

CONFIDENCE INTERVALS FOR DIFFUSION INDEX FORECASTS AND INFERENCE FOR FACTOR-AUGMENTED REGRESSIONS

BY JUSHAN BAI AND SERENA NG¹

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 first to estimate common factors from the panel of data by the method of principal components and then to 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 $\sqrt{T}/N \rightarrow 0$. The conditional mean predicted by the estimated factors is $\min[\sqrt{T}, \sqrt{N}]$ 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 that are valid regardless of the magnitude of N/T and that can also be used when the factors are nonstationary.

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

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.² Several institutions, including the Treasury and the European Central Bank, are studying the empirical properties of these factor forecasts.³ In macroeconomics, [Bernanke, Boivin, and Elias \(2005\)](#) found that the information exploited in factor-augmented vector autoregressions (FAVAR) is important to identify the monetary transmission mechanism properly. In spite of its increasing popularity with practitioners, how to conduct inference in factor-augmented regressions is not well understood. This is a nontrivial problem because the regression model involves “estimated factors.” 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 and the conditional mean, as well as the forecast.

¹The authors acknowledge financial support from NSF Grants SES-0137084 and SES-0136923. We thank three anonymous referees, an editor, and seminar participants at Yale, Columbia, and Princeton for useful comments. This paper was also presented at the Conference on Common Features in London, 2004.

²See, for example, [Stock and Watson \(2001, 2002b\)](#), [Cristadoro, Forni, Reichlin, and Giovanni \(2001\)](#), [Forni, Hallin, Lippi, and Reichlin \(2001\)](#), [Artis, Banerjee, and Marcellino \(2001\)](#), [Banerjee, Marcellino, and Masten \(2004\)](#), and [Shintani \(2002\)](#).

³See, for example, [Angelini, Henry, and Mestre \(2001\)](#).

Suppose information is available on a large number of predictors x_{it} ($i = 1, 2, \dots, N$; $t = 1, 2, \dots, T$) and a smaller set of other observable variables W_t , such as lags of y_t . Consider

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

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} that contain information about F_t . We refer to

$$(2) \quad x_{it} = \lambda_i' F_t + e_{it}$$

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) constitute the *diffusion index forecasting model* of Stock and Watson (2002a). If $h = 1$ and $y_{t+1} = (F_{t+1}', W_{t+1}')'$, (1) is the FAVAR of Bernanke, Boivin, and Elias (2005). 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 shocks that generate comovements in the data.

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

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

where $z_t = (F_t', W_t')'$. However, 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

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

where $\hat{z}_t = (\tilde{F}_t', W_t')'$. We use a tilde for estimates of the factor model (2), while hatted variables are estimated from (1). To be precise, $\hat{\alpha}$ and $\hat{\beta}$ are the least squares estimates obtained from a regression of y_{t+h} on $\tilde{F}_t$ 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 $\hat{\alpha}$ and $\hat{\beta}$ are functions of *estimated regressors* $\tilde{F}_1, \tilde{F}_2, \dots, \tilde{F}_{T-h}$, and that $\hat{y}_{T+h|T}$ itself also depends on $\tilde{F}_T$. Thus, to study the behavior of $\hat{y}_{T+h|T}$ and of the forecast error $\hat{\varepsilon}_{T+h}$, we must examine the statistical properties of the estimated parameters as well as those of the estimated factors. Stock and Watson (2002a) showed that $\hat{y}_{T+h|T}$ is consistent for $y_{T+h|T}$. However, for hypothesis testing, such as to construct standard errors for the impulse response of a FAVAR, to provide a confidence interval for the latent conditional mean,

and to evaluate the uncertainty of a diffusion index forecast, we need the limiting distributions of $(\hat{\alpha}, \hat{\beta})$, $\hat{y}_{T+h|T}$, and $\hat{\varepsilon}_{T+h}$. Our main contribution is to provide these results and clarify the role of N and T in each step. As we will see, it is only when N is large relative to T that the effect of estimated regressors can be completely ignored, and the precise condition on N and T depends on whether interest is in inference of the regression coefficients, the conditional mean, or the forecast. We also provide a covariance matrix estimator that is robust to weak cross-section correlation and heteroskedasticity of unknown form. Supplementary lemmas and simulation results are available in our working paper (Bai and Ng (2004)).

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, $F = (F_1, \dots, F_T)'$ is $T \times r$, r is the number of common factors, $\Lambda = (\lambda_1, \dots, \lambda_N)'$ is $N \times r$, and e is a $T \times N$ error matrix.

