Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions

Jushan Bai* Serena Ng †

July 2005

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 some variable of interest. 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 with the estimated factors. In this paper, we show that the least squares estimates obtained from these factor augmented regressions are $\sqrt{T}$ consistent and asymptotically normal if $\sqrt{T}/N \to 0$. The conditional mean predicted by the estimated factors are $\min[\sqrt{T}, \sqrt{N}]$ consistent and asymptotically normal. Except when T/N goes to zero, inference should take into account the effect of "estimated regressors" on the estimated conditional mean. We present analytical formulas for prediction intervals. 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.

Keywords: Panel data, common factors, generated regressors, cross-section dependence, robust covariance matrix

JEL Classification: C2, C3, C5.

*Department of Economics, NYU, 269 Mercer St, New York, NY 10003 Email: Jushan.Bai@nyu.edu, and School of Economics and Management, Tsinghua University, Beijing, China.

†Department of Economics, University of Michigan, Ann Arbor, MI 48109 Email: Serena.Ng@umich.edu
The authors acknowledge financial support from the NSF (grants SES-0137084, SES-0136923). We thank seminar participants at Columbia, Princeton, and Yale for useful comments. The paper was also presented at the conference on Common Features in London, 2004.

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] Bernanke et al. (2002) showed that the information exploited in factor-augmented autoregressions (FAVAR) is important to properly identify the monetary transmission mechanism. However, the theoretical properties of the estimates obtained from factor augmented regressions are not well understood. In particular, how to conduct inference 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 to enable construction of confidence intervals for the parameters, the conditional mean, as well as the forecast.

Suppose information is available on 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, such as lags of y_t. Let

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

where $h \geq 0$ is the lead time between information available and the dependent variable. The vector F_t is unobservable. Instead of F_t, we observe a panel of data x_{it} which contain 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. If y_t is a scalar, (1) and (2) constitutes the 'diffusion index forecasting model' (DI) of Stock and Watson (2002a). If $h = 1$ and $y_{t+1} = (F'_{t+1}, W'_{t+1})'$, (1) is the FAVAR of Bernanke et al. (2002). Both types of analyses exploit the possibility that information in x_{it} can be summarized in a low dimensional vector, F_t. In economic analysis, F_t can be interpreted as the common factors that generate comovements in the data.

If F_t is observable, and assuming the mean of ε_t conditional on past information is zero,

[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).

the (mean-squared) optimal prediction of y_t is the conditional mean and is given by

$$y_{T+h|T} = E(y_{T+h}|z_T, z_{T-1}, \ldots) = \alpha' F_T + \beta' W_T \equiv \delta' z_T,$$

where $z_t = (F_t', \ W_t')'$. But such a prediction is not feasible because α, β, and F_t are all unobserved. The feasible prediction 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 (2), while hatted variables are estimated from (1). 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, $t = 1, \ldots T-h$. 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 $\widehat{y}_{T+h|T}$ itself also depends on $\widetilde{F}_T$. Thus, to study the behavior of $\widehat{y}_{T+h|T}$ and of the forecast error $\widehat{\varepsilon}_{T+h}$, we must examine the statistical properties of the estimated parameters as well as those of the estimated factors. Stock and Watson (2002a) showed that $\widehat{y}_{T+h|T}$ is consistent for $y_{T+h|T}$. But for hypothesis testing, to construct standard error bands of the impulse response functions of a FAVAR, to provide a confidence interval for the latent conditional mean, and to evaluate the uncertainty of a diffusion index forecasts, we need the limiting distributions of $(\widehat{\alpha}, \widehat{\beta})$, $\widehat{y}_{T+h|T}$, and $\widehat{\varepsilon}_{T+h}$.

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. An overview of the analysis is as follows. If we observe F_t but α is being estimated, the variance of $y_{T+1} - \widehat{y}_{T+1|T}$ is $O_p(T^{-1})$. Section 2 will show that when the factors have to be estimated, $\widehat{\alpha}$ remains $\sqrt{T}$ consistent if $\sqrt{T}/N \to 0$. But estimating the factor process F_t will contribute another $O(N^{-1})$ term to the forecasting error variance. Section 3 then shows 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. On the other hand, the forecast error $\widehat{\varepsilon}_{t+h}$ is asymptotically normal and dominated by the unconditional error variance. We will make precise how to estimate the error covariance matrices so that valid predictive inference can be conducted. 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 considered in Section 4. Simulations are given in Section 5. The main proofs are given in the Appendix.

2 Inference with Estimated Factors

In matrix notation, the factor model is $X = F\Lambda' + e$, where X is a $T \times N$ matrix of stationary data, $F = (F_1, ..., F_T)'$ 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.

Assumption A: Common factors

1. $E\|F_t\|^4 \leq M$ and $\frac{1}{T}\sum_{t=1}^{T} F_t F_t' \overset{p}{\longrightarrow} \Sigma_F$ for a $r \times r$ positive definite (non-random) 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, $\Lambda'\Lambda/N \overset{p}{\longrightarrow} \Sigma_\Lambda$ 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. $E(e_{it}e_{js}) = \sigma_{ij,ts}$, $|\sigma_{ij,ts}| \leq \bar{\sigma}_{ij}$ for all (t, s) and $|\sigma_{ij,ts}| \leq \tau_{ts}$ for all (i, j) such that

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

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, $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \lambda_i e_{it} \overset{d}{\longrightarrow} N(0, \Gamma_t)$, 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 mutually independent groups. Dependence within each group is allowed.

Assumption E: Let $z_t = (F_t' \quad W_t')'$, $E\|z_t\|^4 \leq M$; $E(\varepsilon_{t+h}|y_t, z_t, y_{t-1}, z_{t-1}, \ldots) = 0$ for any $h > 0$; z_t and ε_t are independent of the idiosyncratic errors e_{is} for all i and s. Furthermore,

1. $\frac{1}{T}\sum_{t=1}^{T} z_t z_t' \overset{p}{\longrightarrow} \Sigma_{zz} > 0$

2. $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_t \varepsilon_{t+h} \overset{d}{\longrightarrow} N(0, \Sigma_{zz,\varepsilon})$, where $\Sigma_{zz,\varepsilon} = \text{plim}\, \frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{t+h}^2 z_t z_t')) > 0$.

Assumptions A and B together imply r common factors. Assumption C allows for heteroskedasticity and limited time series and cross section dependence in the idiosyncratic component. The assumptions include e_{it} that are independent for all i and t as a special case. 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. Under Assumption D, $\{\lambda_i\}, \{F_t\}$, and $\{e_{it}\}$ are three mutually independent groups, but within group dependence is allowed. The assumption is standard in factor analysis. Assumption E.1 ensures that the (non-random) population moment matrix Σ_{zz} is full rank so that the regression model is well specified and that the parameters are identifiable. The assumption that ε_t is a martingale difference is appropriate in the context of forecasting. E.2 assumes $z_t\varepsilon_{t+h}$ obeys the central limit theorem.

These assumptions are similar to those of Stock and Watson (2002). But they also allow time-varying factor loadings with small variations such that $\lambda_{it} = \lambda_{i,t-1} + v_{it}/T$ with v_{it} being random variables with finite fourth moments. We assume time-invariant factor loadings.

