Finite Sample Performance of Principal Components Estimators for Dynamic Factor Models:
Asymptotic vs. Bootstrap Approximations*

Mototsugu Shintani† and Zi-Yi Guo‡

This version: September 2012
(Job Market Paper)

Abstract

This paper investigates the finite sample properties of the two-step estimators of dynamic factor models when unobservable common factors are estimated by the principal components methods in the first step. Effects of the number of individual series on both estimation of the factor dynamics and forecasting regression are considered in a Monte Carlo simulation. For less persistent data, asymptotic approximation performs well when the number of the series is at least half the number of time series observations. Otherwise, the estimators of autoregressive parameters are biased downward with the bias becoming larger as the data becomes more persistent. In such a case, bootstrap procedures are more likely to yield a better result. The bootstrap method also works well in out-of-sample forecasting tests with a small number of available series.

Keywords: Bootstrap; Dynamic Factor Model; Out-of-sample Forecasts; Principal Components

JEL classification: C15; C53

*The earlier version of the paper was circulated under the title “Finite sample performance of principal components estimators for dynamic factor models.” The authors would like to thank Todd Clark, Silvia Goncalves, James MacKinnon, Serena Ng, Benoit Perron and seminar participants at Indiana University and University of Montreal for helpful comments and discussions.

†Department of Economics, Vanderbilt University; e-mail: mototsugu.shintani@vanderbilt.edu.
‡Department of Economics, Vanderbilt University; e-mail: ziyi.guo@vanderbilt.edu.
1 Introduction

This paper investigates the finite sample properties of two-step estimators for dynamic factor models when unobservable common factors are estimated by the principal components methods in the first step. The first-step estimation is followed by the estimation of time series models of common factors in the second step. From a simulation experiment, we evaluate the effect of the number of the series ($N$) relative to the time series observations ($T$) on the performance of such a two-step estimator. Furthermore, we propose a simple bootstrap procedure that works well in the case of relatively small $N$.

In modern macroeconomics literature, dynamic stochastic general equilibrium (DSGE) models predict that a small set of driving forces is responsible for covariation in macroeconomic variables. Since dynamic factor models fit well within this framework, they are often employed as an empirical tool in macroeconomic analysis, including construction of business cycle indexes, analysis of stock price dynamics and international business cycle analysis. One way to estimate dynamic factor models is to employ a single-step method based on the exact maximum likelihood, which is the approach used by Stock and Watson (1989) in their analysis of business cycle index. However, the two-step method based on principal components has become increasingly popular in recent empirical studies.

rajczyk and Linton (2006) applied a GARCH model to the common factor in asset return. Shintani (2005, 2008) utilized the nonparametric regression method to investigate nonlinearity in business cycles. The two-step method has been also applied to nonstationary panel data to develop unit root tests by Bai and Ng (2004) and Moon and Perron (2004) among others.

There are several notable advantages of this two-step procedure as an estimation method of dynamic factor models compared to the single-step procedure. First, the method is valid with a large number of cross-sectional observations $N$ with allowance for cross-sectional correlations of idiosyncratic errors. Therefore, it allows an approximate factor structure contrast to the strict factor model which requires uncorrelated errors. Second, it allows very general dynamic structures of latent factors. A basic requirement of the time series property of the common factor is simply a covariance stationarity. In consequence, we can justify estimating various models, such as the class of conditional heteroskedastic models considered by Connor, Korajczyk and Linton (2006) or nonparametric models considered by Shintani (2005, 2008), in the second step estimation. Third, it has an advantage from the viewpoint of computational simplicity. For example, to estimate a time series model of the latent factor with a Markov switching structure, Kim and Nelson (1998) employed a single-step estimator obtained from the MCMC method. However, such a simulation-based estimation usually requires a fairly long computational time, even if the number of variables ($N$) is very small, such as $N = 4$. On the other hand, the two-step method applied by Diebold (1998) with $N$ being around 200 only requires about the same time we need to obtain estimates from a Markov switching model for a single observed series since the time required for the first step (principal components method) is almost negligible.

In applications, inference, or confidence intervals regarding the second-step estimator, is often of main interest. For example, to investigate the persistence of common business cycle
shocks, reporting confidence intervals of the impulse responses or half-lives obtained from the second-step estimation is preferable. However, in general, the limit distribution of the second-step estimator may depend on the estimation error of the factor in the first step. This fact is akin to the issue of generated regressors or the errors in variables. Fortunately, in the context of dynamic factor models, such a problem disappears with the aid of increasing information in the cross-section. For the first-step estimation, the large $N$ asymptotic results obtained by Connor and Korajczyk (1986) and Bai (2003) imply $\sqrt{N}$-consistency of the principal components estimators of common factors up to a scaling factor. Therefore, if $N$ increases at a sufficient rate, we can treat the estimated common factor as if we directly observe the true common factor when conducting inference. However, since this argument is based on the asymptotic theory, it is of practical interest to examine how large $N$ is required for a time series observation $T$ typically available for economic data. For this purpose, we conduct a Monte Carlo experiment designed to assess the properties of the two-step estimator with finite $N$. In addition, we provide the evidence that a simple bootstrap procedure works well in case of small $N$ even if the data is persistent.

There are several simulation results available in the literature on the principal components estimator of dynamic factor models. Stock and Watson (1998) reported the finite sample simulation results on the magnitude of the first-step estimation error of the common factor as well as the performance of an out-of-sample forecast based on the estimated factor relative to that of an infeasible forecast with a true factor. Boivin and Ng (2006) reported similar performance measures in investigating the marginal effect of increasing $N$ when there is a strong cross-sectional correlation of the errors. In addition, Stock and Watson (1998) and Bai and Ng (2002) found that information criteria designed to determine the number of the factors performed well in a finite sample. However, none of these studies directly investigated the effect of $N$ on the estimation of dynamic structure of the common factors.
The remainder of the paper is organized as follows: Section 2 reviews the asymptotic theory of the two-step estimator and investigate the finite sample performance of the estimator in simulation. Section 3 considers a bootstrap approach to improve finite sample performance. Section 4 reports the result of an out-of-sample forecasting. Section 5 provides an empirical illustration of the procedures introduced in the paper. Some concluding remarks are made in Section 6.

2 Asymptotic Approximation

2.1 Review of Large $N$ Asymptotics

In this section, we review the asymptotic properties of a two-step estimator of dynamic factor structure. Let $x_{it}$ be an $i$-th component of $N$-dimensional multiple time series $X_t = (x_{1t}, \ldots, x_{Nt})'$ and $t = 1, \ldots, T$. A natural way to explain the comovement of $x_{it}$'s caused by a single factor, such as productivity shocks, demand shocks or monetary policy shocks, is to use a simple one-factor model

$$x_{it} = \lambda_i^0 f_t^0 + e_{it}$$

for $i = 1, \ldots, N$, where $\lambda_i^0$'s are factor loadings with respect to $i$-th series, $f_t^0$ is a scalar common factor, and $e_{it}$'s are (uncorrelated) idiosyncratic shocks. While the factor $f_t^0$ is not directly observable, it is known that the $f_t^0$ can be consistently estimated (up to normalization) by using the first principal component of the $N \times N$ covariance matrix $X'X$ where $X$ is the $T \times N$ data matrix with $t$-th row $X_t'$, or by using the first eigenvector of the $T \times T$ matrix $XX'$. Since principal components are not scale-invariant, it is common practice to standardized all $x_{it}$’s to have zero sample mean and unit sample variance before applying the principal components method. We denote this common factor estimator by $\hat{f}_t$ with a normalization $T^{-1} \sum_{t=1}^{T} \hat{f}_t^2 = 1$.

There are several ways to generalize this simple model. First, multiple factors can be
included. Second, instead of using the static factor model, a dynamic structure can be introduced by allowing (i) a dynamic data generating process for $f^0_t$, (ii) lags of $f^0_t$ in (1) and (iii) serial correlation in $e_{it}$’s. The factor model with such structure is called as a dynamic factor model and has become popular in macroeconomic analysis after influential works by Sargent and Sims (1977), Geweke (1977) and Stock and Watson (1989). Third, $e_{it}$’s can be correlated across series. In such a case, the model become an approximate factor model as opposed to the strict factor model that does not allow cross-sectional correlation.

In this paper, we consider an explicit factor dynamic structure so that the factor is generated from a linear stationary AR($p$) model,

$$\phi(L)f^0_t = \varepsilon_t$$

(2)

where $\phi(L) = 1 - \phi_1L - \phi_2L^2 - \cdots - \phi_pL^p$ with all roots of $\phi(z) = 0$ lie outside the unit circle, and $\varepsilon_t$ is i.i.d. with $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma^2$ and a finite fourth moment. If $f^0_t$ is available, the AR parameters $\phi = (\phi_1, ..., \phi_p)^T$ can be estimated by ordinary least squares (OLS),

$$\hat{\phi} = \left( \sum_{t=1}^{T} F^0_{t-1} F^0_{t-1}' \right)^{-1} \sum_{t=1}^{T} F^0_{t-1} f^0_t$$

where $F^0_{t-1} = (f^0_{t-1}, ..., f^0_{t-p})'$. However, since $f^0_t$ is not directly observable, we replace $f^0_t$ by $\tilde{f}_t$ and the feasible estimator is

$$\tilde{\phi} = \left( \sum_{t=1}^{T} \tilde{F}_{t-1} \tilde{F}_{t-1}' \right)^{-1} \sum_{t=1}^{T} \tilde{F}_{t-1} \tilde{f}_t$$

where $\tilde{F}_{t-1} = (\tilde{f}_{t-1}, ..., \tilde{f}_{t-p})'$.

Below, we consider the validity of such two-step estimation methods. To this end, we employ the following assumptions on the moment conditions for factors, factor loadings and idiosyncratic errors.
Assumption F (factors): (i) \( E(f_0^0) = 0, \ E((f_0^0)^2) = \Sigma_F = 1, \ E((f_0^0)^4) < \infty, \) and (ii) \( F^0'F^0/T - \Sigma_F = o_p(1) \) where \( F^0 = [f_1^0, \ldots, f_T^0]' \).

Assumption FL (factor loadings): (i) \( |\lambda_i^0| \leq \lambda < \infty, \) and (ii) \( \Lambda^0\Lambda^0/N - \Sigma_\Lambda = o_p(1) \) where \( \Lambda^0 = [\lambda_1^0, \ldots, \lambda_N^0]' \).

Assumption E (errors): For some positive constant \( M < \infty, \)

(i) \( E(e_{it}) = 0, \ E(e_{it}^2) = \sigma_{ei}^2 \leq M, \ E|e_{it}^8| \leq M, \) (ii) \( E(e_{is}e_{it}) = 0 \) for all \( t \neq s, \) and (iii) \( E(e_{it}e_{jt}) = \tau_{ij} \leq M \) for all \( t, i \) and \( j. \)

Assumption E allows cross-sectional correlation and heteroskedasticity but not serial correlation of idiosyncratic error terms. It should be noted that Assumption E can be replaced by a weaker assumption that allows serial correlations of idiosyncratic errors (see Bai, 2003, and Bai and Ng, 2002). In addition, we employ the following assumption on the relation among three random variables.

Assumption I (independence): \( \{f_t^0\}, \{\lambda_i^0\} \) and \( \{e_{it}\} \) are mutually independent.

The following proposition provides the validity of the two-step estimator of AR coefficients and the inference.

**Proposition 1.** Let \( x_{it} \) and \( f_t^0 \) be generated from (1) and (2), respectively, and suppose that assumptions F, FL, E and I are satisfied. If \( N = cT^\alpha \) with \( c > 0 \) and \( \alpha > 1/2 \) as \( T \to \infty, \) then \( \sqrt{T}(\hat{\phi} - \phi) \overset{d}{\to} N(0, \Gamma^{-1}) \) where \( \sigma^2\Gamma = E[F_{t-1}^0F_{t-1}^0]. \)

When the number of the series (\( N \)) increases at a sufficiently fast rate relative to the time series observations (\( T \)), the effect of the estimation error of the factor becomes negligible in the limit distribution of the second-step estimators. This asymptotic equivalence of \( \hat{\phi} \) and \( \tilde{\phi} \) can be derived by a simple application of the result of Bai (2003). The proof is provided in the Appendix.
For the purpose of evaluating the performance of the two-step estimator with small $N$, we conduct a Monte Carlo experiment. The observed series are generated from (1) with the factor loading $\lambda_{it}^0 \sim N(0, 1)$, serially and cross-sectionally uncorrelated homoskedastic error $e_{it} \sim N(0, \sigma_e^2)$, and the factor $f_{0t}$ being generated from AR(1) model

$$f_{0t} = \rho f_{0,t-1} + \varepsilon_t$$

where $|\rho| < 1$ and $\varepsilon_t \sim N(0, 1 - \rho^2)$ so that $f_{0t}$ has a unit variance. Relative size of common component and idiosyncratic error in $x_{it}$ is controlled by changing the ratio $Var(\lambda_{it}^0 f_{0t}^0)/Var(e_{it}) = 1/\sigma_e^2$. We refer to this quantity simply as a signal-to-noise ratio and the values 2.0, 1.0 and 0.5 are considered in the simulation. For the AR parameter, $\rho = 0.5, 0.8$ and 0.9 are considered. The initial value $f_{0t}^0$ is drawn from the unconditional distribution of $f_{0t}^0$, that is $N(0, 1)$. The performance of the two-step estimator is evaluated by the effective coverage rate of the nominal 90% confidence intervals. All the simulation results are based on 1000 replications, thus the standard error of 90% coverage rate in the simulation is about 0.01 ($\approx \sqrt{0.9 \times 0.1/1000}$). In addition to the coverage rates, the median length of the confidence intervals as well as the percentage of the cases in which the true $\rho$ lay to the left and right of the estimated intervals, is reported.