ASSUMPTION A—Common Factors: *The factor process satisfies $E\|F_t\|^4 \leq M$ and $\frac{1}{T} \sum_{t=1}^T F_t F_t' \xrightarrow{p} \Sigma_F > 0$, an $r \times r$ nonrandom matrix.*

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, $N^{-1}\Lambda'\Lambda \xrightarrow{p} \Sigma_\Lambda > 0$, an $r \times r$ nonrandom matrix, as $N \rightarrow \infty$.*

ASSUMPTION C—Time and Cross-Section Weak Dependence and Heteroskedasticity:

1. For all (i, t) , $E(e_{it}) = 0$, $E|e_{it}|^8 \leq M$.
2. There exist $E(e_{it}e_{js}) = \sigma_{ij,ts}$ and $|\sigma_{ij,ts}| \leq \bar{\sigma}_{ij}$ for all (t, s) , and $|\sigma_{ij,ts}| \leq \tau_{ts}$ for all (i, j) such that $\frac{1}{N} \sum_{i,j=1}^N \bar{\sigma}_{ij} \leq M$, $\frac{1}{T} \sum_{t,s=1}^T \tau_{ts} \leq M$, and $\frac{1}{NT} \sum_{i,j,t,s=1}^N |\sigma_{ij,ts}| \leq M$.
3. For every (t, s) , $E|N^{-1/2} \sum_{i=1}^N [e_{is}e_{it} - E(e_{is}e_{it})]|^4 \leq M$.
4. For each t , $\frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i e_{it} \xrightarrow{d} N(0, \Gamma_t)$, where $\Gamma_t = \lim_{N \rightarrow \infty} \frac{1}{N} \times \sum_{i=1}^N \sum_{j=1}^N E(\lambda_i \lambda_j' e_{it} e_{jt})$.

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

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

1. $\frac{1}{T} \sum_{t=1}^T z_t z_t' \xrightarrow{p} \Sigma_{zz} > 0$;
2. $\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t \varepsilon_{t+h} \xrightarrow{d} N(0, \Sigma_{zz,\varepsilon})$, 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, weak time series, and cross-section dependence in the idiosyncratic component, leading to the *approximate factor structure* of Chamberlain and Rothschild (1983). These assumptions are more general than a strict factor model. Assumption **D** is standard in factor analysis and Assumption **E** is standard for regression analysis. These assumptions are similar to those of Stock and Watson (2002a), except that they allow time-varying factor loadings. We only consider time-invariant factor loadings.

2.1. Estimation

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

THEOREM 1—Estimation: Suppose Assumptions **A**–**E** hold. If $\sqrt{T}/N \rightarrow 0$, then

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

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

$$(3) \quad \widehat{\text{Avar}}(\hat{\delta}) = \left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{z}_t \hat{z}_t' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{\varepsilon}_{t+h}^2 \hat{z}_t \hat{z}_t' \right) \left(\frac{1}{T} \sum_{t=1}^{T-h} \hat{z}_t \hat{z}_t' \right)^{-1}.$$

As is well known, the factor model is unidentified because $\alpha'LL^{-1}F_t = \alpha'F_t$ for any invertible matrix L . Theorem 1 is a result that pertains to the difference between $\hat{\alpha}$ and the space spanned by α . Having estimated factors as regressors does not affect consistency of the parameter estimates. Stock and Watson (2002a) showed consistency of $\hat{\delta}$ for δ . Here we establish the rate of convergence and the limiting distribution.

Formula (3) is robust to heteroskedasticity. Under homoskedasticity such that $E(\varepsilon_{t+h}^2 | z_t) = \sigma_{\varepsilon}^2 \forall t$ and letting $\hat{\sigma}_{\varepsilon}^2 = \frac{1}{T} \sum_{t=1}^{T-h} \hat{\varepsilon}_{t+h}^2$, a consistent estimate of $\widehat{\text{Avar}}(\hat{\delta})$ is

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

Theorem 1 is useful in rather broader contexts, because having to conduct inference when the latent common factors are replaced by estimates is not uncommon, an example being the Phillips curve, where the estimated common factors serve as proxies for the unobserved state of the economy.

A new tool in empirical work is the FAVAR, which essentially augments the principal component estimates of the factors to an otherwise standard vector autoregression (VAR).⁴ 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

$$y_{t+1} = \sum_{k=0}^p a_{11}(k)y_{t-k} + \sum_{k=0}^p a_{12}(k)F_{t-k} + \varepsilon_{1t+1},$$

$$F_{t+1} = \sum_{k=0}^p a_{21}(k)y_{t-k} + \sum_{k=0}^p a_{22}(k)F_{t-k} + \varepsilon_{2t+1},$$

where $a_{11}(k)$ and $a_{21}(k)$ are coefficients on y_{t-k} , while $a_{12}(k)$ and $a_{22}(k)$ are coefficients on F_{t-k} . Consider estimation of the FAVAR with F_t replaced by $\tilde{F}_t$.