2.1 Estimation

We first consider the properties of the least squares estimates when principal component estimates of the factors, $\widetilde{F}$, are used as regressors. Let $\widetilde{F} = (\widetilde{F}_1, ..., \widetilde{F}_T)$ be the matrix consisting of r eigenvectors (multiplied by $\sqrt{T}$) associated with the r largest eigenvalues of the matrix $XX'/(TN)$ in decreasing order. Then $\widetilde{\Lambda} = (\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_N)') = X'\widetilde{F}/T$, and $\widetilde{e} = X - \widetilde{F}\widetilde{\Lambda}'$. Also let $\widetilde{V}$ be the $r \times r$ diagonal matrix consisting of the r largest eigenvalues of $XX'/(TN)$, and $H = \widetilde{V}^{-1}(\widetilde{F}'F/T)(\Lambda'\Lambda/N)$. Let $\widehat{\alpha}$ and $\widehat{\beta}$ be the least squares estimates from regressing y_{t+h} on $\widehat{z}_t = (\widetilde{F}_t' \; W_t')'$. Define $\widehat{\delta} = (\widehat{\alpha}' \; \widehat{\beta}')'$, and $\delta = (\alpha' H^{-1} \; \beta')'$.

Theorem 1 *(Estimation) Suppose Assumptions A to E hold. If $\sqrt{T}/N \to 0$, then*

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

where $\Sigma_\delta = \Phi_0'^{-1}\Sigma_{zz}^{-1}\Sigma_{zz,\varepsilon}\Sigma_{zz}^{-1}\Phi_0^{-1}$, with $\Phi_0 = diag(V^{-1}Q\Sigma_\Lambda, I)$ being block diagonal, $V = \operatorname{plim}\widetilde{V}$, $Q = \operatorname{plim}\widetilde{F}'F/T$, and Σ_Λ defined in Assumption B. A consistent estimator for Σ_δ, denoted by $\widehat{Avar(\widehat{\delta})}$ is

$$\widehat{Avar(\widehat{\delta})} = \Big(\frac{1}{T}\sum_{t=1}^{T-h} \widehat{z}_t\widehat{z}_t'\Big)^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T-h} \widehat{\varepsilon}_{t+h}^2 \widehat{z}_t\widehat{z}_t'\Big)\Big(\frac{1}{T}\sum_{t=1}^{T-h} \widehat{z}_t\widehat{z}_t'\Big)^{-1}. \tag{3}$$

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). Having estimated factors as regressors does not affect consistency of the parameter estimates. 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 $\widehat{z}_t$.

Formula (3) 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 $Avar(\widehat{\delta})$ is

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

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 (3) 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 (FAVAR), 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 FAVAR(p) is defined as

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

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 FAVAR with F_t replaced by $\widetilde{F}_t$. Theorem 1 covers

[3] See, for example, Bernanke and Boivin (2003), Bernanke et al. (2002), and Giannone et al. (2002), and Marcellino et al. (2004).

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 *(FAVAR) 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 $\widehat{z}_t = (y_t' \dots y_{t-p}', \widetilde{F}_t', \dots, \widetilde{F}_{t-p}')'$, let $\widehat{Y}_t = (y_t', \widetilde{F}_t')'$ and $\widehat{Y}_{jt}$ be the jth element of $\widehat{Y}_t$. For $j = 1, \dots q + r$, let $\widehat{\delta}_j$ be obtained by least squares from regressing $\widehat{Y}_{jt+1}$ on $\widehat{z}_t$, with $\widehat{\varepsilon}_{jt+1} = \widehat{Y}_{jt+1} - \widehat{\delta}_j'\widehat{z}_t$. Under Assumptions A-E and 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. Since impulse response functions are based upon estimates of the FAVAR, Theorem 2 enables calculation of the standard errors. Although the condition $\sqrt{T}/N \to 0$ is not stringent, it puts discipline on when estimated factors can be used in regression analysis.

The limiting distribution in Theorem 2 is the same as if $\widehat{z}_t$ were observable. It is interesting to compare the result with Pagan (1984), whose model is $y_t = \alpha' x_t^e + \varepsilon_t$, and $x_t = \gamma' z_t + u_t$ with $x_t^e = \gamma' z_t$. Let $\widehat{x}_t = \widehat{\gamma}' z_t$. Thus,

$$y_t = \alpha'\widehat{x}_t + \varepsilon_t + \alpha'(x_t^e - \widehat{x}_t) = \alpha'\widehat{x}_t + \varepsilon_t + \widehat{u}_t$$

where $\widehat{u}_t = -(\widehat{\gamma} - \gamma)' z_t$. Using $\widehat{x}_t$ as regressor, we have $\widehat{\alpha} - \alpha = (\widehat{x}'\widehat{x})^{-1}\widehat{x}'(\varepsilon + \widehat{u})$. But

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{x}_t\widehat{u}_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{x}_t z_t'(\gamma - \widehat{\gamma}) = \widehat{\gamma}'\frac{1}{T}\sum_{t=1}^{T} z_t z_t'\sqrt{T}(\gamma - \widehat{\gamma}) = O_p(1)$$

which is non-negligible, so estimated regressors have an effect on parameter estimation. In our case, the corresponding term is $O_p(\frac{\sqrt{T}}{\min[N,T]})$. If $N = 1$, this term would be $O_p(\sqrt{T})$, much larger than Pagan's $O_p(1)$. This is because in Pagan's model only finite number of parameters γ is estimated in the first stage and we need to estimate T unknown quantities $F_1, ..., F_T$. But if N is large, the corresponding term in our analysis is negligible.

2.2 Prediction Intervals

We first provide some intuition for the appeal of diffusion index forecasts. Consider a simple forecasting equation $y_{t+1} = \alpha F_t + \varepsilon_{t+1}$ where ε_t are iid $(0, \sigma_\varepsilon^2)$. Also assume F_t is an AR(1) process $F_t = \rho F_{t-1} + u_t$ with u_t being iid $(0, \sigma_u^2)$ and being independent of ε_s 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 such that $y_{t+1} = \rho y_t + \eta_{t+1} + \theta\eta_t$, where η_t is white noise and θ depends on other parameters. 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 $\rho y_t + \theta\eta_t$. The forecast error is η_{t+1} and the forecast error variance is $\sigma_\eta^2 = E(\eta_{t+1}^2)$. It can be shown that $\sigma_\eta^2 > \sigma_\varepsilon^2$, 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 $\sigma_\eta^2 - \sigma_\varepsilon^2$, 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.

Suppose the object of interest is the (latent) conditional mean of (1). If y_t is inflation, the estimated conditional mean can be interpreted as an estimate of the expected rate of inflation. We now suggest how a confidence interval for the conditional mean can be constructed. 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 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,

$$\sqrt{N}(\widetilde{F}_t - HF_t) \quad \xrightarrow{d} \quad N\Big(0, V^{-1}Q\Gamma_t Q'V^{-1}\Big) \equiv N\Big(0, Avar(\widetilde{F}_t)\Big),$$

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

Theorem 3 *Let $\widehat{y}_{T+h|T} = \widehat{\delta}'\widehat{z}_T$. Under the assumptions of Theorem 1, and $\sqrt{N}/T \to 0$,*

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

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 for $\widehat{y}_{T+h|T}$ 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 error variance in predicting the conditional mean falls at rate T, and for a given T, it increases with the number of observed predictors through a loss in degrees of freedom. In contrast, the error variance here decreases at rate $\min[N, T]$, and for a given T, forecast efficiency improves with the number of predictors used to estimate F_t. This is because in the present setting, a large N enables more precise estimation of the common factors and thus results in more efficient predictions. 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.

