Confidence Intervals for Diffusion Index Forecasts with a Large Number of Predictors

Jushan Bai* Serena Ng †

August 2004

Abstract

We consider the situation when there is a large number of series, N, each with T observations, and each series has some predictive ability for the variable of interest, y. A methodology of growing interest is to first estimate common factors from the panel of data by the method of principal components, and then augment an otherwise standard regression or forecasting equation with the estimated factors. In this paper, we show that the least squares estimates obtained from these factor augmented regressions are $\sqrt{T}$ consistent if $\sqrt{T}/N \to 0$. The factor forecasts for the conditional mean are $\min[\sqrt{T}, \sqrt{N}]$ consistent, but the effect of "estimated regressors" is asymptotically negligible when T/N goes to zero. We present analytical formulas for predication intervals that take into account the sampling variability of the factor estimates. These formulas are valid regardless of the magnitude of N/T, and can also be used when the factors are non-stationary. The generality of these results is made possible by a covariance matrix estimator that is robust to weak cross-section correlation and heteroskedasticity in the idiosyncratic errors. We provide a consistency proof for this CS-HAC estimator.

*Department of Economics, NYU, 269 Mercer St, New York, NY 10003 Email: Jushan.Bai@nyu.edu.
†Department of Economics, University of Michigan, Ann Arbor, MI 48109 Email: Serena.Ng@umich.edu
The authors acknowledge financial support from the NSF (grants SBR-9896329, SES-0137084, SES-0136923)
We thank seminar participants at Yale University for useful comments

1 Introduction

The use of factors to achieve dimension reduction has been found to be empirically useful in analyzing macroeconomic time series, and adding factors to an otherwise standard regression or forecasting model is being used by an increasing number of researchers[1]. Several institutions, including the Treasury and the European Central Bank, are experimenting with real time use of these factor forecasts.[2] However, the theoretical properties of the method are not fully understood and important issues remain to be addressed. In particular, how to construct confidence intervals remains unknown. This is a nontrivial problem as the regression model involves "estimated regressors." In this paper, we derive the rate of convergence and the limiting distribution of the parameter estimates as well as the forecasts, enabling the construction of prediction confidence intervals.

The object of interest is the h-period ahead forecast of a series y_t. The information available includes a large number of predictors x_{it} ($i = 1, 2, ..., N; t = 1, 2, ..., T$) and a smaller set of other observable variables W_t. For example, W_t might be lags of y_t. We consider a single forecasting equation

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

where h is the forecast horizon. The vector F_t is unobservable. When $h = 0$, we simply have a regression model with a vector of latent regressors. Instead of F_t, we observe a panel of data x_{it} which contains information about F_t. We refer to

$$x_{it} = \lambda_i' F_t + e_{it} \tag{2}$$

as the factor representation of the data, where F_t is a $r \times 1$ vector of common factors, λ_i is the corresponding vector of factor loadings, and e_{it} is an idiosyncratic error. Equations (1) and (2) above constitute what is referred to by Stock and Watson (2002a) as a 'diffusion index forecasting model' (DI). Its defining characteristic is that information about x_{it} is parsimoniously summarized in a low dimensional vector, F_t. In economic analysis, these generate comovements in economic time series.

If F_t is observable, and assuming the mean of ε_t conditional on past information is zero, the (mean-squared) optimal forecast of y_t is the conditional mean and is given by

$$y_{T+h|T} = E(y_{T+h}|z_T) = \alpha' F_T + \beta' W_T \equiv \delta' z_T,$$

[1]See, for example, Stock and Watson (2002b), Stock and Watson (2001), Cristadoro et al. (2001), Forni et al. (2001b), Artis et al. (2001), Banerjee et al. (2004), and Shintani (2002).

[2]See, for example, Angelini et al. (2001).

where $z_t = (F_t', \ W_t')'$. But such a forecast is not feasible because α, β, and F_t are all unobserved. The feasible forecast that replaces the unknown objects by their estimates is:

$$\widehat{y}_{T+h|T} = \widehat{\alpha}'\widetilde{F}_T + \widehat{\beta}'W_T = \widehat{\delta}'\widehat{z}_T,$$

where $\widehat{z}_t = (\widetilde{F}_t', \ W_t')'$. We use a 'tilde' for estimates of the factor model of x_{it}, while hatted variables are estimated from the forecasting equation. To be precise, $\widehat{\alpha}$ and $\widehat{\beta}$ are the least squares estimates obtained from a regression of y_{t+h} on $\widetilde{F}_t$ and W_t, $t = 1, \dots T - h$. The factors, F_t, are estimated from x_{it} by the method of principal components using data up to period T and will be discussed further below.

It is clear that $\widehat{\alpha}$ and $\widehat{\beta}$ are functions of "estimated regressors" $\widetilde{F}_1, \widetilde{F}_2, ..., \widetilde{F}_{T-h}$, and the forecast $\widehat{y}_{T+h|T}$ itself also depends on $\widetilde{F}_T$. Thus, to study the behavior of the forecasts, we must examine the statistical properties of the estimated parameters $(\widehat{\alpha}, \widehat{\beta})$ as well as those of the estimated factors. Stock and Watson (2002a) showed that $(\widehat{\alpha}, \widehat{\beta})$ is consistent for (α, β) and $\widehat{y}_{T+h|T}$ is consistent for $y_{T+h|T}$. To construct confidence intervals, we must obtain the rate of convergence and the limiting distributions of these quantities.

We are specifically interested in the case of large dimensional panels. By a 'large panel', we mean that our theory will allow both N and T to tend to infinity, and N possibly larger than T. We begin in Section 2 with an intuitive discussion of the problem to be investigated and of the results to follow. Section 3 presents the asymptotic theory and discusses how terms necessary for predictive inference can be constructed. A by-product of the present exercise is estimation of the error covariance matrix when heteroskedasticity and cross-section correlation are of unknown form. This is presented in Section 4. Section 5 presents simulation results to assess the adequacy of the asymptotic approximations in finite samples. Empirical applications are considered in Section 6. The analysis is extended to non-stationary factors in Section 7. Proofs are given in the Appendix.

2 Motivation and Overview

We first provide some intuition for the appeal of diffusion index forecasts. For ease of exposition, consider the one-step ahead forecast:

$$y_{t+1} = \alpha F_t + \varepsilon_{t+1}$$

where ε_t are iid $(0, \sigma_\varepsilon^2)$. Furthermore, assume that the scalar series F_t is an AR(1) process

$$F_t = \rho F_{t-1} + u_t$$

where u_t are iid $(0, \sigma_u^2)$ and u_s and ε_t are independent for all t and s. Suppose also for the moment that the model parameters are known.

If F_t is observable, the one-step ahead forecast of y_{t+1} at time t is given by αF_t so that the forecast error is ε_{t+1}, and the forecast error variance is σ_ε^2. If F_t is not observable, then y_t is an unobserved components model. The univariate time series forecast is based on the ARMA representation of y_t. In this case, y_t is an ARMA(1,1) process:

$$y_{t+1} = \rho y_t + z_{t+1} + \theta z_t$$

where z_t is a white noise process. Assuming the infinite past history of y_t $(...., y_{t-2}, y_{t-1}, y_t)$ is available, the one-step ahead forecast of y_{t+1} at time t is $\rho y_t + \theta z_t$. The forecast error is z_{t+1} and the forecast error variance is $\sigma_z^2 = E(z_{t+1}^2)$. It can be shown that $\sigma_z^2 > \sigma_\varepsilon^2$, so smaller forecasting error variance is obtained when F_t is observable. This is not surprising and conforms to the intuition that more information permits a better forecast.

The assumption that F_t is observable is of course not realistic. Nevertheless, if we observe a large number of indicators that have F_t as their common sources of variation, we can exploit this commonality to estimate the process F_t very well by the method of principal components (up to a transformation). This is the essence of the diffusion index forecasting. In the limit when N goes to infinity, the DI forecasts are the same as when F_t is observable. In this example, the reduction in forecast error is $\sigma_z^2 - \sigma_\varepsilon^2$, which is strictly positive. In cases with more complex dynamics and/or when W_t are present, knowledge of F_t can still be expected to yield better forecasts, because one can, in general, do no worse with more information.

In practice, the model parameters are also unknown. Parameter uncertainty contributes an $O(T^{-1})$ term to the forecast error variance. So if we observe F_t but α is being estimated, the variance of $y_{T+1} - \widehat{y}_{T+1|T}$ is simply $\sigma_\varepsilon^2 + O(T^{-1})$. However, when the factors have to be estimated, we first need to show that $\widehat{\alpha}$ remains $\sqrt{T}$ consistent. Furthermore, estimating the factor process F_t will contribute another $O(N^{-1})$ term to the forecasting error variance. One of our findings is that when α and F_t both have to be estimated, the variance of $y_{T+1} - \widehat{y}_{T+1|T}$ is $\sigma_\varepsilon^2 + O(T^{-1}) + O(N^{-1})$. This is less than σ_z^2 when T and N are both large (because $\sigma_\varepsilon^2 < \sigma_z^2$) so one can expect the diffusion index approach to yield better forecasts even when F_t is not observed. Our main contribution is to show that the forecast for the conditional mean is $\min[\sqrt{N}, \sqrt{T}]$ consistent and asymptotically normal, where the precise rate will depend on whether T/N is bounded. In the event when T and N are both large and are such that T/N goes to zero, we can further show that uncertainty in F_t is dominated by parameter uncertainty so that F_t can be treated as though it is observable. We will make precise how to estimate the error covariance matrices so that valid predictive inference can be conducted. The importance of a large N must be stressed, however, because when N is fixed, consistent estimation of the factor process F_t is not possible even if the λ_is are observed. We now turn to the theory underlying these results.

3 Inference with Estimated Factors

In matrix notation, the factor model is $X = F\Lambda' + e$, where X is a $T \times N$ data matrix, $F = (F_1, ..., F_T)'$ is $T \times r$, r is the number of common factors, $\Lambda = (\lambda_1, ..., \lambda_N)'$ is $N \times r$, and e is a $T \times N$ error matrix. The principal component estimates are denoted $\widetilde{F} = (\widetilde{F}_1, ..., \widetilde{F}_T)'$ and $\widetilde{\Lambda} = (\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_N)'$, where $\widetilde{F}$ is the matrix consisting of the r eigenvectors (multiplied by $\sqrt{T}$) associated with the r largest eigenvalues of the matrix $XX'/(TN)$ in decreasing order, and $\widetilde{\Lambda} = X'\widetilde{F}/T$. For future reference, we also let $\widetilde{V}$ be the $r \times r$ diagonal matrix consisting of the r largest eigenvalues of $XX'/(TN)$. We need the following assumptions:

Assumption A: Common factors

1. $E\|F_t\|^4 \leq M$ and $\frac{1}{T}\sum_{t=1}^{T} F_t F_t' \xrightarrow{p} \Sigma_F$ for a $r \times r$ positive definite matrix Σ_F.

Assumption B: Heterogeneous factor loadings

The loading λ_i is either deterministic such that $\|\lambda_i\| \leq M$ or it is stochastic such that $E\|\lambda_i\|^4 \leq M$. In either case, $\Lambda'\Lambda/N \xrightarrow{p} \Sigma_\Lambda$ as $N \to \infty$ for some $r \times r$ positive definite non-random matrix Σ_Λ.

Assumption C: Time and cross-section weak dependence and heteroskedasticity

1. $E(e_{it}) = 0$, $E|e_{it}|^8 \leq M$;

2. $E(e_{it}e_{js}) = \tau_{ij,ts}$, $|\tau_{ij,ts}| \leq \tau_{ij}$ for all (t, s) and $|\tau_{ij,ts}| \leq \pi_{ts}$ for all (i, j) such that

$$\frac{1}{N}\sum_{i,j=1}^{N} \tau_{ij} \leq M, \ \frac{1}{T}\sum_{t,s=1}^{T} \pi_{ts} \leq M, \text{ and } \frac{1}{NT}\sum_{i,j,t,s=1} |\tau_{ij,ts}| \leq M$$

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

4. For each t, $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \lambda_t e_{it} \xrightarrow{d} N(0, \Gamma_t)$, where $\Gamma_t = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} E(\lambda_i {\lambda_j}' e_{it}e_{jt})$.

Assumption D: $\{\lambda_i\}, \{F_t\}$, and $\{e_{it}\}$ are three groups of mutually independent stochastic variables.

Assumption E: Let $z_t = (F_t' \;\; W_t')'$, $E\|z_t\|^4 \le M$, and z_t is independent of the idiosyncratic errors e_{it}. Furthermore,

1. $\frac{1}{T}\sum_{t=1}^T z_t z_t' \xrightarrow{p} \Sigma_{zz} = \begin{bmatrix} \Sigma_{FF} & \Sigma_{FW} \\ \Sigma_{WF} & \Sigma_{WW} \end{bmatrix} > 0;$
2. $\frac{1}{\sqrt{T}}\sum_{t=1}^T z_t \varepsilon_{t+h} \xrightarrow{d} N(0, \text{plim}\, \frac{1}{T}\sum_{t=1}^T (\varepsilon_{t+h}^2 z_t z_t')).$

Assumptions A and B together imply r common factors. Assumption C allows for limited time series and cross section dependence in the idiosyncratic component. Heteroskedasticity in both the time and cross section dimensions is also allowed. Given Assumption C1, the remaining assumptions in C are easily satisfied if the e_{it} are independent for all i and t. The allowance for weak cross-section correlation in the idiosyncratic components leads to the *approximate factor structure* of Chamberlain and Rothschild (1983). It is more general than a *strict factor model* which assumes e_{it} is uncorrelated across i. Assumption D is standard in factor analysis. Assumption E ensures that the forecasting model is well specified and that the parameters of the model can be identified.

3.1 Estimation

We begin by establishing the sampling properties of the least squares estimates when the estimated factors are used as regressors.

Theorem 1 *(Estimation) Suppose Assumptions A to E hold. Let $\widetilde{F}_t$ be the factor estimates obtained by the method of principal components, and let $\widehat{\alpha}$ and $\widehat{\beta}$ be the least squares estimates from a regression of y_{t+h} on $\widehat{z}_t = (\widetilde{F}_t' \; W_t')'$. Let $H = \widetilde{V}^{-1}(\widetilde{F}'F/T)(\Lambda'\Lambda/N)$. If $\sqrt{T}/N \to 0$,*

$$\sqrt{T}\left(\begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix} - \begin{bmatrix} H^{-1\prime}\alpha \\ \beta \end{bmatrix} \right) \xrightarrow{d} N\left(0, Avar\left(\begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix} \right)\right),$$

where

$$Avar\left(\begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix} \right) = \text{plim} \left(\frac{1}{T}\sum_{t=1}^{T-h} \widehat{z}_t\widehat{z}_t' \right)^{-1} \left(\frac{1}{T}\sum_{t=1}^{T-h} \widehat{\varepsilon}_{t+h}^2 \widehat{z}_t\widehat{z}_t' \right) \left(\frac{1}{T}\sum_{t=1}^{T-h} \widehat{z}_t\widehat{z}_t' \right)^{-1}. \tag{3}$$

As is well known, the factor model is unidentified because $\alpha' L L^{-1} F_t = \alpha' F_t$ for any invertible matrix L. Theorem 1 is a result pertaining to the difference between $\widehat{\alpha}$ and the space spanned by α. Consistency of the parameter estimates follows from the fact that the averaged squared deviations between $\widetilde{F}_t$ and HF_t vanish as N and T both tend to infinity, see Bai and Ng (2002). The consequence of having generated regressors in the forecasting equation has no effect on consistency of the parameter estimates. Letting $\widehat{\delta} = (\widehat{\alpha}' \ \widehat{\beta}')'$, and $\delta = (\alpha' H^{-1} \ \beta')'$, Theorem 1 can be compactly stated as

$$\sqrt{T}(\widehat{\delta} - \delta) \xrightarrow{d} N(0, Avar(\widehat{\delta})).$$

Stock and Watson (2002a) showed consistency of $\widehat{\delta}$ for δ. Here we establish the rate of convergence and the limiting distribution. Asymptotic normality of $\widehat{\delta}$ follows from that fact that $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_t \varepsilon_{t+h}$ obeys a central limit theorem. Because $\widetilde{F}_t$ is close to F_t, the same asymptotic result holds when z_t is replaced by $\widehat{z}_t$.