Table 1 shows coverage of confidence intervals based on the infeasible estimator $\hat{\rho}$ that cannot be obtained in practice but can be obtained in the simulation, since the factor $f_{0t}^0$ is known. Because we cannot expect feasible two-step estimator $\hat{\rho}$ to perform better than such an infeasible estimator, this result serves as a benchmark to evaluate the performance of the two-step estimator. Since the main objective of this experiment is to evaluate the effect of small $N$ in the two-step estimation, it is desirable not to have a result influenced by small $T$. Table 1 suggests that coverage of conventional asymptotic confidence intervals is very
accurate for sample sizes $T = 100$ and 200. In addition to asymptotic intervals, bootstrap methods are also considered. Two standard bootstrap confidence intervals, Efron’s percentile confidence intervals and Hall’s percentile-$t$ intervals, are computed for all cases with number of bootstrap replication $B = 499$. For a near unit root case ($\rho = 0.9$), Hansen’s (1999) grid and grid-$t$ intervals are also reported since they are known to perform well in such a case. The bootstrap intervals turn out to have a good coverage except for the percentile intervals in the near unit root case. In addition, for all AR parameter values, the percentile intervals are found to have some downward biases. Nevertheless, with an exception of the percentile intervals, our results show that both asymptotic and bootstrap confidence intervals are near exact central confidence intervals. Therefore, the result for $\hat{\rho}$ in our simulation format seems to be a satisfactory benchmark case for evaluating the performance of the feasible two-step estimator.

Table 2 reports coverage of asymptotic confidence intervals based on the two-step estimator $\hat{\rho}$. The theoretical result in the previous section predicts that the coverage probability should be close to 0.90 provided sufficiently large $N$ relative to $T$. The values of $N/T$ we consider are 0.1, 0.2, 0.5, 1.0 and 2.0. Corresponding $N$s are 10, 20, 50, 100 and 200 for $T = 100$ and 20, 40, 100, 200 and 400 for $T = 200$. For $\rho = 0.5$ case, when $N/T \geq 0.5$, coverage is quite accurate regardless of signal to noise ratio or sample size. However, compared to the benchmark asymptotic result in Table 1, they are somewhat biased downward. When the AR parameter becomes as large as $\rho = 0.8$, good coverage is only obtained for $N/T \geq 0.5$ with signal-to-noise ratio being 2.0. Otherwise, the confidence intervals are not only biased downwards but also show poor coverage probability. One interesting observation with $\rho = 0.8$ case is that, when $N/T < 0.5$, lengths of intervals are much wider than those for $N/T \geq 0.5$

---

1It should be noted that constant and time trend are not included in our simulation. As reported in Hansen (1999) and others, inclusion of constant and trend, in general, leads to poorer coverage and significant bias with a sample size under consideration. In order to incorporate these cases in the simulation, we would need to use a larger sample size to eliminate the effect of finite $T$. 

8
case, with coverage being quite smaller than the nominal rate. This fact implies that the presence of bias is a major possible source of poor coverage. As expected, in general, the performance deteriorates as signal-to-noise ratio becomes smaller. However, such an effect is small in case of large $N$ and modest AR roots.

In summary, the asymptotic interval works well in terms of the coverage rate when $N$ is as large as half of $T$ and when the AR parameter is not close to unity. In contrast, when $N/T$ is as small as 0.1 or 0.2, there is a downward bias in the confidence interval which results in a poor coverage property. The effect of this downward bias becomes more severe as the AR parameter approaches to unity. This finding is somewhat similar to the small sample bias problem in the OLS estimator of AR models associated with finite $T$ in the sense that the magnitude of the downward bias increases as $\rho$ increases (see Kendall, 1954).

In the next section, we consider the possibility of improving the performance of the two-step estimator for small $N$ by introducing a bootstrap procedure.

3 Bootstrap approximation

3.1 Bootstrapping Two-step Estimators

The main purpose of this section is to construct a confidence interval of the two-step estimator based on a bootstrap method and determine whether it provides better coverage probability compared to the asymptotic interval. In the previous section, we found that the presence of small sample bias associated with finite $N$ was likely the main source of poor coverage. For a simple AR(1) estimator $\hat{\rho}$ without factor estimation, Kendall (1954) has shown that the dominant term of finite sample bias is of $O(T^{-1})$. From the proof of Proposition 1 in the Appendix, the dominant term from the estimation error of the factor $\tilde{\rho}$ is of order $1/\min\{N, T\}$. Therefore, if $N$ diverges at a rate slower than $T$ (so $1/2 < \alpha < 1$), bias from the factor estimation dominates Kendall’s bias. One approach to reduce the bias
is to use the explicit bias function analytically derived in the two-step estimator. Here, we use an alternative approach of bias reduction based on the bootstrap. For simplicity, in this section, we assume cross-sectionally uncorrelated errors ($\tau_{ij} = 0$ for all $i \neq j$). However, such an assumption, as well as the absence of serial correlation and time-varying variance, can be relaxed by modifying the bootstrap procedure (e.g., block bootstrap and wild bootstrap).\footnote{Recent papers by Goncalves and Perron (2010) and Yamamoto (2010) also apply bootstrap procedures to improve the finite sample performance of dynamic factor models. The former paper employs a wild bootstrap procedure in factor-augmented regression models of Bai and Ng (2006), while the latter studies factor-augmented vector autoregressive (FAVAR) models of Bernanke, Boivin and Eliasz (2005).}

We first introduce a simple procedure for bias correction.

**Bootstrap bias correction**

1. Estimate factors and factor loadings using principal components method and obtain residuals $\tilde{e}_{it} = x_{it} - \tilde{\lambda}_i \tilde{f}_t$.

2. Recenter $\tilde{e}_{it}$, $\tilde{\lambda}_i$ and $\tilde{f}_t$ around zero. Generate $x_{it}^* = \lambda_i^* \tilde{f}_t + e_{it}^*$ for $t = 1, \ldots, T$ by first drawing $\lambda_i^*$ from $\tilde{\lambda}_i$ and then drawing $e_{it}^*$ for $t = 1, \ldots, T$ from $\tilde{e}_{jt}$ given $\lambda_i^* = \tilde{\lambda}_j$. Repeat the same procedure $N$ times to generate all $x_{it}^*$s for $i = 1, \ldots, N$.

3. Apply the principal components method to $x_{it}^*$ and estimate $\tilde{f}_t^*$.

4. Compute bootstrap AR coefficient estimate $\hat{\rho}^*$ from $\tilde{f}_t^*$.

5. Repeat steps 2 to 4 $B$ times to obtain the bootstrap bias estimate $\hat{\text{bias}} = B^{-1} \sum_{b=1}^{B} \hat{\rho}_b^* - \hat{\rho}$ where $\hat{\rho}_b^*$ is the $b$-th bootstrap AR estimate and $\hat{\rho}$ is the AR estimate from $\tilde{f}_t$.

Beran and Srivastava (1985) have established the validity of applying the bootstrap procedure to the principal component analysis. Our procedure slightly differs from theirs in that we resample $x_{it}^*$ using estimated factor model in step 2. Note that our procedure employs the simplest form of bootstrap bias correction based on a constant bias function. While this
form of bias correction seems to be the one most frequently used (e.g., Kilian, 1998), the performance of the bias-corrected estimator may be improved by introducing linear or non-linear bias functions in the correction (see MacKinnon and Smith, 1998). It should also be noted that the procedure above is designed to evaluate the bias from small $N$ in the principal components method rather than the bias from small $T$ in the autoregression. In order to incorporate the latter, or Kendall’s bias, we may combine the procedure with bootstrapping autoregressive models. This possibility is considered in the construction of the bootstrap confidence interval below.

**Bootstrap confidence interval**

1. Estimate factors and factor loadings using principal components method and obtain residuals $\tilde{e}_{it} = x_{it} - \tilde{\lambda}_i \tilde{f}_t$.

2. Compute the AR coefficient estimate $\tilde{\rho}$ from $\tilde{f}_t$ and obtain residuals $\tilde{e}_t = \tilde{f}_t - \tilde{\rho} \tilde{f}_{t-1}$.

3. Recenter $\tilde{e}_t$ around zero if necessary and generate $\tilde{e}_t^*$ by resampling from $\tilde{e}_t$. Then generate pseudo factors using either $f_t^* = \tilde{\rho} f_{t-1}^* + \tilde{\epsilon}_t^*$ or $f_t^* = \tilde{\rho}_{BC} f_{t-1}^* + \tilde{\epsilon}_t^*$ where $\tilde{\rho}_{BC} = \tilde{\rho} - \text{bias}$.

4. Recenter $\tilde{e}_{it}$, $\tilde{\lambda}_i$ and $\tilde{f}_t$ around zero. Generate $x_{it}^* = \lambda_i^* f_{it}^* + e_{it}^*$ for $t = 1, ..., T$ by first drawing $\lambda_i^*$ from $\tilde{\lambda}_i$ and then drawing $e_{it}^*$ for $t = 1, ..., T$ from $\tilde{e}_{jt}$ given $\lambda_i^* = \tilde{\lambda}_j$. Repeat the same procedure $N$ times to generate all $x_{it}^*$’s for $i = 1, ..., N$.

5. Apply principal components method to $x_{it}^*$ and estimate $\tilde{f}_t^*$.

6. Compute bootstrap AR coefficient estimate $\tilde{\rho}^*$ or $t$ statistic $t(\tilde{\rho}^*)$ from $\tilde{f}_t^*$.

7. Repeat steps 3 to 6 $B$ times to obtain the empirical distribution of the statistic and construct the confidence interval.
The procedure above involves a combination of bootstrapping principal components (Beren and Srivastava, 1985) and bootstrapping the residuals in autoregressive models (Freedman, 1984, and Bose, 1988). Note that, in step 3, either original AR parameter estimates or bias corrected estimates $\tilde{\rho}_{BC}$ can be used to generate bootstrap common factors. This alternative choice does not have any effects on the limit distribution of the bootstrap estimator as in Kilian’s (1998) argument. Also note that $\tilde{\rho}_{BC}$ can be obtained either by using the procedure employed before or a combination of that procedure with a bootstrapping autoregressive model. The former approach only considers the principal components bias while the latter takes both the principal components bias and Kendall’s bias into account.

Finally, we consider some theoretical justification of using the bootstrap procedure in the two-step estimation of AR coefficients and the inference. The following proposition provides the asymptotic validity of the bootstrap procedure.

**Proposition 2.** Let all the assumptions of Proposition 1 are satisfied and the bootstrap data $\{X^*\}$ are generated as described in Bootstrap Confidence Interval. If $N = cT^\alpha$ with $c > 0$ and $\alpha > 1/2$ as $T \to \infty$, then $\sup_{x \in \mathbb{R}^p} |P^*(\sqrt{T}(\tilde{\phi}^* - \phi) \leq x) - P(\sqrt{T}(\tilde{\phi} - \phi) \leq x)| \xrightarrow{P} 0$.

When the number of the series ($N$) increases at a sufficiently fast rate relative to the time series observations ($T$), the limiting distribution of the bootstrap estimator $\tilde{\phi}^*$ is asymptotically equivalent to that of $\tilde{\phi}$.\(^3\) The proof is provided in the Appendix.

### 3.2 Performance of Bootstrap Confidence Intervals

We now apply the proposed bootstrap procedures to the case for $N/T < 0.5$ where the conventional asymptotic confidence intervals did not perform well in the previous section. Table 3 reports coverage of both asymptotic confidence intervals with bias correction and bootstrap

---

\(^3\)In general, signs of the coefficients in the factor forecasting regression cannot be identified and Goncalves and Perron (2010) argue the consistency of their bootstrap procedure in renormalized parameter space. Here, our interest is the AR coefficients in univariate AR models, rather than the forecasting regression coefficients, so that the sign identification is not an issue.
confidence intervals of the two-step estimator $\hat{\rho}$ for $\rho = 0.5$ and $\rho = 0.8$ cases. A simple bias correction procedure improves over the conventional asymptotic interval without bias correction. Especially, when the signal-to-noise ratio is 2.0, they are nearly central confidence intervals. However, their coverage rates are still below the nominal rate. Both percentile and percentile-$t$ confidence intervals based on the full bootstrap procedure also improve over the asymptotic intervals. For $T = 100$ cases, coverage rates are below nominal rate with percentile interval somewhat closer to the nominal rate. Most notable improvement is found in $T = 200$ case with signal-to-noise ratio being 2.0 and $\rho = 0.5$. For both bias corrected and standard bootstrap procedures, near-exact central confidence intervals are obtained where downward bias in the asymptotic intervals was observed in the previous section.