When the objective is forecasting, one would be more interested in the distribution of the forecast error. Since $y_{T+h} = y_{T+h|T} + \varepsilon_{T+h}$, it follows that the forecasting error

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

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

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

We state this result as a corollary.

Corollary 1 *Under the assumptions of Theorem 3, and assume* ε_t *is iid* $N(0, \sigma_\varepsilon^2)$*, then the forecasting error* $\widehat{\varepsilon}_{T+h}$ *is*

$$\widehat{\varepsilon}_{T+h} \sim N(0, \sigma_\varepsilon^2 + var(\widehat{y}_{T+h|T}))$$

Assuming a constant conditional variance σ_ε^2, a predictive interval for y_{T+h} can be obtained upon replacing σ_ε^2 by its consistent estimator $\frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_t^2$. The above formula extends the textbook definition of forecast uncertainty that only allows for estimation error in $\widehat{\delta}$ (as in Greene (2003), Chapter 6), to also permit using $\widetilde{F}_t$ as regressors. Therefore, $var(\widehat{y}_{T+h|T})$ reflects model parameter uncertainty and regressor uncertainty. Note that in large samples, $\text{var}(\widehat{\varepsilon}_{T+h})$ is dominated by σ_ε^2, just as when all predictors are observed. However, $\text{var}(\widehat{y}_{T+h|T})$ vanishes at rate $\min[T, N]$, rather than the usual rate of T. Nonetheless, if we ignore $\text{var}(\widehat{y}_{T+h|T})$, σ_ε^2 will under-estimate the true forecast uncertainty for a given T and N. If ε_t is only conditionally normal with conditional heteroskedasticity, then σ_ε^2 should be replaced by the conditional variance, which may be modeled by an ARCH or GARCH process.

To conduct inference for either the conditional mean or the forecasting error, a consistent estimator for $B_T^2 = var(\widehat{y}_{T+h|T})$ is required. 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 the $r \times r$ matrix $\widetilde{\Gamma}_t$ can be one of the following:

$$\widetilde{\Gamma}_t = \frac{1}{N}\sum_{i=1}^{N}\widetilde{e}_{it}^2\widetilde{\lambda}_i\widetilde{\lambda}_i' \tag{5a}$$

$$\widetilde{\Gamma}_t = \widetilde{\sigma}_e^2\frac{1}{N}\sum_{i=1}^{N}\widetilde{\lambda}_i\widetilde{\lambda}_i' \tag{5b}$$

$$\widetilde{\Gamma} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\widetilde{\lambda}_i\widetilde{\lambda}_j'\frac{1}{T}\sum_{t=1}^{T}\widetilde{e}_{it}\widetilde{e}_{jt} \tag{5c}$$

with $n/\min[N, T] \to 0$ in (5c), where $\widetilde{e}_{it} = x_{it} - \widetilde{\lambda}_i'\widetilde{F}_t$. The various specifications of $\widetilde{\Gamma}_t$ accommodate flexible error structures in the factor model. Both (5a) and (5b) assume that e_{it} is cross-sectionally uncorrelated with e_{jt}. Consistency of both estimators was shown in our earlier work. The estimator (5b) 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 (5a) and (5b) 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 (5a) and (5b) 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 (5c). Consistency of $\widetilde{\Gamma}_t$ will be established below and it requires nontrivial arguments. 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 to be discussed below.

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

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

and the 95% confidence interval for the variable y_{T+h} is

$$\left(\widehat{y}_{T+h|T} - 1.96\sqrt{\widehat{\sigma}_\varepsilon^2 + \widehat{var}(\widehat{y}_{T+h|T})}, \quad \widehat{y}_{T+h|T} + 1.96\sqrt{\widehat{\sigma}_\varepsilon^2 + \widehat{var}(\widehat{y}_{T+h|T})}\right),$$

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.

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.

3 Covariance Matrix Estimator: the CS-HAC

The CS-HAC estimator defined in (5c) is robust to cross-section correlation and cross-section heteroskedasticity but requires the assumption of covariance stationarity which is not necessary for (5a) and (5b), since these assume cross-sectionally uncorrelated idiosyncratic errors. To understand the problem, it helps to first consider the cross-section regression $y_i = \beta_i'\lambda_i + e_i$, where λ_i is a $r \times 1$ vector of observed regressors. Then $\Gamma = \frac{1}{N}\Lambda'\Omega\Lambda$, where $\Omega = \lim_{N\to\infty} E(ee')$ with $e = (e_1, ..., e_N)'$, and Λ is the $N \times r$ regressor matrix. Here, the "natural" estimator, $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j'\widehat{e}_i\widehat{e}_j$, is not consistent because it is equal to $(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i\widehat{e}_i)(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i\widehat{e}_i)'$ and $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i\widehat{e}_i$ converges to a random vector. The problem is analogous to inconsistency of the unweighted sum of T sample autocovariances as a long run variance estimator in a time series context. As time series data have a natural ordering, it is possible to consider a kernel estimator that truncates at $M < T$ lags with $M \to \infty$ and $M/T \to 0$.

Cross-section data have no natural ordering. It is only in special cases such as the one considered in Conley (1999) that a truncated sum can be justified. Neither economic theory nor intuition is usually of much help in obtaining a 'mixing condition' type ordering of the data. More generally any cross-section permutation of the data is an equally valid representation of information available, and the different orderings also cannot be ranked. The common practice in cross-section regressions is to assume $E(e_ie_j) = 0$ $i \neq j$, so that $\Omega = \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\lambda_i\lambda_i'e_i^2$.

A third alternative is available if we have observations on the cross-section units over time. The basic intuition is as follows. If covariance stationarity holds, the time series observations will allow us to consistently estimate the cross-section correlations provided T is large. Furthermore, the covariance matrix of interest is of dimension $(r \times r)$, much smaller than N, and can be estimated with $n < N$ observations. An estimator along these lines was considered in Driscoll and Kraay (1998) who showed in a panel context that using all N cross-section units will yield an inconsistent estimate of the covariance matrix. They require $n = n(T)$, as their residuals are based upon estimators that are $\sqrt{T}$ consistent. 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 λ_i in our analysis is not observed. To consistently estimate Γ_t, we require Γ_t not to depend on t, see Assumption C4, so that we can use observations from other periods in addition to period t to estimate Γ. This covariance stationarity rules out time series (unconditional) heteroskedasticity.

Theorem 4 *Suppose Assumptions A-D hold. In addition, $E(e_{it}e_{jt}) = \sigma_{ij}$ for all t, and Γ_t does not depend on t, denoted by Γ. Let $n = n(N, T)$ and define*

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

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

The object of interest is the $r \times r$ matrix Γ, not the $N \times N$ covariance matrix for e_{it}.[4] Accordingly, consistent estimation of Γ is possible using only $n < N$ pairs of $(\widetilde{\lambda}_i, \widetilde{e}_i)$. Use of all N pairs is undesirable because the sampling variability induced by the 'estimated regressors' $\widetilde{\lambda}_i$ will be excessive. This reason for using $n < N$ observations is distinct from the truncation required in estimation of the long run variance, which is justified by the 'mixing' properties of the data.

