s}^{-1}(x)) \rightarrow F_{\infty,s}(F_{\infty,s}^{-1}(x)) = x.$$

6 Simulation

We now report results from a simulation experiment to analyze the properties of the normal asymptotic intervals as well as their bootstrap counterpart analyzed above.

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

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

where F_t is drawn from a standard normal distribution independently over time. Because we are interested in constructing intervals for y_{T+h} (which depend on the distribution of ε_{T+h} and require homoskedasticity and independence over time), the regression error ε_{t+h} will be homoskedastic and independent over time with expectation 0 and variance 1. To analyze the effects of deviations from normality, we have considered five distributions for ε_{t+h} :

$$\begin{aligned} \text{Normal:} & \quad \varepsilon_t \sim N(0, 1) \\ \chi^2(2) & : \quad \varepsilon_t \sim \frac{1}{2} [\chi^2(2) - 2] \\ \text{Uniform} & : \quad \varepsilon_t \sim \sqrt{12} \left(U(0, 1) - \frac{1}{2} \right) \\ \text{Exponential:} & \quad \varepsilon_t \sim \text{Exp}(1) - 1 \\ \text{Mixture} & : \quad \varepsilon_t \sim \frac{1}{\sqrt{10}} [pN(-1, 1) + (1-p)N(9, 1)], \quad p \sim B(.9), \end{aligned}$$

but we will report results only for the normal and mixture distributions as they illustrate our conclusions. Results for the other cases are available from the authors upon request. The particular mixture distribution we are using was 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 $(T \times N)$ 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 .

As in Gonçalves and Perron (2011), we consider two values for the coefficient, either $\alpha = 0$ or 1. When $\alpha = 0$, the OLS estimator of α is unbiased. When $\alpha = 1$, the OLS estimator of α is biased unless $\frac{\sqrt{T}}{N} \rightarrow 0$.

We consider asymptotic and bootstrap confidence intervals at a nominal level of 95%. We use Algorithms 1 and 2 described above to generate the bootstrap data with $B = 499$ bootstrap replications. We consider the same two types of intervals analyzed above (symmetric percentile- t and equal-tailed percentile- t). We report experiments based on 5,000 replications and with two values for T (50 and 200) and 5 values for N (25, 50, 100, 150, and 200). The forecast horizon h , is set to 4 as in Bai and Ng (2006).

6.1 Results

We report graphically four sets of results: those for our quantities of interest y_{T+h} and the conditional mean $\hat{y}_{T+h|T}$, and those for their components, the rotated factor at the end of the sample HF_T and the rotated coefficient $H^{-1}\alpha$ as these help understand some of the behavior. In the first two cases, we report the frequency of times the true parameter is to the left or right of the corresponding 95% confidence interval. For the rotated factor and coefficient, we only report the total rejection rate as the side of rejection is not identified because of the sign indeterminacy associated with H . Each figure has two rows corresponding to $T = 50$ and $T = 200$.

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 each of the other three quantities. On the other hand, the results of y_{T+h} are dominated by the behavior of ε_{T+h} . Given our single factor model, the asymptotic variance of the conditional mean is:

$$Var(\hat{y}_{T+h|T}) = \frac{1}{T} \tilde{F}_T^2 Var(\sqrt{T}(\hat{\delta} - \delta)) + \frac{1}{N} \delta^2 Var(\sqrt{N}(\tilde{F}_T - HF_t))$$

and in our setup, $Var(\sqrt{N}(\tilde{F}_T - HF_t)) = 3$ and $Var(\sqrt{T}(\hat{\delta} - \delta)) = 1$. So even in the worst scenario where $N = 25$ and $T = 50$, this variance is only 0.14 (the unconditional expectation of \tilde{F}_T^2 is 1) compared to 1 in all setups for ε_{T+h} . This allows us to easily separate the contribution of the conditional mean from the contribution of ε in the forecasts of y_{T+h} .

Forecast of y_{T+h}

We report four figures for this quantity. The first two correspond to case 1 where $\alpha = 0$ with ε either normal or from the mixture, while the second two figures provide the same information when $\alpha = 1$.

As mentioned above, the behavior of forecast intervals are dominated by the distribution of ε_{T+h} so the value of α is not very important. This means that Figures 1 and 3 are almost identical and so are Figures 2 and 4.

Figures 1 and 3 show that under normality, inference for y_{T+h} is quite accurate, and it is essentially unaffected by the values of N and T as predicted as it is dominated by the behavior of ε .

Figures 2 and 4 reveal that when the errors are drawn from a mixture of normals, we see overrejections with asymptotic theory, and these almost exclusively come from the right side. The equal-tailed intervals improve the under rejection on the left side to a large extent and reduce the over-rejection on the right side.

In the results we do not report, the results are similar with the other non-normal distributions. We get under rejection with uniform errors (symmetrically on both sides), while in the case of χ^2 (2) and exponential errors, we get overall rejection rates that are essentially correct, but all rejections are on the right side because of the asymmetry in the distributions. Again, the equal-tailed bootstrap intervals perform best because they capture some of the asymmetry.

Conditional mean, $\hat{y}_{T+h|T}$

The results for the conditional mean are presented in Figures 5 and 6. Because the results are quite insensitive to the distribution of ε_{T+h} , we only report results for the normal distribution. Figure 5 reports results for case 1 where $\alpha = 0$ and Figure 6 does the same for $\alpha = 1$.

With $\alpha = 0$, asymptotic intervals underreject the true conditional mean. This can be explained by the fact that the true variance only depends on the variance of the coefficient, but when estimating it, the second term in (7) contributes because $\hat{\alpha}$ is obviously not 0 in finite samples. Thus, the variance is over estimated, and the t statistics are smaller than they should be. We see this underrejection disappearing as the sample grows as expected.

In Figure 6, we see that with $\alpha = 1$, there is some overrejection with asymptotic theory (a 95% confidence interval would cover the true conditional mean roughly 91% of the time for $N = T = 50$), but this disappears as N and T increase. This is due to a large bias in the estimation of the coefficient documented by Gonçalves and Perron (2011) and that we discuss below. This bias comes from the estimation of the factors and is reduced with an increasing N . The bootstrap corrects some of these over rejections, in particular the symmetric intervals.

Factor, HF_T

Figure 7 presents results for inference on the (rotated) factor at the end of the sample. Obviously, this is independent of both the value of the parameter α and of the distribution of ε so we only report one figure. As mentioned above, the side of rejection is not well-defined as it depends on the value of H . A left-side rejection with a positive H would be a right-side rejection with a negative H . Since the sign of H is arbitrary, the side of rejection is also arbitrary. Thus, only total rejections are meaningful, and this is the quantity we report.

The first thing to note is that there are notable size distortions, in particular when $T = 50$. The true rotated factor is not in the confidence interval between 11 and 12% of the times instead of the nominal 5% rate. For this smaller value of T , rejection rates are not sensitive to the value of N . For the larger value of T , size distortions are much smaller, and we do see a slight decrease as N increases.