THEOREM 2—FAVAR: *Consider a p th order vector autoregression in q observable variables y_t and r factors, $\tilde{F}_t$, estimated by the method of principal components. Let $\hat{z}_t = (y'_t, \dots, y'_{t-p}, \tilde{F}'_t, \dots, \tilde{F}'_{t-p})'$ and $\hat{Y}_t = (y'_t, \tilde{F}'_t)'$. Also let $\hat{Y}_{jt}$ be the j th element of $\hat{Y}_t$. For $j = 1, \dots, q + r$, let δ_j be the coefficient vector in the $\hat{Y}_{jt}$ equation and let $\hat{\delta}_j$ be obtained by least squares from regressing $\hat{Y}_{jt+1}$ on $\hat{z}_t$. Let $\hat{\varepsilon}_{jt+1} = \hat{Y}_{jt+1} - \hat{\delta}'_j \hat{z}_t$. Under Assumptions A–E and if $\sqrt{T}/N \rightarrow 0$ as $N, T \rightarrow \infty$,*

$$\sqrt{T}(\hat{\delta}_j - \delta_j)$$

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

Theorem 2 states that the parameter estimates for those equations with $\tilde{F}_{t+1}$ as regressors are also $\sqrt{T}$ consistent, just as when y_{t+1} is the dependent variable considered in Theorem 1. Since impulse response functions are based on estimates of the FAVAR, Theorem 2 enables calculation of the standard errors. Although the condition $\sqrt{T}/N \rightarrow 0$ (or equivalently $\sqrt{T}/\min[N, T] \rightarrow 0$) is not stringent, it puts discipline on when estimated factors can be used in regression analysis.

Pagan (1984) considered the model $y_t = \alpha' x_t^e + \varepsilon_t$ and $x_t = \gamma' z_t + u_t$ with $x_t^e = \gamma' z_t$. Let $\hat{\gamma}$ be the least squares of x_t on z_t . Write $y_t = \alpha' \hat{x}_t + \varepsilon_t + \alpha'(x_t^e -$

⁴See, for example, Bernanke and Boivin (2003), Bernanke, Boivin, and Elias (2005), Giannone, Reichlin, and Sala (2002), and Favero, Marcellino, and Neglia (2005).

$\widehat{x}_t) = \alpha' \widehat{x}_t + \varepsilon_t + \widehat{u}_t$, where $\widehat{x}_t = \widehat{\gamma}' z_t$ and $\widehat{u}_t = -(\widehat{\gamma} - \gamma)' z_t$. The least squares regression of y_t on $\widehat{x}_t$ yields $\widehat{\alpha} - \alpha = (\widehat{x}' \widehat{x})^{-1} \widehat{x}' (\varepsilon + \widehat{u})$. However,

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

which is nonnegligible, so estimated regressors have an effect on parameter estimation. In our case, the corresponding term is $O_p(\frac{\sqrt{T}}{\min[N, T]})$.⁵ Because N and T are large by assumption, our term becomes negligible. The limiting distributions in Theorems 1 and 2 are thus the same as if $\widehat{z}_t$ was the true regressor.

2.2. Prediction Intervals

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 to construct a confidence interval for the conditional mean. Note that the equation

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

has two components, which arise from estimating δ and F_t . Under Assumptions A–D, Bai (2003) showed that if $\sqrt{N}/T \rightarrow 0$, then for each t ,

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

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

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

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

where $\text{var}(\widehat{y}_{T+h|T}) = \frac{1}{T} \widehat{z}_T' \text{Avar}(\widehat{\delta}) \widehat{z}_T + \frac{1}{N} \widehat{\alpha}' \text{Avar}(\widetilde{F}_T) \widehat{\alpha}$.

⁵If $N = 1$, this term would be $O_p(\sqrt{T})$, much larger than Pagan's $O_p(1)$. This is because in Pagan's model, only a finite number of parameters γ are estimated in the first stage and we need to estimate T unknown quantities $F_1, \dots, F_T$.

Because the two terms in $\text{var}(\hat{y}_{T+h|T})$ vanish at different rates, the overall convergence rate for $\hat{y}_{T+h|T}$ is $\min[\sqrt{N}, \sqrt{T}]$. In a standard setting, $\text{var}(\hat{y}_{T+h|T})$ 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 , efficiency improves with the number of predictors used to estimate F_t . This is because in the present setting, a large N enables better 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 of standard factor analysis, which assumes a fixed N . With a fixed N , consistent estimation of the factor space is not possible however large T becomes.