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 large samples, $\widetilde{\Gamma}$ will converge to the same Γ whichever n series we pick.

4 Finite Sample Properties

To assess the finite sample properties of the procedures, simulated data are generated as:

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

where $e_t = (e_{1t}, ...e_{Nt})'$ and $v_t = (v_{1t}, ..., v_{Nt})'$, and u_{jt} and v_{it} are mutually uncorrelated $N(0, \sigma_v^2)$ random variables and $\bar{\Omega}^{1/2}(b)$ is the Choleski decomposition of $\bar{\Omega}(b)$, an $N \times N$ Toeplitz matrix whose j-th main diagonal is b^j if $j \leq 10$, and zero otherwise.[5] By design, the cross-section correlation 'dies out' if the units are spatially far apart, much like an AR(1) process. We draw λ_i (once) from the uniform distribution with support on [0,1]. Four

[4] Note that $\widetilde{\Gamma}$ is not directly estimating Γ. This is because 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 $Q\Gamma_t Q' = QH'H'^{-1}\Gamma_t H^{-1}HQ$, the matrix H is effectively canceled out.

[5] The results are 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.

variations of the DGP are considered. In DGP 1 (homoskedasticity and cross-sectionally uncorrelated errors), we set $b = 0$ and $\sigma_v^2 = 1$. In DGP 2 (heteroskedasticity and cross-sectionally uncorrelated errors), we set $b = 0$ but $\sigma_v^2(i)$ is uniformly distributed on $(.5, 1.5)$. In DGP 3 (homoskedasticity and cross-sectionally correlated errors), we let $b = .5, \sigma_v^2(i) = 1$, and in DGP 4 (heteroskedasticity and cross-sectionally correlated errors), $b = 0.5$ and σ_v^2 is again uniformly distributed on (.5,1.5).

In the simulations, $r = 2$ and is assumed known. The series to be forecasted is

$$y_{t+4} = 1 + F_{1t} + F_{2t} + \varepsilon_{t+4}.$$

That is, $h = 4$, $W_t = 1$, $\alpha = (1, 1)'$, and $\beta = 1$. The simulation design is similar to Stock and Watson (2002a), but allows stronger cross-section correlations in e_{it}. Three types of confidence intervals will be presented:

(A): (5b) +(4) ;   (B): (5a) + (3) ;   (C): (5c) + (3).

For the sake of comparison, we also consider the coverage rates that would obtain when F_t is known (and thus the standard errors omit terms involving $Avar(\widetilde{F}_t)$). This is labelled (D).

The coverage rates are reported in Table 1 for (i) the estimated conditional mean, $\widehat{y}_{T+h|T}$; and (ii) the diffusion index forecast, $\widehat{y}_{T+h}$. The coverage rates are generally close to the nominal rate of .95, though three results are noteworthy. First, when N is small, the coverage of (C) is too low for DGPs 1 and 2, with CS-HAC when in fact there is no cross-section correlation. Second, for DGPs 3 and 4, the coverage of (A) and (B) are always too low since these ignore the correlation in the errors. Coverage is improved using the CS-HAC, see (C). Third, the coverage rates for $y_{T+h|T}$ are more sensitive to the relation between N and T than for y_{T+h}. This is in accord with theory, because the error in y_{T+h} is dominated by ε_{t+h}, whereas the error in $y_{T+h|T}$ is induced by the error in estimating F_t and the parameters.

4.1 Empirical Application

To illustrate, we use as predictors the 150 series as in Stock and Watson (2002b).[6] We consider $h = 12$ period ahead forecast of the annual growth rate of industrial production, DIP, ie $y_{t+12} = DIP = \log(IP_{t+12}) - \log(IP_t)$. For the sake of comparison, we first consider the autoregressive forecast $\widehat{\beta}'W_{1969:1}$, where W_t, are the lags of DIP plus a constant. We first select the order of this autoregression using the BIC. The diffusion index model then

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

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. 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.

Because the 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 $\widehat{r} = \text{argmax}_{k=0,\ldots kmax} \log \widetilde{\sigma}^2(k) + k \cdot g(N,T)$, where $\widetilde{\sigma}^2(k) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{e}_{it}^2$, see Bai and Ng (2002). We report results for $g(N,T) = (N+T)\frac{\log(NT)}{NT}$. This penalty tends to select a smaller number of estimated factors, but we correct for cross-section correlation in the idiosyncratic errors.[7]

The average mean-squared error for the diffusion index and AR forecasts are 24.95 and 26.46, respectively. Figures 1a,b present the series to be forecasted, along with the 95% prediction interval as suggested by the diffusion index and the AR forecasts, respectively.

5 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) $E\|u_t\|^{4+\delta} \leq M$ 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.

[7]Additional results with $g_1(N,T) = \frac{\log(\min[N,T])}{\min[N,T]}$ are given in the working paper version of the paper

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. But it also rules out lagged dependent variable. When F_t is I(1), y_t is also I(1), implying cointegration between the dependent variable and F_t. Lagged dependent variable will be asymptotically multicollinear with F_t. We therefore assume the absent of lagged dependent variable. As a result, the following mixture normality is a reasonable assumption:

$$D_T^{-1} \sum_{t=1}^{T} z_t \varepsilon_{t+h} \xrightarrow{d} MN(0, \sigma_\varepsilon^2 \Omega) \tag{6}$$

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 $D_T = T I_{r+p}$ if W_t is also I(1), and $D_T = (T I_r, \sqrt{T} I_p)$ 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 $XX'/(T^2 N)$, 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 (6) hold.*

(i) Let $\widehat{\alpha}$ and $\widehat{\beta}$ be the least squares estimators from a regression of y_{t+h} on $\widehat{z}_t = (\widetilde{F}_t' \; W_t')'$. Again denote $\widehat{\delta} = (\widehat{\alpha}' \; \widehat{\beta}')'$ and $\delta = (\alpha' H^{-1} \; \beta')'$. As $N, T \to \infty$ with $\sqrt{T}/N \to 0$,

$$(D_T^{-1} \widehat{z}' \widehat{z} D_T^{-1})^{1/2} D_T (\widehat{\delta} - \delta) \xrightarrow{d} N\Big(0, \sigma_\varepsilon^2 I\Big) \tag{7}$$

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

$$\frac{\widehat{y}_{T+h|T} - y_{T+h|T}}{C_T} \xrightarrow{d} N(0, 1) \tag{8}$$

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 $(D_T^{-1}\widehat{z}'\widehat{z}D_T^{-1})^{1/2}$, 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 $I(0)$[8]. 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.

6 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. We also suggest how the covariance matrix of cross-correlated errors can be consistently estimated.

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

Appendix: Proofs

The following identity is used in the proof of Lemma A1 below, see Bai and Ng (2002):

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

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* $\delta_{NT}^2 = \min[N,T]$, *and* $H = \widetilde{V}^{-1}(\widetilde{F}'F/T)(\Lambda'\Lambda/N)$. *Under Assumptions A-E,*
(i) $\frac{1}{T}\sum_{t=1}^{T}\|\widetilde{F}_t - HF_t\|^2 = O_p(\delta_{NT}^{-2})$;
(ii) $\frac{1}{T}\sum_{t=1}^{T}(\widetilde{F}_t - HF_t)z_t' = O_p(\delta_{NT}^{-2})$;
(iii) $\frac{1}{T}\sum_{t=1}^{T}(\widetilde{F}_t - HF_t)\widehat{z}_t' = O_p(\delta_{NT}^{-2})$;
(vi) $\frac{1}{T}\sum_{t=1}^{T}(\widetilde{F}_t - HF_t)\varepsilon_{t+h} = O_p(\delta_{NT}^{-2})$

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. The proof for (iv) is similar to (ii), with ε_t replacing z_t.