Again, the bootstrap succeeds in correcting these distortions to a large extent. In particular, symmetric percentile- t intervals are quite accurate and provide reliable inference for all values of N and T . The equal-tailed intervals are somewhat less accurate, but they still offer a notable improvement over asymptotic theory.

Coefficient, $H^{-1}\alpha$

Finally, we report the results for the coefficients in figures 8 ($\alpha = 0$) and 9 ($\alpha = 1$). These results are in line with those of Gonçalves and Perron (2011). Asymptotic theory works well in all cases when $\alpha = 0$ in figure 8. The bootstrap has some problems with $T = 50$, but its performance improves markedly for $T = 200$.

When $\alpha = 1$, a bias appears, and this gets reflected in over-rejections (the rejection rate is 32.1% instead of the nominal 5% with $N = T = 50$) which disappears as we increase N . The bootstrap corrects most of this bias and reduces the rejection rate to about 11% with $N = T = 50$. The two types of intervals are remarkably close.

One last interesting observation is that the behavior of the individual components does not translate directly to the prediction intervals. For example, the asymptotic confidence intervals for both the factor and coefficients tend to over reject with $\alpha = 1$, but the interval for the conditional mean is reasonably accurate. This is because, contrary to what the theory assumes, there is a negative correlation between the two components of the forecast error in finite samples. This negative correlation arises from the presence of \tilde{F}_T in the first term and $(\tilde{F}_T - HF_T)$ in the second term. This means that the variance of the sum of the two terms is less than the sum of the variances as is assumed in the asymptotic approximation.

In summary, the bootstrap is particularly useful in two situations. The first one is in constructing prediction intervals for y_{T+h} when the distribution of the innovation is non-Gaussian and its variance is large relative to the uncertainty around the conditional mean. The second one is in constructing confidence intervals for the conditional mean. With a non-zero coefficient on the factors, the OLS estimator is severely biased unless N is large, and the bootstrap can correct this bias as documented by Gonçalves and Perron (2011).

7 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. For the realization, the importance of the bootstrap is in providing 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 hold 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

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

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

- (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 2 *The term $O_{P^*} (1/T^{3/4})$ 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 $\max_{1 \leq s \leq T} \|F_s\| = O_P (T^{1/4})$ using the fourth order moment assumption on F_s (cf. Assumption 1(a)). In general, if $E \|F_s\|^q \leq M$ for all s , then $\max_{1 \leq s \leq T} \|F_s\| = O_P (T^{1/q})$ and we will obtain a term of order $O_{P^*} (1/T^{1-1/q})$.*

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

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

where $T^{-1} \sum_{s=1}^T \left\| \tilde{F}_s^* - H^* \tilde{F}_s \right\|^2 = O_{P^*}(\delta_{NT}^{-2})$ by Lemma 3.1 of Gonçalves and Perron (2011) (note that this lemma only requires Conditions A*(b), A*(c), and B*(d), which correspond to our Condition $\mathcal{A}.1(a)$, $\mathcal{A}.2(a)$ and $\mathcal{A}.5$). For b_t^* , we have that (ignoring H^* , which is $O_{P^*}(1)$),

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

where $b_{1t}^* = O_P(1/\delta_{NT} \sqrt{T})$ using the fact that $T^{-1} \sum_{s=1}^T \left\| \tilde{F}_s - H F_s \right\|^2 = O_P(\delta_{NT}^{-2})$ 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 $\mathcal{A}.1(b)$. For b_{2t}^* , note that (ignoring $H = O_P(1)$),

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

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

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

for M sufficiently large. For (b), we follow exactly the proof of Bai (2003) and use the second part of Condition $\mathcal{A}.2$ to bound $T^{-1} \sum_{s=1}^T \zeta_{st}^{*2} = O_{P^*}(1/N)$ 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 $T^{-1} \sum_{s=1}^T \tilde{F}_s \eta_{st}^* = N^{-1} H^* \sum_{i=1}^N \tilde{\lambda}_i e_{it}^* = O_{P^*}(1/\sqrt{N})$ 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 $T^{-1} \sum_{s=1}^T \tilde{F}_s \zeta_{st}^* = O_{P^*}(1/\sqrt{NT})$ for each t and we use Condition $\mathcal{A}.5$ to bound $T^{-1} \sum_{s=1}^T \zeta_{st}^{*2} = O_{P^*}(1/N)$ 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 $\sqrt{N}/T^{3/4} \rightarrow 0$ whereas the second and the fourth terms are $O_{P^*}(1/\delta_{NT}) = o_{P^*}(1)$ as $N, T \rightarrow \infty$.

Thus, we have that

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

given the definition of H^* and the fact that $\tilde{V} = \frac{\tilde{\Lambda}' \tilde{\Lambda}}{N}$. Since $\det(\Gamma_t^*) > \epsilon > 0$ 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 $\tilde{V} \rightarrow^P V$ under our assumptions imply that $\Pi_t^* \rightarrow^P \Pi_t \equiv V^{-1} Q \Gamma_t Q' V^{-1}$. 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 $\Phi^* = \text{diag}(H^*, I_q)$, it follows that

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

where the remainder is

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

First, we argue that

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

where B_T^* is the asymptotic variance of $\hat{y}_{T+h|T}^* - y_{T+h|T}^*$, i.e. $B_T^* = a \text{Var}^* \left(\hat{y}_{T+h|T}^* - y_{T+h|T}^* \right) = \frac{1}{T} \hat{z}_T' \Sigma_\delta \hat{z}_T + \frac{1}{N} \hat{\alpha}' \Pi_T \hat{\alpha}$. To show (27), we follow the arguments of Bai and Ng (2003, proof of their Theorem 3) and show that (1) $Z_{1T}^* = \sqrt{T} \left(\Phi^{*'} \hat{\delta}^* - \hat{\delta} \right) \rightarrow^{d^*} N(-c \Delta_\delta, \Sigma_\delta)$; (2) $Z_{2T}^* = \sqrt{N} \left(H^{*-1} \tilde{F}_T^* - \tilde{F}_T \right) \rightarrow^{d^*} N(0, \Pi_T)$; (3) Z_{1T}^* and Z_{2T}^* are asymptotically independent (conditional on the original sample). Condition (1) follows from Gonçalves and Perron (2011) under Assumptions 1-6; (2) follows from Corollary 4.1 provided $\sqrt{N}/T^{11/12} \rightarrow 0$ 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+h}^* .