When the objective is to forecast y_{T+h} , 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

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

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

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

COROLLARY 1: *Under the assumptions of Theorem 3 and assuming ε_t is $N(0, \sigma_\varepsilon^2)$, then the forecasting error $\hat{\varepsilon}_{T+h}$ is*

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

A confidence interval can be obtained upon replacing σ_ε^2 by its consistent estimate, $\frac{1}{T} \sum_{i=1}^T \hat{\varepsilon}_i^2$. The above formula extends the textbook definition of forecast uncertainty, which allows only for estimation error in $\hat{\delta}$ (e.g., Greene (2003, Chapter 6)), to also permit using $\tilde{F}_t$ as regressors. Therefore, $\text{var}(\hat{y}_{T+h|T})$ reflects both parameter uncertainty and regressor uncertainty. However, it vanishes at rate $\min[T, N]$, rather than the usual rate of T . In large samples, $\text{var}(\hat{\varepsilon}_{T+h})$ is dominated by σ_ε^2 , just as when all predictors are observed, but if we ignore $\text{var}(\hat{y}_{T+h|T})$, σ_ε^2 alone will underestimate the true forecast uncertainty for a given T and N .

It is clear from Theorem 3 and Corollary 1 that a consistent estimate of $\text{var}(\hat{y}_{T+h|T})$ is required to proceed with inference. In view of (5), an estimate of $\text{Avar}(\tilde{F}_t)$ is⁶

$$\widehat{\text{Avar}}(\tilde{F}_t) = \tilde{V}^{-1} \tilde{F}_t' \tilde{V}^{-1},$$

⁶ This formula is obtained by first substituting $\tilde{F}$ for F and noting that $\tilde{Q} = \tilde{F}'\tilde{F}/T$ is an r -dimensional identity matrix by construction ($\tilde{Q}$ is an estimate for QH' whose limit is an identity).

where the $r \times r$ matrix $\tilde{\Gamma}_t$ can be one of

$$(5a) \quad \tilde{\Gamma}_t = \frac{1}{N} \sum_{i=1}^N \tilde{e}_{it}^2 \tilde{\lambda}_i \tilde{\lambda}_i',$$

$$(5b) \quad \tilde{\Gamma}_t = \tilde{\sigma}_e^2 \frac{1}{N} \sum_{i=1}^N \tilde{\lambda}_i \tilde{\lambda}_i',$$

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

with $n/\min[N, T] \rightarrow 0$ in (5c), where $\tilde{e}_{it} = x_{it} - \tilde{\lambda}_i' \tilde{F}_t$. The various specifications of $\tilde{\Gamma}_t$ 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 , which is estimated by $\tilde{\sigma}_e^2 = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \tilde{e}_{it}^2$. 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. For small cross-section correlation in the errors, constraining them to be zero could sometimes be desirable because sampling variability from estimating them could generate nontrivial efficiency loss. The estimators (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 $\tilde{F}_t$ can be estimated by (5c). Consistency of $\tilde{\Gamma}_t$ 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 (cross-section and heteroskedastic autocorrelation consistent), 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 estimators for $\text{Avar}(\hat{\delta})$ and $\text{Avar}(\tilde{F}_T)$ are given, prediction intervals can be easily constructed. For example, the 95% confidence interval for the conditional mean $y_{T+h|T}$ is

$$(\hat{y}_{T+h|T} - 1.96\sqrt{\widehat{\text{var}}(\hat{y}_{T+h|T})}, \hat{y}_{T+h|T} + 1.96\sqrt{\widehat{\text{var}}(\hat{y}_{T+h|T})})$$

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

$$(\hat{y}_{T+h|T} - 1.96\sqrt{\hat{\sigma}_e^2 + \widehat{\text{var}}(\hat{y}_{T+h|T})}, \hat{y}_{T+h|T} + 1.96\sqrt{\hat{\sigma}_e^2 + \widehat{\text{var}}(\hat{y}_{T+h|T})}).$$