When $\rho = 0.8$, percentile intervals show some downward bias but this is not surprising since such a bias was also present in the benchmark case. As in the case of the asymptotic interval result, the performance gets worse in general as the signal-to-noise ratio decreases.

Finally, we consider our bootstrap procedure for factor estimation combined with the grid bootstrap of Hansen (1999). This can be done by replacing $\tilde{\rho}$ to generate bootstrap factors in step 3 by a set of possible values of the parameter. Since the grid bootstrap is known to work well globally in the AR parameter space, we apply this method to $\rho = 0.9$ case. Table 4 reports coverage of grid bootstrap confidence intervals along with bias-corrected and standard bootstrap intervals. The result shows that the grid bootstrap method combined with a bootstrapping first-step factor estimation performs better than other methods while coverage is somewhat conservative.

4 Out-of-Sample Forecasting

One of the most important areas of applying the estimated common factor is the use of such extracted information from many variables in forecasting macroeconomic time series.
This aspect is emphasized in Stock and Watson (1998, 1999, 2002) and Marcellino, Stock and Watson (2003). In this section, we investigate the finite sample performance of tests regarding the out-of-sample forecasting applied to the estimated factors. We consider the following $h$-period ahead forecasting equation with a factor included as a regressor,

$$y_{t+h} - y_t = \beta f_t^0 + u_{t+h}$$

(4)

where $y_{t+h}$ is the scalar time series variable of interest and $u_{t+h}$ is the forecast error. By a similar argument in the estimation of the factor dynamics, a two-step method is required to run the forecasting regression. Indeed, such a two-step procedure has been proven to provide a first-order efficient forecast by Stock and Watson (1998).

To evaluate the predictive ability of the factor, the sample is divided into first $R$ observations and last $P$ observations. Both factor and forecasting equation are estimated using the following recursive scheme. First, the factor is estimated by the principal component method using normalized $x_{it}$’s from period 1 to $R$. The estimated factor is then used in forecasting regression to obtain the forecast of $y_{R+h}$. For the second forecast $y_{R+h+1}$, the data is again standardized and the factors and forecasting models are reestimated using the observations from 1 to $R+1$. This procedure is repeated $P^* \equiv P - h + 1$ times to obtain $P^*$ of $h$-period ahead forecasts. We consider as a competing model a simple random walk forecast

$$y_{t+h} - y_t = u_{t+h}.$$  

(5)

Such a simple forecasting model is often considered an alternative to the fundamental model in the literature of forecasting foreign exchange rates (e.g., Meese and Rogoff, 1983 and Mark, 1995).

Let $\hat{u}_{1,t+h}$ and $\hat{u}_{2,t+h}$ be the realized forecast errors from the factor forecast and the random walk forecast, respectively. The sample mean square forecast error of each forecasting model
is given by $MSE_1 = P^{-1}_s \sum_{t=R}^{T-h} \hat{u}_{1,t+h}^2$ and $MSE_2 = P^{-1}_s \sum_{t=R}^{T-h} \hat{u}_{2,t+h}^2$. If the factor forecast outperforms the random walk forecast in prediction accuracy, we expect $MSE_1$ to be smaller than $MSE_2$. To test the null hypothesis of equal forecast accuracy, we consider the following three tests.

The first test utilizes the ratio of mean square forecast error,

$$U = \frac{MSE_1}{MSE_2}$$

which is often called as Theil’s $U$ statistic. The second test is that proposed by Diebold and Mariano (1995) based on the difference between mean square forecast errors $\tilde{d} = MSE_2 - MSE_1$. The DM test statistic is given by

$$DM = \sqrt{P^*} \frac{\tilde{d}}{\sqrt{\tilde{\omega}_d}}$$

where $\tilde{\omega}_d$ is the long-run variance estimator of $\tilde{d}$. The third test we consider is the encompassing test proposed by Harvey, Leybourne and Newbold (1998). Their test statistic is given by

$$R_1 = \sqrt{P^*} \frac{\tilde{c}}{\sqrt{\tilde{\omega}_c}}$$

where $\tilde{c} = P^{-1}_s \sum_{t=R}^{T-h} \hat{u}_{1,t+h}(\hat{u}_{1,t+h} - \hat{u}_{2,t+h})$ and $\tilde{\omega}_c$ is the long-run variance estimator of $\tilde{c}$. If the two models are non-nested, the theoretical result by West (1996) implies that the standardized $U$, $DM$ and $R_1$ follow asymptotically normal under the null hypothesis of equal forecast accuracy. However, as pointed out by Clark and McCracken (2001, 2005), when two models are nested as in our setting, the limit distribution under the null will be non-standard. Furthermore, when the forecast horizon $h$ is greater than one, the limit distribution depends on the nuisance parameters. For this reason, in such a case, Clark and McCracken (2005) have recommended employing the bootstrap method to obtain the critical values.

To construct the bootstrap critical values, the bootstrap sample of the regressor in the
forecasting regression is usually generated from the estimated autoregressive model of the regressor (e.g., Mark, 1995). However, when such an autoregressive model is estimated from the estimated factor, the findings from the previous sections imply the presence of downward bias with small $N$. Therefore, it seems reasonable to combine our bootstrap procedure of factor estimation with the bootstrap critical value for testing equal out-of-sample forecasting accuracy in the following way.

**Bootstrap critical value of out-of-sample forecast test**

1. Using the full sample $T = R + P$, compute factor estimates $\tilde{f}_t$, factor loading estimates $\tilde{\lambda}_i$, and the AR coefficient estimate $\tilde{\rho}$.

2. Generate the pairs $(\Delta y_t^*, \epsilon_t^*)$ by resampling from $(\Delta y_t, \epsilon_t)$.

3. Generate bootstrap factors $f_t^*$ and observed variables $x_{it}^* = \lambda_i^* f_t^* + \epsilon_{it}^*$ using the procedure described in the previous section.

4. Based on first $R$ observations, normalize $x_{it}^*$ and estimate $\tilde{f}_t^*$ by the principal components method. Run the forecasting regression $y_{t+h}^* - y_t^*$ on $\tilde{f}_t^*$ and obtain the bootstrap factor forecast error $\hat{u}_{1,R+h}^*$ as well as $\hat{u}_{2,R+h}^* = y_{R+h}^* - y_R^*$.

5. Run the forecasting regression recursively, using $R + 1$ observations to $T$ observations with each time $x_{it}^*$ being renormalized and the factors $\tilde{f}_t^*$ being reestimated. Compute the test statistic of interest based on $P^*$ pairs of bootstrap forecast errors.

6. Repeat steps 4 and 5 $B$ times to obtain the empirical distribution of the statistic for the critical value.
To investigate the effect of small $N$ in the two-step forecast, we generate data from

$$
x_{it} = \lambda_i^0 f_t^0 + e_{it}
$$

$$
\Delta y_t = \beta f_t^0 + u_t
$$

$$
f_t^0 = \rho f_{t-1}^0 + \varepsilon_t
$$

with $\rho = 0.8$ and $u_t \sim N(0, 1)$. Other parameter settings are the same as the one used in the previous section. We can evaluate the empirical size by setting $\beta = 0$ and the empirical power of the test by choosing the non-zero value. We can expect increasing power with increasing $\beta$.

Following Clark and McCracken (2002), we set $\beta = 0.25$ to evaluate the power. The long-run variances required for the DM test and encompassing test are estimated using Bartlett kernel with $h - 1$ lag truncation parameter. The forecast horizons we consider are $h = 1, 2,$ and $4$. Sample sizes are $P = 50$ with $R = 50$ and $P = 50$ with $R = 100$. The number of series $N$ is 10. The bootstrap replications are $B = 499$. The nominal significance level of the test is set to 10%.

Table 5 shows the result of the experiment. The first three columns are the out-of-sample forecast result based on the true factor which is available only in the simulation. This result can be considered as a benchmark similar to the one we considered for estimation of factor dynamics in the previous sections. Compared to the benchmark case, the empirical size of the two-step forecast turn out surprisingly well even if $N$ is as small as 10. Some size distortion is observed in case of longer horizon ($h = 4$).

In respect to the power of the test, Theil’s $U$ and the encompassing test perform uniformly better than the DM test for both infeasible and two-step forecasts. As expected, the power of the test of the two-step forecast is somewhat smaller compared to the benchmark case. However, the difference in power is much smaller between infeasible and feasible two-step forecasts than the difference among the three tests of the out-of-sample forecast.
5 Empirical Example: US Diffusion Index

In this section, we apply our bootstrap procedure to the analysis of a diffusion index based on a dynamic factor model. Stock and Watson (1998, 2002) extracted common factors from 215 U.S. monthly macroeconomic time series and reported that the forecasts based on such diffusion indexes outperformed the conventional forecasts. We use the same data source (and transformations) as Stock and Watson, but we only use balanced panels and construct cross-sectional subsamples for the purpose of applying our procedure to the data with different number of series. The series in a list provided in Appendix B of Stock and Watson (2002) are divided into 14 categories (real output and income; employment and hours; real retail, manufacturing and trade sales; consumption; housing starts and sales; real inventories and inventory-sales ratios; orders and unfilled orders; stock prices; exchange rates; interest rates; money and credit quantity aggregates; price indexes; average hourly earnings; and miscellaneous). After constructing a balanced panel using all available observations, we take the first five series from each category to construct a subsample. The second subsample is similarly constructed by taking the first two variables from each category. The numbers of the series in the full balanced panel and the two subsamples are $N = 159, 63$ and $27$, respectively.\footnote{The number of the series in the full balanced panel differs from that of Stock and Watson (2002) due to the different treatment of outliers. The numbers of the series in subsamples are not multiples of five and two because only a limited number of the series are available for some categories.} All the series are from 1959:3 to 1998:12 giving a maximum number of time series observation $T = 478$. Note that for all three cases, $N/T$ is less than 0.5, so that the bootstrap method is likely more appropriate than the asymptotic approximation in the two-step estimation. Diffusion indexes obtained in the first step (by applying the principal components method) are shown in Figure 1. In the figure, diffusion indexes are rescaled to have the same drift and variance as the (log of) industrial production; the NBER recessionary episodes are also shown in the shaded area. As the asymptotic theory predicts, we observe
that the difference between the indexes based on \( N = 27 \) and \( N = 159 \) seems to be larger than the difference between the \( N = 63 \) and \( N = 159 \) cases.

In the next step, we estimate the dynamic structure of three diffusion indexes using the AR(1) specification. Table 6 reports the point estimate \( \tilde{\rho} \), bias-corrected estimate \( \tilde{\rho}_{BC} \), and 90% confidence intervals based on various methods introduced in section 4. The bias-corrected estimates and the bootstrap intervals, except for the grid bootstrap intervals, are computed with the number of bootstrap replication \( B = 799 \). The grid bootstrap intervals are constructed using 399 bootstrap replications at each of 50 gridpoints. One notable observation from this empirical exercise is that the size of the bootstrap bias correction is substantial for all three cases with the size largest for the \( N = 27 \) case and smallest for the \( N = 159 \) case. Another interesting point to note is that the asymptotic confidence intervals are quite different from the rest of the intervals, while there is a very small difference among the bootstrap intervals. These are consistent with our finding in the Monte Carlo section.

Finally, we examine the performance of diffusion indexes in forecasting inflation. Inflation \((y_t)\) is defined using the consumer price index and is assumed to have a unit root. We use 12 months as a forecasting horizon \((h)\) and the number of observations for the first in-sample forecasting regression \((R)\) is set to 120. The diffusion index forecast is constructed using (4) and is compared to a simple random walk forecast given by (5). Following the procedure described in section 5, the out-of-sample forecasting performance is evaluated using a bootstrap with the number of bootstrap replications \( B = 499 \). Table 7 shows three test statistics for the null hypothesis of equal forecasting accuracy along with the bootstrap \( p \)-values. For all data sets, the null hypothesis is rejected at the 5% level of significance. This insensitivity to the sample size is also consistent with our simulation result.
6 Conclusion

In this paper, we have examined the finite sample properties of the two-step estimators of dynamic factor models when unobservable common factors are estimated by the principal components methods in the first step. Both the estimation of the factor dynamics and that of forecasting regression models were considered in simulation. As a result of the experiment, we found that the estimation of AR parameter is downward-biased for small $N$, with the bias becoming larger as the true parameter approaches unity. This finding resembles the bias problem of AR estimator for small $T$. However, the bias caused by small $N$ is also present in the large $T$ case. When there is a possibility of such a downward bias, a bootstrap procedure proposed in the paper is found to solve the problem to a certain degree.

The rule of thumb for practical implementation obtained from our experiment is as follows.

(1) When $N$ is at least half the size of $T$, the first-order asymptotic result is reliable provided that data is less persistent.