Since $Avar(\widehat{\delta})$ is the probability limit of (3), it can be consistently estimated as follows:

$$\widehat{Avar(\widehat{\delta})} = \left(\frac{1}{T} \sum_{t=1}^{T-h} \widehat{z}_t \widehat{z}_t' \right)^{-1} \left[\frac{1}{T} \sum_{t=1}^{T-h} \widehat{\varepsilon}_{t+h}^2 \widehat{z}_t \widehat{z}_t' \right] \left(\frac{1}{T} \sum_{t=1}^{T-h} \widehat{z}_t \widehat{z}_t' \right)^{-1} \tag{4a}$$

$$\widehat{Avar(\widehat{\delta})} = \widehat{\sigma}_\varepsilon^2 \left[\frac{1}{T} \sum_{t=1}^{T-h} \widehat{z}_t \widehat{z}_t' \right]^{-1}. \tag{4b}$$

Formula (4a) is the White-Eicker estimate of asymptotic variance and is robust to heteroskedasticity. However, if we assume homoskedasticity so that $E(\varepsilon_{t+h}^2 | z_t) = \sigma_\varepsilon^2 \ \forall t$, a consistent estimate of $Avar(\widehat{\delta})$ can be obtained using (4b), where $\widehat{\sigma}_\varepsilon^2 = \frac{1}{T} \sum_{t=1}^{T-h} \widehat{\varepsilon}_{t+h}^2$. As stated, the asymptotic variance is valid when $z_t \varepsilon_{t+h}$ is serially uncorrelated. Extension of (4a) to allow for serial correlation in $z_t \varepsilon_{t+h}$ is straightforward. As shown in Newey and West (1987) and Andrews (1991), a heteroskedastic-autocorrelation consistent variance covariance (HAC) matrix that converges to the population covariance matrix can be constructed provided the bandwidth is chosen appropriately. It is noted, however, when ε_t is serially correlated, $y_{T+h|T}$ defined earlier will cease to be the conditional mean, given past information.

Theorem 1 is useful in rather broader contexts, as having to conduct inference when the latent common factors are replaced by estimates is not uncommon. The estimated common factors are natural proxies for the unobserved state of the economy. In Phillips curve regressions, y_{t+h} would be inflation, W_t would be lags of inflation, and Theorem 1 provides the inferential theory for assessing the trade-off between inflation and the state of the economy.

A new tool in empirical work is factor-augmented vector autoregressions (FVAR), which amounts to including the principal component estimates of the factors to an otherwise standard VAR.[3] More specifically, if y_t is a vector of q series, and F_t is a vector of r factors, a

[3] See, for example, Bernanke and Boivin (2002), Bernanke et al. (2002), and Giannone et al. (2002), and Marcellino et al. (2004).

FVAR(p) is defined as

$$\begin{array}{rcl} y_{t+1} & = & \sum_{k=0}^{p} a_{11}(k)y_{t-k} + \sum_{k=0}^{p} a_{12}(k)F_{t-k} + v_{1t+1} \\ F_{t+1} & = & \sum_{k=0}^{p} a_{21}(k)y_{t-k} + \sum_{k=0}^{p} a_{22}(k)F_{t-k} + v_{2t+1}, \end{array}$$

where $a_{11}(k)$ and $a_{21}(k)$ are coefficients on lags of y_{t+1}, while $a_{12}(k)$ and $a_{22}(k)$ are coefficients on lags of F_{t-k}. Consider estimation of the FVAR with F_t replaced by $\widetilde{F}_t$. Theorem 1 covers estimation of those equations of the VAR with y_{t+1}on the left hand side, W_t and $\widetilde{F}_t$ on the right hand side, where in the present context, W_t are the lags of y_t. The following theorem provides the limiting distribution of $\widehat{\delta}_j$ for those equations with $\widetilde{F}_{t+1}$ on the left hand side.

Theorem 2 *(FVAR) Consider a p-th order vector autoregression in q observable variables y_t and r factors, $\widetilde{F}_t$, estimated by the method of principal components. Let $\widehat{z}_t = (y_t \ \dots y_{t-p}, \widetilde{F}_t, \dots, \widetilde{F}_{t-p})'$, and let $\widehat{z}_{jt}$ be the j-th element of $\widehat{z}_t$. For $j = 1, \dots q + r$, let $\widehat{\delta}_j$ be obtained by least squares from regressing $\widehat{z}_{jt+1}$ on $\widehat{z}_t$, with $\widehat{\varepsilon}_{jt+1} = \widehat{z}_{jt+1} - \widehat{\delta}_j'\widehat{z}_t$. If $\sqrt{T}/N \to 0$ as $N, T \to \infty$,*

$$\sqrt{T}(\widehat{\delta}_j - \delta_j) \xrightarrow{d} N\left(0, \text{plim}(\frac{1}{T}\sum_{t=1}^{T} \widehat{z}_t\widehat{z}_t')^{-1} \left(\frac{1}{T}\sum_{t=1}^{T}(\widehat{\varepsilon}_{jt})^2 \widehat{z}_t\widehat{z}_t'\right)(\frac{1}{T}\sum_{t=1}^{T} \widehat{z}_t\widehat{z}_t')^{-1}\right).$$

Theorem 2 states that the parameter estimates for these equations remain $\sqrt{T}$ consistent provided $\sqrt{T}/N \to 0$. Although this condition is not stringent, it puts discipline on when estimated factors can be used in regression analysis. Having N and T both large is not enough. Once this condition is granted, the expression for asymptotic variance is the same whether y_t or $\widetilde{F}_t$ is the regressand (compare with Theorem 1). Thus, if homoskedasticity is assumed, as is common in the VAR literature, the asymptotic variance can be evaluated using (4b). Since impulse response functions are based upon estimates of the FVAR, Theorem 2 enables calculation of the standard errors.

3.2 Forecasting

Suppose now (1) is the forecasting equation and the objective is the forecast error distribution. From

$$(\widehat{y}_{T+h|T} - y_{T+h|T}) = (\widehat{\delta} - \delta)'\widehat{z}_T + \alpha' H^{-1}(\widetilde{F}_T - HF_T),$$

we see that the forecast error has two components. The first term arises from having to estimate α and β. Theorem 1 makes clear that what this error is asymptotically. The second term arises from having to estimate F_t. Under Assumptions A-D, Bai (2003) showed that if $\sqrt{N}/T \to 0$, then for each t,

$$\begin{aligned} \sqrt{N}(\widetilde{F}_t - HF_t) &\xrightarrow{d} N\Big(0, V^{-1}Q\Gamma_t Q'V^{-1}\Big) \\ &\equiv N\Big(0, Avar(\widetilde{F}_t)\Big), \end{aligned} \tag{5}$$

where $Q = \text{plim}\, \widetilde{F}'F/T$, $V = \text{plim}\, \widetilde{V}$, and $\Gamma_t = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} E(\lambda_i \lambda_j' e_{it} e_{jt})$.

We are now in a position to state the asymptotic properties of the DI forecasts.

Theorem 3 *Let $\widehat{y}_{T+h|T} = \widehat{\delta}'\widehat{z}_T$ be the feasible h-step ahead forecast of y_{T+h}. Under the assumptions of Theorem 1,*

$$\frac{(\widehat{y}_{T+h|T} - y_{T+h|T})}{B_T} \xrightarrow{d} N(0,1)$$

where $B_T^2 = \frac{1}{T}\widehat{z}_T'\, Avar(\widehat{\delta})\, \widehat{z}_T + \frac{1}{N}\widehat{\alpha}'\, Avar(\widetilde{F}_T)\, \widehat{\alpha}$.

Because the two terms in B_T^2 vanish at different rates, the overall convergence rate is $\min[\sqrt{T}, \sqrt{N}]$. More precisely, it depends on whether or not T/N is bounded. $\sqrt{T}$ convergence to the normal distribution follows from considering the limit distribution of

$$\sqrt{T}(\widehat{y}_{T+h|T} - y_{T+h|T}) = \sqrt{T}(\widehat{\delta} - \delta)'\widehat{z}_T + (T/N)^{1/2}\alpha' H^{-1}\sqrt{N}(\widetilde{F}_T - HF_T).$$

When T/N is bounded, the estimation error associated with $\widehat{\delta}$ and $\widetilde{F}_t$ both contribute to the asymptotic forecast error variance. However, the cost of having to estimate F_t is negligible when $T/N \to 0$ because $\sqrt{N}(\widetilde{F}_t - HF_t)$ is $O_p(1)$. Intuitively, when N is large, the factors can be estimated so precisely that estimation error can be ignored. On the other hand, when N/T is bounded, the convergence rate is $\sqrt{N}$. This follows from the fact that

$$\sqrt{N}(\widehat{y}_{T+h|T} - y_{T+h|T}) = (\sqrt{N/T})\sqrt{T}(\widehat{\delta} - \delta)'\widehat{z}_T + \alpha' H^{-1}\sqrt{N}(\widetilde{F}_T - HF_T).$$

If $N/T \to 0$, the error from having to estimate δ is dominated by the error from having to estimate F_t.

In a standard setting, the forecast error variance falls at rate T, and for a given T, it increases with the number of predictors through a loss in degrees of freedom. In contrast, the error variance of the factor forecasts decreases at rate $\min[N, T]$, and for a given T, forecast efficiency improves with the number of predictors. This is because in the present setting, a large N enables more precise estimation of the common factors and thus results in more efficient forecasts. This property of the factor estimates is also in sharp contrast to that obtained in standard factor analysis that assumes a fixed N. With the sample size fixed in one dimension, consistent estimation of the factor space is not possible however large T becomes.

In view of (5), an estimate of $Avar(\widetilde{F}_t)$ (for any given t) can be obtained by first substituting $\widetilde{F}$ for F, and noting that $\widetilde{Q} = \widetilde{F}'\widetilde{F}/T$ is an r-dimensional identity matrix by construction ($\widetilde{Q}$ is an estimate for QH' whose limit is an identity). We can then consider the estimator

$$\widehat{Avar(\widetilde{F}_t)} = \widetilde{V}^{-1}\widetilde{\Gamma}_t\widetilde{V}^{-1},$$

where $\widetilde{\Gamma}_t$ can be one of the following:

$$\widetilde{\Gamma}_t = \frac{1}{N}\sum_{i=1}^{N}\widetilde{e}_{it}^2\widetilde{\lambda}_i\widetilde{\lambda}_i' \tag{6a}$$

$$\widetilde{\Gamma}_t = \widetilde{\sigma}_e^2\frac{1}{N}\sum_{i=1}^{N}\widetilde{\lambda}_i\widetilde{\lambda}_i' \tag{6b}$$

$$\widetilde{\Gamma}_t = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\widetilde{\lambda}_i\widetilde{\lambda}_j'\frac{1}{T}\sum_{t=1}^{T}\widetilde{e}_{it}\widetilde{e}_{jt}. \tag{6c}$$

The various specifications of $\widetilde{\Gamma}_t$ accommodate flexible error structures in the factor model. Both (6a) and (6b) assume that e_{it} is cross-sectionally uncorrelated with e_{jt}. Consistency of both estimators was shown in our earlier work. The estimator (6b) further assumes $E(e_{it}^2) = \sigma_e^2$ for all i and t. Under regularity conditions, $\widetilde{\sigma}_e^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{e}_{it}^2 \xrightarrow{p} \sigma_e^2$. Although (6a) and (6b) both assume the idiosyncratic errors are cross-sectionally uncorrelated, it is not especially restrictive because much of the cross-correlation in the data is presumably captured by the common factors. At an empirical level, allowing for cross-section correlation in the errors would entail estimation of $N(N-1)/2$ additional parameters. Because N is large by assumption, sampling variability could generate non-trivial efficiency loss. For small cross-section correlation in the errors, constraining them to be zero could sometimes be desirable. The estimators defined in (6a) and (6b) are useful even if residual cross-correlation is genuinely present.

When it is deemed inappropriate to assume zero cross-section correlation in the errors, the asymptotic variance of $\widetilde{F}_t$ can be estimated by (6c). Consistency of $\widetilde{\Gamma}_t$ will be established below and it requires nontrivial argument. Suffice it to note for now that the estimator, which we will refer to as CS-HAC, is robust to cross-section correlation and heteroskedasticity in e_{it} of unknown form, but requires covariance stationarity with $E(e_{it}e_{jt}) = \sigma_{ij}$ for all t, and that $n = n(N,T)$ satisfies the conditions of Theorem 4 (see Section 5 below). Loosely speaking, covariance stationarity of e_{it} implies that Γ_t does not depend on t so that the residuals from other periods, not just t, can be used to estimate Γ_t. This, however, is not sufficient, as we will also require that $\frac{n}{\min[N,T]} \to 0$ to avoid excess sampling variability of $\widetilde{\lambda}_i$ on $\widetilde{\Gamma}_t$.

Once appropriate estimators for $Avar(\widehat{\delta})$ and $Avar(\widetilde{F}_T)$ are chosen, the above results allow us to construct prediction intervals. This exercise is straightforward given asymptotic normality of the forecasts errors. For example, the 95% confidence interval for the $y_{T+h|T}$ is

$$\left(\widehat{y}_{T+h|T} - 1.96\sqrt{\widehat{var}(\widehat{y}_{T+h|T})}, \quad \widehat{y}_{T+h|T} + 1.96\sqrt{\widehat{var}(\widehat{y}_{T+h|T})}\right),$$

where $\widehat{var}(\widehat{y}_{T+h|T})$ is equal to B_T^2, as defined in Theorem 3, with $Avar(\widehat{\delta})$ and $Avar(\widetilde{F}_t)$ replaced by their consistent estimates.