Proof of Theorem 1. Adding and subtracting terms, the model can be written as:

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

In matrix notation: $Y = \widehat{z}\delta + \varepsilon + (FH' - \widetilde{F})H^{-1\prime}\alpha$, where $Y = (y_{h+1}, ..., y_T)'$, $\varepsilon = (\varepsilon_{h+1}, ..., \varepsilon_T)'$, and $\widehat{z} = (\widehat{z}_1, ..., \widehat{z}_{T-h})'$. The OLS estimator is $\widehat{\delta} = (\widehat{z}'\widehat{z})^{-1}\widehat{z}'Y$. Thus,

$$\widehat{\delta} - \delta = (\widehat{z}'\widehat{z})^{-1}\widehat{z}'\varepsilon + (\widehat{z}'\widehat{z})^{-1}\widehat{z}'(FH' - \widetilde{F})H^{-1}\alpha, \quad \text{or}$$

$$\sqrt{T}(\widehat{\delta} - \delta) = (T^{-1}\widehat{z}'\widehat{z})^{-1}T^{-1/2}\widehat{z}'\varepsilon + (T^{-1}\widehat{z}'\widehat{z})^{-1}[T^{-1/2}\widehat{z}'(FH' - \widetilde{F})]H^{-1}\alpha.$$

The second term on the right hand side is $o_p(1)$. This follows from $T^{-1/2}\widehat{z}'(FH' - \widetilde{F}) = O_p(T^{1/2}/\min(N, T)) = o_p(1)$ if $\sqrt{T}/N \to 0$, by Lemma A1. For the first term, $T^{-1/2}\widehat{z}'\varepsilon = T^{-1/2}(\varepsilon'\widehat{F}, \varepsilon'W)'$. Now $T^{-1/2}\widetilde{F}'\varepsilon = T^{-1/2}HF'\varepsilon + T^{-1/2}(\widetilde{F} - FH')'\varepsilon$. The second term is $o_p(1)$ if $\sqrt{T}/N \to 0$ by Lemma A1. That is, $T^{-1/2}\widehat{z}'\varepsilon = T^{-1/2}(\varepsilon'FH', \varepsilon'W)' + o_p(1) = T^{-1/2}\Phi z'\varepsilon + o_p(1)$, where Φ is a block diagonal matrix $\Phi = diag(H, I)$. Thus,

$$\begin{aligned}\sqrt{T}(\widehat{\delta} - \delta) &= (T^{-1}\widehat{z}'\widehat{z})^{-1}T^{-1/2}\widehat{z}'\varepsilon + o_p(1) \\ &= (T^{-1}\widehat{z}'\widehat{z})^{-1}\Phi T^{-1/2}z'\varepsilon + o_p(1)\end{aligned} \tag{A.2}$$

Since $z'\varepsilon/\sqrt{T}\xrightarrow{d}N(0, \Sigma_{zz,\varepsilon})$ by Assumption E2, the above is asymptotically normal. The asymptotic variance matrix is the probability limit of

$$\left(\frac{\widehat{z}'\widehat{z}}{T}\right)^{-1}\Phi\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t+h}^2 z_t z_t'\right)\Phi'\left(\frac{\widehat{z}'\widehat{z}}{T}\right)^{-1}, \qquad \text{where} \qquad \Phi = \begin{bmatrix} H & 0 \\ 0 & I \end{bmatrix} \tag{A.3}$$

Define $H_0 = \operatorname{plim} H = V^{-1}Q\Sigma_\Lambda$, $\Phi_0 = \operatorname{plim} \Phi = diag(H_0, I)$. Now $T^{-1}\widehat{z}'\widehat{z} = \Phi(T^{-1}z'z)\Phi' + o_p(1) \xrightarrow{p} \Phi_0\Sigma_{zz}\Phi_0'$. The asymptotic variance or the limit of (A.3) is

$$\Sigma_\delta = (\Phi_0\Sigma_{zz}\Phi_0')^{-1}(\Phi_0\Sigma_{zz,\varepsilon}\Phi_0')(\Phi_0\Sigma_{zz}\Phi_0')^{-1} = \Phi_0'^{-1}\Sigma_{zz}^{-1}\Sigma_{zz,\varepsilon}\Sigma_{zz}^{-1}\Phi_0^{-1}.$$

Since $HF_t = \widetilde{F}_t + o_p(1)$ and $z_t = (F_t', W_t')'$, we have $\Phi(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t+h}^2 z_t z_t')\Phi' = (\frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_{t+h}^2\widehat{z}_t\widehat{z}_t') + o_p(1)$. Therefore, $\widehat{\Sigma}_\delta = (T^{-1}\widehat{z}'\widehat{z})^{-1}(T^{-1}\sum_{t=1}^{T}\widehat{\varepsilon}_{t+h}^2\widehat{z}_t\widehat{z}_t')(T^{-1}\widehat{z}'\widehat{z})^{-1}$ is a consistent estimator for Σ_δ. This completes the proof of Theorem 1.

Proof of Theorem 2. Without loss of generality, consider a FAVAR(1). For FAVAR(1), Y_t and z_t coincides, i.e., $Y_t = z_t = (y_t' \;\; F_t')'$. The infeasible FAVAR 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 FAVAR expressed in terms of $\widetilde{F}_t$ is

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

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 FAVAR 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

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

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,

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

For $j \le q$, $u^1_{j,t+1}$ is the jth component of ε_{1t+1}. This case is treated in (A.2) and the limiting variance is shown to be, compare with (A.3)

$$\Big(\frac{1}{T}\sum_{t=1}^{T}\widehat{z}_t\widehat{z}_t'\Big)^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}(u^1_{jt+1})^2\widehat{z}_t\widehat{z}_t'\Big)\Big(\frac{1}{T}\sum_{t=1}^{T}\widehat{z}_t\widehat{z}_t'\Big)^{-1}. \tag{A.4}$$

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

For $j = q+1, ..., q+r$, u^1_{jt+1} is the kth component ($k = j-q$) of $H\varepsilon_{2t+1}$, which can be written as $\iota_k' H\varepsilon_{2t+1}$, where ι_k is a vector of 0's with the kth element being 1. Note that $\iota_k' H\varepsilon_{2t+1}$ is a linear combination of the components of $\varepsilon_{2\varepsilon}$. The analysis of (A.2) in the previous proof implies that the limiting variance is given by (A.4) with $u^1_{jt+1} = \iota_k' H\varepsilon_{2t+1}$.

Proof of Theorem 3. Begin by rewriting

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

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 $B_T^2 = \frac{1}{T}\widehat{z}_T' Avar(\widehat{\delta})\widehat{z}_T + \frac{1}{N}\widehat{\alpha}' Avar(\widetilde{F}_T)\widehat{\alpha}$, which is an estimate for the variance of $\widehat{y}_{T+h|T} - y_{T+h|T}$. Thus $(\widehat{y}_{T+h|T} - y_{T+h|T})/B_T \xrightarrow{d} N(0,1)$.