(2) When $N$ is smaller than half the size of $T$, it safe to use the bootstrap method combined with bootstrapping first-step estimation. If the data is more persistent, the grid bootstrap method combined with bootstrapping first-step estimation yields better results.

(3) Two-step forecasting procedure using bootstrap method works well even when $N$ is small.

Using the large number of series in the dynamic factor analysis is becoming a very popular approach in applications. The finding of this paper implies that the distortion from using the number of observation typically available to the researcher is not severe as long as the number of series is greater than the half of the number of time series observation. Finally, it would be interesting to extend the experiment to allow for cross-sectional correlation, multivariate factors, and nonlinear factor dynamics.
Appendix: Proofs

Proof of Proposition 1.

The principal components estimator \( \tilde{F} = [\tilde{f}_1, \ldots, \tilde{f}_T]' \) is the first eigenvector of the \( T \times T \) matrix \( XX' \) with normalization \( T^{-1} \sum_{t=1}^{T} \tilde{f}_t^2 = 1 \), where

\[
X = \begin{bmatrix} X_1' \\ \vdots \\ X_T' \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{N1} \\ \vdots & \ddots & \vdots \\ x_{1T} & \cdots & x_{NT} \end{bmatrix}.
\]

By definition, \((1/TN)XX' \tilde{F} = \tilde{F}v_{NT}\) where \(v_{NT}\) is the largest eigenvalue of \((1/TN)XX'\).

Let \( \xi_{st} = N^{-1} \sum_{i=1}^{N} e_{is} e_{it} - E(e_{is} e_{it}) \), \( \eta_{st} = N^{-1} f_0^t \sum_{i=1}^{N} \lambda_i^0 e_{it} \), and \( \xi_{st} = N^{-1} f_0^t \sum_{i=1}^{N} \lambda_i^0 e_{is} \).

Following the proof of Theorem 5 in Bai (2003), the estimation error of the factor can be decomposed as

\[
\tilde{f}_t - H_N T f_0^t = J_N T^{-1} \sum_{s=1}^{T} \tilde{f}_s \xi_{st} + J_N T^{-1} \sum_{s=1}^{T} \tilde{f}_s \eta_{st} + J_N T^{-1} \sum_{s=1}^{T} \tilde{f}_s \xi_{st}
\]

\[
= O_P \left( N^{-1/2} \delta_{NT}^{-1} \right) + O_P \left( N^{-1/2} \right) + O_P \left( N^{-1/2} \delta_{NT}^{-1} \right) = O_P \left( N^{-1/2} \right)
\]

where \( H_N \) = \((\bar{F}' F_0^0 / T)(A^0 A_0^0 / N) J_N \), \( J_N = (v_{NT} - T^{-1} \sigma_e^2)^{-1} \), \( \sigma_e^2 = N^{-1} \sum_{i=1}^{N} \sigma_{ei}^2 \) and \( \delta_{NT} = \min \{ \sqrt{N} \sqrt{T} \} \).

In this proof, we show the case for \( p = 1 \) and \( \phi_1 = \rho \). If \( f_0^{l-1} \) is observable,

\[
\sqrt{T}(\tilde{p} - \rho) = \sqrt{T} \left( \sum_{t=1}^{T} (f_0^{l-1})^2 \right)^{-1/2} \sum_{t=1}^{T} f_0^{l-1} \epsilon_t = T^{-1/2} \sum_{t=1}^{T} f_0^{l-1} \epsilon_t + o_P(1)
\]

since \( T^{-1} \sum_{t=1}^{T} (f_0^{l-1})^2 - 1 = o_P(1) \). If \( f_0^{l-1} \) is replaced with the factor estimator,

\[
\sqrt{T}(\tilde{p} - \rho) = \sqrt{T} \left( \sum_{t=1}^{T} (\tilde{f}_t^{l-1})^2 \right)^{-1/2} \sum_{t=1}^{T} \tilde{f}_t^{l-1} (\tilde{f}_t - \rho \tilde{f}_t^{l-1})
\]

\[
= T^{-1/2} \sum_{t=1}^{T} \tilde{f}_t^{l-1} (\tilde{f}_t - \rho \tilde{f}_t^{l-1}) + o_P(1)
\]

since \( T^{-1} \sum_{t=1}^{T} \tilde{f}_t^2 - 1 = o_P(1) \). By decomposing the dominant term, we have

\[
T^{-1/2} \sum_{t=1}^{T} \tilde{f}_t^{l-1} (\tilde{f}_t - \rho \tilde{f}_t^{l-1})
\]

\[
= T^{-1/2} H_N \sum_{t=1}^{T} \tilde{f}_t^{l-1} \epsilon_t + T^{-1/2} \sum_{t=1}^{T} \tilde{f}_t^{l-1} \left\{ \tilde{f}_t - H_N f_0^t - \rho \left( \tilde{f}_t^{l-1} - H_N f_0^{l-1} \right) \right\}
\]

\[
= T^{-1/2} H_N \sum_{t=1}^{T} f_0^{l-1} \epsilon_t + T^{-1/2} \sum_{t=1}^{T} \tilde{f}_t^{l-1} \left( \tilde{f}_t - H_N f_0^t \right)
\]

\[
- T^{-1/2} \rho \sum_{t=1}^{T} \tilde{f}_t^{l-1} \left( \tilde{f}_t^{l-1} - H_N f_0^{l-1} \right) + T^{-1/2} H_N \sum_{t=1}^{T} \left( \tilde{f}_t^{l-1} - H_N f_0^{l-1} \right) \epsilon_t
\]
The leading term is dominant since the other three terms are all \(o_P(1)\) which will be shown below. From Bai’s (2003) Lemma A.3, we have \(\text{plim } v_{NT} = \Sigma_{\Lambda} \Sigma_F = v\) and
\[
\text{plim } H_{NT}^2 = \text{plim } H_{NT}^2 = (\tilde{F}'F^0/T)(\Lambda_0^0\Lambda_0^0/N)^2(F^0\tilde{F}/T)H_{NT}^2 = \Sigma_{\Lambda} v^{-2} = \Sigma_{\Lambda}(\Sigma_{\Lambda} \Sigma_F)^{-1} = \Sigma_F^{-1} = 1.
\]
Using \(H_{NT}^2 - 1 = o_P(1)\), the dominant term can be written as
\[
T^{-1/2} (H_{NT}^2 - 1) \sum_{t=1}^T f_{t-1}^0 \varepsilon_t + T^{-1/2} \sum_{t=1}^T f_{t-1}^0 \varepsilon_t = \sqrt{T} (\hat{\rho} - \rho) + o_P(1).
\]

Lemma B.2 of Bai (2003) implies \(T^{-1} \sum_{t=1}^T \tilde{f}_{t-1} - H_{NT} f_{t-1}^0 = O_P(\delta_{NT}^{-2})\).

Lemma A.1(iv) of Bai and Ng (2006) implies \(T^{-1} H_{NT} \sum_{t=1}^T (\tilde{f}_{t-1} - H_{NT} f_{t-1}^0) \varepsilon_t = O_P(\delta_{NT}^{-2})\). To show that the remaining term is \(o_P(1)\), we decompose it as
\[
T^{-1} \sum_{t=1}^T \tilde{f}_{t-1} - H_{NT} f_{t-1}^0
\]
\[
= T^{-1} \sum_{t=1}^T \left( \tilde{f}_{t-1} - H_{NT} f_{t-1}^0 \right) \left( \tilde{f}_t - H_{NT} f_t^0 \right) + T^{-1} H_{NT} \sum_{t=1}^T f_{t-1}^0 \left( \tilde{f}_t - H_{NT} f_t^0 \right)
\]
\[
= A + B.
\]

From Lemma A.1 of Bai (2003), we have \(T^{-1} \sum_{t=1}^T \left( \tilde{f}_t - H_{NT} f_t^0 \right)^2 = O_P(\delta_{NT}^{-2})\) and
\[
A \leq \left( T^{-1} \sum_{t=1}^T \left( \tilde{f}_{t-1} - H_{NT} f_{t-1}^0 \right)^2 \right)^{1/2} \left( T^{-1} \sum_{t=1}^T \left( \tilde{f}_t - H_{NT} f_t^0 \right)^2 \right)^{1/2} = O_P(\delta_{NT}^{-2})
\]

By decomposing \(B\), we have
\[
B = J_N \left[ T^{-2} \sum_{t=1}^T \sum_{s=1}^T \tilde{f}_s f_{t-1}^0 \zeta_{st} + T^{-2} \sum_{t=1}^T \sum_{s=1}^T \tilde{f}_s f_{t-1}^0 \eta_{st} + T^{-2} \sum_{t=1}^T \sum_{s=1}^T \tilde{f}_s f_{t-1}^0 \xi_{st} \right]
\]
\[
= J_N [B_1 + B_2 + B_3].
\]

We first show
\[
B_1 = T^{-2} \sum_{t=1}^T \sum_{s=1}^T \left( \tilde{f}_s - H_{NT} f_s^0 \right) f_{t-1}^0 \zeta_{st} + T^{-2} \sum_{t=1}^T \sum_{s=1}^T H_{NT} f_s^0 f_{t-1}^0 \zeta_{st} = O_P((NT)^{-1/2}).
\]

For the first term of \(B_1\),
\[
T^{-2} \sum_{t=1}^T \sum_{s=1}^T \left( \tilde{f}_s - H_{NT} f_s^0 \right) f_{t-1}^0 \zeta_{st}
\]
\[
\leq \left( T^{-1} \sum_{s=1}^T \left( \tilde{f}_s - H_{NT} f_s^0 \right)^2 \right)^{1/2} \left( T^{-1} \sum_{t=1}^T \left( \tilde{f}_{t-1} - H_{NT} f_{t-1}^0 \right)^2 \right)^{1/2}
\]
\[
= O_P(\delta_{NT}^{-1}) \cdot O_P((NT)^{-1/2}) = O_P((NT)^{-1/2} \delta_{NT}^{-1})
\]

since
\[
T^{-1} \sum_{t=1}^T f_{t-1}^0 \zeta_{st} = \left( (NT)^{-1} \sum_{t=1}^T \sum_{i=1}^N f_{t-1}^0 [e_{is} e_{it} - E(e_{is} e_{it})] \right) = O_P((NT)^{-1/2})
\]

22
from the independence and zero mean conditions. For the second term of \( B_1 \),

\[
T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} H_{NT} f_s^0 f_{t-1}^0 \xi_{st} \leq H_{NT} \left( T^{-1} \sum_{s=1}^{T} f_{s}^{02} \right)^{1/2} \left( T^{-1} \sum_{s=1}^{T} \left( T^{-1} \sum_{t=1}^{T} f_{t-1}^0 \xi_{st} \right)^2 \right)^{1/2} = O_P((NT)^{-1/2}) = O_P((NT)^{-1/2}).
\]

Next, we show

\[
B_2 = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left( \tilde{f}_s - H_{NT} f_s^0 \right) f_{t-1}^0 \eta_{st} + T^{-2} \sum_{t=1}^{T} H_{NT} f_s^0 f_{t-1}^0 \eta_{st} = O_P((NT)^{-1/2}).
\]

For the first term of \( B_2 \),

\[
T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left( \tilde{f}_s - H_{NT} f_s^0 \right) f_{t-1}^0 \eta_{st} \leq \left( T^{-1} \sum_{s=1}^{T} \left( \tilde{f}_s - H_{NT} f_s^0 \right)^2 \right)^{1/2} \left( T^{-1} \sum_{s=1}^{T} \left( T^{-1} \sum_{t=1}^{T} f_{t-1}^0 \eta_{st} \right)^2 \right)^{1/2} = O_P(\delta^{-1}_{NT}) \cdot O_P((NT)^{-1/2}) = O_P((NT)^{-1/2})
\]

since

\[
T^{-1} \sum_{t=1}^{T} f_{t-1}^0 \eta_{st} = (NT)^{-1} f_s^0 \sum_{t=1}^{T} \sum_{i=1}^{N} f_{t-1}^0 \lambda_i^0 e_{it} = O_P((NT)^{-1/2})
\]

from the independence and zero mean conditions. For the second term of \( B_2 \),

\[
T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} H_{NT} f_s^0 f_{t-1}^0 \eta_{st} = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} H_{NT} f_s^0 f_{t-1}^0 N^{-1} f_s^0 \sum_{i=1}^{N} \lambda_i^0 e_{it} = H_{NT} \left( T^{-1} \sum_{s=1}^{T} f_{s}^{02} \right) (NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} \lambda_i^0 f_{t-1}^0 e_{it} = O_P((NT)^{-1/2}).
\]

Finally, \( B_3 = O_P((NT)^{-1/2}) \) can be similarly obtained by using

\[
T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} H_{NT} f_s^0 f_{t-1}^0 \xi_{st} = T^{-2} \sum_{t=1}^{T} \sum_{i=1}^{N} f_s^0 \lambda_i^0 e_{is} = H_{NT} \left( T^{-1} \sum_{s=1}^{T} f_{s}^{02} \right) (TN)^{-1} \sum_{s=1}^{T} \sum_{i=1}^{N} \lambda_i^0 f_s^0 e_{is} = O_P((NT)^{-1/2})
\]

Therefore, \( T^{-1} \sum_{t=1}^{T} \tilde{f}_t - H_{NT} f_t^0 \) is \( O_P(\delta^{-2}_{NT}) \). Combining all results yields

\[
\sqrt{T} (\tilde{\rho} - \rho) - \sqrt{T} (\hat{\rho} - \rho) = O_P(\sqrt{T} \delta^{-2}_{NT}) \text{ which is } o_P(1) \text{ given } \sqrt{T} \delta^{-2}_{NT} \to 0 \text{ or } N/\sqrt{T} \to \infty.
\]
Proof of Proposition 2.