Although the conditional mean is a useful benchmark for the theoretical properties of forecasts, it is not observable. Thus, in practice, forecast comparisons are inevitably made in terms of y_{T+h}. Since $y_{T+h} = y_{T+h|T} + \varepsilon_{T+h}$, it follows that

$$\widehat{y}_{T+h|T} - y_{T+h} = (\widehat{y}_{T+h|T} - y_{T+h|T}) + \varepsilon_{T+h}.$$

So if ε_t is normally distributed, $\widehat{y}_{T+h|T} - y_{T+h}$ is also approximately normal with

$$\mathrm{var}(\widehat{y}_{T+h|T} - y_{T+h}) = \sigma_\varepsilon^2 + var(\widehat{y}_{T+h|T}),$$

which in large samples will be dominated by σ_ε^2, since $\mathrm{var}(\widehat{y}_{T+h|T})$ vanishes at rate $\min[T,N]$. The result that σ_e^2 dominates in large samples, which is standard in the forecasting literature, continues to hold when the factors are estimated. It should, however, be stressed that the error arising from using $\widetilde{F}_t$ is asymptotically negligible only if Theorems 1 and 3 hold. It is thus essential that N and T are both large, with $\sqrt{T}/N \to 0$.

Theorem 3 fills an important void in the diffusion index forecasting literature, as it goes beyond the consistency result to establish asymptotic normality. The result has uses beyond forecasting, as it provides the basis of testing economic hypothesis that involves fundamental factors. Observed variables are often used in place of the latent factors when testing various theories of asset returns. Using Theorem 3, tests can be developed to determine whether the observables are good proxies for the latent factors. An application was considered in Bai and Ng (2004). That analysis, which amounts to assessing the in-sample predictability of the latent factors, makes use of the results presented here, with h set to zero.

4 Covariance Matrix Estimator: the CS-HAC

The CS-HAC estimator introduced earlier is robust to cross-section correlation and cross-section heteroskedasticity. As a general matter, correcting for cross-section correlation is not an easy task because unlike time series data, a natural ordering of cross-section data rarely arises. Exceptions are spatial models and analysis in which economic distance can be meaningfully defined as in Conley (1999). More generally, neither economic theory nor intuition can be expected to be of much help in obtaining a 'mixing condition' type ordering of the data. Since any permutation of the data is an equally valid representation of information available, the different orderings also cannot be ranked. Instead of truncating terms 'far from' an observation, the common practice in cross-section regressions is to impose restrictions on the off-diagonal elements, or to parameterize Ω in terms of a finite number of parameters. Both approaches serve the purpose of reducing the number of unknowns in Ω from $O(N^2)$ to something more manageable.

A third alternative is to make use of the availability of observations on the cross-section units over time. The basic intuition is as follows. Under covariance stationarity, time series observations allow us to consistently estimate the cross-section correlation matrix. Thus, although the cross-section regressions do not permit consistent estimation of the covariance matrix of interest, this is possible with T large. An estimator along these lines was considered in Driscoll and Kraay (1998). Their estimator is consistent if information from some $n < N$ terms are used, with $n = n(T)$. They place no other restriction on n, nor do they limit the amount of cross-section correlation. In their setup, the regressors are observable.

We also seek to estimate the covariance matrix from panel data, but our analysis is complicated by the fact that λ_i is not observed, and consistent estimation of the factor space necessitates that the cross-section correlation in e_{it} is weak. The notion of cross-section correlation, as defined in Chamberlain and Rothschild (1983) puts bounds on the eigenvalues of Ω. Assumption C restates the condition in terms of the column sum of a matrix. A key condition for "weak" cross section correlation is $\frac{1}{N}\sum_{i=1}^N \sum_{j=1}^N |\sigma_{ij}| \leq M$, where $\sigma_{ij} = E(e_{it}e_{jt})$.

Theorem 4 *Suppose Assumptions A-D hold. Let $n = n(N,T)$ and define*

$$\widetilde{\Gamma}_t = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\widetilde{\lambda}_i\widetilde{\lambda}_j'\frac{1}{T}\sum_{t=1}^{T}\widetilde{e}_{it}\widetilde{e}_{jt}.$$

Then $\left\|\widetilde{\Gamma}_t - H^{-1\prime}\Gamma_t H^{-1}\right\| \xrightarrow{p} 0$ *if* $\frac{n}{\min[N,T]} \to 0$.

In the factor model setup, $\widetilde{e}_{it}$ are the regression residuals associated with the regressors $\widetilde{\lambda}_i$, which are the principal component estimates. Accordingly, the number of correlated pairs we can consider, n, is primarily determined by the convergence rate of the factor estimates. Importantly, the estimator $\widetilde{\Gamma}_t$ is inconsistent if $n = N$ because use of too many $\widetilde{\lambda}_i$ will introduce excess variability to $\widetilde{\Gamma}_t$. Note that $\widetilde{\Gamma}_t$ is an estimate for $H^{-1\prime}\Gamma_t H^{-1}$ not for Γ_t. The end result is correct because the asymptotic variance depends on $Q\Gamma_t Q'$. We use $\widetilde{\lambda}_i$ to estimate $H^{-1\prime}\lambda_i$, and we also estimate QH' instead of Q, where Q is the limit of $\widetilde{F}'F/T$. From $Q\Gamma_t Q' = QH'H'^{-1}\Gamma_t HH^{-1}Q$, the matrix H is effectively canceled out.

The conditions that $n/N \to 0$ and $n/T \to 0$ are not restrictive. The simple rule we use in the simulations below is $n = \min[\sqrt{N}, \sqrt{T}]$. Once n is defined, an estimator can be constructed upon picking n out of N series from the sample. In practice, we first randomly select n series to obtain $\widetilde{\Gamma}_t^{(1)}$. Another n series is picked randomly to obtain $\widetilde{\Gamma}_t^{(2)}$), and so forth. Averaging over $\widetilde{\Gamma}_t^{(k)}$, $k = 1, \ldots K$ gives $\widetilde{\Gamma}_t$. For the DGPs considered below, the results are not sensitive to K. We report results for $K = \min[\sqrt{N}, \sqrt{T}]$.

5 Finite Sample Properties

We now use simulations to assess the finite sample properties of the procedures. Data are generated as follows:

$$\begin{aligned} x_{it} &= \lambda_i' F_t + e_{it}, \quad i = 1, \ldots N, t = 1, \ldots, T \\ F_{jt} &= \rho_j F_{jt-1} + (1-\rho_j^2)^{1/2} u_{jt} \quad j = 1, \ldots, r \\ e_{it} &= (1+b^2)v_{it} + b v_{i+1,t} + b v_{i-1,t}. \\ \rho_j &= (.8)^j, \end{aligned}$$

where u_{jt} and v_{it} are mutually uncorrelated $N(0,1)$ random variables.[4] Cross section correlation is allowed when $b \neq 0$. We draw λ_i from the standard normal distribution. In the simulations, we set $r = 2$ and assume that it is known. The series to be forecasted is

$$y_{t+h} = 1 + F_{1t} + F_{2t} + \varepsilon_{t+h}.$$

That is, $W_t = 1\ \forall t$, α is the unit vector, and β equals 1. The simulation design follows Stock and Watson (2002a) closely. Configurations that include additional W_t series yield similar results and will not be presented.

[4]The results are very similar if the innovation variance of u_t is not scaled by $1-\rho_j^2$. The scaling is enables us to control the size of the common to the idiosyncratic component.

Our main interest is in the coverage of the confidence intervals. Three types of confidence intervals will be presented:

$$\text{Model (A): (6b) +(4b) ;} \quad \text{Model (B): (6a) + (4a) ;} \quad \text{Model (C): (6c) + (4a).}$$

For each model, the coverage rates are reported for (i) the diffusion index forecast for the conditional mean, $\widehat{y}_{T+h|T}$; (ii) the infeasible forecast of the conditional mean $\widehat{y}^0_{T+h|T}$; (iii) the diffusion index forecast for y_{T+h}, and (iv) the infeasible forecast y^0_{T+h}. By infeasible forecast, we mean that F_t is used, and estimation of the factors is not necessary. A comparison of the feasible and infeasible forecasts gives an indication of the error arising from the estimation of F_t.

The results are presented in Tables 1, 2, and 3 respectively. The top panel are coverage rates when the forecasting model is correctly specified (in terms of the number of factors). Overall, the coverage rates are excellent. The probability that $y_{T+h|T}$ or y_{T+h} lies inside the estimated prediction intervals is always close to the nominal coverage rate of .95, even when N and T are only 50.

The idiosyncratic errors are cross-sectionally uncorrelated when $b = 0$, in which case all three estimators of $Avar(\widetilde{F}_t)$ are valid. Although (6c) should be less efficient, comparing the results in Table 1 and 2 with those in 3 reveal that estimating the cross-section correlation when none is present seems to have little effect on coverage. In the simulations, the errors are homoskedastic by design. The results using the heteroskedastic robust estimator in Tables 2 and 3 are also similar to those in Table 1 with homoskedasticity imposed.

When $b \neq 0$, use of (6c) is appropriate. Omitting cross-section correlation tends to weaken coverage marginally. This should not be taken as indication that cross-section correlation in the errors does not need to be dealt with. In situations when the cross-correlation is more prevalent, the effect will be amplified.

The bottom panel of Tables 1 to 3 consider situations when too few factors are used. In these cases, the coverage for the conditional mean is well below .95. This problem is not specific to diffusion index forecasting, however, as inference cannot be expected to be correct when the object of interest is misspecified. Nonetheless, the coverage for y_{T+h} remains accurate because the misspecification in the conditional mean leads to a correspondingly larger unconditional prediction error variance. Inference on y_{T+h} is not significantly affected by whether the error comes from the conditional mean, or from the residual component.

6 Empirical Application

Although diffusion index forecasts have been found to yield improvements over simple models, a major shortcoming is that only point forecasts are available. There exist no tools to assess uncertainty around the forecasts. With the distribution of the forecast errors presented in the previous section, it is now possible to compute prediction intervals.

To illustrate, we use as predictors the 150 series as in Stock and Watson (2002b).[5] We consider $h = 12$ period ahead forecast of the annual growth rate of industrial production, DIP, and inflation, DP. Thus, y_{t+12} is either $DIP = \log(IP_{t+12}) - \log(IP_t)$, or $DP = \log(PUNEW_{t+12}) - \log(PUNEW_t)$. For W_t, we include lags of the monthly first difference of the series, plus a constant. The forecasting exercise begins by estimating the factors using data on x_{it} from 1959:1 to 1969:1. We then obtain $\widehat{\alpha}$ and $\widehat{\beta}$ from a regression of y_t on $\widetilde{F}_{t-12}$ and W_{t-12}, for t=1959:1 to 1969:1. The forecast for $y_{1970:1}$ is computed as $\widehat{\alpha}'\widetilde{F}_{1969:1} + \widehat{\beta}'W_{1969:1}$. The sample is then extended by one month, the factors and all the parameters are re-estimated, and the forecast for $y_{1970:2}$ is formed. The procedure is repeated until the forecast for 1996:12 is made in 1995:12.

For the sake of comparison, we also consider the autoregressive forecast $\widehat{\beta}'W_{1969:1}$. We first select the order of this autoregression using the BIC. The diffusion index model then augments this autoregression with the estimated factors. If the factors have no useful information, α should be zero, and the autoregressive forecast will be the optimal forecast.

Because the two series to be forecasted are one of the x_{it}s, the number of factors in y_t is the same as the number of common factors in the panel of data. This is determined using $\widehat{r} = \text{argmax}_{k=0,\ldots kmax}\, IC_P(k)$ where

$$IC_P(k) = \log \widetilde{\sigma}^2(k) + k \cdot g(N,T),$$

where $\widetilde{\sigma}^2(k) = \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \widetilde{e}_{it}^2$. In Bai and Ng (2002), we showed that any penalty satisfying $g(N,T) \to 0$ and $\min[N,T]g(N,T) \to \infty$ is theoretically valid. Stock and Watson (2002b) used $g_1(N,T) = \frac{\log(\min[N,T])}{\min[N,T]}$. This penalty tends to favor a larger number of factors than $g_2(N,T) = (N+T)\frac{\log(NT)}{NT}$, an equally valid penalty except in the unusual case that $N = \exp(T)$. Obviously, the larger the number of factors, the less likely will the errors be cross-sectionally correlated. Thus, we consider two sets of confidence intervals. Configuration A uses $g_1(N,T)$ with $Avar(\widetilde{F}_t)$ specified by (6a). Configuration B uses $g_2(N,T)$ with $Avar(\widetilde{F}_t)$ specified by (6c). In both cases, (4a) is used for $Avar(\widehat{\delta})$. As it turns out, the

[5]The data are taken from Mark Watson's web site http://www.princeton.edu/ mwatson.

results are quite similar, with results for configuration B slightly better. We will only report results for configuration B. It uses a smaller number of estimated factors, but correct for cross-section correlation in the idiosyncratic errors.

Industrial Production Figure 1a presents the autoregressive (AR) and the diffusion index forecasts for industrial production. Because DIP is only mildly serially correlated, the AR forecast (broken line) is roughly constant. The diffusion index forecast (dotted line) is more volatile, but tracks the actual data more closely. The average mean-squared error for the diffusion index and AR forecasts are 24.95 and 26.46, respectively. Figures 2a and 2b present the series to be forecasted, along with the 95% prediction interval as suggested by the diffusion index and the AR forecasts, respectively. The mean length of the confidence interval is 17.17 for the diffusion index model, and is 20.48 for the AR model. This agrees with the visual impression that the confidence interval is narrower when the factors are used.

Inflation The inflation forecasts are presented in Figures 3. As inflation displays stronger persistence, the AR forecast mirrors lagged inflation. The factors add information beyond what is in lagged inflation, reducing the MSE from 5.09 to 3.98. The data along with the 95% prediction interval are given in Figure 4. The prediction interval for the diffusion index forecasts are again tighter, with an average length of 5.19 compared to 7.41.

A notable feature of the two applications considered is the reduced adherence of the factor forecasts to the lags of the data, even when the autoregressive structure is built in. This illustrates that diffusion index forecasts add information in the large panel not contained in the history of the series itself, and in a very parsimonious way.

7 Non-Stationary Factors

The preceding analysis can be extended to nonstationary factors. Although nonstationary factors imply different rates of convergence for the estimated model parameters, we will now show that for the purpose of constructing confidence intervals for forecasts, the formula for stationary factors remains valid, at least under conditional homoskedasticity.

Assume again that the forecasting equation is $y_{t+h} = \alpha' F_t + \beta' W_t + \varepsilon_{t+h}$, and the data have a factor representation $x_{it} = \lambda_i' F_t + e_{it}$. Instead of assuming F_t is covariance stationary, we now assume

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

where u_t is a sequence of I(0) processes. To analyze this case of non-stationary factors, all previous assumptions are maintained, except for the following:

Assumption A': (1) $E\|u_t\|^{4+\delta} \le M$ and $\frac{1}{T^2}\sum_{t=1}^{T} F_t F_t' \xrightarrow{d} \Sigma_F$, where Σ_F is positive definite (random) matrix with probability 1, and (2) ε_t is an iid sequence with zero mean and variance σ_ε^2, where ε_s is independent of $z_t = (F_t', W_t')'$ for all t and s.

Assumption $A'(1)$ rules out cointegration among the components of F_t, although the results are applicable for this case. Cointegration among F_t is equivalent to the existence of both I(1) and I(0) factors, see Bai (2004). This case would require more complicated notation and will not be presented to simplify the exposition.