Proof of Condition \mathcal{A} for the wild bootstrap. Starting with $\mathcal{A}.1$, note that $\mathcal{A}.1(a)$ was verified in Gonçalves and Perron (2011). We verify $\mathcal{A}.1(b)$ 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

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

which implies that

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N |\tilde{e}_{it}|^2 &\leq 3 \frac{1}{N} \sum_{i=1}^N |e_{it}|^2 + 3 \frac{1}{N} \sum_{i=1}^N \|\lambda_i\|^2 \|H^{-1}\|^2 \|\tilde{F}_T - H F_T\|^2 \\ &\quad + 3 \frac{1}{N} \sum_{i=1}^N \|\tilde{\lambda}_i - H^{-1'} \lambda_i\|^2 \|\tilde{F}_T\|^2. \end{aligned}$$

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 $\|\tilde{F}_T - H F_T\|^2 = O_P(1/N) = o_P(1)$; and the third term is $O_P(1)$ given Lemma C.1.(ii) of Gonçalves and Perron (2011) and the fact that $\|\tilde{F}_T\|^2 = O_P(1)$. Next, we verify $\mathcal{A}.2$. Part (a) was verified already by Gonçalves and Perron (2011), so we only need to check part (b) for $t = T$. Following the proof of Theorem 4.1 in Gonçalves and Perron (2011) (condition $\mathcal{A}^*(c)$), we have that

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

where the first factor in (28) can be bounded by an argument similar to that used above to bound $\frac{1}{N} \sum_{i=1}^N \tilde{e}_{iT}^2$, and the second factor can be bounded by Lemma C.1 (iii) of Gonçalves and Perron (2011). $\mathcal{A}.3$ follows by an argument similar to that used by Gonçalves and Perron (2011) to verify Condition $\mathcal{B}^*(b)$. In particular,

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

under our assumptions. Conditions $\mathcal{A}.4$ and $\mathcal{A}.5$ correspond to Gonçalves and Perron's (2011) 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 $\eta_{iT} \sim \text{i.i.d. } (0, 1)$ across i , note that

$$\Gamma_T^* = \text{Var}^* \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{\lambda}_i e_{iT}^* \right) = \frac{1}{N} \sum_{i=1}^N \tilde{\lambda}_i \tilde{\lambda}_i' \tilde{e}_{iT}^2 \rightarrow^P Q \Gamma_T Q',$$

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

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \ell' \Gamma_T^{*-1/2} \tilde{\lambda}_i e_{iT}^* = \frac{1}{\sqrt{N}} \sum_{i=1}^N \underbrace{\ell' \Gamma_T^{*-1/2} \tilde{\lambda}_i \tilde{e}_{iT} \eta_{iT}}_{=\omega_{iT}^*} \rightarrow^{d^*} N(0, 1),$$

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

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

Thus, it suffices to verify Lyapunov's condition, i.e. for some $r > 1$, $\frac{1}{N^r} \sum_{i=1}^N E^* |\omega_{iT}^*|^{2r} \rightarrow^P 0$. 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. $\Gamma_T^* = \frac{1}{N} \sum_{i=1}^N \tilde{\lambda}_i \tilde{\lambda}_i' \tilde{e}_{iT}^2 \rightarrow^P Q \Gamma_T Q'$, 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 $\hat{z}_T^{*'} \hat{\Sigma}_\delta^* \hat{z}_T^* - \hat{z}_T' \Sigma_\delta \hat{z}_T \rightarrow^{P^*} 0$, and $\hat{\alpha}^{*'} \hat{\Sigma}_{\tilde{F}_T}^* \hat{\alpha}^* - \hat{\alpha}' \hat{\Sigma}_{\tilde{F}_T} \hat{\alpha} \rightarrow^{P^*} 0$, 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,

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

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

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

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

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

We can write

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

where $\Phi = \text{diag}(H, I_q)$. This implies that

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

Similarly,

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

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 $s_{T+h}^* \rightarrow^{d^*} 1 - F_\varepsilon(-x\sigma_\varepsilon)$, in probability. Since $s_{T+h} \rightarrow^d 1 - F_\varepsilon(-x\sigma_\varepsilon)$ 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, 2011. "Principal Components Estimation and Identification of the Factors", manuscript, Columbia University.
- [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] 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.
- [8] Gonçalves, S. and B. Perron, 2011, "Bootstrapping Factor-Augmented Regressions", manuscript, Université de Montréal.
- [9] 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.
- [10] Wolf, M. and D. Wunderli, 2012. "Bootstrap joint prediction regions", manuscript. University of Zurich.