The bootstrap principal components estimator \( \tilde{F}^* = [\tilde{f}_{1i}^*, \ldots, \tilde{f}_{Ti}^*] \) is the first eigenvector of the \( T \times T \) matrix \( X^*X^{*T} \) with normalization \( T^{-1}\sum_{t=1}^{T} \tilde{f}_t^2 = 1 \), where the bootstrap sample is given by

\[
X^* = \begin{bmatrix} X_1^* & \cdots & X_N^* \\ \vdots & \ddots & \vdots \\ X_T^* & \cdots & X_{NT}^* \end{bmatrix} = \begin{bmatrix} x_{11}^* & \cdots & x_{N1}^* \\ \vdots & \ddots & \vdots \\ x_{1T}^* & \cdots & x_{NT}^* \end{bmatrix}.
\]

Analogous to the original version, we have \((1/TN)X^*X^{*T}\tilde{F}^* = \tilde{F}^*V_{NT}^*\), where \( V_{NT}^* \) is the largest eigenvalue of \((1/TN)X^*X^{*T}\). Let \( \zeta_{st}^* = N^{-1}\sum_{i=1}^{N} e_{is}^* e_{it}^* - E^*(e_{is}^*e_{it}^*) \), \( \eta_{st}^* = N^{-1}f_{si}^* \sum_{i=1}^{N} \lambda_{is}^* e_{it}^* \), and \( \xi_{st}^* = N^{-1}f_{ti}^* \sum_{i=1}^{N} \lambda_{is}^* e_{it}^* = \eta_{is}^* \). As the proof of proposition 1, the estimation error of the factor can be decomposed as

\[
\tilde{f}_t - H_{NT}^*f_t^* = J_{NT}^*T^{-1}\sum_{s=1}^{T} \tilde{f}_s^* \zeta_{st}^* + J_{NT}^*T^{-1}\sum_{s=1}^{T} \tilde{f}_s^* \eta_{st}^* + J_{NT}^*T^{-1}\sum_{s=1}^{T} \tilde{f}_s^* \xi_{st}^*.
\]

where \( H_{NT}^* = (\tilde{F}^*F/T)(\Lambda^*/\Lambda^*/N)J_{NT}^* \), \( J_{NT}^* = (v_{NT}^* - T^{-1}\sigma_{e}^2)^{-1} \), \( \sigma_{e}^2 = N^{-1}\sum_{s=1}^{N} \sigma_{ei}^2 \), and \( \sigma_{ei}^2 = E^*(e_{is}^2) \). In this proof, we only show the case for \( p = 1 \) and \( \phi_1 = \rho \). From Lemma C.1 of Goncalves and Perron (2010), under our assumption F, FL and E, we have (i) \( T^{-1}\sum_{t=1}^{T} |\tilde{f}_t - H_{NT}f_t^0|^4 = O_P(1) \), (ii) \( N^{-1}\sum_{i=1}^{N} \lambda_i - H_{NT}^{-1}\lambda_i|^4 = O_P(1) \), (iii) \( (NT)^{-1}\sum_{i=1}^{N} \sum_{t=1}^{T} e_{it}^2 = O_P(1) \). This result implies that \( E^*(\epsilon_{it}^4) = O_P(1) \),

\[
E^* \lambda_i^4 = N^{-1}\sum_{i=1}^{N} \lambda_i^4 \leq 8N^{-1}(\sum_{i=1}^{N} \lambda_i - H_{NT}^{-1}\lambda_i|^4 + \sum_{i=1}^{N} |H_{NT}^{-1}\lambda_i|^4) = O_P(1),
\]

and

\[
E^* \epsilon_{it}^4 = T^{-1}\sum_{t=1}^{T} (\tilde{f}_t - \rho \tilde{f}_{t-1})^4 = T^{-1}\sum_{t=1}^{T} (\tilde{f}_t - H_{NT}f_t^0 + H_{NT}f_t^0 - \rho (\tilde{f}_{t-1} - H_{NT}f_{t-1}^0) - \rho H_{NT}f_{t-1}^0)^4 \leq 4^{3}T^{-1}\sum_{t=1}^{T} [(\tilde{f}_t - H_{NT}f_t^0)^4 + (H_{NT}f_t^0)^4 + \rho^4(\tilde{f}_{t-1} - H_{NT}f_{t-1}^0)^4 + (\rho H_{NT}f_{t-1}^0)^4] = O_P(1).
\]

The bootstrap estimation error can be approximated as

\[
\sqrt{T}(\rho^* - \rho) = \sqrt{T} \left( \sum_{t=1}^{T} \frac{(\tilde{f}_{t-1}^*)^2}{T} \right)^{-1} \sum_{t=1}^{T} (\tilde{f}_t^* - \rho \tilde{f}_{t-1}^*) = T^{-1/2} \sum_{t=1}^{T} (\tilde{f}_t^* - \rho \tilde{f}_{t-1}^*) + o_P(1).
\]

24
since $T^{-1} \sum_{t=1}^{T} (\tilde{f}_{t-1}^{*})^2 - 1 = o_p(1)$. By decomposing the dominant term, we have

\[
T^{-1/2} \sum_{t=1}^{T} \tilde{f}_{t-1}^{*} \left( \tilde{f}_{t}^{*} - \tilde{\rho} \tilde{f}_{t-1}^{*} \right) = T^{-1/2} \sum_{t=1}^{T} \tilde{f}_{t-1}^{*} \left( \tilde{f}_{t}^{*} - H_{NT}^{*} \tilde{f}_{t-1}^{*} - \tilde{\rho} \left( \tilde{f}_{t-1}^{*} - H_{NT}^{*} \tilde{f}_{t-1}^{*} \right) \right) + T^{-1/2} H_{NT}^{*} \sum_{t=1}^{T} \tilde{f}_{t-1}^{*} \varepsilon_{t}^{*}.
\]

The leading term can be written as

\[
T^{-1/2} (H_{NT}^{2*} - 1) \sum_{t=1}^{T} f_{t-1}^{*} \varepsilon_{t}^{*} + T^{-1/2} \sum_{t=1}^{T} f_{t-1}^{*} \varepsilon_{t}^{*} = T^{-1/2} \sum_{t=1}^{T} f_{t-1}^{*} \varepsilon_{t}^{*} + o_p(1).
\]

The last equality follows from the fact that $v_{NT}^{*} = v^{*} + o_p(1)$ where $v^{*} = \Sigma_{\lambda}^{*} \Sigma_{F}^{*}$, $\Sigma_{\lambda}^{*} = \tilde{\Lambda}_{\lambda}/N$ and $\Sigma_{F}^{*} = \tilde{F}^{*} \tilde{F}/T = 1$, and $H_{NT}^{2*} - 1 = o_p(1)$ because

\[
H_{NT}^{2*} = (\tilde{F}^{*} \tilde{F}^{*}/T)(\Lambda^{*} \Lambda^{*}/N)(\tilde{F}^{*} \tilde{F}^{*}/T)J_{NT}^{2} = \Sigma_{F}^{*} - 1 + o_p(1).
\]

In what follows, we show that three remaining terms are all $o_p(1)$. We first decompose the second term as

\[
T^{-1} \sum_{t=1}^{T} \tilde{f}_{t-1}^{*} (\tilde{f}_{t-1}^{*} - H_{NT}^{*} \tilde{f}_{t-1}^{*}) = T^{-1} \sum_{t=1}^{T} \left( \tilde{f}_{t-1}^{*} - H_{NT}^{*} \tilde{f}_{t-1}^{*} \right)^2 + T^{-1} H_{NT}^{*} \sum_{t=1}^{T} \tilde{f}_{t-1}^{*} (\tilde{f}_{t-1}^{*} - H_{NT}^{*} \tilde{f}_{t-1}^{*})
\]

and show that both terms are $O_p(\delta_{NT}^{-2})$. We have

\[
A^{*} = J_{NT}^{2*} \left[ T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{f}_{s}^{*} \zeta_{st}^{*} + T^{-1} \sum_{s=1}^{T} \tilde{f}_{s}^{*} \eta_{st}^{*} + T^{-1} \sum_{s=1}^{T} \tilde{f}_{s}^{*} \zeta_{st}^{*} \right] \leq 3J_{NT}^{2*} \left[ T^{-1} \sum_{t=1}^{T} T^{-2} \sum_{s=1}^{T} \tilde{f}_{s}^{*} \zeta_{st}^{*} + T^{-1} \sum_{t=1}^{T} T^{-2} \sum_{s=1}^{T} \tilde{f}_{s}^{*} \eta_{st}^{*} + T^{-1} \sum_{t=1}^{T} T^{-2} \sum_{s=1}^{T} \tilde{f}_{s}^{*} \zeta_{st}^{*} \right] = 3J_{NT}^{2*} \left[ A_{1}^{*} + A_{2}^{*} + A_{3}^{*} \right].
\]

First,

\[
A_{1}^{*} \leq T^{-1} \sum_{t=1}^{T} T^{-2} \sum_{s=1}^{T} \tilde{f}_{s}^{*}^{2} \sum_{s=1}^{T} \zeta_{st}^{*2} = (T^{-1} \sum_{s=1}^{T} \tilde{f}_{s}^{*}^{2}) (T^{-2} \sum_{s=1}^{T} \zeta_{st}^{*2}) = O_p(\delta_{NT}^{-2})
\]

where the last equality follows from

\[
T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} E^{*} \zeta_{st}^{*2} = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} E^{*} \sum_{i=1}^{N} (e_{is}^{*} e_{it}^{*} - E^{*}(e_{is}^{*} e_{it}^{*}))^2 = (NT)^{-2} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{i=1}^{N} Var^{*}(e_{is}^{*} e_{it}^{*}) = O_p(N^{-1})
\]
provided $E^*(e^*_{it}) = O_P(1)$. Second,

$$A^*_2 = T^{-1} \sum_{t=1}^T (T^{-1} \sum_{s=1}^T \hat{f}^2_s N^{-1} f^*_t \sum_{i=1}^N \lambda^*_i e^*_{it})^2 = T^{-1} \sum_{t=1}^T (T^{-1} \sum_{s=1}^T \hat{f}^2_s f^*_t)^2 (N^{-1} \sum_{i=1}^N \lambda^*_i e^*_{it})^2$$

$$\leq T^{-1} \sum_{s=1}^T (T^{-1} \sum_{s=1}^T \hat{f}^2_s)(T^{-1} \sum_{s=1}^T \hat{f}^2_s)(N^{-1} \sum_{i=1}^N \lambda^*_i e^*_{it})^2$$

$$= (T^{-1} \sum_{s=1}^T \hat{f}^2_s)(T^{-1} \sum_{s=1}^T \hat{f}^2_s)[(T^{-1} \sum_{i=1}^N \lambda^*_i e^*_{it})^2] = O_P(\delta^{-2}_{NT})$$

which follows from $T^{-1} \sum_{s=1}^T \hat{f}^2_s = O_P(1)$ using Theorem 4.1 of Freedman (1984) and from

$$T^{-1} \sum_{t=1}^T E^*(N^{-1} \sum_{i=1}^N \lambda^*_i e^*_{it})^2 = N^{-2} T^{-1} \sum_{t=1}^T \sum_{i=1}^N E^*(\lambda^*_i e^*_{it}^2) = N^{-2} T^{-1} \sum_{t=1}^T \sum_{i=1}^N E^*(\lambda^*_i)^2 (\sum_{i=1}^N e^*_{it}^2)$$

$$\leq N^{-2} T^{-1} \left( \sum_{i=1}^N E^*(\lambda^*_i)^4 \right)^{1/2} \left( \sum_{i=1}^N \sum_{t=1}^T \sum_{i=1}^N e^*_{it}^2 \right)^{1/2}$$

$$\leq N^{-1} (N^{-1} \sum_{i=1}^N E^*(\lambda^*_i)^4)^{1/2} (N^{-1} \sum_{i=1}^N T^{-1} E^* \sum_{t=1}^T \sum_{i=1}^N e^*_{it}^2)^{1/2} = O_P(N^{-1})$$

provided $E^*(\lambda^*_i)^4 = O_P(1)$. Third,

$$A^*_3 = (T^{-1} \sum_{t=1}^T f^*_t \hat{f}^2)(T^{-1} \sum_{s=1}^T \hat{f}^2_s N^{-1} \sum_{i=1}^N \lambda^*_i e^*_{is})^2$$

$$\leq (T^{-1} \sum_{t=1}^T f^*_t \hat{f}^2)(T^{-1} \sum_{s=1}^T \hat{f}^2_s)[(T^{-1} \sum_{s=1}^T \sum_{i=1}^N \lambda^*_i e^*_{is})^2] = O_P(\delta^{-2}_{NT})$$