Assumption $A'(2)$ imposes conditional homoskedasticity on ε_t. As a result, the following mixture normality is a reasonable assumption:

$$D_T^{-1}\sum_{t=1}^{T} z_t \varepsilon_{t+h} \xrightarrow{d} MN(0, \sigma_\varepsilon^2 \Omega) \qquad (7)$$

where $MN(0, \sigma_\varepsilon^2\Omega)$ is shorthand notation for conditional normal distribution with covariance matrix $\sigma_\varepsilon^2\Omega$, conditional on Ω, where Ω is the limiting random matrix of $D_T^{-1} z'z D_T^{-1}$ where $D_T = TI_{r+p}$ if W_t is also I(1), and $D_T = (TI_r, \sqrt{T}I_p)$ if W_t is I(0). If some components of W_t are I(1), and others are I(0), D_T is adjusted accordingly. By definition, if $\xi \sim MN(0, \sigma_\varepsilon^2\Omega)$, then $\sigma_\varepsilon^{-1}\Omega^{-1/2}\xi \sim N(0, I)$.

Let $\widetilde{F}$ be a $T \times r$ matrix consisting of r eigenvectors (multiplied by T) of the matrix $XX'/(T^2N)$, corresponding to the first r largest eigenvalues (in deceasing order). Let $\widetilde{V}$ be the diagonal matrix consisting of these eigenvalues. Define $\widetilde{\Lambda} = X'\widetilde{F}/T^2$ and $H = \widetilde{V}^{-1}(\widetilde{F}'F/T^2)(\Lambda'\Lambda/N)$.

Theorem 5 *Suppose assumptions A', B-E and (7) hold.*

(i) Let $\widehat{\alpha}$ and $\widehat{\beta}$ be the least squares estimators from a regression of y_{t+h} on $\widehat{z}_t = (\widetilde{F}_t' \ W_t')'$. As $N, T \to \infty$ with $\sqrt{T}/N \to 0$,

$$(D_T^{-1}\widehat{z}'\widehat{z}D_T^{-1})^{1/2} D_T \left(\begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix} - \begin{bmatrix} H'^{-1}\alpha \\ \beta \end{bmatrix} \right) \xrightarrow{d} N\left(0, \sigma_\varepsilon^2 I\right) \qquad (8)$$

where $\widehat{z} = (\widehat{z}_1, ..., \widehat{z}_{T-h})'$.

(ii) Let $\widehat{y}_{T+h|T} = \widehat{\alpha}'\widetilde{F}_T + \widehat{\beta}'W_T$ be the feasible h-step ahead forecast of y_{T+h}. Under the assumptions of Theorem 5

$$\frac{\widehat{y}_{T+h|T} - y_{T+h|T}}{C_T} \xrightarrow{d} N(0,1) \qquad (9)$$

where $C_T^2 = \widehat{\sigma}_\varepsilon^2 \widehat{z}_T' (\widehat{z}'\widehat{z})^{-1}\widehat{z}_T + \frac{1}{N}\widehat{\alpha}' \widetilde{V}^{-1}\widetilde{\Gamma}_t\widetilde{V}^{-1}\widehat{\alpha}$.

The theorem shows that $\widehat{\alpha}$ converges to $H'^{-1}\alpha$ at rate T and $\widehat{\beta}$ converges to β at rate $\sqrt{T}$ when W_t is I(0). These are the same rates as known F. Of course, for known F, we will directly estimate α instead of $H'^{-1}\alpha$. When the estimator is weighted by the random matrix $(D_T^{-1}\widetilde{z}'\widetilde{z}D_T^{-1})^{1/2}$, the limiting distribution is normal. The unweighted limiting distribution is mixture normal.

The forecast error variance once again has two components. The first term of C_T^2 comes from the estimation of δ and is $O_p(T^{-1})$. The second term comes from the estimation of F_t and is $O_p(N^{-1})$. If T/N is bounded, both errors remain asymptotically (unless $T/N \to 0$) and the convergence rate is $\sqrt{T}$. If T/N is unbounded, asymptotic normality continues to hold, but convergence is at rate $\sqrt{N}$. The overall convergence rate of $\widehat{y}_{T+h|T}$ to $y_{T+h|T}$ is $\min[\sqrt{N}, \sqrt{T}]$, as in the case of I(0) regressors.

If F_t is observed, it is known that it has to be normalized differently depending on whether it is $I(1)$ or $I(0)$[6]. Although less obvious, the triple $(\widetilde{V}, \widetilde{F}, \widetilde{\Lambda})$ also has to be normalized differently, depending on the stationarity property of $\widetilde{F}_t$. One would then expect confidence intervals for stationary and non-stationary factors to be constructed differently. However, the expression $\frac{(\widehat{y}_{T+h|T} - y_{T+h|T})}{B_T}$ in Theorem 3 under homoskedasticity and $\frac{(\widehat{y}_{T+h|T} - y_{T+h|T})}{C_T}$ in Theorem 5 are in fact mathematically identical. As shown in the Appendix, this is because C_T^2 is invariant to normalization. Although Theorem 5 is stated under the assumption of conditional homoskedasticity, the forecast confidence intervals derived for stationary common factors are also valid for nonstationary factors. The practical implication is that knowledge concerning the stationarity property of F_t is not essential for predictive inference.

8 Conclusion

The factor approach to forecasting is extremely useful in situations when a large number of indicator or predictor variables are present. The factors provide a significant reduction in the number of variables entering the forecasting equation while exploiting information in all available data. This latter aspect is important because it is by using information in all data available that permits consistent estimation of the factors. This paper contributes to the small but growing literature on factor forecasting by (i) showing that the conditional mean forecasts are $\min[\sqrt{N}, \sqrt{T}]$ consistent, and (ii) presenting formulas to permit predictive inference. As a by product, we suggest how the covariance matrix of cross-correlated errors can be consistently estimated.

[6]Different scalings are used to derive proper rates of convergence and suitable limiting distributions.

Appendix

We make use of the following identity throughout:

$$\widetilde{F}_t - HF_t = \widetilde{V}^{-1}\Big(\frac{1}{T}\sum_{s=1}^{T}\widetilde{F}_s\gamma_{st} + \frac{1}{T}\sum_{s=1}^{T}\widetilde{F}_s\zeta_{st} + \frac{1}{T}\sum_{s=1}^{T}\widetilde{F}_s\eta_{st} + \frac{1}{T}\sum_{s=1}^{T}\widetilde{F}_s\xi_{st}\Big), \qquad \text{(A.1)}$$

where $\gamma_{st} = E(\frac{1}{N}\sum_{i=1}^{N} e_{is}e_{it})$, $\zeta_{st} = \frac{1}{N}\sum_{i=1}^{N} e_{is}e_{it} - \gamma_{st}$, $\eta_{st} = \frac{1}{N}\sum_{i=1}^{N}\lambda_i' F_s e_{it}$, and $\xi_{st} = \frac{1}{N}\sum_{i=1}^{N}\lambda_i' F_t\ e_{is}$. Note that M will represent a general positive constant, not depending on N and T and not necessarily the same in different expressions.

Lemma A1 *Let* $z_t' = (F_t' \quad W_t')'$, *and* $\widehat{z}_t = (\widetilde{F}_t' \quad W_t')'$. *Let* $\delta_{NT}^2 = \min[N,T]$. *Under Assumptions A-E,*
(i) $\delta_{NT}^2(\frac{1}{T}\sum_{t=1}^{T}\|\widetilde{F}_t - HF_t\|^2) = O_p(1)$;
(ii) $\frac{1}{T}\sum_{t=1}^{T}(\widetilde{F}_t - HF_t)z_t' = O_p(\delta_{NT}^{-2})$;
(iii) $\frac{1}{T}\sum_{t=1}^{T}(\widetilde{F}_t - HF_t)\widehat{z}_t' = O_p(\delta_{NT}^{-2})$;
(vii) $\frac{1}{T}\sum_{t=1}^{T}(\widetilde{F}_t - HF_t)\varepsilon_{t+h} = O_p(\delta_{NT}^{-2})$.

Proof: Part (i) is proved in Bai and Ng (2002). Consider (ii). ¿From A.1,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}(\widetilde{F}_t - HF_t)z_t' &= \widetilde{V}^{-1}\Big[T^{-2}\sum_{t=1}^{T}[\sum_{s=1}^{T}\widetilde{F}_s\gamma_{st}]z_t' \\ &\quad + T^{-2}\sum_{t=1}^{T}[\sum_{s=1}^{T}\widetilde{F}_s\zeta_{st}]z_t' + T^{-2}\sum_{t=1}^{T}[\sum_{s=1}^{T}\widetilde{F}_s\eta_{st}]z_t' + T^{-2}\sum_{t=1}^{T}[\sum_{s=1}^{T}\widetilde{F}_s\xi_{st}]z_t'\Big] \\ &= \widetilde{V}^{-1}[I + II + III + IV],\end{aligned}$$

We begin with I. We have

$$T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_s z_t'\gamma_{st} = T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\widetilde{F}_s - HF_s)z_t'\gamma_{st} + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}HF_s z_t'\gamma_{st}.$$

The first term is bounded by

$$T^{-1/2}\Big(\frac{1}{T}\sum_{s=1}^{T}\|\widetilde{F}_s - HF_s\|^2\Big)^{1/2}\Big(T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}|\gamma_{st}|^2 T^{-1}\sum_{t=1}^{T}\|z_t\|^2\Big)^{1/2} = O_p(T^{-1/2}\delta_{NT}^{-1})$$

by part (i) and Assumption C. Note that Assumption C implies $|\gamma_{st}| \le M$, $\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}|\gamma_{st}| \le M$ and $\frac{1}{T}\sum_{s=1}^{T}\sum_{s=1}^{T}|\gamma_{st}|^2 \le M$. The expected value of the second term is bounded by (ignore H)

$$T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}|\gamma_{st}|(E\,\|F_s\|^2)^{1/2}(E\,\|z_t\|^2)^{1/2} \le MT^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}|\gamma_{st}| = O(T^{-1})$$

by Assumption C and E.1. Thus, $(I) = O_p(T^{-1/2}\delta_{NT}^{-1})$.

For (II),

$$T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_s\zeta_{st}z_t' = T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}HF_s\zeta_{st}z_t' + T^{-2}\sum_{t=1}^{T}(\widetilde{F}_s - HF_s)\zeta_{st}z_t'.$$

The first term can be written as $H\frac{1}{\sqrt{NT}}\frac{1}{T}\sum_{t=1}^{T}m_t z_t'$, where $m_t = \frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s[e_{is}e_{it} - E(e_{is}e_{it})]$. But $E\|m_t\|^2 < M$ by Assumptions C3, and $E\|m_t z_t'\| \le (E(\|m_t\|^2 E(\|z_t\|^2))^{1/2} \le M$. Thus, $\frac{1}{T}\sum_{t=1}^{T}m_t z_t' = O_p(1)$, and the first term is $O_p(1/\sqrt{NT})$. For the second term,

$$\left\|T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\widetilde{F}_s - HF_s)\zeta_{st}z_t'\right\| \le \Big(\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{F}_s - HF_s\right\|^2\Big)^{1/2}\Big(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{T}\sum_{t=1}^{T}\zeta_{st}z_t'\right\|^2\Big)^{1/2}.$$

But $\frac{1}{T}\sum_{t=1}^{T}\zeta_{st}z_t' = \frac{1}{\sqrt{N}}\frac{1}{T}\sum_{t=1}^{T}(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it} - E(e_{is}e_{it})])z_t' = O_p(N^{-1/2})$. Combining the results, $(II) = O_p(1/\sqrt{NT}) + O_p(\delta_{NT}^{-1})\cdot O_p(N^{-1/2}) = O_p(N^{-1/2}\delta_{NT}^{-1})$.

For (III), we have

$$T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{F}_s z_t'\eta_{st} = T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}HF_s z_t'\eta_{st} + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\widetilde{F}_s - HF_s)z_t'\eta_{st}.$$

The first term on the right hand side can be rewritten as

$$T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}HF_s z_t'\eta_{st} = H(\frac{1}{T}\sum_{s=1}^{T}F_sF_s')\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\lambda_i z_t' e_{it},$$

which is $O_p(1)O_p(\frac{1}{\sqrt{NT}})$. The treatment of the second term is similar to that of the second term of (II). The proof for (IV) is similar to (III). Thus,

$$I + II + III + IV = O_p(\frac{1}{\sqrt{T}\delta_{NT}}) + O_p(\frac{1}{\sqrt{N}\delta_{NT}}) + O_p(\frac{1}{\sqrt{NT}}) = O_p(\frac{1}{\min[N,T]}) = O_p(\delta_{NT}^{-2})$$

proving part (ii). Next, consider part (iii). Let $\bar{z}_t = (HF_t', W_t')'$. Then $T^{-1}\sum_{t=1}^{T}(\widetilde{F}_t - HF_t)\widehat{z}_t' = T^{-1}\sum_{t=1}^{T}(\widetilde{F}_t - HF_t)\bar{z}_t' + T^{-1}\sum_{t=1}^{T}(\widetilde{F}_t - HF_t)(\widehat{z}_t - \bar{z}_t)'$. ¿From $\widehat{z}_t - \bar{z}_t = ((\widetilde{F}_t - HF_t)', 0)'$, the second term is $O_p(\delta_{NT}^{-2})$ by part (i). The first term is $O_p(\delta_{NT}^{-2})$ by part (iii) in view of the definition of $\bar{z}_t$ and z_t. Finally, the proof for (iv) is similar to (ii), with ε_t replacing z_t.