To prove Theorem 4, we need additional results.

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

Proof of (i). From the identity

$$\widetilde{\lambda}_i - H^{-1\prime}\lambda_i = T^{-1}HF'\underline{e}_i + T^{-1}\widetilde{F}'(F - \widetilde{F}H^{-1\prime})\lambda_i + T^{-1}(\widetilde{F} - FH')'\underline{e}_i, \tag{A.5}$$

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 $O_p([\min[N,T]]^{-1})$ following Lemma B.1 of Bai (2003), replacing e_{it} with $\frac{1}{n}\sum_{i=1}^n \lambda_i e_{it}$. For part (ii), the expression is equal to (c) multiplied by n, thus it is bounded by $O_p(n/\min[N,T]])$.

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 $\widetilde{\lambda}_i - H^{-1\prime}\lambda_i$ in (A.5) 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 $\|\sum_{i=1}^n \sigma_{ij}\lambda_i\| = O_p(1)$. (c) is $O_p(\min[N,T]^{-1})$ following Lemma B.1 of Bai (2003), replacing e_{it} by $\sum_{i=1}^n \sigma_{ij}e_{it} = O_p(1)$ in view of $\sum_{i=1}^n |\sigma_{ij}| \le M$. □

Proof of Theorem 4. Let $\sigma_{ij} = E(e_{it}e_{jt})$, and $\widetilde{\sigma}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\widetilde{e}_{it}\widetilde{e}_{jt}$. Let $\Gamma_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\lambda_i\lambda_j'$. The limit of Γ_n exists by Assumption C. By definition,

$$\Gamma = \lim_{n\to\infty}\Gamma_n.$$

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

$$\widetilde{\Gamma} - H^{-1\prime}\Gamma H^{-1} = \widetilde{\Gamma} - H^{-1\prime}\bar{\Gamma}_n H^{-1} + H^{-1\prime}(\bar{\Gamma}_n - \Gamma_n)H^{-1} + H^{-1\prime}(\Gamma_n - \Gamma)H^{-1}. \tag{A.6}$$

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

(i) $\bar{\Gamma}_n - \Gamma_n \xrightarrow{p} 0$.

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

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

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 $n/N \to 0$.

Consider 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,

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

where (A.a) and (A.b) are the two terms in parenthesis that are added to B. Now

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

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

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

It follows from A=(A.a)(A.b) that, if $\sqrt{n}/T \to 0$,

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

For B, it is bounded in norm by

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

by Lemma A2(ii). Thus, $B \to 0$ if $\frac{n}{\min[T,N]} \to 0$. Analogously, $III \to 0$ if $\frac{n}{\min[T,N]} \to 0$.

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 because $(a+b)^2 \leq 2a^2 + 2b^2$,

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

Thus summing over t and divided by T,

$$\begin{aligned}IV &\leq 2\Big(\frac{1}{T}\sum_{t=1}^{T}\Big\|\widetilde{F}_t\Big\|^2\Big)\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\lambda_i(H^{-1\prime}\lambda_i - \widetilde{\lambda}_i)'\Big\|^2 \\ &+2\|H^{-1}\|^2\Big(\frac{1}{n}\sum_{i=1}^{n}\|\lambda_i\|^2\Big)^2\cdot n\cdot\frac{1}{T}\sum_{t=1}^{T}\Big\|HF_t - \widetilde{F}_t\Big\|^2 = a + b.\end{aligned}$$

By Lemma A2, $a \to 0$ if $\sqrt{n}/T \to 0$. And $b = O_p(n)O_p(\min[N,T]^{-1}) \to 0$ if $n/T \to 0$ and $n/N \to 0$. □

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

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

Using $\widetilde{\lambda}_i\widetilde{\lambda}_j' - H^{-1\prime}\lambda_i\lambda_j H^{-1} = (\widetilde{\lambda}_i - H^{-1\prime}\lambda_i)\widetilde{\lambda}_j' + H^{-1\prime}\lambda_i(\widetilde{\lambda}_j - H^{-1\prime}\lambda_j)'$, we write II

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

By Lemma A3,

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

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. □.

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 (4) 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).[9] 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 $\widehat{\alpha}'\widetilde{V}^{-1}\widetilde{\Gamma}\widetilde{V}^{-1}\widehat{\alpha}$. 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 estimated regression coefficient when $\widetilde{F}^n$ is the regressor, and likewise for $\widehat{\alpha}^s$). Furthermore,

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

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 $\widetilde{\Gamma}^n = \widetilde{\Gamma}^s / T$ in view of $\widetilde{\lambda}_i^n = \widetilde{\lambda}_i^s / \sqrt{T}$, see equations (5a)-(5c). From these relationships, it is easy to see that

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

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.

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. 2003, Monetary Policy in a Data Rich Environment, *Journal of Monetary Economics* **50:3**, 525–546.

Bernanke, B., Boivin, J. and Eliasz, P. 2002, Factor Augmented Vector Autoregressions (FVARs) and the Analysis of Monetary Policy, *Quarterly Journal of Economics.*

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.

Greene, W. 2003, *Econometric Analysis*, 5 edn, Prentice Hall, New Jersey.

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.

Pagan, A. 1984, Econometric Issues in the Analysis of Regressions with Generated Regressors, *International Economic Review* **25**, 221–47.

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.