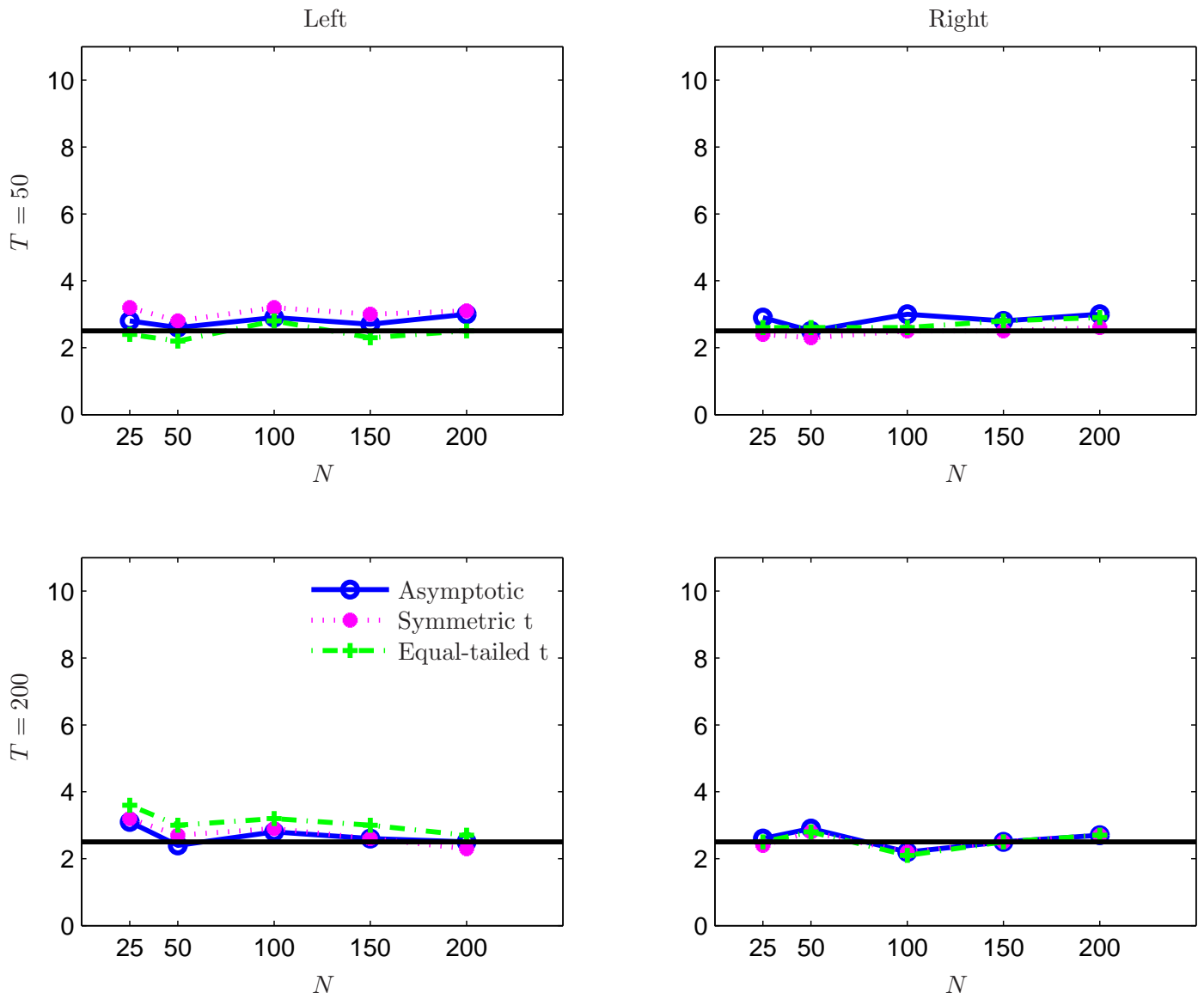


Figure 1: Left and right-hand side rejection rates for y_{T+h} , ε_{T+h} is normal, $\alpha = 0$

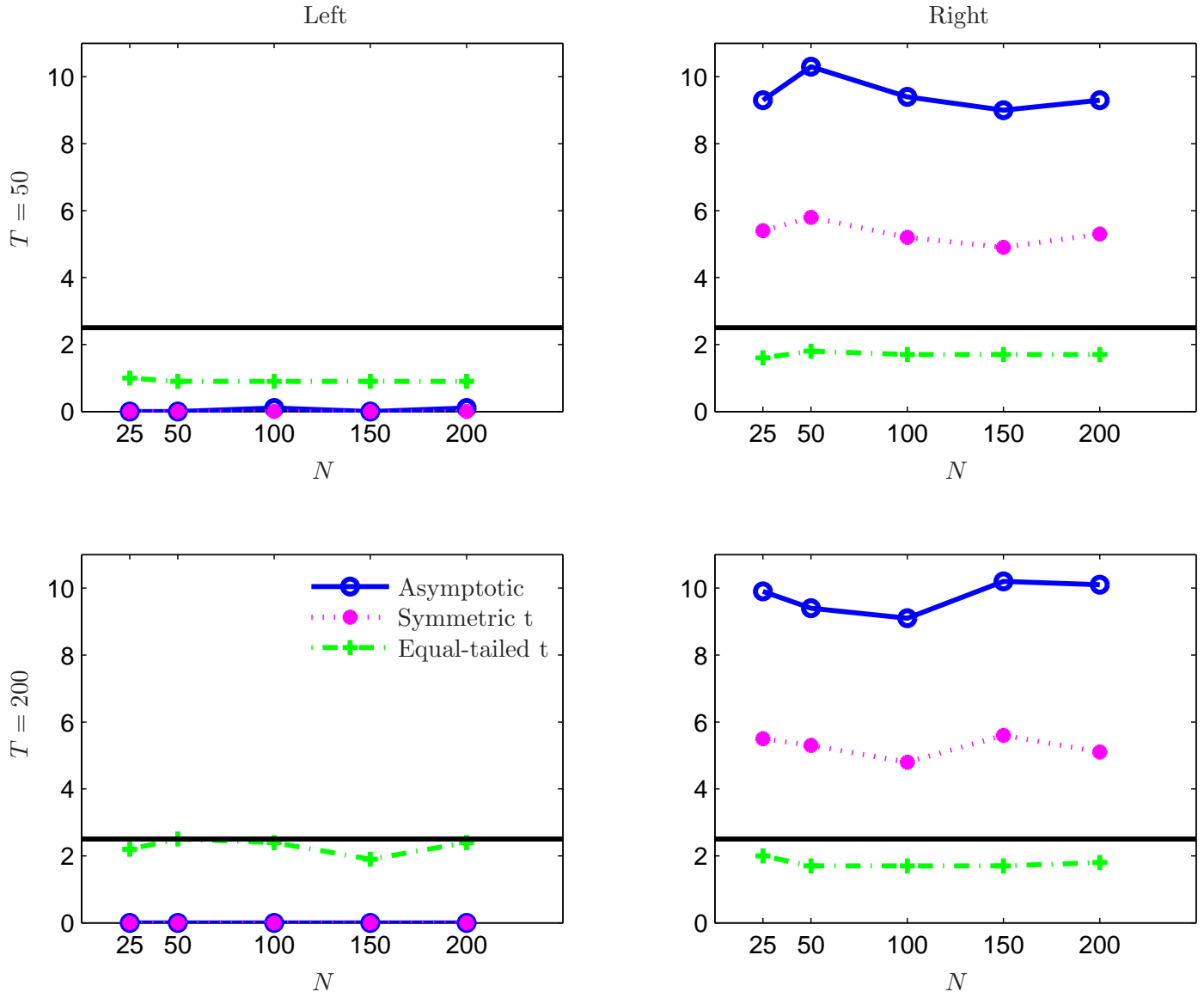


Figure 2: Left and right-hand side rejection rates for y_{T+h} , ε_{T+h} is mixture, $\alpha = 0$