Proof of Theorem 1

Adding and subtracting terms, the forecasting model can be written as:

$$\begin{aligned} y_{t+h} &= \alpha' F_t + \beta' W_t + \varepsilon_{t+h} \\ &= \alpha' H^{-1}\widetilde{F}_t + \beta' W_t + \varepsilon_{t+h} + \alpha' H^{-1}(HF_t - \widetilde{F}_t). \end{aligned}$$

This implies, for $Y = (y_h, y_{h+1}, ..., y_T)'$, $\varepsilon = (\varepsilon_h, ..., \varepsilon_T)'$, and $\widehat{z} = (\widehat{z}_1, ..., \widehat{z}_{T-h})'$,

$$Y = \widehat{z} \begin{bmatrix} H^{-1\prime}\alpha \\ \beta \end{bmatrix} + \varepsilon + (FH' - \widetilde{F})H^{-1\prime}\alpha. \tag{A.2}$$

Consider the regression $y_{t+h} = \alpha'\widetilde{F}_t + \beta' W_t + error$. The least squares estimates are

$$\begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix} = (\widehat{z}'\widehat{z})^{-1}\widehat{z}'Y.$$

Replacing Y by the right hand side of (A.2)

$$\begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix} - \begin{bmatrix} H^{-1\prime}\alpha \\ \beta \end{bmatrix} = (\widehat{z}'\widehat{z})^{-1}\widehat{z}'\varepsilon + (\widehat{z}'\widehat{z})^{-1}\widehat{z}'(FH' - \widetilde{F})H^{-1}\alpha.$$

Or

$$\sqrt{T}\left(\begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix} - \begin{bmatrix} H^{-1\prime}\alpha \\ \beta \end{bmatrix}\right) = \left(\frac{\widehat{z}'\widehat{z}}{T}\right)^{-1}\frac{\widehat{z}'\varepsilon}{\sqrt{T}} + \left(\frac{\widehat{z}'\widehat{z}}{T}\right)^{-1}\frac{\widehat{z}'(FH' - \widetilde{F})}{\sqrt{T}}H^{-1}\alpha.$$

The second term on the right hand side is $o_p(1)$. This follows from $T^{-1/2}\widehat{z}'(FH' - \widetilde{F}) = O_p(T^{1/2}\delta_{NT}^{-2}) = O_p(T^{1/2}/\min(N,T)) = o_p(1)$ if $\sqrt{T}/N \to 0$, by Lemma A1. Consider the first term.

$$\frac{\widehat{z}'\varepsilon}{\sqrt{T}} = \begin{bmatrix} \frac{\widetilde{F}'\varepsilon}{\sqrt{T}} \\ \frac{W'\varepsilon}{\sqrt{T}} \end{bmatrix} = \begin{bmatrix} \frac{(\widetilde{F}-HF')\varepsilon}{\sqrt{T}} + \frac{HF'\varepsilon}{\sqrt{T}} \\ \frac{W'\varepsilon}{\sqrt{T}} \end{bmatrix}.$$

By Lemma A1, $\frac{(\widetilde{F}-FH')'\varepsilon}{\sqrt{T}} \xrightarrow{p} 0$ if $\sqrt{T}/N \to 0$. Therefore,

$$\begin{aligned} \sqrt{T}\left(\begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix} - \begin{bmatrix} H^{-1\prime}\alpha \\ \beta \end{bmatrix}\right) &= \left(\frac{\widehat{z}'\widehat{z}}{T}\right)^{-1}\begin{bmatrix} \frac{HF'\varepsilon}{\sqrt{T}} \\ \frac{W'\varepsilon}{\sqrt{T}} \end{bmatrix} + o_p(1) \\ &= \left(\frac{\widehat{z}'\widehat{z}}{T}\right)^{-1}\begin{bmatrix} H & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} F'\varepsilon \\ W'\varepsilon \end{bmatrix}\frac{1}{\sqrt{T}} + o_p(1) \\ &= \left(\frac{\widehat{z}'\widehat{z}}{T}\right)^{-1}\begin{bmatrix} H & 0 \\ 0 & I \end{bmatrix} z'\varepsilon/\sqrt{T} + o_p(1). \end{aligned}$$

Since $z'\varepsilon/\sqrt{T} \xrightarrow{d} N(0, \text{plim}\,\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t+h}^2 z_t z_t')$ by Assumption D2, the above is asymptotically normal. The asymptotic variance matrix is the probability limit of

$$\left(\frac{\widehat{z}'\widehat{z}}{T}\right)^{-1}\begin{bmatrix} H & 0 \\ 0 & I \end{bmatrix}\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t+h}^2 z_t z_t'\right)\begin{bmatrix} H' & 0 \\ 0 & I \end{bmatrix}\left(\frac{\widehat{z}'\widehat{z}}{T}\right)^{-1}.$$

Since $HF_t = \widetilde{F}_t + o_p(1)$ and $z_t = (F_t', W_t')'$, the product of the middle three matrices is $(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t+h}^2\widehat{z}_t\widehat{z}_t') + o_p(1)$. The asymptotic variance is thus given by (3), proofing Theorem 1.

Proof of Theorem 2

Consider augmenting an q variable VAR in y_t with r factors. Without loss of generality, consider a FVAR(1). Define $z_t = (y_t' \;\; F_t')'$. The infeasible FVAR is $z_{t+1} = Az_t + \varepsilon_{t+1}$, or

$$\begin{pmatrix} y_{t+1} \\ F_{t+1} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} y_t \\ F_t \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t+1} \\ \varepsilon_{2t+1} \end{pmatrix}.$$

Left multiplying the second block equations by H and then adding and subtracting terms, the FVAR expressed in terms of $\widetilde{F}_t$ is

$$\begin{aligned} \begin{pmatrix} y_{t+1} \\ \widetilde{F}_{t+1} \end{pmatrix} &= \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} y_t \\ \widetilde{F}_t \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t+1} \\ H\varepsilon_{2t+1} \end{pmatrix} + \begin{pmatrix} -b_{12}(HF_t - \widetilde{F}_t) \\ b_{21}(HF_t - \widetilde{F}_t) \end{pmatrix} + \begin{pmatrix} 0_{m\times r} \\ -(HF_{t+1} - \widetilde{F}_{t+1}) \end{pmatrix} \\ &= \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} y_t \\ \widetilde{F}_t \end{pmatrix} + u^1_{t+1} + u^2_{t+1} + u^3_{t+1} \end{aligned}$$

where $b_{11} = a_{11}$, $b_{12} = a_{12}H^{-1}$, $b_{21} = Ha_{21}$, and $b_{22} = Ha_{22}H^{-1}$. Let $\widehat{z}_t = (y_t', \; \widetilde{F}_t')'$. The j-th equation of the feasible FVAR is thus

$$\widehat{z}_{jt+1} = \delta_j'\widehat{z}_t + u^1_{jt+1} + u^2_{jt+1} + u^3_{jt+1}.$$

The least squares estimator for δ_j is

$$\sqrt{T}(\widehat{\delta}_j - \delta_j) = \Big(\frac{1}{T}\sum_{t=1}^T \widehat{z}_t\widehat{z}_t'\Big)^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^T \widehat{z}_t(u^1_{jt+1} + u^2_{jt+1} + u^3_{jt+1})\Big).$$

By Lemma A1, $\frac{1}{\sqrt{T}}\sum_{t=1}^T \widehat{z}_t u^2_{jt+1} = O_p(\frac{\sqrt{T}}{\min[N,T]})$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^T \widehat{z}_t u^3_{jt+1} = O_p(\frac{\sqrt{T}}{\min[N,T]})$. Thus,

$$\begin{aligned} \sqrt{T}(\widehat{\delta}_j - \delta_j) &= \Big(\frac{1}{T}\sum_{t=1}^T \widehat{z}_t\widehat{z}_t'\Big)^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^T \widehat{z}_t u^1_{jt+1}\Big) + o_p(1) \\ &\xrightarrow{d} N\Big(0, \text{plim}(\frac{1}{T}\sum_{t=1}^T \widehat{z}_t\widehat{z}_t')^{-1}\Big(\frac{1}{T}\sum_{t=1}^T (u^1_{jt+1})^2\widehat{z}_t\widehat{z}_t'\Big)(\frac{1}{T}\sum_{t=1}^T \widehat{z}_t\widehat{z}_t')^{-1}\Big). \end{aligned}$$

This can be consistently estimated with upon replacing u^1_{jt+1} by $\widehat{u}^1_{jt+1} = \widehat{z}_{jt+1} - \widehat{\delta}_j'\widehat{z}_t$.

Proof of Theorem 3

Begin by rewriting

$$\begin{aligned} \widehat{y}_{T+h|T} - y_{T+h|T} &= \widehat{\alpha}'\widetilde{F}_T + \widehat{\beta}'W_T - \alpha'F_T - \beta'W_T \\ &= (\widehat{\alpha} - H^{-1\prime}\alpha)'\widetilde{F}_T + \alpha'H^{-1}(\widetilde{F}_T - HF_T) + (\widehat{\beta} - \beta)'W_T. \end{aligned}$$

Equivalently,

$$\begin{aligned}
\widehat{y}_{T+h|T} - y_{T+h|T} &= \begin{bmatrix} \widehat{\alpha} - H^{-1\prime}\alpha \\ \widehat{\beta} - \beta \end{bmatrix}' \begin{bmatrix} \widetilde{F}_T \\ W_T \end{bmatrix} + \alpha' H^{-1}(\widetilde{F}_T - HF_T) \\
&= \widetilde{z}_T'(\widehat{\delta} - \delta) + \alpha' H^{-1}(\widetilde{F}_T - HF_T) \\
&= \frac{1}{\sqrt{T}} \widetilde{z}_T' \sqrt{T}(\widehat{\delta} - \delta) + \frac{1}{\sqrt{N}} \alpha' H^{-1} \sqrt{N}(\widetilde{F}_T - HF_T)
\end{aligned}$$

Thus, if T/N is bounded, $\sqrt{T}(\widehat{y}_{T+h|T} - y_{T+h|T}) = O_p(1)$ and is asymptotically normal because $\sqrt{T}(\widehat{\delta}-\delta)$ and $\sqrt{N}(\widehat{F}_T - HF_T)$ are asymptotically normal. Similarly, if N/T is bounded, then $\sqrt{N}(y_{T+h|T} - y_{T+h|T}) = O_p(1)$ and is asymptotically normal. Furthermore, note that $\sqrt{T}(\widehat{\delta} - \delta)$ and $\sqrt{N}(\widetilde{F}_T - HF_T)$ are asymptotically independent because the limiting distribution of $\sqrt{T}(\widehat{\delta} - \delta)$ is determined by $(\varepsilon_1, ..., \varepsilon_T)$ and the limiting distribution of $\sqrt{N}(\widetilde{F}_T - HF_T)$ is determined by cross-section disturbances at period T, e_{iT} for $i = 1, 2..., N$. Due to this asymptotic independence, the sum of the variances of the right hand side terms is an estimate for the variance of $\widehat{y}_{T+h|T} - y_{T+h|T}$. Let $B_T^2 = \frac{1}{T}\widetilde{z}_T' Avar(\widehat{\delta})\widehat{z}_T + \frac{1}{N}\widehat{\alpha}' Avar(\widetilde{F}_T)\widehat{\alpha}$, which is an estimate for the variance of $\widehat{y}_{T+h|T} - y_{T+h|T}$. Thus $(\widehat{y}_{T+h|T} - y_{T+h|T})/B_T \xrightarrow{d} N(0,1)$.

Proof of Theorem 4

Let $\sigma_{ij} = E(e_{it}e_{jt})$, and $\widetilde{\sigma}_{ij} = \frac{1}{T}\sum_{t=1}^T \widetilde{e}_{it}\widetilde{e}_{jt}$. Recall,

$$\Gamma_t = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}\lambda_i\lambda_j'.$$

The proposed estimator is $\widetilde{\Gamma}_t = \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n \widetilde{\sigma}_{ij}\widetilde{\lambda}_i\widetilde{\lambda}_j$. Also let $\bar{\Gamma}_t = \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n \widetilde{\sigma}_{ij}\lambda_i\lambda_j'$. It follows that

$$\widetilde{\Gamma}_t - H^{-1\prime}\Gamma_t H^{-1} = \widetilde{\Gamma}_t - H^{-1\prime}\bar{\Gamma}_t H^{-1} + H^{-1\prime}(\bar{\Gamma}_t - \Gamma_t)H^{-1}.$$

We will show (i) that $\bar{\Gamma}_t - \Gamma_t \xrightarrow{p} 0$ if $\frac{n}{N} \to 0$ and $\frac{n}{T} \to 0$, and (ii) that $\widetilde{\Gamma}_t - H^{-1\prime}\bar{\Gamma}_t H^{-1} = O_p(T^{-1/2}) + O_p(\min[N,T]^{-1})$.

(i) $\bar{\Gamma}_t - \Gamma_t \xrightarrow{p} 0$**.** ¿From $\widetilde{e}_{it} = x_{it} - \widetilde{c}_{it}$ and $e_{it} = x_{it} - c_{it}$, where $c_{it} = \lambda_i' F_t$ and $\widetilde{c}_{it} = \widetilde{\lambda}_i'\widetilde{F}_t$, we have $\widetilde{e}_{it} = e_{it} - (c_{it} - \widetilde{c}_{it})$. Thus,

$$\widetilde{e}_{it}\widetilde{e}_{jt} = e_{it}e_{jt} - e_{it}(c_{jt} - \widetilde{c}_{jt}) - e_{jt}(c_{it} - \widetilde{c}_{it}) + (c_{it} - \widetilde{c}_{it})(c_{jt} - \widetilde{c}_{jt}).$$

It follows that

$$\begin{aligned}\bar{\Gamma}_t - \Gamma_t &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{T}\sum_{t=1}^{T}(e_{it}e_{jt}-\sigma_{ij})\lambda_i\lambda_j' - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{T}\sum_{t=1}^{T}e_{it}(c_{jt}-\widetilde{c}_{jt})\lambda_i\lambda_j' \\ &- \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{T}\sum_{t=1}^{T}e_{jt}(c_{it}-\widetilde{c}_{it})\lambda_i\lambda_j' + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{T}\sum_{t=1}^{T}(c_{it}-\widetilde{c}_{it})(c_{jt}-\widetilde{c}_{jt})\lambda_i\lambda_j' \\ &= I + II + III + IV.\end{aligned}$$

We will now show that $I\xrightarrow{p}0$ as $T\to\infty$; II and III tend to zero if $\sqrt{n}/T\to 0$; IV tends to zero if $n/T\to 0$ and $n/N\to 0$.

We begin with I. Define $\xi_t = n^{-1/2}\sum_{i=1}^{n}\lambda_i e_{it}$. Then $I=\frac{1}{T}\sum_{t=1}^{T}[\xi_t\xi_t' - E(\xi_t\xi_t')]$. Each element of the $r\times r$ matrix $\xi_t\xi_t' - E(\xi_t\xi_t')$ is a zero mean process, thus each entry of I is $O_p(T^{-1/2})$.

Now consider II. Rewrite $c_{jt}-\widetilde{c}_{jt} = (H^{-1\prime}\lambda_j - \widetilde{\lambda}_j)'\widetilde{F}_t + \lambda_j' H^{-1}(HF_t - \widetilde{F}_t)$. We will use the fact that each term is a scalar and thus equals to its transpose and is commutable with any vector or matrix and hence λ_i. Rewrite II accordingly,

$$\begin{aligned} II &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{T}\sum_{t=1}^{T}e_{it}(H^{-1\prime}\lambda_j-\widetilde{\lambda}_j)'\widetilde{F}_t\lambda_i\lambda_j' + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{T}\sum_{t=1}^{T}e_{it}(HF_t-\widetilde{F}_t)'H^{-1\prime}\lambda_j\lambda_i\lambda_j{}' \\ &= A+B \\ &= \Big(\frac{1}{n}\sum_{i=1}^{n}\lambda_i\frac{1}{T}\sum_{t=1}^{T}e_{it}\widetilde{F}_t'\Big)\Big(\sum_{j=1}^{n}(H^{-1\prime}\lambda_j-\widetilde{\lambda}_j)\lambda_j'\Big) + B \\ &= (A.a)(A.b) + B.\end{aligned}$$

Now $\|A.a\| = \Big\|\frac{1}{T}\sum_{t=1}^{T}(\frac{1}{n}\sum_{i=1}^{n}\lambda_i e_{it})\widetilde{F}_t\Big\|$. Thus,

$$\|A.a\| \le \Big(\frac{1}{T}\sum_{t=1}^{T}\Big\|\frac{1}{n}\sum_{i=1}^{n}\lambda_i e_{it}\Big\|^2\Big)^{1/2}\Big(\frac{1}{T}\sum_{t=1}^{T}\Big\|\widetilde{F}_t\Big\|^2\Big)^{1/2} = O_p(n^{-1/2})\cdot O_p(1)$$

because $\frac{1}{n}\sum_{i=1}^{n}\lambda_i e_{it} = O_p(n^{-1/2})$ and $\frac{1}{T}\sum_{t=1}^{T}\Big\|\widetilde{F}_t\Big\|^2 = O_p(1)$. For A.b, by Lemma A2 below

$$\|A.b\| = \Big\|n\frac{1}{n}\sum_{j=1}^{n}(H^{-1\prime}\lambda_j-\widetilde{\lambda}_j)\lambda_j'\Big\| = n\Big[\cdot O_p(\frac{1}{\sqrt{nT}}) + O_p(\frac{1}{\min[N,T]})\Big].$$

It follows from A=(A.a)(A.b) that

$$A = O_p(n^{-1/2})n\Big[O_p(\frac{1}{\sqrt{nT}}) + O_p(\frac{1}{\min[N,T]})\Big] = O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{\sqrt{n}}{\min[N,T]}) \to 0$$

if $\sqrt{n}/T \to 0$.