Theorem 3 fills an important void in the diffusion index forecasting literature, because it goes beyond the consistency result to establish asymptotic

normality. The result 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. In [Bai and Ng \(2006\)](#), we developed tests to determine whether the observed variables are good proxies for the latent factors. Part of 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) because they assume cross sectionally uncorrelated idiosyncratic errors. To understand the problem, it helps to consider first the cross-section regression $y_i = \beta' \lambda_i + e_i$, where λ_i is a $r \times 1$ vector of observed regressors. The covariance of $\hat{\beta}$ involves $\Gamma = \lim_{N \rightarrow \infty} \frac{1}{N} \Lambda' \Omega \Lambda$, where $\Omega = E(ee')$ with $e = (e_1, \dots, e_N)'$, and Λ is the $N \times r$ regressor matrix. Here, the “natural” estimator,

$$\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j' \hat{e}_i \hat{e}_j = \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i \hat{e}_i \right) \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i \hat{e}_i \right)',$$

is inconsistent because $\frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i \hat{e}_i$ 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. Whereas time series data have a natural ordering, it is possible to construct heteroskedastic and autocorrelation consistent (HAC) estimators as in [Newey and West \(1987\)](#) and [Andrews \(1991\)](#).

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. The common practice in cross-section regressions is to assume $E(e_i e_j) = 0$, $i \neq j$, so that $\Omega = \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \lambda_i \lambda_i' e_i^2$.

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 $(r \times r)$ and can be estimated with $n < N$ observations. An estimator along these lines was considered in [Driscoll and Kraay \(1998\)](#). 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 estimate Γ_t consistently, we require Γ_t not to depend on t (see Assumption C.4) so that we can also use observations outside period t to estimate Γ .

THEOREM 4: *Suppose Assumptions A–D hold. In addition, $E(e_{it}e_{jt}) = \sigma_{ij}$ for all t , so $\Gamma_t = \Gamma$ not depending on t . Let $\tilde{\Gamma}$ be defined as in (5c). Then $\|\tilde{\Gamma} - H^{-1}\Gamma H^{-1}\| \xrightarrow{p} 0$ if $\frac{n}{\min[N, T]} \rightarrow 0$.*

The conditions that $n/N \rightarrow 0$ and $n/T \rightarrow 0$ are not restrictive. The simple rule we use in the simulations below is $n = \min[\sqrt{N}, \sqrt{T}]$.

4. FINITE SAMPLE PROPERTIES

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

$$\begin{aligned} x_{it} &= \lambda_i' F_t + e_{it} & (i = 1, \dots, N, t = 1, \dots, T), \\ F_{jt} &= \rho_j F_{jt-1} + (1 - \rho_j^2)^{1/2} u_{jt} & (j = 1, \dots, r, \rho_j = (0.8)^j), \\ e_t &= v_t \bar{\Omega}(b)^{1/2}, \end{aligned}$$

where $e_t = (e_{1t}, \dots, e_{Nt})'$, $v_t = (v_{1t}, \dots, v_{Nt})'$, 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 $\bar{\Omega}(b)$, which is an $N \times N$ Toeplitz matrix whose j th main diagonal is b^j if $j \leq 10$ and is zero otherwise.⁷ 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 variations of the data generating program (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 $(0.5, 1.5)$. In DGP 3 (homoskedasticity and cross sectionally correlated errors), we let $b = 0.5$ and $\sigma_v^2(i) = 1$. In DGP 4 (heteroskedasticity and cross sectionally correlated errors), we set $b = 0.5$ and σ_v^2 is again uniformly distributed on $(0.5, 1.5)$.

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} .

⁷The results are similar if the innovation variance of u_t is not scaled by $1 - \rho_j^2$. The scaling enables us to control the size of the common to the idiosyncratic component.

Three types of confidence intervals will be presented:

$$\text{A: (5b) + (4);} \quad \text{B: (5a) + (3);} \quad \text{C: (5c) + (3).}$$

When constructing (5c), we use the first $n = \min[\sqrt{N}, \sqrt{T}]$ series to construct the CS-HAC. 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 $\text{Avar}(\tilde{F}_t)$). This is labelled D.

The coverage rates are reported in Table I for (i) the estimated conditional mean $\hat{y}_{T+h|T}$ and (ii) the diffusion index forecast $\hat{y}_{T+h}$. The coverage rates are generally close to the nominal rate of 0.95, although three results are noteworthy. First, when N is small, the coverage of C is too low for DGPs 1 and 2. This can be attributed to the use of CS-HAC when in fact there is no cross-section correlation. Second, for DGPs 3 and 4, the coverage of A and B is always too low since these types 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 relationship 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.

5. SUMMARY

We first establish the rate of convergence and the limiting distribution for the estimated parameters of the forecasting model and of the factor-augmented regressions when the factors are unobserved but estimated. We then derive similar results for the predicted conditional mean and for the forecasting error. For predictive inference, we suggest how the covariance matrix of cross-correlated and heteroskedastic errors can be consistently estimated.