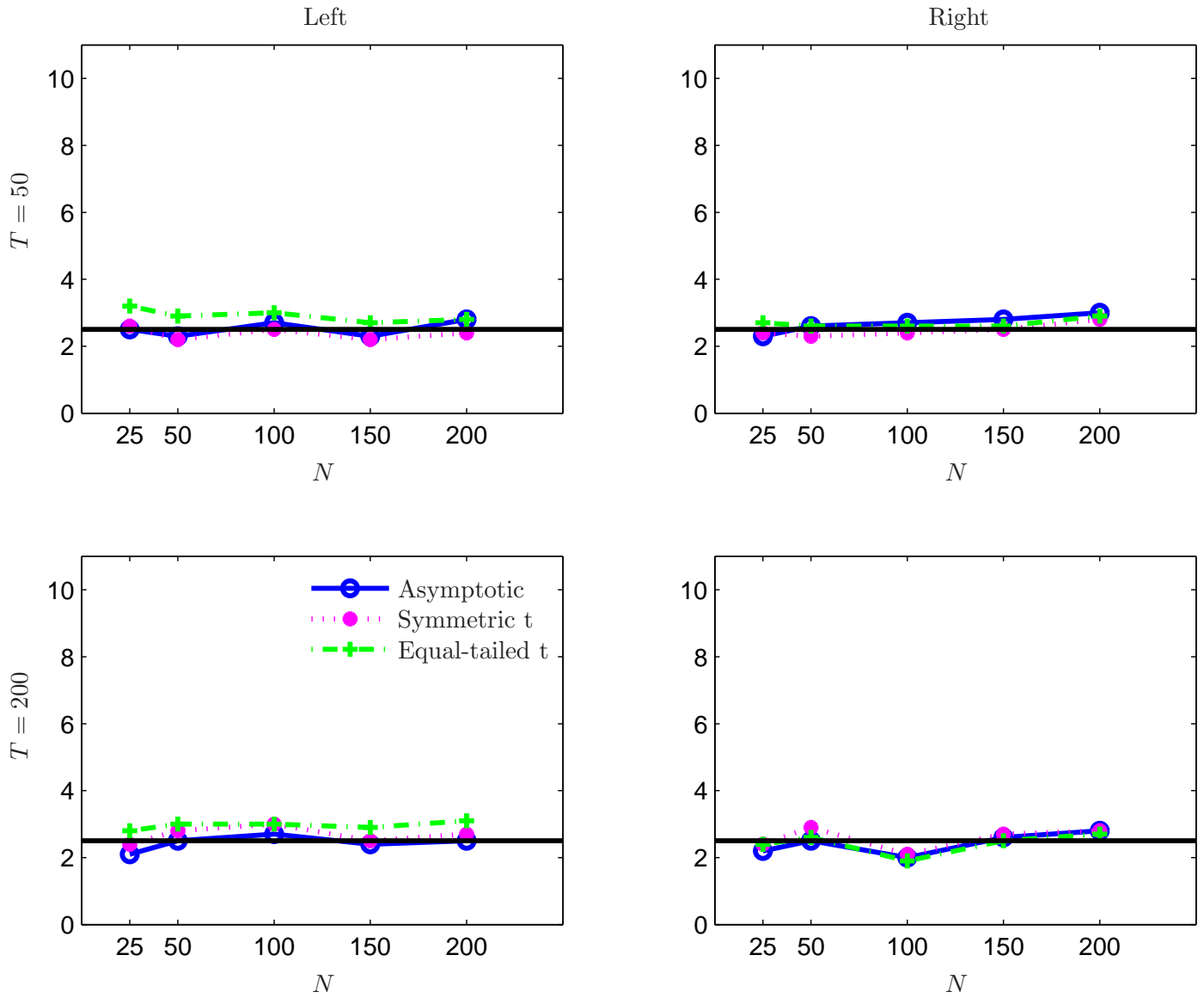


Figure 3: Left and right-hand side rejection rates for y_{T+h} , ε_{T+h} is normal, $\alpha = 1$

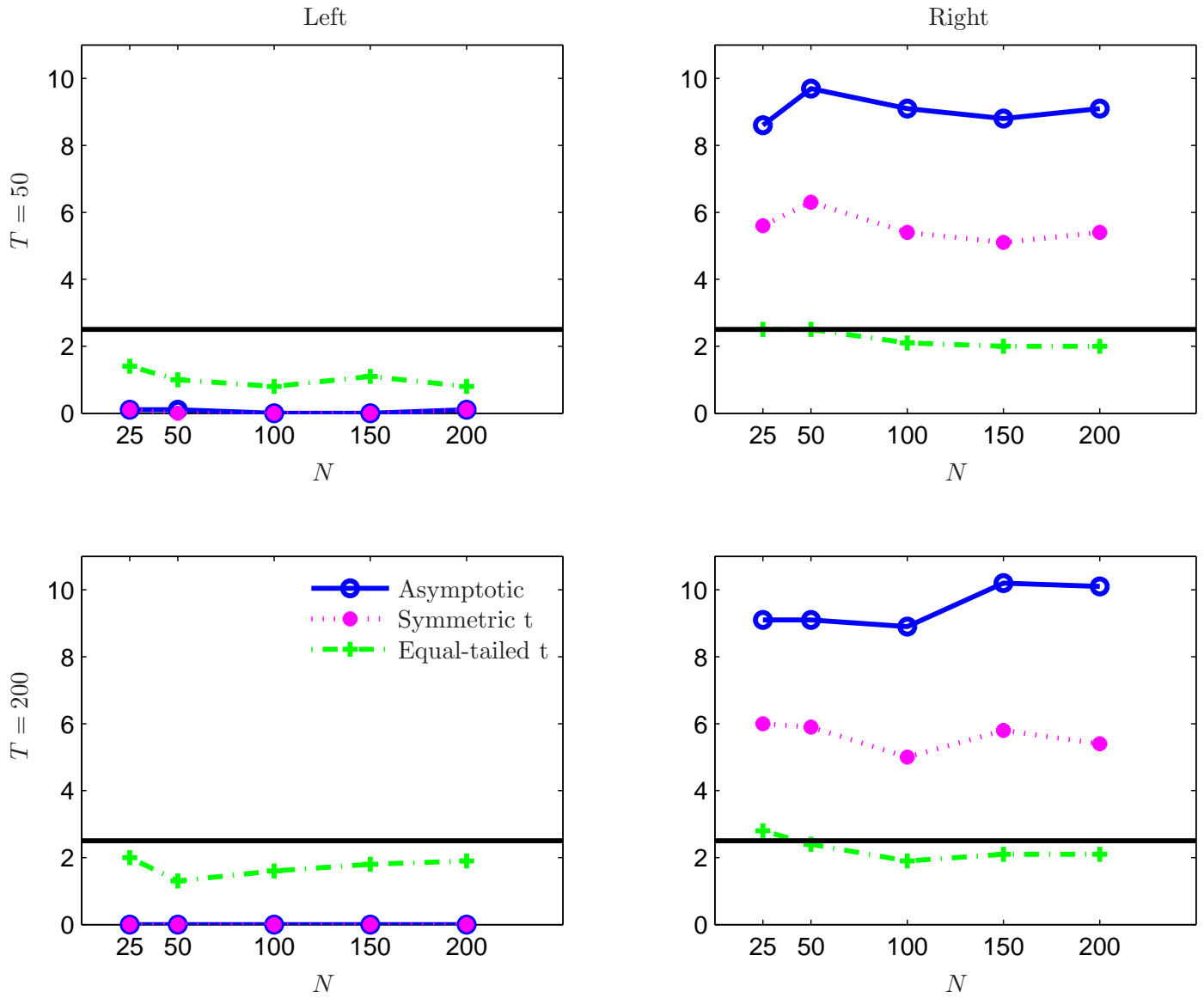


Figure 4: Left and right-hand side rejection rates for y_{T+h} , ε_{T+h} is mixture, $\alpha = 1$