For B, it is bounded in norm by

$$\Big\|\frac{1}{T}\sum_{t=1}^{T}(\sum_{i=1}^{n}\lambda_i e_{it})(HF_t-\widetilde{F}_t)'\Big\|\Big(\frac{1}{n}\sum_{j=1}^{n}\|\lambda_j\|^2\Big)\|H\| = O_p(n^{1/2}\delta_{NT}^{-2})O_p(1)$$

by Lemma A2(ii) below. Thus, $B \to 0$ if $\sqrt{n}/T \to 0$. Analogously, $III \to 0$ if $\sqrt{n}/T \to 0$.

For IV, note first that this term can be written as

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{T}\sum_{t=1}^{T}(c_{it}-\widetilde{c}_{it})(c_{jt}-\widetilde{c}_{jt})\lambda_i\lambda_j' = \frac{1}{T}\sum_{t=1}^{T}\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(c_{it}-\widetilde{c}_{it})\lambda_i\Big\|^2.$$

Using $c_{it}-\widetilde{c}_{it} = (H^{-1\prime}\lambda_i-\widetilde{\lambda}_i)'\widetilde{F}_t + \lambda_i' H^{-1}(HF_t-\widetilde{F}_t)$, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(c_{it}-\widetilde{c}_{it})\lambda_i = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(H^{-1\prime}\lambda_i-\widetilde{\lambda}_i)'\widetilde{F}_t\lambda_i + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\lambda_i'H^{-1}(HF_t-\widetilde{F}_t)\lambda_i,$$

and

$$\begin{aligned}\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(c_{it}-\widetilde{c}_{it})\lambda_i\Big\|^2 \;&\le\; 2\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\lambda_i(H^{-1\prime}\lambda_i-\widetilde{\lambda}_i)'\Big\|^2\|\widetilde{F}_t\|^2 \\ &\quad +2\|H^{-1}\|^2\Big(\frac{1}{n}\sum_{i=1}^{n}\|\lambda_i\|^2\Big)^2\cdot n\cdot\|F_t-HF_t\|^2\end{aligned}$$

Thus

$$\begin{aligned} IV \;&\le\; 2\Big(\frac{1}{T}\sum_{t=1}^{T}\Big\|\widetilde{F}_t\Big\|^2\Big)\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\lambda_i(H^{-1\prime}\lambda_i-\widetilde{\lambda}_i)'\Big\|^2 \\ &\quad +2\|H^{-1}\|^2\Big(\frac{1}{n}\sum_{i=1}^{n}\|\lambda_i\|^2\Big)^2\cdot n\cdot\frac{1}{T}\sum_{t=1}^{T}\Big\|HF_t-\widetilde{F}_t\Big\|^2 \\ &= O_p(1)\Big\|\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\lambda_i(H^{-1\prime}\lambda_i-\widetilde{\lambda}_i)'\Big\|^2 + O_p(n)O_p(\min[N,T]^{-1})O_p(1) \\ &= a+b\end{aligned}$$

By Lemma A2 below, $a \to 0$ if $\sqrt{n}/T \to 0$. Furthermore, $b \to 0$ if $n/N \to 0$ and $n/T \to 0$.□

ii. $\widetilde{\Gamma}_t - H^{-1\prime}\bar{\Gamma}_t H^{-1} \xrightarrow{p} 0$**.** By the definition of $\widetilde{\Gamma}_t$ and $\bar{\Gamma}_t$, we have

$$\begin{aligned}
\widetilde{\Gamma}_t - H^{-1\prime}\bar{\Gamma}_t H^{-1} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\widetilde{\sigma}_{ij}(\widetilde{\lambda}_i\widetilde{\lambda}_j' - H^{-1\prime}\lambda_i\lambda_j' H^{-1}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(\widetilde{\sigma}_{ij}-\sigma_{ij})(\widetilde{\lambda}_i\widetilde{\lambda}_j' - H^{-1\prime}\lambda_i\lambda_j' H^{-1}) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}(\widetilde{\lambda}_i\widetilde{\lambda}_j' - H^{-1\prime}\lambda_i\lambda_j' H^{-1}) \\
&= I + II.
\end{aligned}$$

We begin with II. Now

$$\widetilde{\lambda}_i\widetilde{\lambda}_j' - H^{-1\prime}\lambda_i\lambda_j H^{-1} = (\widetilde{\lambda}_i - H^{-1\prime}\lambda_i)\widetilde{\lambda}_j' + H^{-1\prime}\lambda_i(\widetilde{\lambda}_j - H^{-1\prime}\lambda_j)'.$$

Thus,

$$\begin{aligned}
II &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}(\widetilde{\lambda}_i - H^{-1\prime}\lambda_i)\widetilde{\lambda}_j' + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\lambda_i H^{-1}(\widetilde{\lambda}_j - H^{-1\prime}\lambda_j)' \\
&= a + b.
\end{aligned}$$

By Lemma A3 below,

$$\begin{aligned}
|a| &\leq \left(\frac{1}{n}\sum_{j=1}^{n}\left\|\widetilde{\lambda}_j\right\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{j=1}^{n}\left\|\sum_{i=1}^{n}\sigma_{ij}(\widetilde{\lambda}_i - H^{-1\prime}\lambda_i)\right\|^2\right)^{1/2} \\
&= O_p(1)\Big[O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{\min[N,T]})\Big] \to 0.
\end{aligned}$$

Similarly, $b = O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{\min[N,T]})$. The proof of I being $o_p(1)$ is analogous to that of part (i). This completes the proof of Theorem 4. □.

Lemma A2 *(i)* $\frac{1}{n}\sum_{j=1}^{n}(H^{-1\prime}\lambda_i - \widetilde{\lambda}_i)\lambda_i' = O_p((nT)^{-1/2}) + O_p(\min[N,T]^{-1})$.
(ii) The $r \times r$ *matrix* $\frac{1}{T}\sum_{t=1}^{T}[(HF_t - \widetilde{F}_t)(\sum_{i=1}^{n}\lambda_i' e_{it})] = O_p(\frac{\sqrt{n}}{\min[N,T]})$.

Proof of (i). From the identity

$$\widetilde{\lambda}_i - H^{-1\prime}\lambda_i = T^{-1}HF'\underline{e}_i + T^{-1}\widetilde{F}'(F - \widetilde{F}H^{-1\prime})\lambda_i + T^{-1}(\widetilde{F} - FH')'\underline{e}_i,$$

where $\underline{e}_i = (e_{i1}, e_{i2}, \ldots e_{iT})'$, we have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(\widetilde{\lambda}_i - H^{-1\prime}\lambda_i)\lambda_i' &= T^{-1}HF'(\frac{1}{n}\sum_{i=1}^{n}\underline{e}_i\lambda_i') + T^{-1}\widetilde{F}'(F - \widetilde{F}H^{-1\prime})(\frac{1}{n}\sum_{i=1}^{n}\lambda_i\lambda_i') \\
&\quad + T^{-1}(\widetilde{F} - FH')'(\frac{1}{n}\sum_{i=1}^{n}\underline{e}_i\lambda_i') = a + b + c.
\end{aligned}$$

Now (a) equals $H\frac{1}{Tn}(\sum_{i=1}^n \sum_{t=1}^T F_t\lambda_i' e_{it}) = O_p(\frac{1}{\sqrt{nT}})$. (b) equals $T^{-1}\widetilde{F}'(F - \widetilde{F}H^{-1\prime}) \cdot O_p(1) = O_p(\min[N,T]^{-1})$ by Lemma B.3 of Bai (2003). (c) is $O_p([\sqrt{n}\min[N,T]]^{-1})$ following Lemma B.1 of Bai (2003), replacing e_{it} with $\frac{1}{n}\sum_{i=1}^n \lambda_i e_{it} = O_p(\frac{1}{\sqrt{n}})$. The lemma follows since (c) is dominated by (a) and (b). For part (ii), the expression is equal to (c) multiplied by n, thus it is equal to $O_p(\sqrt{n}/\min[N,T]])$.

Lemma A3 *For each j, $\sum_{i=1}^n \sigma_{ij}(\widetilde{\lambda}_i - H^{-1\prime}\lambda_i) = O_p(T^{-1/2}) + O_p(\min[N,T]^{-1})$.*

Using the expression for $\widetilde{\lambda}_i - H^{-1\prime}\lambda_i$ above, we have

$$\begin{aligned}\sum_{i=1}^n \sigma_{ij}(\widetilde{\lambda}_i - H^{-1\prime}\lambda_i) &= T^{-1}H'(\sum_{i=1}^n \sigma_{ij}F'\underline{e}_i) \\ &\quad + T^{-1}\widetilde{F}'(F - \widetilde{F}H^{-1})(\sum_{i=1}^n \sigma_{ij}\lambda_i) + T^{-1}(\widetilde{F} - FH)'(\sum_{i=1}^n \sigma_{ij}\underline{e}_i) \\ &= (a) + (b) + (c).\end{aligned}$$

Now (a) is $O_p(T^{-1/2})$ because $\frac{1}{T}F'\underline{e}_i = \frac{1}{T}\sum_{t=1}^T F_t e_{it}$ is $O_p(T^{-1/2})$ for each i, and by Assumption C, $|\sum_{i=1}^n \sigma_{ij}| \le M$. (b) is $O_p(\min[N,T]^{-1})$ because $T^{-1}\widetilde{F}'(F - \widetilde{F}H^{-1}) = O_p(\min[N,T]^{-1})$ and $\|\sum_{i=1}^n \sigma_{ij}\lambda_i\| \le M$. (c) is $O_p(\min[N,T]^{-1})$ following the argument of Lemma A.2 (ii), replacing $\sum_{i=1}^n \lambda_i e_{it} = O_p(\sqrt{n})$ with $\sum_{i=1}^n \sigma_{ij}e_{it} = O_p(1)$. □

Proof of Theorem 5

The argument for Theorem 5 is almost identical to that of Theorems 1 and 3. The details are omitted. We next argue that it is not necessary to know if the underlying factors are I(0) or I(1), as far as prediction interval is concerned. The expression C_T^2 is equal to B_T^2 when (4b) is used in estimating $Avar(\widehat{\delta})$ of Theorem 3. Nevertheless, the triple $(\widetilde{V}, \widetilde{F}, \widetilde{\Lambda})$ in Theorem 5 are estimated (or are scaled) differently, depending on whether F_t is I(1) or I(0).[7] It might appear that it is essential to know the stationarity property of F_t. It turns out that C_T^2 is invariant to different scalings. First consider the first term of C_T^2, which is $\widehat{z}_T'(\widehat{z}'\widehat{z})^{-1}\widehat{z}_T$. From $\widehat{z}_t = (\widetilde{F}_t', W_t')'$, it is clear that $\widetilde{F}_t$ appears twice in the numerator and twice in the denominator, thus immune to scaling. Next consider $\widehat{\alpha}'\widetilde{V}^{-1}\widetilde{\Gamma}_t\widetilde{V}^{-1}\widehat{\alpha}$. Given a data matrix X, let $(\widetilde{V}^s, \widetilde{F}^s, \widetilde{\Lambda}^s)$ be the estimated triple assuming F_t to be I(0), and let $(\widetilde{V}^n, \widetilde{F}^n, \widetilde{\Lambda}^n)$ be the corresponding triple assuming F_t to be I(1). Then $(\widetilde{V}^n, \widetilde{F}^n, \widetilde{\Lambda}^n) = (\widetilde{V}^s/T, \sqrt{T}\widetilde{F}^s, \widetilde{\Lambda}^s/\sqrt{T})$, by the definition of the estimation procedures. This implies that $\widehat{\alpha}^n = \widehat{\alpha}^s/\sqrt{T}$ (note $\widehat{\alpha}^n$ is the

[7] Different scalings are used to derive proper rates of convergence and suitable limiting distributions.

estimated regression coefficient when $\widetilde{F}^n$ is the regressor, and likewise for $\widehat{\alpha}^s$). Furthermore, the panel residuals $\widetilde{e}_{it}$ are invariant to scalings because $\widetilde{F}^n\widetilde{\Lambda}^{n\prime}$ is equal to $\widetilde{F}^s\widetilde{\Lambda}^{s\prime}$, it follows that $\widetilde{\Gamma}_t^n = \widetilde{\Gamma}_t^s/T$ in view of $\widetilde{\lambda}_i^n = \widetilde{\lambda}_i^s/\sqrt{T}$, see equations (6a)-(6c). From these relationships, it is easy to see that

$$\widehat{\alpha}^{n\prime}(\widetilde{V}^n)^{-1}\widetilde{\Gamma}_t^n(\widetilde{V}^n)^{-1}\widehat{\alpha}^n = \widehat{\alpha}^{s\prime}(\widetilde{V}^s)^{-1}\widetilde{\Gamma}_t^s(\widetilde{V}^s)^{-1}\widehat{\alpha}^s.$$

Thus, C_T^2 is the same whether F_t is assumed to be I(0) or I(1). The above argument is valid for F_t being I(2) or other processes. This result has the practical implication that forecasting confidence intervals derived for I(0) common factors are valid for nonstationary factors.