Dept. of Economics, New York University, 269 Mercer Street, New York, NY 10003, U.S.A.; and School of Economics and Management, Tsinghua University, Beijing, China; Jushan.Bai@nyu.edu

and

Dept. of Economics, University of Michigan, Ann Arbor, MI 48109, U.S.A.; Serena.Ng@umich.edu

Manuscript received August, 2004; final revision received August, 2005.

APPENDIX: PROOFS

LEMMA A.1: Let $z'_t = (F'_t, W'_t)'$ and $\hat{z}_t = (\tilde{F}'_t, W'_t)'$. Let $\delta_{NT}^2 = \min[N, T]$ and $H = \tilde{V}^{-1}(\tilde{F}'F/T)(A'A/N)$. Under Assumptions A–E:

- (i) $\frac{1}{T} \sum_{t=1}^T \|\tilde{F}_t - HF_t\|^2 = O_p(\delta_{NT}^{-2})$;
- (ii) $\frac{1}{T} \sum_{t=1}^T (\tilde{F}_t - HF_t)z'_t = O_p(\delta_{NT}^{-2})$;
- (iii) $\frac{1}{T} \sum_{t=1}^T (\tilde{F}_t - HF_t)\hat{z}'_t = O_p(\delta_{NT}^{-2})$;

TABLE I
COVERAGE RATES, $h = 4, r = 2$

Method		A: (5b) + (4)		B: (5a) + (3)		C: (5c) + (3)		D: F Known	
N	T	$\hat{y}_{T+h T}$	$\hat{y}_{T+h}$	$\hat{y}_{T+h T}$	$\hat{y}_{T+h}$	$\hat{y}_{T+h T}$	$\hat{y}_{T+h}$	$\hat{y}_{T+h T}$	$\hat{y}_{T+h}$
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: $b = 0, \sigma_v^2(i) \sim U(0.5, 1.5) \forall i$									
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: $b = 0.5, \sigma_e^2(i) = 1 \forall i$									
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: $b = 0.5, \sigma_e^2(i) \sim U(0.5, 1.5) \forall i$									
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

$$(vi) \quad \frac{1}{T} \sum_{t=1}^T (\tilde{F}_t - HF_t) \varepsilon_{t+h} = O_p(\delta_{NT}^{-2}).$$

PROOF: See the working version of this paper.

Q.E.D.

PROOF OF THEOREM 1: Adding and subtracting terms, the model can be written as

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

In matrix notation, $Y = \hat{z} \delta + \varepsilon + (FH' - \tilde{F})H^{-1}\alpha$, where $Y = (y_{h+1}, \dots, y_T)'$, $\varepsilon = (\varepsilon_{h+1}, \dots, \varepsilon_T)'$, and $\hat{z} = (\hat{z}_1, \dots, \hat{z}_{T-h})'$. The ordinary least squares estimator is $\hat{\delta} = (\hat{z}'\hat{z})^{-1}\hat{z}'Y$. Thus,

$$\hat{\delta} - \delta = (\hat{z}'\hat{z})^{-1}\hat{z}'\varepsilon + (\hat{z}'\hat{z})^{-1}\hat{z}'(FH' - \tilde{F})H^{-1}\alpha$$

or

$$\begin{aligned} \sqrt{T}(\hat{\delta} - \delta) &= (T^{-1}\hat{z}'\hat{z})^{-1}T^{-1/2}\hat{z}'\varepsilon + (T^{-1}\hat{z}'\hat{z})^{-1}[T^{-1/2}\hat{z}'(FH' - \tilde{F})]H^{-1}\alpha. \end{aligned}$$

The second term on the right is $o_p(1)$ because $T^{-1/2}\hat{z}'(FH' - \tilde{F}) = O_p(T^{1/2}/\min(N, T)) = o_p(1)$ when $\sqrt{T}/N \rightarrow 0$, by Lemma A.1. For the first term, $T^{-1/2}\hat{z}'\varepsilon = T^{-1/2}(\varepsilon'\tilde{F}, \varepsilon'W)'$. Now $T^{-1/2}\tilde{F}'\varepsilon = T^{-1/2}HF'\varepsilon + T^{-1/2}(\tilde{F} - FH')'\varepsilon$, where the second term is $o_p(1)$ when $\sqrt{T}/N \rightarrow 0$ by Lemma A.1. Thus, $T^{-1/2}\hat{z}'\varepsilon = T^{-1/2}(\varepsilon'FH', \varepsilon'W)' + o_p(1) = T^{-1/2}\Phi z'\varepsilon + o_p(1)$, where $\Phi = \text{diag}(H, I)$ is a block diagonal matrix. Thus,