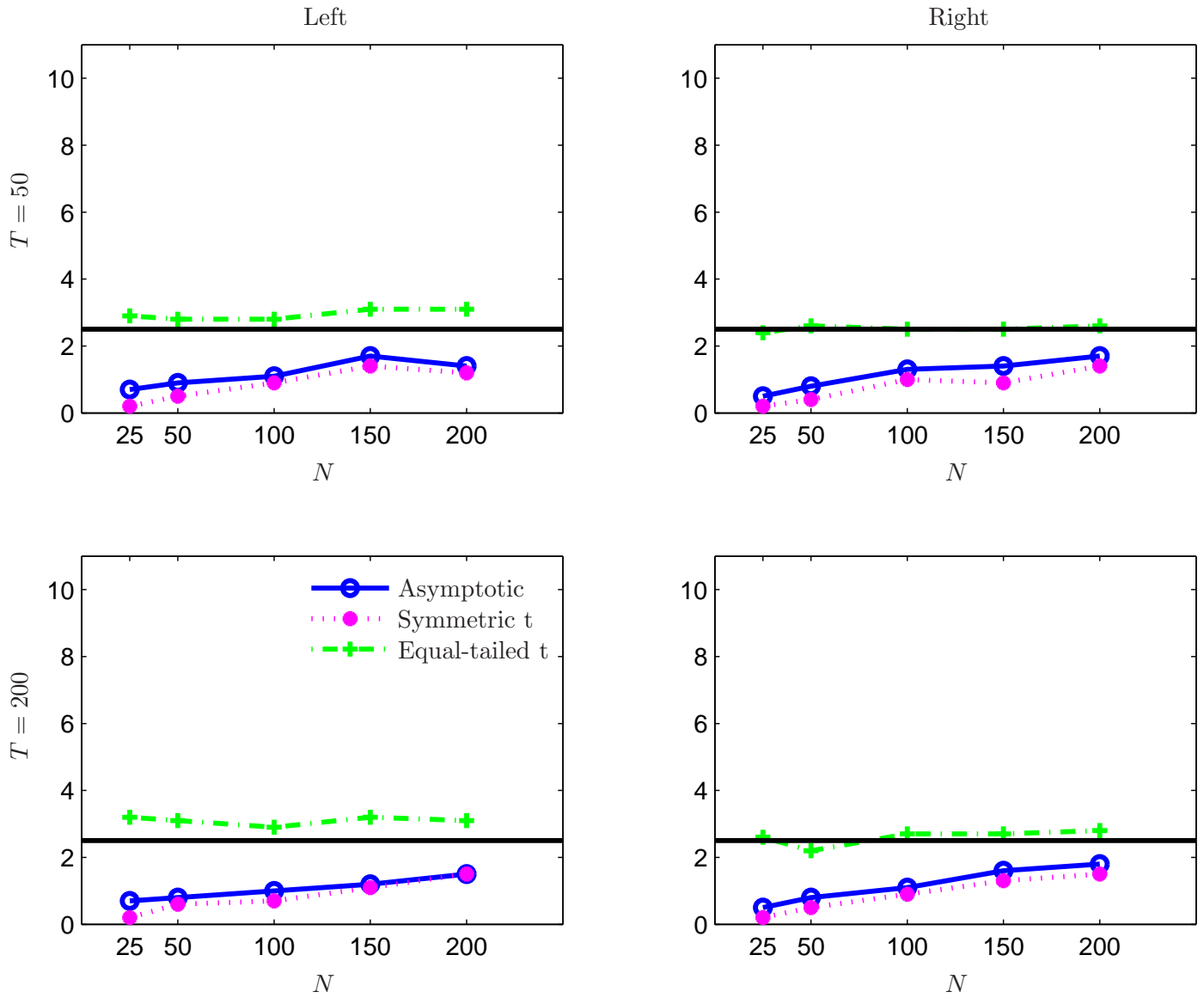


Figure 5: Left and right-hand side rejection rates for $y_{T+h|T}$, ε_{T+h} is normal, $\alpha = 0$

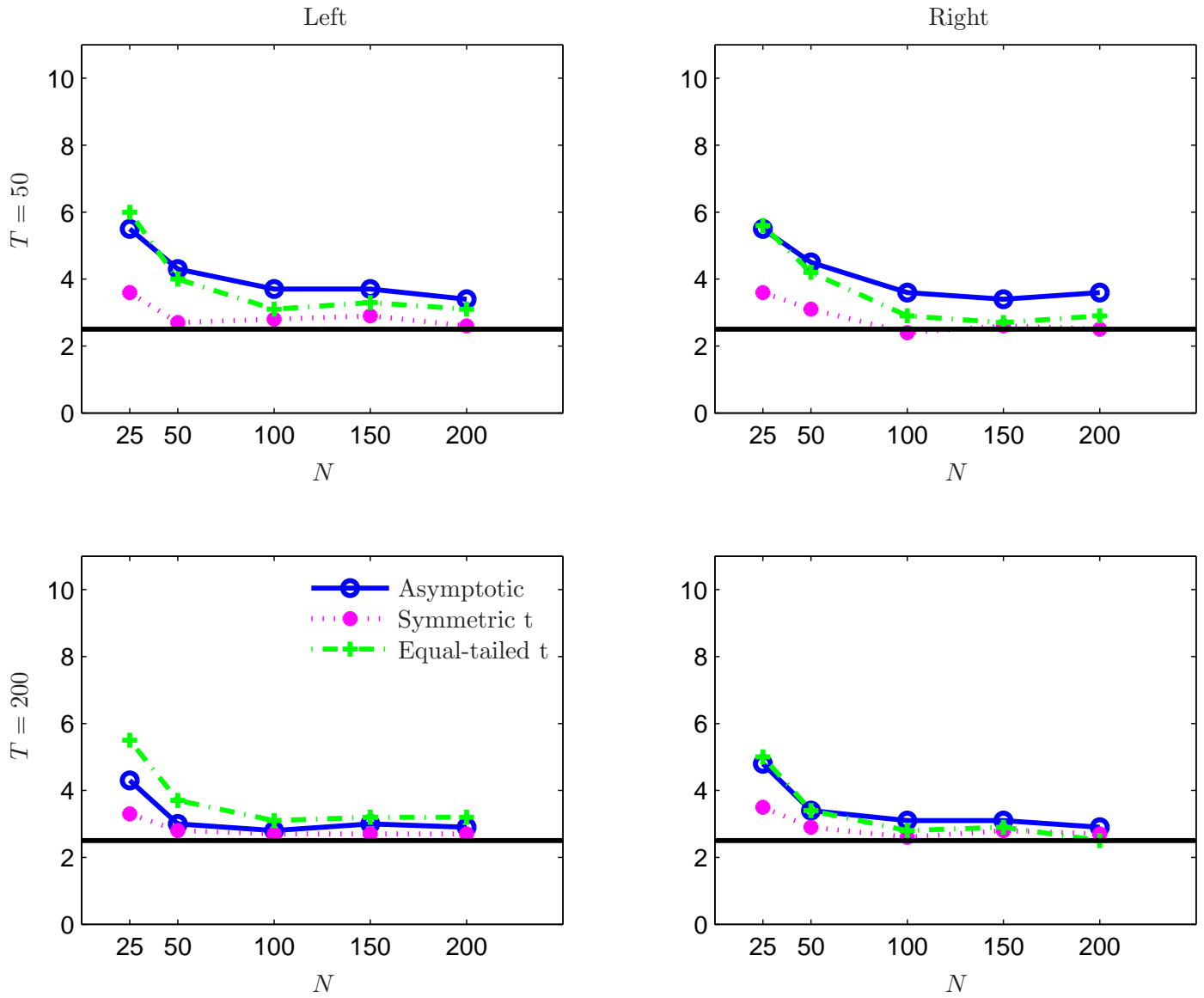


Figure 6: Left and right-hand side rejection rates for $y_{T+h|T}$, ε_{T+h} is normal, $\alpha = 1$