Table 1: Coverage Rates, $h = 4$, $r = 2$

Method		A: (5b)+(4)		B: (5a)+(3)		C: (5c)+(3)		D: F known	
N	T	$\widehat{y}_{T+h\|T}$	$\widehat{y}_{T+h}$	$\widehat{y}_{T+h\|T}$	$\widehat{y}_{T+h}$	$\widehat{y}_{T+h\|T}$	$\widehat{y}_{T+h}$	$\widehat{y}_{T+h\|T}$	$\widehat{y}_{T+h}$
DGP 1: $b = 0$, $\sigma_v^2(i) = 1\ \forall i$									
50	50	0.95	0.94	0.93	0.94	0.93	0.94	0.91	0.93
100	50	0.93	0.94	0.91	0.94	0.91	0.93	0.91	0.94
200	50	0.93	0.93	0.91	0.93	0.91	0.93	0.89	0.93
50	100	0.95	0.95	0.95	0.95	0.93	0.95	0.93	0.94
50	200	0.94	0.96	0.92	0.96	0.88	0.95	0.95	0.95
200	100	0.94	0.94	0.94	0.94	0.93	0.94	0.94	0.93
100	200	0.96	0.95	0.94	0.95	0.93	0.95	0.94	0.95
200	200	0.96	0.95	0.95	0.95	0.94	0.95	0.94	0.95
100	400	0.97	0.95	0.96	0.95	0.94	0.95	0.94	0.95
DGP 2: $b = 0$, $\sigma_v^2(i) \sim U(.5, 1.5)\ \forall i$									
50	50	0.95	0.94	0.93	0.94	0.93	0.94	0.91	0.93
100	50	0.94	0.94	0.92	0.94	0.92	0.94	0.91	0.94
200	50	0.93	0.93	0.91	0.93	0.91	0.93	0.89	0.93
50	100	0.95	0.95	0.94	0.95	0.92	0.95	0.93	0.94
50	200	0.93	0.95	0.91	0.95	0.87	0.95	0.95	0.95
200	100	0.94	0.94	0.93	0.94	0.93	0.94	0.94	0.93
100	200	0.96	0.95	0.94	0.95	0.94	0.95	0.94	0.95
200	200	0.95	0.95	0.94	0.95	0.94	0.95	0.94	0.95
100	400	0.98	0.96	0.96	0.96	0.93	0.96	0.94	0.95
DGP 3: $b = .5$, $\sigma_e^2(i) = 1\ \forall i$									
50	50	0.82	0.94	0.81	0.94	0.86	0.94	0.91	0.93
100	50	0.84	0.94	0.83	0.93	0.88	0.94	0.91	0.94
200	50	0.87	0.93	0.85	0.93	0.90	0.93	0.89	0.93
50	100	0.85	0.95	0.84	0.95	0.89	0.95	0.93	0.94
50	200	0.73	0.95	0.69	0.95	0.78	0.95	0.95	0.95
200	100	0.89	0.94	0.87	0.94	0.93	0.94	0.94	0.93
100	200	0.83	0.95	0.80	0.95	0.92	0.96	0.94	0.95
200	200	0.86	0.95	0.83	0.95	0.93	0.95	0.94	0.95
100	400	0.80	0.95	0.76	0.95	0.94	0.95	0.94	0.95
DGP 4: $b = .5$, $\sigma_e^2(i) \sim U(.5, 1.5)\ \forall i$									
50	50	0.82	0.94	0.80	0.93	0.85	0.94	0.91	0.93
100	50	0.85	0.93	0.83	0.93	0.89	0.94	0.91	0.94
200	50	0.86	0.93	0.85	0.93	0.91	0.93	0.89	0.93
50	100	0.83	0.95	0.81	0.95	0.90	0.95	0.93	0.94
50	200	0.65	0.94	0.63	0.94	0.69	0.94	0.95	0.95
200	100	0.89	0.94	0.87	0.94	0.92	0.94	0.94	0.93
100	200	0.83	0.95	0.80	0.95	0.90	0.96	0.94	0.95
200	200	0.85	0.95	0.83	0.95	0.93	0.95	0.94	0.95
100	400	0.80	0.95	0.77	0.95	0.93	0.96	0.94	0.95

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

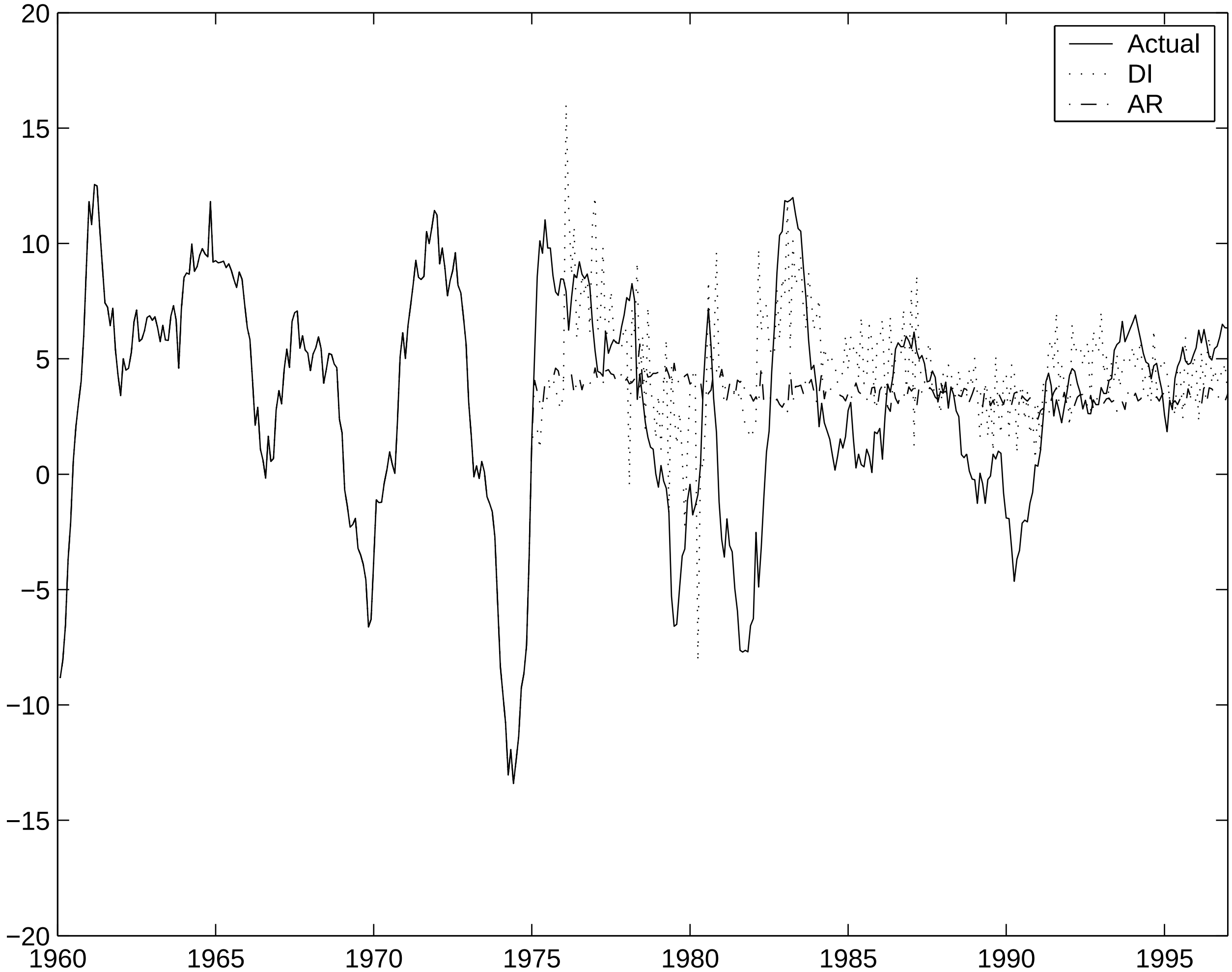

Figure 2a: Diffusion Index Forecast and Confidence Intervals: Growth Rate of Industrial Production