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

Since $z'\varepsilon/\sqrt{T} \xrightarrow{d} N(0, \Sigma_{zz, \varepsilon})$ by Assumption E.2, the above is asymptotically normal. The asymptotic variance matrix is the probability limit of

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

Define $H_0 = \text{plim } H = V^{-1}Q\Sigma_\Lambda$ and $\Phi_0 = \text{plim } \Phi = \text{diag}(H_0, I)$. Now $T^{-1}\hat{z}'\hat{z} = \Phi(T^{-1}z'z)\Phi' + o_p(1) \xrightarrow{p} \Phi_0\Sigma_{zz}\Phi_0'$. The limiting variance or the limit of (A.2)

is

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

Since $HF_t = \tilde{F}_t + o_p(1)$ 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, $\widehat{\text{Avar}}(\widehat{\delta}) = (T^{-1} \widehat{z}' \widehat{z})^{-1} (T^{-1} \sum_{t=1}^T \widehat{\varepsilon}_{t+h}^2 \times \widehat{z}_t \widehat{z}_t') (T^{-1} \widehat{z}' \widehat{z})^{-1}$ is a consistent estimator for Σ_{δ} . This completes the proof of Theorem 1. Q.E.D.

PROOF OF THEOREM 2: Without loss of generality, consider a FAVAR(1). For FAVAR(1), Y_t and z_t coincide, 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 $\tilde{F}_t$ is

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

where $b_{11} = a_{11}$, $b_{12} = a_{12}H^{-1}$, $b_{21} = Ha_{21}$, and $b_{22} = Ha_{22}H^{-1}$. Let $\widehat{z}_t = (y_t', \tilde{F}_t')'$ and $\widehat{z} = (\widehat{z}_1', \dots, \widehat{z}_T')'$. The j th equation is $\widehat{z}_{jt+1} = \delta_j' \widehat{z}_t + u_{jt+1}^1 + u_{jt+1}^2 + u_{jt+1}^3$. Thus,

$$\begin{aligned}\sqrt{T}(\widehat{\delta}_j - \delta_j) &= (T^{-1} \widehat{z}' \widehat{z})^{-1} T^{-1/2} \widehat{z}' (u_j^1 + u_j^2 + u_j^3) \\ &= (T^{-1} \widehat{z}' \widehat{z})^{-1} T^{-1/2} \widehat{z}' u_j^1 + o_p(1),\end{aligned}$$

where the second equality follows, by Lemma A.1, from $T^{-1/2} \widehat{z}' u_j^2 = O_p(\sqrt{T}/\min[N, T])$ and $T^{-1/2} \widehat{z}' u_j^3 = O_p(\sqrt{T}/\min[N, T])$. Now $T^{-1/2} \widehat{z}' u_j^1 = T^{-1/2} \times \sum_{t=1}^T \widehat{z}_t u_{jt+1}^1$. For $j \leq q$, u_{jt+1}^1 is the j th component of ε_{t+1} . This case is treated in (A.1) and the limiting variance is, compare with (3) and (A.2),

$$(A.3) \quad \text{plim} \left(\frac{1}{T} \sum_{t=1}^T \widehat{z}_t \widehat{z}_t' \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^T (u_{jt+1}^1)^2 \widehat{z}_t \widehat{z}_t' \right) \left(\frac{1}{T} \sum_{t=1}^T \widehat{z}_t \widehat{z}_t' \right)^{-1},$$

which can be consistently estimated upon replacing u_{jt+1}^1 by $\hat{u}_{jt+1}^1 = Y_{jt+1} - \hat{\delta}_j' \hat{z}_t$.

For $j = q + 1, \dots, q + r$, u_{jt+1}^1 is the k th 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 k th 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.1) in the previous proof implies that the limiting variance is given by (A.3) with $u_{jt+1}^1 = \iota_k' H \varepsilon_{2t+1}$. Again, replacing u_{jt+1}^1 by $\hat{u}_{jt+1}^1$ gives the same limiting variance. Q.E.D.