Table 1: Coverage Rates and MSE:
$\widehat{Avar(\widehat{\delta})} = \widehat{\sigma}_{\varepsilon}^2 \left[\frac{1}{T}\sum_{t=1}^{T} \widehat{z}_t \widehat{z}_t'\right]^{-1}$,
$\widetilde{\Gamma}_t = \widetilde{\sigma}_e^2 \frac{1}{N} \sum_{i=1}^{N} \widetilde{\lambda}_i \widetilde{\lambda}_i \quad \forall\, t.$

				Coverage Probability				MSE			
N	T	b	k	$\widehat{y}_{T+h\|T}$	$\widehat{y}^0_{T+h\|T}$	$\widehat{y}_{T+h}$	$\widehat{y}^0_{T+h}$	$\widehat{y}_{T+h\|T}$	$\widehat{y}^0_{T+h\|T}$	$\widehat{y}_{T+h}$	$\widehat{y}^0_{T+h}$
50	50	0.00	2	0.94	0.93	0.93	0.92	0.15	0.09	1.17	1.15
100	50	0.00	2	0.94	0.92	0.94	0.94	0.12	0.09	1.09	1.07
200	50	0.00	2	0.95	0.92	0.93	0.93	0.09	0.08	1.16	1.16
50	100	0.00	2	0.95	0.92	0.94	0.94	0.10	0.04	1.17	1.09
50	200	0.00	2	0.96	0.94	0.96	0.95	0.07	0.02	1.07	1.03
200	100	0.00	2	0.96	0.94	0.95	0.94	0.05	0.04	1.07	1.07
100	200	0.00	2	0.96	0.94	0.95	0.94	0.04	0.02	1.04	1.02
200	200	0.00	2	0.95	0.94	0.95	0.95	0.03	0.02	1.03	1.03
100	400	0.00	2	0.97	0.95	0.96	0.96	0.03	0.01	0.95	0.91
50	50	0.50	2	0.91	0.93	0.94	0.92	0.23	0.09	1.22	1.15
100	50	0.50	2	0.93	0.92	0.94	0.94	0.16	0.09	1.12	1.07
200	50	0.50	2	0.94	0.92	0.93	0.93	0.10	0.08	1.16	1.16
50	100	0.50	2	0.93	0.92	0.94	0.94	0.15	0.04	1.24	1.09
50	200	0.50	2	0.94	0.94	0.96	0.95	0.13	0.02	1.14	1.03
200	100	0.50	2	0.96	0.94	0.95	0.94	0.06	0.04	1.08	1.07
100	200	0.50	2	0.95	0.94	0.95	0.94	0.07	0.02	1.09	1.02
200	200	0.50	2	0.96	0.94	0.96	0.95	0.04	0.02	1.03	1.03
100	400	0.50	2	0.97	0.95	0.96	0.96	0.06	0.01	0.99	0.91
50	50	0.00	1	0.55	0.93	0.90	0.92	1.13	0.09	2.22	1.15
100	50	0.00	1	0.52	0.92	0.92	0.94	0.99	0.09	1.97	1.07
200	50	0.00	1	0.53	0.92	0.93	0.93	0.90	0.08	2.01	1.16
50	100	0.00	1	0.52	0.92	0.94	0.94	1.05	0.04	2.14	1.09
50	200	0.00	1	0.50	0.94	0.94	0.95	0.93	0.02	2.01	1.03
200	100	0.00	1	0.44	0.94	0.94	0.94	1.01	0.04	2.11	1.07
100	200	0.00	1	0.43	0.94	0.94	0.94	0.94	0.02	2.03	1.02
200	200	0.00	1	0.38	0.94	0.95	0.95	0.90	0.02	1.80	1.03
100	400	0.00	1	0.40	0.95	0.96	0.96	0.86	0.01	1.78	0.91
50	50	0.50	1	0.57	0.93	0.91	0.92	1.15	0.09	2.24	1.15
100	50	0.50	1	0.54	0.92	0.92	0.94	1.00	0.09	1.99	1.07
200	50	0.50	1	0.53	0.92	0.93	0.93	0.91	0.08	2.02	1.16
50	100	0.50	1	0.55	0.92	0.93	0.94	1.09	0.04	2.21	1.09
50	200	0.50	1	0.53	0.94	0.94	0.95	0.96	0.02	2.05	1.03
200	100	0.50	1	0.47	0.94	0.94	0.94	1.01	0.04	2.11	1.07
100	200	0.50	1	0.45	0.94	0.93	0.94	0.96	0.02	2.07	1.02
200	200	0.50	1	0.39	0.94	0.96	0.95	0.91	0.02	1.81	1.03
100	400	0.50	1	0.43	0.95	0.96	0.96	0.86	0.01	1.79	0.91

Table 2: Coverage Rates and MSE:
$$\widehat{Avar(\widehat{\delta})} = \left(\frac{1}{T}\sum_{t=1}^{T}\widehat{z}_t\widehat{z}_t'\right)^{-1}\left[\frac{1}{T}\sum_{t=1}^{T-h}\widehat{\varepsilon}_{t+h}^2\widehat{z}_t\widehat{z}_t'\right]\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{z}_t\widehat{z}_t'\right)^{-1},$$
$$\widetilde{\Gamma}_t = \widetilde{\sigma}_e^2\frac{1}{N}\sum_{i=1}^{N}\widetilde{\lambda}_i\widetilde{\lambda}_i \quad \forall\, t.$$

				Coverage Probability				MSE			
N	T	b	k	$\widehat{y}_{T+h\|T}$	$\widehat{y}^0_{T+h\|T}$	$\widehat{y}_{T+h}$	$\widehat{y}^0_{T+h}$	$\widehat{y}_{T+h\|T}$	$\widehat{y}^0_{T+h\|T}$	$\widehat{y}_{T+h}$	$\widehat{y}^0_{T+h}$
50	50	0.00	2	0.92	0.85	0.93	0.92	0.15	0.09	1.17	1.15
100	50	0.00	2	0.92	0.85	0.94	0.94	0.12	0.09	1.09	1.07
200	50	0.00	2	0.94	0.86	0.93	0.92	0.09	0.08	1.16	1.16
50	100	0.00	2	0.93	0.89	0.94	0.94	0.10	0.04	1.17	1.09
50	200	0.00	2	0.93	0.91	0.96	0.95	0.07	0.02	1.07	1.03
200	100	0.00	2	0.95	0.90	0.95	0.94	0.05	0.04	1.07	1.07
100	200	0.00	2	0.94	0.92	0.95	0.94	0.04	0.02	1.04	1.02
200	200	0.00	2	0.94	0.92	0.95	0.95	0.03	0.02	1.03	1.03
100	400	0.00	2	0.95	0.94	0.96	0.96	0.03	0.01	0.95	0.91
50	50	0.50	2	0.88	0.85	0.94	0.92	0.23	0.09	1.22	1.15
100	50	0.50	2	0.91	0.85	0.94	0.94	0.16	0.09	1.12	1.07
200	50	0.50	2	0.93	0.86	0.93	0.92	0.10	0.08	1.16	1.16
50	100	0.50	2	0.92	0.89	0.94	0.94	0.15	0.04	1.24	1.09
50	200	0.50	2	0.92	0.91	0.96	0.95	0.13	0.02	1.14	1.03
200	100	0.50	2	0.95	0.90	0.95	0.94	0.06	0.04	1.08	1.07
100	200	0.50	2	0.92	0.92	0.94	0.94	0.07	0.02	1.09	1.02
200	200	0.50	2	0.94	0.92	0.96	0.95	0.04	0.02	1.03	1.03
100	400	0.50	2	0.94	0.94	0.96	0.96	0.06	0.01	0.99	0.91
50	50	0.00	1	0.51	0.85	0.90	0.92	1.13	0.09	2.22	1.15
100	50	0.00	1	0.50	0.85	0.92	0.94	0.99	0.09	1.97	1.07
200	50	0.00	1	0.51	0.86	0.93	0.92	0.90	0.08	2.01	1.16
50	100	0.00	1	0.48	0.89	0.94	0.94	1.05	0.04	2.14	1.09
50	200	0.00	1	0.46	0.91	0.94	0.95	0.93	0.02	2.01	1.03
200	100	0.00	1	0.42	0.90	0.94	0.94	1.01	0.04	2.11	1.07
100	200	0.00	1	0.40	0.92	0.94	0.94	0.94	0.02	2.03	1.02
200	200	0.00	1	0.34	0.92	0.95	0.95	0.90	0.02	1.80	1.03
100	400	0.00	1	0.34	0.94	0.96	0.96	0.86	0.01	1.78	0.91
50	50	0.50	1	0.52	0.85	0.90	0.92	1.15	0.09	2.24	1.15
100	50	0.50	1	0.52	0.85	0.92	0.94	1.00	0.09	1.99	1.07
200	50	0.50	1	0.50	0.86	0.93	0.92	0.91	0.08	2.02	1.16
50	100	0.50	1	0.51	0.89	0.94	0.94	1.09	0.04	2.21	1.09
50	200	0.50	1	0.49	0.91	0.94	0.95	0.96	0.02	2.05	1.03
200	100	0.50	1	0.45	0.90	0.94	0.94	1.01	0.04	2.11	1.07
100	200	0.50	1	0.42	0.92	0.93	0.94	0.96	0.02	2.07	1.02
200	200	0.50	1	0.36	0.92	0.96	0.95	0.91	0.02	1.81	1.03
100	400	0.50	1	0.39	0.94	0.96	0.96	0.86	0.01	1.79	0.91

Table 3: Coverage Rates and MSE, $h = 4$:

$$\widehat{Avar(\widehat{\delta})} = \left(\frac{1}{T}\sum_{t=1}^{T}\widehat{z}_t\widehat{z}_t'\right)^{-1}\left[\frac{1}{T}\sum_{t=1}^{T-h}\widehat{\varepsilon}_{t+h}^2\widehat{z}_t\widehat{z}_t'\right]\left(\frac{1}{T}\sum_{t=1}^{T}\widehat{z}_t\widehat{z}_t'\right)^{-1},$$

$$\widetilde{\Gamma}_t = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\widetilde{\lambda}_i\widetilde{\lambda}_j'\frac{1}{T}\sum_{t=1}^{T}\widetilde{e}_{it}\widetilde{e}_{jt} \quad \forall\, t.$$

				Coverage Probability				MSE			
N	T	b‘	k	$\widehat{y}_{T+h\|T}$	$\widehat{y}^0_{T+h\|T}$	$\widehat{y}_{T+h}$	$\widehat{y}^0_{T+h}$	$\widehat{y}_{T+h\|T}$	$\widehat{y}^0_{T+h\|T}$	$\widehat{y}_{T+h}$	$\widehat{y}^0_{T+h}$
50	50	0.00	2	0.91	0.85	0.93	0.92	0.15	0.09	1.17	1.15
100	50	0.00	2	0.91	0.85	0.94	0.94	0.12	0.09	1.09	1.07
200	50	0.00	2	0.94	0.86	0.93	0.92	0.09	0.08	1.16	1.16
50	100	0.00	2	0.92	0.89	0.94	0.94	0.10	0.04	1.17	1.09
50	200	0.00	2	0.92	0.91	0.96	0.95	0.07	0.02	1.07	1.03
200	100	0.00	2	0.95	0.90	0.95	0.94	0.05	0.04	1.07	1.07
100	200	0.00	2	0.94	0.92	0.95	0.94	0.04	0.02	1.04	1.02
200	200	0.00	2	0.94	0.92	0.95	0.95	0.03	0.02	1.03	1.03
100	400	0.00	2	0.95	0.94	0.96	0.96	0.03	0.01	0.95	0.91
50	50	0.50	2	0.88	0.85	0.94	0.92	0.23	0.09	1.22	1.15
100	50	0.50	2	0.90	0.85	0.94	0.94	0.16	0.09	1.12	1.07
200	50	0.50	2	0.92	0.86	0.93	0.92	0.10	0.08	1.16	1.16
50	100	0.50	2	0.91	0.89	0.94	0.94	0.15	0.04	1.24	1.09
50	200	0.50	2	0.90	0.91	0.96	0.95	0.13	0.02	1.14	1.03
200	100	0.50	2	0.94	0.90	0.95	0.94	0.06	0.04	1.08	1.07
100	200	0.50	2	0.91	0.92	0.94	0.94	0.07	0.02	1.09	1.02
200	200	0.50	2	0.93	0.92	0.96	0.95	0.04	0.02	1.03	1.03
100	400	0.50	2	0.93	0.94	0.96	0.96	0.06	0.01	0.99	0.91
50	50	0.00	1	0.51	0.85	0.90	0.92	1.13	0.09	2.22	1.15
100	50	0.00	1	0.50	0.85	0.92	0.94	0.99	0.09	1.97	1.07
200	50	0.00	1	0.51	0.86	0.93	0.92	0.90	0.08	2.01	1.16
50	100	0.00	1	0.47	0.89	0.94	0.94	1.05	0.04	2.14	1.09
50	200	0.00	1	0.46	0.91	0.94	0.95	0.93	0.02	2.01	1.03
200	100	0.00	1	0.42	0.90	0.94	0.94	1.01	0.04	2.11	1.07
100	200	0.00	1	0.40	0.92	0.94	0.94	0.94	0.02	2.03	1.02
200	200	0.00	1	0.34	0.92	0.95	0.95	0.90	0.02	1.80	1.03
100	400	0.00	1	0.35	0.94	0.96	0.96	0.86	0.01	1.78	0.91
50	50	0.50	1	0.52	0.85	0.91	0.92	1.15	0.09	2.24	1.15
100	50	0.50	1	0.52	0.85	0.92	0.94	1.00	0.09	1.99	1.07
200	50	0.50	1	0.50	0.86	0.93	0.92	0.91	0.08	2.02	1.16
50	100	0.50	1	0.50	0.89	0.93	0.94	1.09	0.04	2.21	1.09
50	200	0.50	1	0.48	0.91	0.94	0.95	0.96	0.02	2.05	1.03
200	100	0.50	1	0.44	0.90	0.94	0.94	1.01	0.04	2.11	1.07
100	200	0.50	1	0.41	0.92	0.93	0.94	0.96	0.02	2.07	1.02
200	200	0.50	1	0.36	0.92	0.96	0.95	0.91	0.02	1.81	1.03
100	400	0.50	1	0.41	0.94	0.96	0.96	0.86	0.01	1.79	0.91

References

Andrews, D. W. K. 1991, Heteroskedastic and Autocorrelation Consistent Matrix Estimation, *Econometrica* **59**, 817–854.

Angelini, E., Henry, J. and Mestre, R. 2001, Diffusion Index-based Inflation Forecasts for the Euro Area, European Central Bank, WP 61.

Artis, M., Banerjee, A. and Marcellino, M. 2001, Factor Forecasts for the U.K., CEPR Discussion Paper 3119.

Bai, J. 2003, Inferential Theory for Factor Models of Large Dimensions, *Econometrica* **71:1**, 135–172.

Bai, J. 2004, Estimating Cross-Section Common Stochastic Trends in Non-Stationary Panel Data, *Journal of Econometrics* **122**, 137–183.

Bai, J. and Ng, S. 2002, Determining the Number of Factors in Approximate Factor Models, *Econometrica* **70:1**, 191–221.

Bai, J. and Ng, S. 2004, Evaluating Latent and Observed Factors in Macroeconomics and Finance, mimeo, University of Michigan.

Banerjee, A., Marcellino, M. and Masten, I. 2004, Forecasting Macroeconomic Variables for the Acceding Countries, IGIER WP 260.

Bernanke, B. and Boivin, J. 2002, Monetary Policy in a Data Rich Environment, *Journal of Monetary Economics*.

Bernanke, B., Boivin, J. and Eliasz, P. 2002, Factor Augmented Vector Autoregressions (FVARs) and the Analysis of Monetary Policy, mimeo, Columbia University.