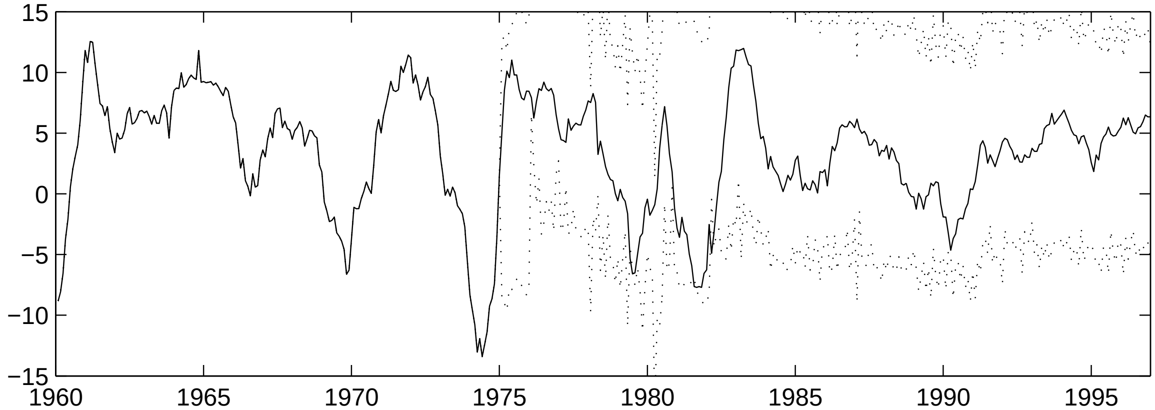

Figure 2b: AR Forecast and Confidence Intervals: Industrial Production

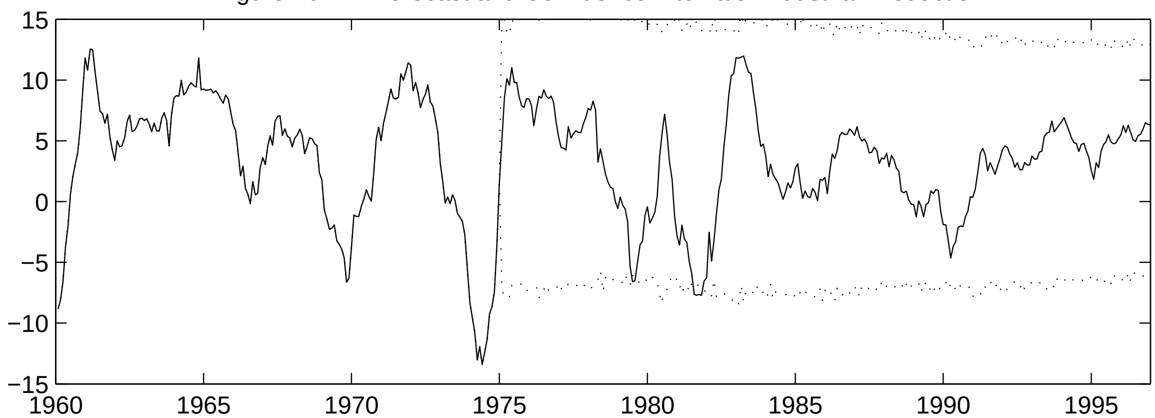

Figure 3: 12−Month Ahead Forecast: Inflation

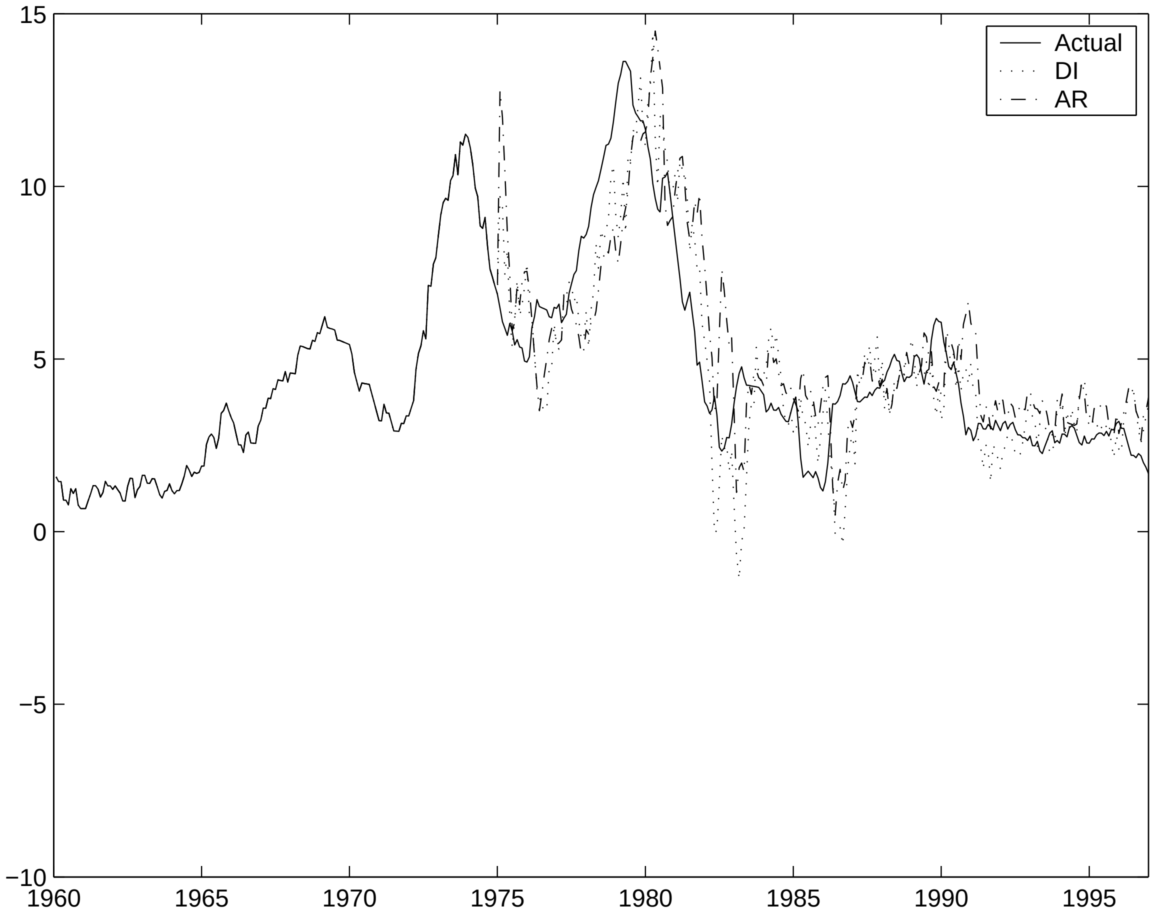

Figure 4a: Diffusion Index Forecast and Confidence Intervals: Inflation

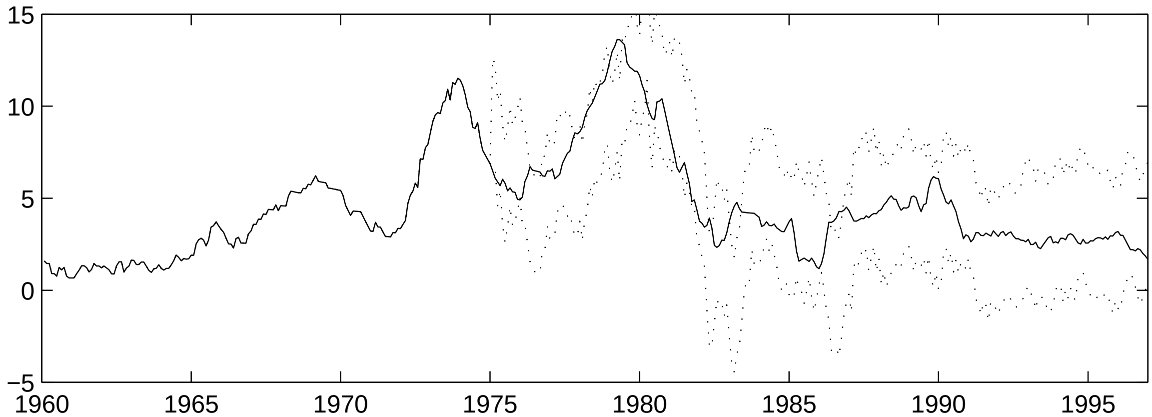

Figure 4b: AR Forecast and Confidence Intervals: Inflation