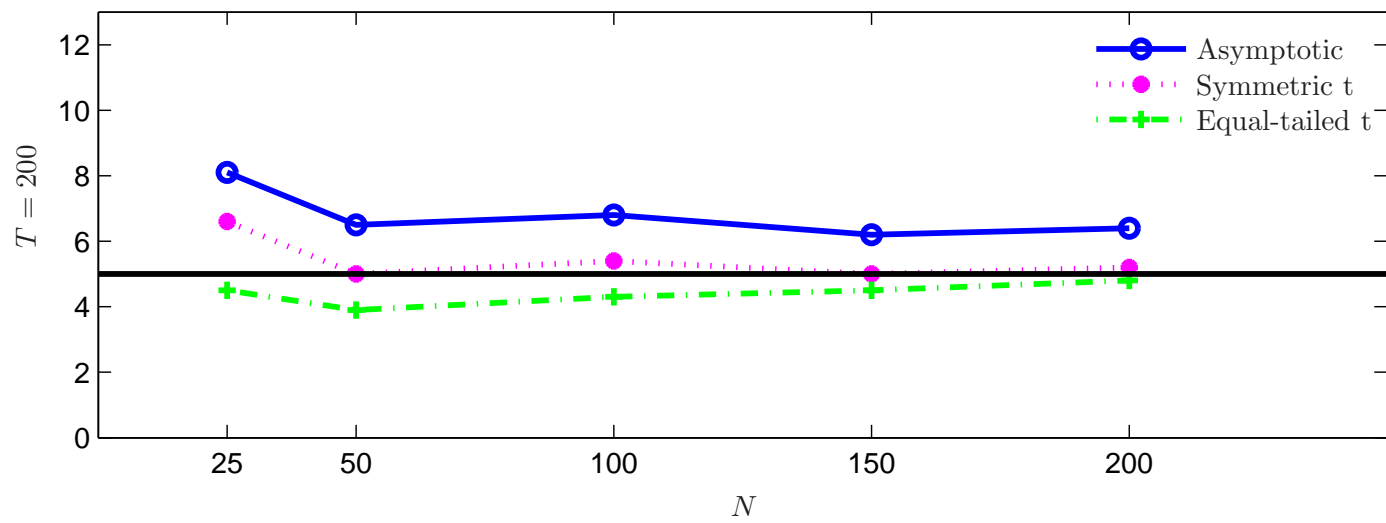
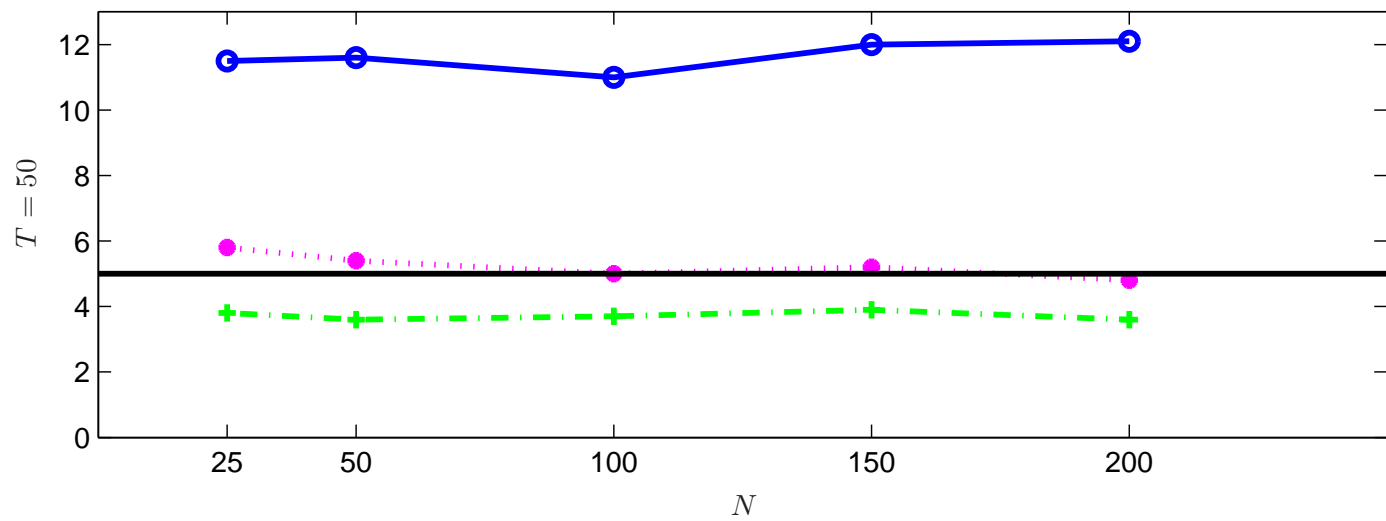