Therefore, $A^* = O_P(\delta^{-2}_{NT})$. For $B^*$, we have

$$B^* = J^*_NT \left[ T^{-2} \sum_{t=1}^T \sum_{s=1}^T \hat{f}^*_s f^* t_{t-1} \xi^*_{st-1} + T^{-2} \sum_{t=1}^T \sum_{t=1}^T \hat{f}^*_s f^* t_{t-1} \eta^*_{st-1} + T^{-2} \sum_{t=1}^T \sum_{s=1}^T \hat{f}^*_s f^* t_{t-1} \xi^*_{st-1} \right]$$

$$= J^*_NT [B^*_1 + B^*_2 + B^*_3].$$

First,

$$B^*_1 \leq [T^{-1} \sum_{s=1}^T (\hat{f}^*_s)^2]^{1/2} [T^{-1} \sum_{s=1}^T (f^* s t_{t-1})^2]^{1/2}$$

$$= [T^{-1} \sum_{s=1}^T (T^{-1} \sum_{s=1}^T f^* s t_{t-1} \xi^*_{st-1})]^{1/2}$$

$$= [T^{-1} \sum_{s=1}^T (T^{-1} \sum_{i=1}^N \xi^*_{is} e^*_{it-1} - E^*(e^*_{it-1})])^{1/2} = O_P(\delta^{-2}_{NT})$$
since

\[
T^{-1}\sum_{s=1}^{T}(T^{-1}\sum_{t=1}^{T}f_{s-1}^{*}N^{-1}\sum_{i=1}^{N}[e_{is}^{*}e_{it-1}^{*} - E^{*}(e_{is}^{*}e_{it-1}^{*})])^{2}
\]

\[
= N^{-2}T^{-2}\sum_{t=1}^{T}E^{*}\{\sum_{s=1}^{T}\sum_{i=1}^{T}f_{s-1}^{*}E^{*}(e_{is}^{*}e_{it-1}^{*}) - E^{*}(e_{is}^{*}e_{it-1}^{*})\}^{2}N^{-1}\sum_{j=1}^{N}[e_{jt}^{*}e_{jt-1}^{*} - E^{*}(e_{jt}^{*}e_{jt-1}^{*})]\}
\]

\[
= N^{-2}T^{-3}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{T}E^{*}(f_{s-1}^{*}f_{t}^{*})\sum_{i=1}^{N}\sum_{j=1}^{N}E^{*}\{[e_{is}^{*}e_{it-1}^{*} - E^{*}(e_{is}^{*}e_{it-1}^{*})][e_{jt}^{*}e_{jt-1}^{*} - E^{*}(e_{jt}^{*}e_{jt-1}^{*})]\}
\]

\[
= N^{-2}T^{-3}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{T}E^{*}(f_{s}^{*}f_{t}^{*})\sum_{i=1}^{N}\sum_{j=1}^{N}Cov^{*}(e_{is}^{*}e_{it-1}^{*}, e_{jt}^{*}e_{jt-1}^{*})
\]

\[
\leq (NT)^{-1}\{T^{-1}\sum_{s=1}^{T}[E^{*}(f_{s}^{*})^{2}]^{1/2}[T^{-1}\sum_{s=1}^{T}((NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}Var^{*}(e_{is}^{*}e_{it-1}^{*})^{2}]^{1/2}
\]

\[
\leq (NT)^{-1}[T^{-1}\sum_{s=1}^{T}E^{*}(f_{s}^{*})^{4}]^{1/2}[T^{-1}\sum_{s=1}^{T}(NT)^{-1}[\sum_{i=1}^{N}\sum_{j=1}^{N}Var^{*2}(e_{is}^{*}e_{it-1}^{*})^{1/2} = O_{P}((NT)^{-1}).
\]

Second,

\[
B_{2}^{*} = T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(f_{s}^{*}f_{t-1}^{*})\eta_{st-1}^{*} = T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\tilde{f}_{s}^{*} - H_{NT}f_{s}^{*})f_{t-1}^{*}\eta_{st-1}^{*} + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}H_{NT}f_{s}^{*}f_{t-1}^{*}\eta_{st-1}^{*}
\]

\[
= B_{21}^{*} + B_{22}^{*}
\]

where

\[
B_{21}^{*} = T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\tilde{f}_{s}^{*} - H_{NT}f_{s}^{*})f_{t-1}^{*}\eta_{st-1}^{*} = T^{-2}\sum_{t=1}^{T}[\tilde{f}_{s}^{*} - H_{NT}f_{s}^{*}](\sum_{t=1}^{T}f_{t-1}^{*}\eta_{st-1}^{*})]
\]

\[
\leq [T^{-1}\sum_{s=1}^{T}(\tilde{f}_{s}^{*} - H_{NT}f_{s}^{*})^{2}]^{1/2}[T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}f_{t-1}^{*}\eta_{st-1}^{*}^{2}]^{1/2}
\]

\[
\leq [T^{-1}\sum_{s=1}^{T}(\tilde{f}_{s}^{*} - H_{NT}f_{s}^{*})^{2}]^{1/2}[T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}(\tilde{f}_{t-1}^{*})^{2}][T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\eta_{st-1}^{2}]^{1/2}
\]

\[
= [T^{-1}\sum_{s=1}^{T}(\tilde{f}_{s}^{*} - H_{NT}f_{s}^{*})^{2}]^{1/2}[(T^{-1}\sum_{t=1}^{T}f_{s}^{2})(T^{-1}\sum_{t=1}^{T}f_{t}^{2})(T^{-1}\sum_{t=1}^{T}(N^{-1}\sum_{i=1}^{N}\lambda_{i}^{*}c_{it-1}^{*})^{2})^{1/2}
\]

\[
= O_{P}(\delta_{NT}^{-1}) = [T^{-1}\sum_{t=1}^{T}(N^{-1}\sum_{i=1}^{N}\lambda_{i}^{*}c_{it-1}^{*})^{2}]^{1/2} = O_{P}(\delta_{NT}^{-2}).
\]

since \(T^{-1}\sum_{t=1}^{T}(N^{-1}\sum_{i=1}^{N}\lambda_{i}^{*}c_{it-1}^{*})^{2} = O_{P}(N^{-1})\) can be obtained similarly as in the case
of $T^{-1} \sum_{t=1}^{T} (N^{-1} \sum_{i=1}^{N} \lambda_{it}^* e_{it})^2 = O_p(N^{-1})$ and
\[
B_{22} = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} f_{s}^* f_{t-s}^* \eta_{st-1}^2 = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} f_{s}^* f_{t-s}^* (N^{-1} f_{s}^* \sum_{i=1}^{N} \lambda_{it}^* e_{it-1})
\]
\[
= (T^{-1} \sum_{s=1}^{T} f_{s}^2) [(NT)^{-1} T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} f_{t-1}^* \lambda_{it}^* e_{it-1}]
\]
\[
= O_p(\delta_{NT}^2)
\]
since
\[
E^* (N^{-1}T^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} f_{t-1}^* \lambda_{it}^* e_{it-1})^2 = (NT)^{-2} [E^* \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{i=1}^{N} f_{t-1}^* f_{s}^* \lambda_{it}^* \lambda_{is}^* e_{it-1} e_{is-1}]
\]
\[
= (NT)^{-2} E^* \sum_{t=1}^{T} \sum_{i=1}^{N} f_{t-1}^* \lambda_{it}^* e_{it-1}^2
\]
\[
\leq (NT)^{-2} (T^{-1} E^* \sum_{t=1}^{T} f_{t-1}^4) (T^{-1} E^* \sum_{t=1}^{T} (N^{-1} \sum_{i=1}^{N} \lambda_{it}^2 e_{it-1}^2)^2)
\]
\[
\leq (NT)^{-1} (T^{-1} E^* \sum_{t=1}^{T} f_{t-1}^4) [(NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} E^* \lambda_{it}^4 E^* e_{it-1}^4]
\]
\[
= O_p(\delta_{NT}^4).
\]
Therefore, $B_2^2 = O_p(\delta_{NT}^2)$. Third,
\[
B_3^* = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} (\tilde{f}_{s}^* - H_{NT} f_{s}^*) f_{t-s}^* \xi_{st-1}^* = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} (\tilde{f}_{s}^* - H_{NT} f_{s}^*) f_{t-1}^* \xi_{st-1}^* + T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} H_{NT} f_{s}^* f_{t-1}^* \xi_{st-1}^*
\]
\[
= B_{31}^* + B_{32}^*
\]
where
\[
B_{31}^* = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} (\tilde{f}_{s}^* - H_{NT} f_{s}^*) f_{t-1}^* \xi_{st-1}^* = T^{-2} \sum_{s=1}^{T} (\tilde{f}_{s}^* - H_{NT} f_{s}^*) \sum_{t=1}^{T} f_{t-1}^* \xi_{st-1}^*
\]
\[
\leq [T^{-1} \sum_{s=1}^{T} (\tilde{f}_{s}^* - H_{NT} f_{s}^*)^2]^{1/2} [T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} f_{t-1}^* \xi_{st-1}^2]^{1/2}
\]
\[
\leq [T^{-1} \sum_{s=1}^{T} (\tilde{f}_{s}^* - H_{NT} f_{s}^*)^2]^{1/2} [T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \xi_{st-1}^2]^{1/2} [T^{-1} \sum_{t=1}^{T} \eta_{st-1}^2]^{1/2}
\]
\[
= [T^{-1} \sum_{s=0}^{T-1} (\tilde{f}_{s}^* - H_{NT} f_{s}^*)^2]^{1/2} [T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \eta_{st-1}^2]^{1/2}
\]
\[
= O_p(\delta_{NT}^2)
\]
and
\[
B_{32}^* = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} f_{s}^* f_{t-1}^* \xi_{st-1}^* = T^{-1} \sum_{s=1}^{T} f_{s}^* (T^{-1} \sum_{t=1}^{T} f_{t-1}^* \xi_{st-1}^*)
\]
\[
= T^{-1} \sum_{s=1}^{T} f_{s}^* (T^{-1} \sum_{t=1}^{T} f_{t-1}^2 \sum_{i=1}^{N} \lambda_{it}^* e_{is}) = (T^{-1} \sum_{t=1}^{T} f_{t-1}^2) [(NT)^{-1} \sum_{s=1}^{T} \sum_{i=1}^{N} f_{s}^* \lambda_{it}^* e_{is}]
\]
\[
= O_p(\delta_{NT}^2)
\]
because \( E^*[\mathcal{N}(NT)^{-1}\sum_{s=1}^{T} \sum_{t=1}^{N} f_s^* \lambda_t^* e_t^*]^2 = O_P(\delta_{NT}^{-2}) \) can be obtained similarly as
\( E^*[\mathcal{N}(NT)^{-1}\sum_{s=1}^{T} \sum_{t=1}^{N} f_s^* \lambda_t^* e_{s-1}^*]^2 = O_P(\delta_{NT}^{-4}) \). Therefore, we have \( B_1^* = O_P(\delta_{NT}^{-2}) \) and \( B_2^* = O_P(\delta_{NT}^{-2}) \). Combining the results for \( A^* \) and \( B^* \) yields
\[ T^{-2}\sum_{t=1}^{T}(\tilde{f}_{t-1}^* - H_{NT}^* f_t^*) = O_P(\delta_{NT}^{-2}) \] for the second term.

Next, for the third term, we have
\[ T^{-1}\sum_{t=1}^{T}(\tilde{f}_{t-1}^* - H_{NT}^* f_t^*) = T^{-1}\sum_{t=1}^{T}(\tilde{f}_{t-1}^* - H_{NT}^* f_t^*) + T^{-1}H_{NT}^* f_{t-1}^* \]
\[ \leq T^{-1}\sum_{t=1}^{T}(\tilde{f}_{t-1}^* - H_{NT}^* f_t^*)^2T^{-1}\sum_{t=1}^{T}(\tilde{f}_{t-1}^* - H_{NT}^* f_t^*)^2]^{1/2} \]
\[ + T^{-1}H_{NT}^* f_{t-1}^* \]
\[ = T^{-1}H_{NT}^* f_{t-1}^* + O_P(\delta_{NT}^{-2}) = O_P(\delta_{NT}^{-2}) \]

because \( T^{-1}\sum_{t=1}^{T}(\tilde{f}_{t-1}^* - H_{NT}^* f_t^*) f_{t-1}^* = O_P(\delta_{NT}^{-2}) \) can be obtained similarly as in the case of
\( T^{-1}\sum_{t=1}^{T}(\tilde{f}_{t-1}^* - H_{NT}^* f_t^*) f_{t-1}^* = O_P(\delta_{NT}^{-2}) \).