Chamberlain, G. and Rothschild, M. 1983, Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets, *Econometrica* **51**, 1305–1324.

Conley, T. 1999, GMM Estimation with Cross-Section Dependence, *Journal of Econometrics* **92**, 1–45.

Cristadoro, R., Forni, M., Reichlin, L. and Giovanni, V. 2001, A Core Inflation Index for the Euro Area, manuscript, www.dynfactor.org.

Driscoll, H. and Kraay, A. 1998, Consistent Covariance Matrix Estimation with Spatially-Dependent Panel Data, *Review of Economics and Statistics* **80:4**, 549–560.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. 2001b, Do Financial Variables Help in Forecasting Inflation and Real Activity in the Euro Area, manuscript, www.dynfactor.org.

Giannone, D., Reichlin, L. and Sala, L. 2002, Tracking Greenspan: Systematic and Unsystematic Monetary Policy Revisited, manuscript, www.dynfactor.org.

Marcellino, M., Favero, C. and Neglia, F. 2004, The Empirical Analysis of Monetary Policy With Large Datasets, *Journal of Applied Econometrics*.

Newey, W. and West, K. 1987, A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* **55**, 703–708.

Shintani, M. 2002, Nonlinear Analysis of Business Cycles Using Diffusion Indexes: Applications to Japan and the U.S., mimeo, Vanderbilt University.

Stock, J. H. and Watson, M. W. 2001, Forecasting Output and Inflation: the Role of Asset Prices, *Journal of Economic Literature* **47:1**, 1–48.

Stock, J. H. and Watson, M. W. 2002a, Forecasting Using Principal Components from a Large Number of Predictors, *Journal of the American Statistical Association* **97**, 1167–1179.

Stock, J. H. and Watson, M. W. 2002b, Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business and Economic Statistics* **20:2**, 147–162.

Figure 1: 12−Step Ahead Forecast: Growth Rate of Industrial Production

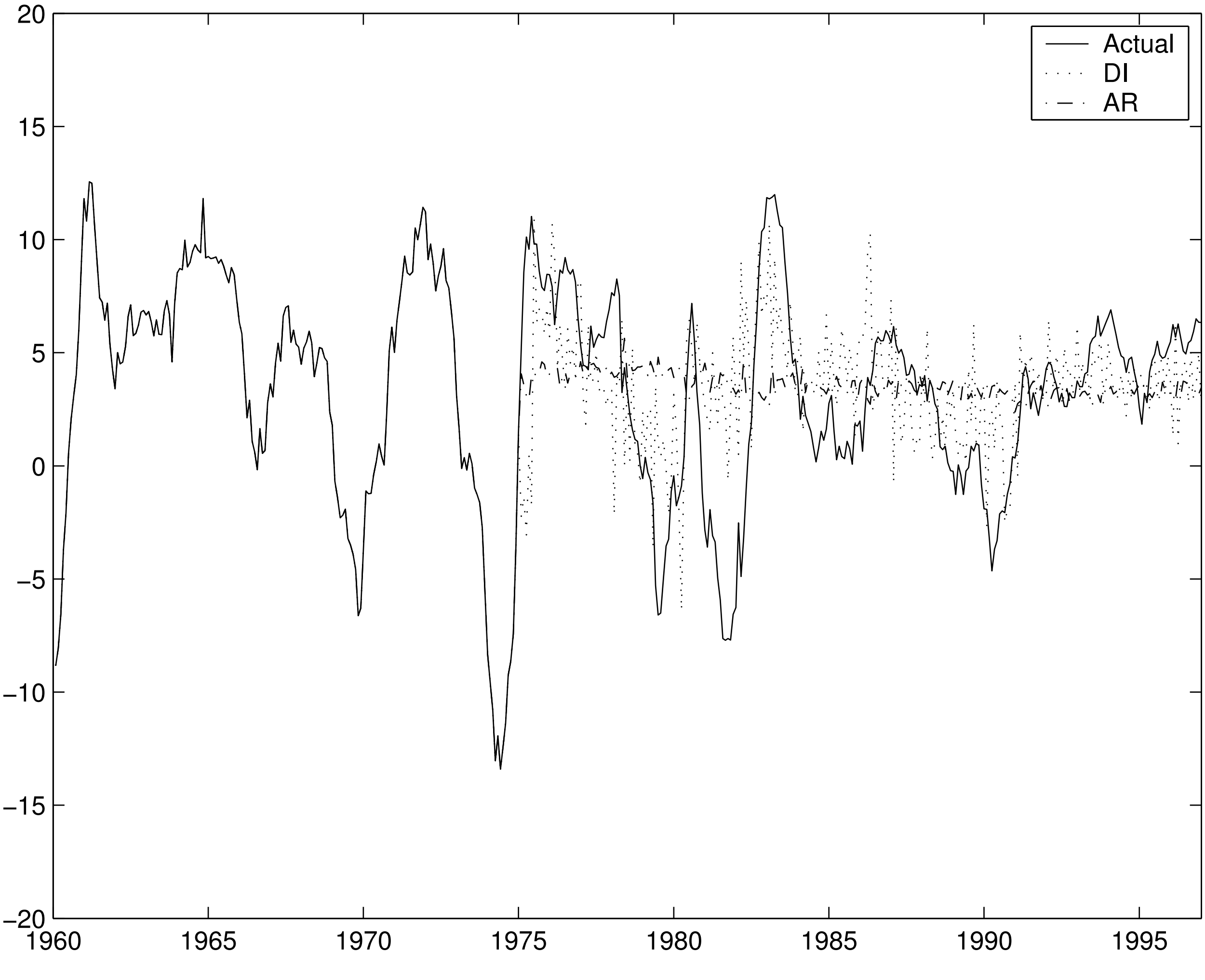

Figure 2a: Diffusion Index Forecast and Confidence Intervals: Growth Rate of Industrial Production

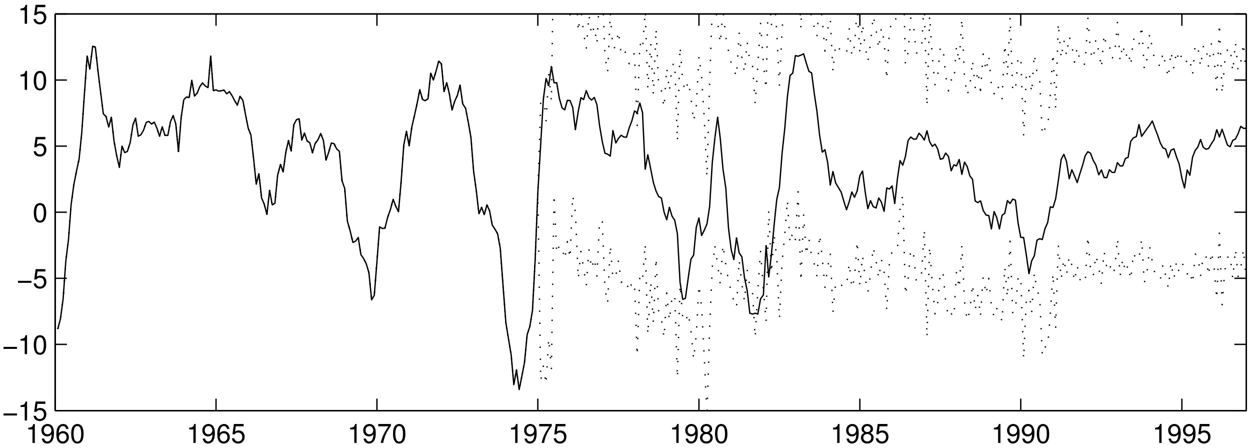

Figure 2b: AR Forecast and Confidence Intervals: Inflation

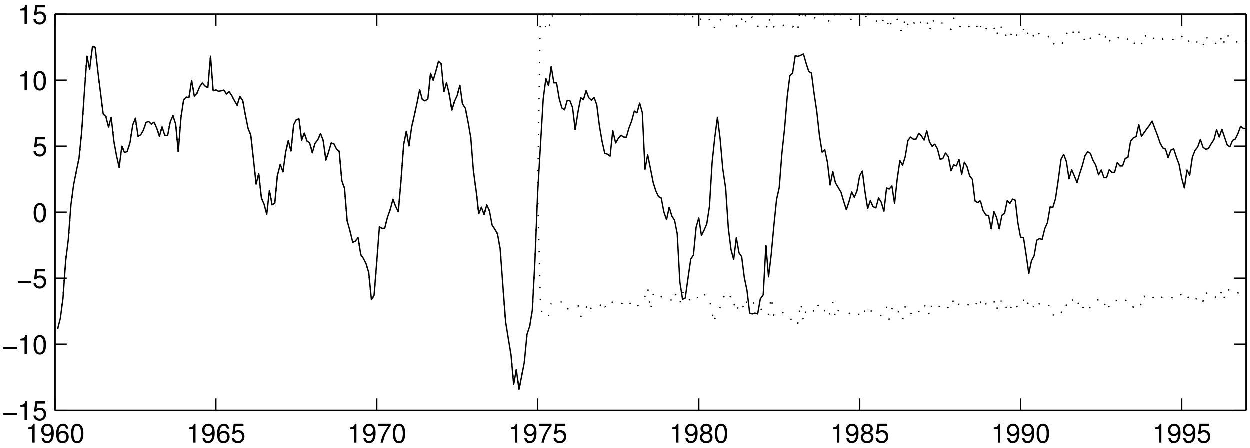

Figure 3: 12−Month Ahead Forecast: Inflation

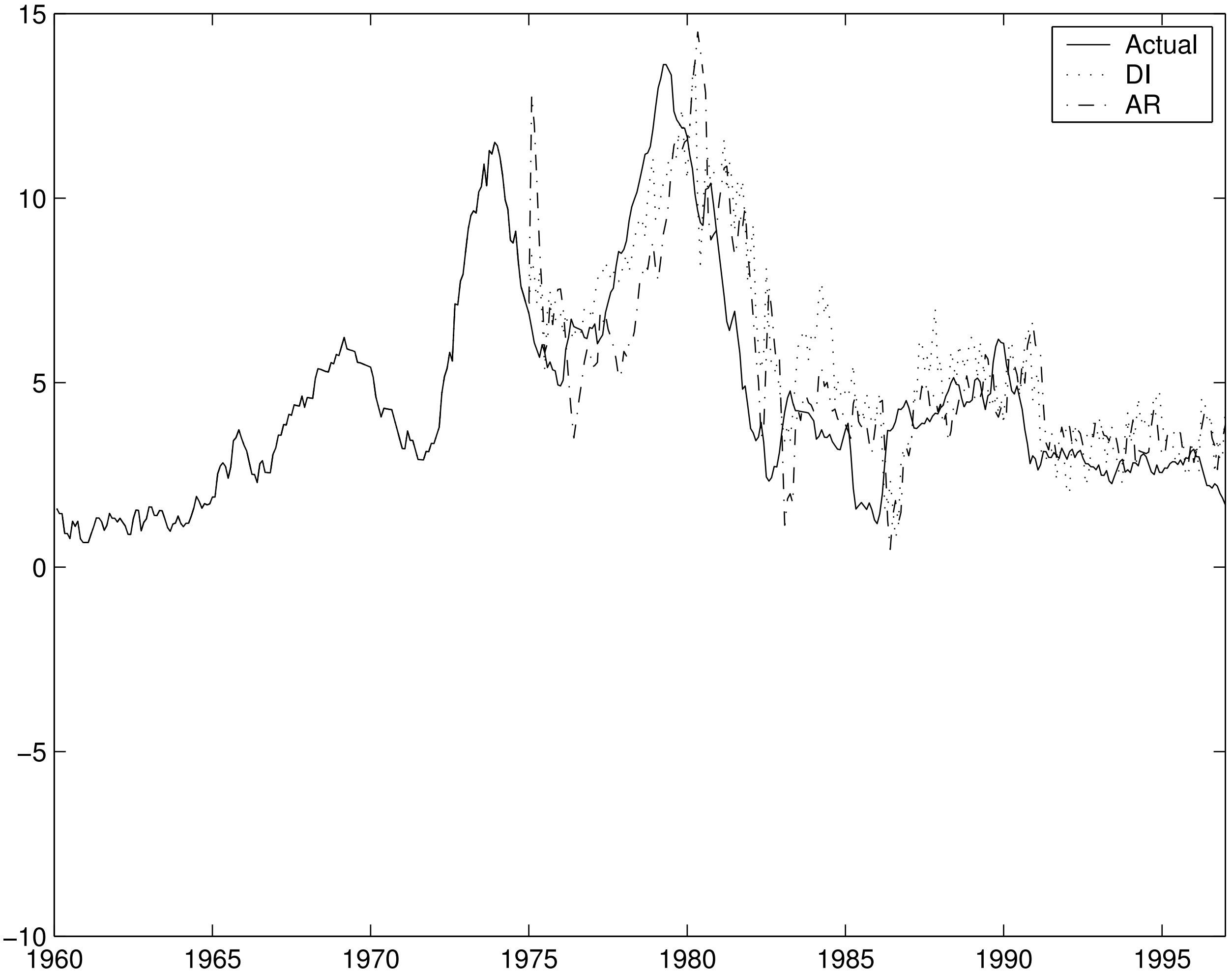

Figure 4a: Diffusion Index Forecast and Confidence Intervals: Inflation

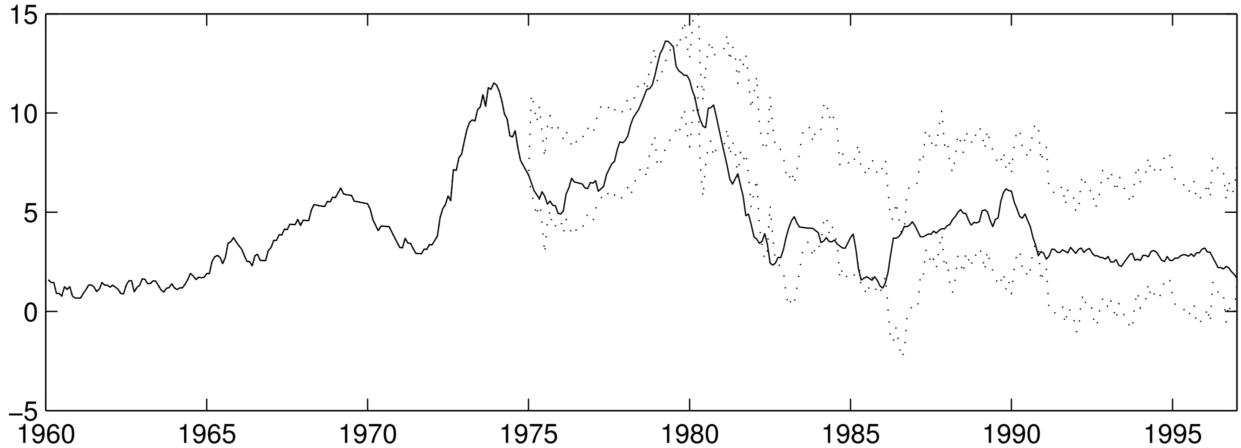

Figure 4b: AR Forecast and Confidence Intervals: Inflation