Figure 7: Total rejection rates for HF_T

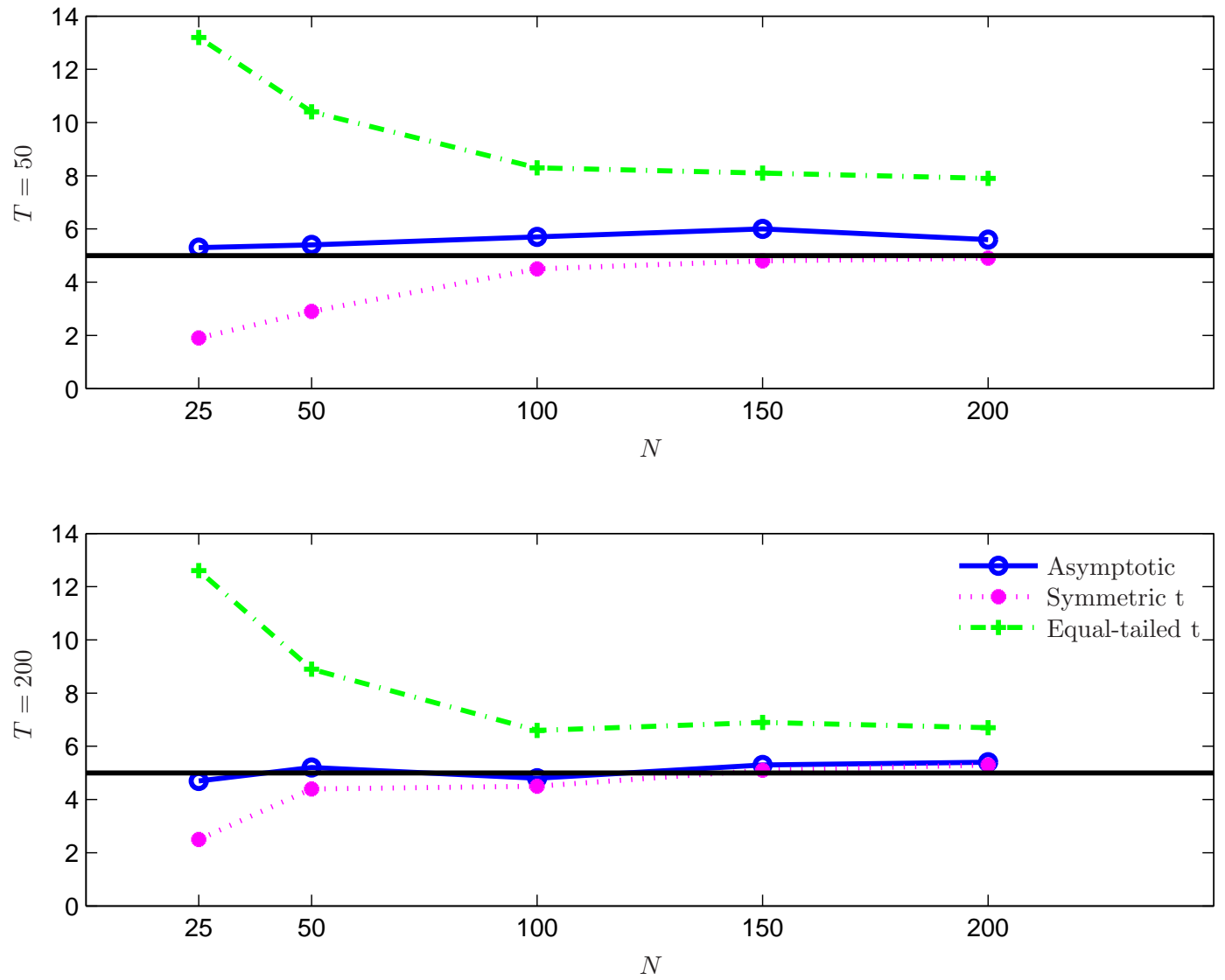


Figure 8: Total rejection rates for $\delta, \alpha = 0$

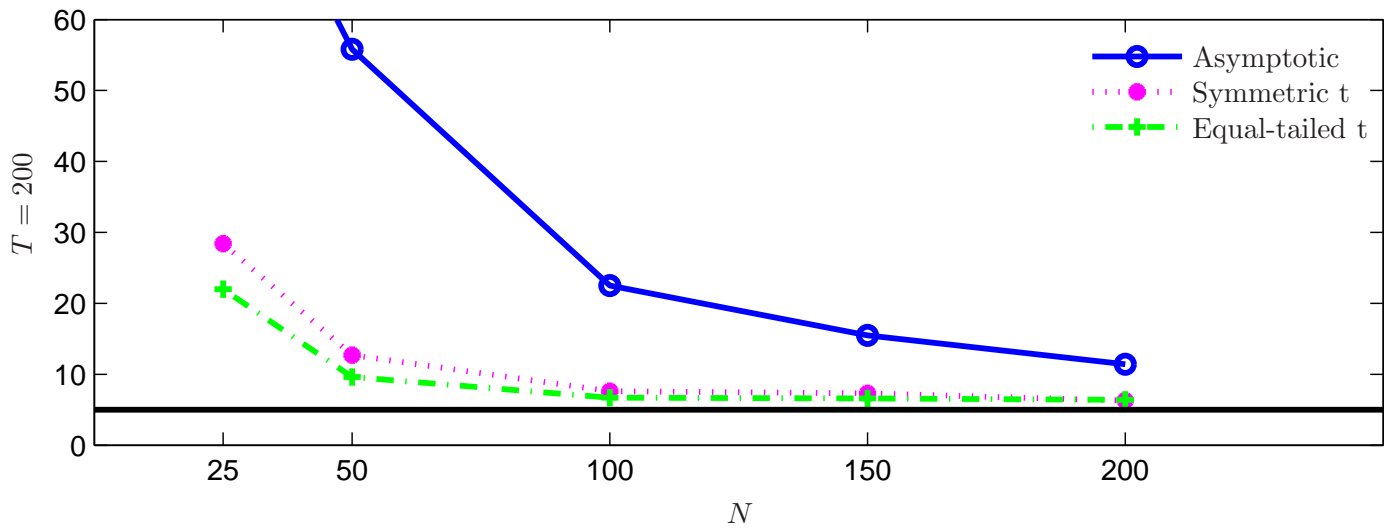
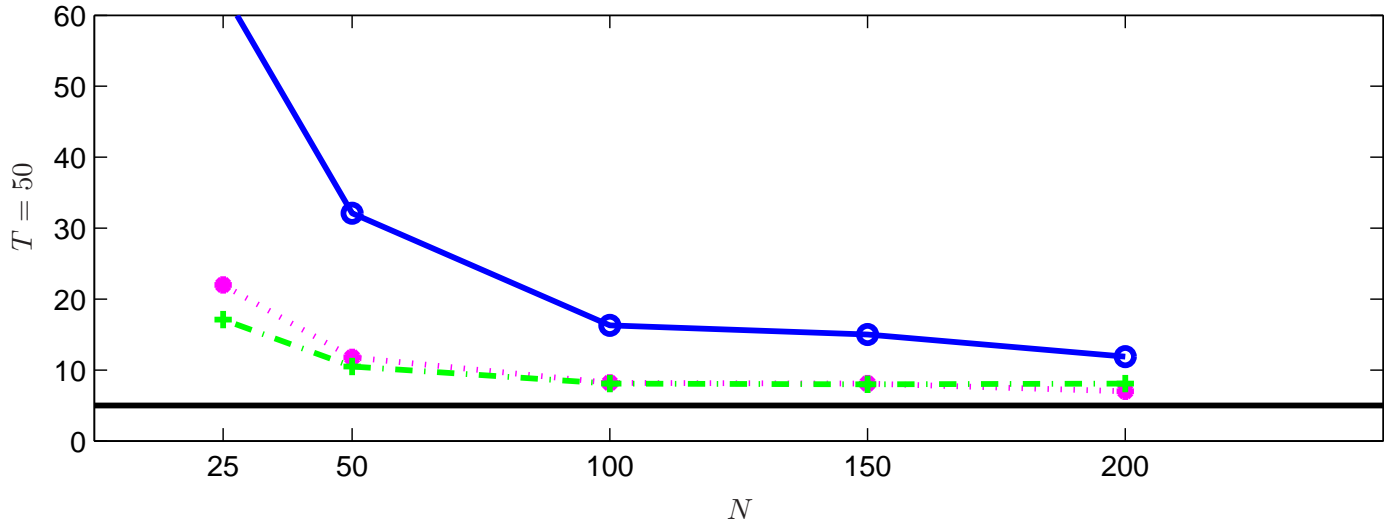


Figure 9: Total rejection rates for $\delta, \alpha = 1$