Next, we decompose the last term as
\[ T^{-1}\sum_{t=1}^{T}(\tilde{f}_{t-1}^* - H_{NT}^* f_t^*) \xi_{st-1}^* = T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}(\tilde{f}_{t-1}^* - H_{NT}^* f_t^*) \xi_{st-1}^* + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T} \xi_{st-1}^* \xi_{t}^* \]
\[ + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T} \xi_{st-1}^* \xi_{t}^* \]
\[ = C^* + D^* + F^* \]

and show that all three terms are \( O_P(\delta_{NT}^{-2}) \). For \( C^* \), we have
\[ C^* = T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\tilde{f}_{t-1}^* - H_{NT}^* f_t^*) \xi_{st-1}^* \xi_{t}^* + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T} H_{NT}^* f_t^* \xi_{st-1}^* \xi_{t}^* = C_1^* + C_2^* \]
where
\[ C_1^* = T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\tilde{f}_{t-1}^* - H_{NT}^* f_t^*) \xi_{st-1}^* \xi_{t}^* = T^{-2}\sum_{t=1}^{T}(\tilde{f}_{t-1}^* - H_{NT}^* f_t^*) \xi_{st-1}^* \xi_{t}^* \]
\[ \leq [T^{-1}\sum_{t=1}^{T}(\tilde{f}_{t-1}^* - H_{NT}^* f_t^*)^2]^{1/2} [T^{-1}\sum_{t=1}^{T}(\tilde{f}_{t-1}^* - H_{NT}^* f_t^*)^2]^{1/2} \]
\[ \leq [T^{-1}\sum_{t=1}^{T}(\tilde{f}_{t-1}^* - H_{NT}^* f_t^*)^2]^{1/2} [T^{-1}\sum_{t=1}^{T}(\tilde{f}_{t-1}^* - H_{NT}^* f_t^*)^2]^{1/2} = O_P(\delta_{NT}^{-2}) \]
provided \( E^* \varepsilon_t^4 = o_P(1) \), and

\[
C_2^* = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} f_s^* \zeta_{st-1}^* \varepsilon_t^* = T^{-1} \sum_{t=1}^{T} \varepsilon_t^* (T^{-1} \sum_{s=1}^{T} f_s^* \zeta_{st-1}^*)
\]

\[
\leq (T^{-1} \sum_{t=1}^{T} \varepsilon_t^2)^{1/2} [T^{-1} \sum_{s=1}^{T} (T^{-1} \sum_{t=1}^{T} f_s^* \zeta_{st-1}^*)^2]^{1/2} = o_P(\delta_{NT}^{-2})
\]

because \( T^{-1} \sum_{t=1}^{T} (T^{-1} \sum_{s=1}^{T} f_s^* \zeta_{st-1}^*)^2 = o_P((NT)^{-1}) \) can be obtained similarly as in the case of \( T^{-1} \sum_{t=1}^{T} (T^{-1} \sum_{s=1}^{T} f_s^* \zeta_{st-1}^*)^2 = o_P((NT)^{-1}) \). Therefore, we have \( C^* = o_P(\delta_{NT}^{-2}) \). For \( D^* \), we have

\[
D^* = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} (f_s^* - H_{NT}^* f_s^*) \eta_{st-1}^* \varepsilon_t^* + T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} H_{NT}^* f_s^* \eta_{st-1}^* \varepsilon_t^* = D_1^* + D_2^*
\]

where

\[
D_1^* = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} (f_s^* - H_{NT}^* f_s^*) \eta_{st-1}^* \varepsilon_t^* = T^{-2} \sum_{s=1}^{T} (f_s^* - H_{NT}^* f_s^*) (\sum_{t=1}^{T} \eta_{st-1}^* \varepsilon_t^*)
\]

\[
\leq (T^{-1} \sum_{s=1}^{T} (f_s^* - H_{NT}^* f_s^*)^2)^{1/2} (T^{-1} \sum_{s=1}^{T} (\sum_{t=1}^{T} \eta_{st-1}^* \varepsilon_t^*)^2)^{1/2}
\]

\[
\leq (T^{-1} \sum_{s=1}^{T} (f_s^* - H_{NT}^* f_s^*)^2)^{1/2} (T^{-1} \sum_{s=1}^{T} (\sum_{t=1}^{T} \eta_{st-1}^* \varepsilon_t^*)^2)^{1/2} = o_P(\delta_{NT}^{-2})
\]

and

\[
D_2^* = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} f_s^* \eta_{st-1}^* \varepsilon_t^* = T^{-1} \sum_{t=1}^{T} f_s^* (T^{-1} \sum_{t=1}^{T} N^{-1} f_s^* \sum_{i=1}^{N} \lambda_i^* \varepsilon_{it-1}^* \varepsilon_t^*)
\]

\[
= (T^{-1} \sum_{s=1}^{T} f_s^*) [(NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} \lambda_i^* \varepsilon_{it-1}^* \varepsilon_t^*] = o_P(\delta_{NT}^{-2}),
\]

since

\[
E^*[(NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} \lambda_i^* \varepsilon_{it-1}^* \varepsilon_t^*]^2 = (NT)^{-2} E^* \left[ \sum_{t=1}^{T} \sum_{i=1}^{T} \sum_{j=1}^{N} \lambda_i^* \lambda_j^* \varepsilon_{it-1}^* \varepsilon_j^* \varepsilon_t^* \varepsilon_i^* \right]
\]

\[
= (NT)^{-1} E^* \left[ \sum_{t=1}^{T} \varepsilon_t^2 (N^{-1} \sum_{i=1}^{N} \lambda_i^2 \varepsilon_{it-1}^2) \right]
\]

\[
\leq (NT)^{-1} E^* \left[ (T^{-1} \sum_{t=1}^{T} \varepsilon_t^4)^{1/2} (T^{-1} \sum_{t=1}^{T} (N^{-1} \sum_{i=1}^{N} \lambda_i^2 \varepsilon_{it-1}^2)^2)^{1/2} \right]
\]

\[
\leq (NT)^{-1} E^* \left[ (T^{-1} \sum_{t=1}^{T} \varepsilon_t^4)^{1/2} ((NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} \lambda_i^4 \varepsilon_{it-1}^4)^{1/2} \right] = o_P((NT)^{-1}).
\]
Therefore, \( D^* = O_{p^*}(\delta_{NT}^{-2}) \). For \( F^* \), we have,

\[
\begin{align*}
F^* &= T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{f}_s \xi_{st-1} \varepsilon_{it}^* = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} (\tilde{f}_s - H_{NT} f_s^*) \xi_{st-1} \varepsilon_{it}^* + T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} H_{NT} f_s^* \xi_{st-1} \varepsilon_{it}^* \\
&= F_1^* + F_2^*
\end{align*}
\]

where

\[
\begin{align*}
F_1^* &= T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} (f_s^* - H_{NT} f_s^*) \xi_{st-1} \varepsilon_{it}^* = T^{-2} \sum_{s=1}^{T} (f_s^* - H_{NT} f_s^*) (\sum_{s=1}^{T} \xi_{st-1} \varepsilon_{it}^*) \\
&\leq [T^{-1} \sum_{s=1}^{T} (f_s^* - H_{NT} f_s^*)^2]^{1/2} [T^{-1} \sum_{s=1}^{T} \xi_{st-1} \varepsilon_{it}^*]^2]^{1/2} \\
&\leq [T^{-1} \sum_{s=1}^{T} (f_s^* - H_{NT} f_s^*)^2]^{1/2} [T^{-1} \sum_{t=1}^{T} \xi_{st-1}^2] (T^{-1} \sum_{t=1}^{T} \varepsilon_{it}^2)]^{1/2} \\
&= O_{p^*}(\delta_{NT})
\end{align*}
\]

and

\[
\begin{align*}
F_2^* &= T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} f_s^* \xi_{st-1} \varepsilon_{it}^* = T^{-1} \sum_{t=1}^{T} (T^{-1} \sum_{s=1}^{T} N^{-1} f_s^* \sum_{i=1}^{N} \lambda_i^* \varepsilon_{is}^*) \\
&\leq (T^{-1} \sum_{t=1}^{T} f_{t-1}^2)^{1/2} [T^{-1} \sum_{t=1}^{T} \varepsilon_{it}^2 ((NT)^{-1} \sum_{s=1}^{T} \sum_{i=1}^{N} f_s^* \lambda_i^* \varepsilon_{is}^*^2)]^{1/2} \\
&= (T^{-1} \sum_{s=1}^{T} f_{t-1}^2)^{1/2} (T^{-1} \sum_{t=1}^{T} \varepsilon_{it}^2)^{1/2} [(NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} f_s^* \lambda_i^* \varepsilon_{is}^*] \\
&= O_{p^*}(\delta_{NT}^{-2})
\end{align*}
\]

Therefore, \( F^* = O_{p^*}(\delta_{NT}^{-2}) \). Combining the results for \( C^* \), \( D^* \) and \( F^* \) yields \( T^{-1} \sum_{t=1}^{T} (\tilde{f}_t^* - H_{NT} f_{t-1}) \varepsilon_{it}^* = O_{p^*}(\delta_{NT}) \) for the last term.

Finally, we have

\[
\sqrt{T} (\tilde{p}^* - \hat{p}) = T^{-1/2} \sum_{t=1}^{T} \tilde{f}_{t-1} \varepsilon_{it}^* + o_{p^*}(1)
\]

and apply the bootstrap central limit theorem to the dominant term. Since \( E^*[f_{t-1}^* \varepsilon_{it}^* | f_{t-2}^* \varepsilon_{it-1}^* , \ldots] = 0 \), we can use the central limit theorem for the martingale difference sequence under the bootstrap probability measure and thus \( P^*(\sqrt{T} (\tilde{p}^* - \hat{p}) \leq x) \) approaches normal distribution function with variance \( E^*(f_{t-2}^* \varepsilon_{it}^2) = T^{-1} \sum_{t=1}^{T} \tilde{f}_{t-2}^2 \varepsilon_{it}^2 \) under the bootstrap probability measure. Combining it with \( T^{-1} \sum_{t=1}^{T} \tilde{f}_{t-1}^2 \varepsilon_{it}^2 \rightarrow P^* E(f_{t-1}^2 \varepsilon_{it}^2) = \Gamma^{-1} \), we have \( P^*(\sqrt{T} (\tilde{p}^* - \hat{p}) \leq x) = P(\sqrt{T} (\tilde{p} - \hat{p}) \leq x) \rightarrow P 0 \) for any \( x \). By using Polya’s theorem, we have the uniform convergence result given by

\[
\sup_{x \in \mathbb{R}} [P^*(\sqrt{T} (\tilde{p}^* - \hat{p}) \leq x) - P(\sqrt{T} (\tilde{p} - \hat{p}) \leq x)] \rightarrow P 0.
\]
References


Table 1  
Coverage of Nominal 90% Confidence Intervals  
(Infeasible Benchmark)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>100</th>
<th></th>
<th>200</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>Asy 0.89 [0.285] (0.05, 0.06)</td>
<td></td>
<td>0.90 [0.202] (0.04, 0.06)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Per 0.90 [0.286] (0.03, 0.07)</td>
<td></td>
<td>0.90 [0.203] (0.03, 0.07)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Per-t 0.90 [0.286] (0.06, 0.05)</td>
<td></td>
<td>0.89 [0.203] (0.05, 0.06)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>Asy 0.90 [0.200] (0.04, 0.07)</td>
<td></td>
<td>0.89 [0.141] (0.04, 0.07)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Per 0.88 [0.208] (0.01, 0.11)</td>
<td></td>
<td>0.89 [0.145] (0.01, 0.10)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Per-t 0.90 [0.200] (0.05, 0.06)</td>
<td></td>
<td>0.89 [0.142] (0.06, 0.06)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>Asy 0.90 [0.148] (0.04, 0.07)</td>
<td></td>
<td>0.90 [0.103] (0.04, 0.06)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Per 0.85 [0.164] (0.00, 0.15)</td>
<td></td>
<td>0.87 [0.109] (0.00, 0.13)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Per-t 0.89 [0.147] (0.05, 0.05)</td>
<td></td>
<td>0.89 [0.103] (0.06, 0.05)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grid 0.91 [0.154] (0.05, 0.04)</td>
<td></td>
<td>0.89 [0.106] (0.06, 0.05)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grid-t 0.91 [0.154] (0.05, 0.05)</td>
<td></td>
<td>0.89 [0.106] (0.05, 0.05)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Results are based on 1,000 replications. Numbers in brackets are the median lengths of confidence intervals. Pairs of numbers in parentheses are the frequencies of the true value lay to the left and right of the intervals, respectively. Asy: asymptotic interval. Per: Efron’s percentile interval. Per-t: Hall’s percentile-t interval. Grid: Hansen’s grid interval. Grid-t: Hansen’s grid-t interval.
Table 2
Coverage of Nominal 90% Confidence Intervals
(Asymptotic)

<table>
<thead>
<tr>
<th>( T )</th>
<th>( N )</th>
<th>( N/T )</th>
<th>Signal-to-noise ratio ( \rho = 0.5 )</th>
<th>Signal-to-noise ratio ( \rho = 0.8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>0.1</td>
<td>0.80 [0.296]</td>
<td>(0.01, 0.19)</td>
</tr>
<tr>
<td>20</td>
<td>0.2</td>
<td>0.88 [0.293]</td>
<td>(0.02, 0.10)</td>
<td>0.85 [0.294]</td>
</tr>
<tr>
<td>50</td>
<td>0.5</td>
<td>0.89 [0.290]</td>
<td>(0.02, 0.09)</td>
<td>0.88 [0.291]</td>
</tr>
<tr>
<td>100</td>
<td>1.0</td>
<td>0.88 [0.290]</td>
<td>(0.03, 0.09)</td>
<td>0.88 [0.290]</td>
</tr>
<tr>
<td>200</td>
<td>2.0</td>
<td>0.89 [0.289]</td>
<td>(0.03, 0.08)</td>
<td>0.89 [0.289]</td>
</tr>
<tr>
<td>200</td>
<td>20</td>
<td>0.1</td>
<td>0.88 [0.206]</td>
<td>(0.02, 0.10)</td>
</tr>
<tr>
<td>40</td>
<td>0.2</td>
<td>0.88 [0.204]</td>
<td>(0.02, 0.09)</td>
<td>0.88 [0.205]</td>
</tr>
<tr>
<td>100</td>
<td>0.5</td>
<td>0.90 [0.203]</td>
<td>(0.03, 0.07)</td>
<td>0.89 [0.203]</td>
</tr>
<tr>
<td>200</td>
<td>1.0</td>
<td>0.88 [0.203]</td>
<td>(0.03, 0.09)</td>
<td>0.88 [0.203]</td>
</tr>
<tr>
<td>400</td>
<td>2.0</td>
<td>0.88 [0.204]</td>
<td>(0.04, 0.08)</td>
<td>0.88 [0.204]</td>
</tr>
</tbody>
</table>

| 100   | 10    | 0.1   | 0.60 [0.235] | (0.00, 0.40) | 0.40 [0.248] | (0.00, 0.59) | 0.20 [0.267] | (0.00, 0.81) |
| 20    | 0.2   | 0.77 [0.222] | (0.01, 0.22) | 0.67 [0.230] | (0.00, 0.33) | 0.47 [0.243] | (0.00, 0.53) |
| 50    | 0.5   | 0.86 [0.213] | (0.01, 0.13) | 0.83 [0.216] | (0.01, 0.16) | 0.76 [0.223] | (0.00, 0.24) |
| 100   | 1.0   | 0.85 [0.212] | (0.02, 0.13) | 0.85 [0.214] | (0.02, 0.14) | 0.81 [0.218] | (0.01, 0.17) |
| 200   | 2.0   | 0.87 [0.209] | (0.02, 0.11) | 0.87 [0.210] | (0.01, 0.12) | 0.86 [0.212] | (0.01, 0.13) |
| 200   | 20    | 0.1   | 0.75 [0.153] | (0.00, 0.25) | 0.58 [0.159] | (0.00, 0.42) | 0.33 [0.168] | (0.00, 0.67) |
| 40    | 0.2   | 0.84 [0.149] | (0.01, 0.15) | 0.79 [0.152] | (0.00, 0.21) | 0.64 [0.157] | (0.00, 0.36) |
| 100   | 0.5   | 0.88 [0.145] | (0.01, 0.11) | 0.87 [0.147] | (0.01, 0.12) | 0.82 [0.149] | (0.01, 0.18) |
| 200   | 1.0   | 0.87 [0.145] | (0.03, 0.11) | 0.86 [0.146] | (0.02, 0.12) | 0.85 [0.147] | (0.02, 0.13) |
| 400   | 2.0   | 0.88 [0.145] | (0.02, 0.10) | 0.87 [0.145] | (0.00, 0.11) | 0.86 [0.146] | (0.02, 0.12) |

Notes: Results are based on 1,000 replications. See notes in Table 1.
Table 3
Coverage of Nominal 90% Confidence Intervals
(Standard Bootstrap)

<table>
<thead>
<tr>
<th>T</th>
<th>N</th>
<th>N/T</th>
<th>Signal-to-noise ratio</th>
<th>2.0</th>
<th>1.0</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\rho = 0.5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>0.1</td>
<td>Bc</td>
<td>0.84 [0.295] (0.08, 0.08)</td>
<td>0.83 [0.300] (0.01, 0.10)</td>
<td>0.77 [0.306] (0.05, 0.18)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Per</td>
<td>0.88 [0.311] (0.03, 0.09)</td>
<td>0.85 [0.321] (0.02, 0.13)</td>
<td>0.76 [0.334] (0.02, 0.22)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Per-t</td>
<td>0.85 [0.302] (0.09, 0.06)</td>
<td>0.83 [0.310] (0.09, 0.08)</td>
<td>0.79 [0.322] (0.07, 0.13)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bc</td>
<td>0.84 [0.293] (0.08, 0.08)</td>
<td>0.83 [0.295] (0.08, 0.09)</td>
<td>0.82 [0.298] (0.07, 0.11)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Per</td>
<td>0.86 [0.296] (0.05, 0.09)</td>
<td>0.86 [0.303] (0.04, 0.10)</td>
<td>0.85 [0.312] (0.03, 0.12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Per-t</td>
<td>0.84 [0.292] (0.08, 0.08)</td>
<td>0.83 [0.295] (0.09, 0.08)</td>
<td>0.82 [0.302] (0.09, 0.10)</td>
</tr>
<tr>
<td>200</td>
<td>20</td>
<td>0.1</td>
<td>Bc</td>
<td>0.88 [0.205] (0.06, 0.06)</td>
<td>0.87 [0.207] (0.07, 0.07)</td>
<td>0.85 [0.210] (0.06, 0.09)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Per</td>
<td>0.89 [0.209] (0.04, 0.07)</td>
<td>0.89 [0.214] (0.04, 0.07)</td>
<td>0.87 [0.221] (0.03, 0.11)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Per-t</td>
<td>0.88 [0.206] (0.06, 0.06)</td>
<td>0.87 [0.209] (0.07, 0.06)</td>
<td>0.86 [0.214] (0.07, 0.07)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bc</td>
<td>0.89 [0.204] (0.06, 0.05)</td>
<td>0.89 [0.205] (0.06, 0.05)</td>
<td>0.88 [0.206] (0.05, 0.07)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Per</td>
<td>0.89 [0.206] (0.05, 0.06)</td>
<td>0.89 [0.207] (0.04, 0.07)</td>
<td>0.89 [0.211] (0.04, 0.07)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Per-t</td>
<td>0.89 [0.204] (0.06, 0.06)</td>
<td>0.88 [0.205] (0.06, 0.06)</td>
<td>0.87 [0.207] (0.06, 0.07)</td>
</tr>
<tr>
<td>400</td>
<td>40</td>
<td>0.2</td>
<td>Bc</td>
<td>0.89 [0.153] (0.07, 0.05)</td>
<td>0.88 [0.159] (0.06, 0.06)</td>
<td>0.85 [0.167] (0.04, 0.11)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Per</td>
<td>0.93 [0.168] (0.02, 0.06)</td>
<td>0.92 [0.184] (0.01, 0.07)</td>
<td>0.86 [0.208] (0.00, 0.14)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Per-t</td>
<td>0.88 [0.152] (0.08, 0.04)</td>
<td>0.89 [0.162] (0.07, 0.05)</td>
<td>0.87 [0.179] (0.06, 0.07)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bc</td>
<td>0.88 [0.148] (0.07, 0.05)</td>
<td>0.86 [0.156] (0.07, 0.07)</td>
<td>0.86 [0.156] (0.07, 0.07)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Per</td>
<td>0.91 [0.155] (0.03, 0.06)</td>
<td>0.91 [0.176] (0.01, 0.08)</td>
<td>0.91 [0.176] (0.01, 0.08)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Per-t</td>
<td>0.87 [0.146] (0.07, 0.06)</td>
<td>0.87 [0.157] (0.08, 0.06)</td>
<td>0.87 [0.157] (0.08, 0.06)</td>
</tr>
</tbody>
</table>

Notes: Results are based on 1,000 replications. See notes in Table 1. BC: simple bias corrected interval.
Table 4
Coverage of Nominal 90% Confidence Interval (Grid Bootstrap)

<table>
<thead>
<tr>
<th>T</th>
<th>N</th>
<th>N/T</th>
<th>Signal-to-noise ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.0</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.81 [0.204]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.10, 0.09)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.87 [0.323]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.00, 0.13)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.75 [0.230]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.20, 0.05)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.94 [0.243]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.06, 0.00)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.94 [0.241]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.06, 0.00)</td>
</tr>
<tr>
<td>20</td>
<td>0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.84 [0.187]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.10, 0.06)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.83 [0.199]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.08, 0.09)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.91 [0.254]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.00, 0.09)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.96 [0.242]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.04, 0.00)</td>
</tr>
</tbody>
</table>

Notes: Results are based on 1,000 replications. See notes in Table 1.
### Table 5
Size and Power of Nominal 10% Tests of Out-of-Sample Forecast

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$R$</th>
<th>$P$</th>
<th>$N$</th>
<th>$\beta = 0$</th>
<th>$\beta = 0.25$</th>
<th>$\beta = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$U$</td>
<td>$DM$</td>
<td>$Enc$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$h = 1$</td>
<td>$h = 2$</td>
<td>$h = 4$</td>
</tr>
<tr>
<td>0.8</td>
<td>50</td>
<td>50</td>
<td>10</td>
<td>0.09</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.09</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.09</td>
<td>0.10</td>
<td>0.11</td>
</tr>
<tr>
<td>0.8</td>
<td>100</td>
<td>50</td>
<td>10</td>
<td>0.09</td>
<td>0.10</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.10</td>
<td>0.10</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.10</td>
<td>0.10</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Notes: Results are based on 1,000 replications. $U$: Theil’s U statistic. $DM$: Diebold-Mariano test. $Enc$: Encompassing test.
### Table 6
**US Diffusion Index, AR(1) Model**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>159</td>
<td>63</td>
<td>27</td>
</tr>
<tr>
<td>Point estimate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\rho}$</td>
<td>0.656</td>
<td>0.496</td>
<td>0.733</td>
</tr>
<tr>
<td>$\hat{\rho}_{BC}$</td>
<td>0.693</td>
<td>0.566</td>
<td>0.926</td>
</tr>
<tr>
<td>90% interval</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asy</td>
<td>(0.600, 0.713)</td>
<td>(0.431, 0.561)</td>
<td>(0.683, 0.784)</td>
</tr>
<tr>
<td>Bc</td>
<td>(0.636, 0.750)</td>
<td>(0.500, 0.631)</td>
<td>(0.875, 0.977)</td>
</tr>
<tr>
<td>Per</td>
<td>(0.628, 0.745)</td>
<td>(0.481, 0.625)</td>
<td>(0.718, 0.966)</td>
</tr>
<tr>
<td>Per-t</td>
<td>(0.641, 0.751)</td>
<td>(0.507, 0.643)</td>
<td>(0.895, 1.049)</td>
</tr>
<tr>
<td>Grid</td>
<td>(0.636, 0.759)</td>
<td>(0.495, 0.665)</td>
<td>(0.872, 0.919)</td>
</tr>
<tr>
<td>Grid-t</td>
<td>(0.636, 0.759)</td>
<td>(0.495, 0.665)</td>
<td>(0.871, 0.919)</td>
</tr>
</tbody>
</table>

Notes: Sample period is 1959:3-1998:12 ($T = 478$).

### Table 7
**Forecasting Inflation Using Diffusion Index**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>159</td>
<td>63</td>
<td>27</td>
</tr>
<tr>
<td>$U$</td>
<td>0.918</td>
<td>0.891</td>
<td>0.909</td>
</tr>
<tr>
<td>$p$-value</td>
<td>0.000 (0.010)</td>
<td>0.000 (0.002)</td>
<td>0.002 (0.011)</td>
</tr>
<tr>
<td>$DM$</td>
<td>1.965</td>
<td>2.236</td>
<td>1.911</td>
</tr>
<tr>
<td>$p$-value</td>
<td>0.024 (0.025)</td>
<td>0.004 (0.013)</td>
<td>0.024 (0.028)</td>
</tr>
<tr>
<td>$Enc$</td>
<td>2.798</td>
<td>3.080</td>
<td>2.917</td>
</tr>
<tr>
<td>$p$-value</td>
<td>0.016 (0.003)</td>
<td>0.004 (0.001)</td>
<td>0.012 (0.002)</td>
</tr>
</tbody>
</table>

Notes: 12-month-ahead forecasts. Sample period is 1959:3-1998:12 ($T = 478$). Test statistics with bootstrap $p$-values. Numbers in parentheses are $p$-values from normal approximation.
Figure 1. US Diffusion Index

Note: Rescaled using industrial production.