PROOF OF THEOREM 3: Begin by rewriting

$$\begin{aligned} \hat{y}_{T+h|T} - y_{T+h|T} &= \hat{\alpha}' \tilde{F}_T + \hat{\beta}' W_T - \alpha' F_T - \beta' W_T \\ &= (\hat{\alpha} - H^{-1} \alpha)' \tilde{F}_T + \alpha' H^{-1} (\tilde{F}_T - H F_T) + (\hat{\beta} - \beta)' W_T \\ &= \hat{z}_T' (\hat{\delta} - \delta) + \alpha' H^{-1} (\tilde{F}_T - H F_T) \\ &= T^{-1/2} \hat{z}_T' [\sqrt{T} (\hat{\delta} - \delta)] + N^{-1/2} \alpha' H^{-1} [\sqrt{N} (\tilde{F}_T - H F_T)]. \end{aligned}$$

Both $\sqrt{T} (\hat{\delta} - \delta)$ and $\sqrt{N} (\tilde{F}_T - H F_T)$ are asymptotically normal. They are also asymptotically independent because the limit of $\sqrt{T} (\hat{\delta} - \delta)$ is determined by $(\varepsilon_1, \dots, \varepsilon_T)$ and that of $\sqrt{N} (\tilde{F}_T - H F_T)$ is determined by e_{it} for $i = 1, 2, \dots, N$. Noting that $T^{-1/2} (\hat{z}_T - z_T) = o_p(1)$, an estimate for the variance of $T^{-1/2} z_T' \sqrt{T} (\hat{\delta} - \delta)$ is $\frac{1}{T} z_T' \text{Avar}(\hat{\delta}) z_T$, which in turn is estimated by $\frac{1}{T} \hat{z}_T' \text{Avar}(\hat{\delta}) \hat{z}_T$. Similarly, an estimate for the variance of the second term is $\frac{1}{N} \alpha' H^{-1} \text{Avar}(\tilde{F}_T) H^{-1} \alpha$, which is in turn estimated by $\frac{1}{N} \hat{\alpha}' \text{Avar}(\tilde{F}_T) \hat{\alpha}$. Due to asymptotic independence, an estimate for the forecasting error variance is $\text{var}(\hat{y}_{T+h|T}) = \frac{1}{T} \hat{z}_T' \text{Avar}(\hat{\delta}) \hat{z}_T + \frac{1}{N} \hat{\alpha}' \text{Avar}(\tilde{F}_T) \hat{\alpha}$; thus, $(\hat{y}_{T+h|T} - y_{T+h|T}) / \text{var}(\hat{y}_{T+h|T})^{1/2} \xrightarrow{d} N(0, 1)$. Q.E.D.

The proof of the following lemma is in the working version of this paper.

LEMMA A.2: *There exists (i) $\frac{1}{n} \sum_{j=1}^n (H^{-1'} \lambda_i - \tilde{\lambda}_i) \lambda_i' = O_p((nT)^{-1/2}) + O_p(\min[N, T]^{-1})$. (ii) The $r \times r$ matrix $\frac{1}{T} \sum_{t=1}^T [(H F_t - \tilde{F}_t) (\sum_{i=1}^n \lambda_i' e_{it})] = O_p(\frac{n}{\min[N, T]})$.*

PROOF OF THEOREM 4: Let $\sigma_{ij} = E(e_{it} e_{jt})$ and $\tilde{\sigma}_{ij} = \frac{1}{T} \sum_{t=1}^T \tilde{e}_{it} \tilde{e}_{jt}$. 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,

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

The proposed estimator is $\tilde{\Gamma} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \tilde{\sigma}_{ij} \tilde{\lambda}_i \tilde{\lambda}_j'$. Also let $\tilde{\Gamma}_n = \frac{1}{n} \times$

$\sum_{i=1}^n \sum_{j=1}^n \tilde{\sigma}_{ij} \lambda_i \lambda'_j$. It follows that

$$(A.4) \quad \begin{aligned} \tilde{\Gamma} - H^{-1'} \Gamma H^{-1} \\ = \tilde{\Gamma} - H^{-1'} \bar{\Gamma}_n H^{-1} + H^{-1'} (\bar{\Gamma}_n - \Gamma_n) H^{-1} + H^{-1'} (\Gamma_n - \Gamma) H^{-1}. \end{aligned}$$

The last term converges to zero since $\Gamma_n - \Gamma \rightarrow 0$. We will show (i) that $\bar{\Gamma}_n - \Gamma_n \xrightarrow{p} 0$ if $\frac{n}{N} \rightarrow 0$ and $\frac{n}{T} \rightarrow 0$, and (ii) that $\tilde{\Gamma} - H^{-1'} \bar{\Gamma}_n H^{-1} = O_p(T^{-1/2}) + O_p(\min[N, T]^{-1})$.

(i) $\bar{\Gamma}_n - \Gamma_n \xrightarrow{p} 0$: From $\tilde{e}_{it} = x_{it} - \tilde{c}_{it}$ and $e_{it} = x_{it} - c_{it}$, where $c_{it} = \lambda'_i F_t$ and $\tilde{c}_{it} = \tilde{\lambda}'_i \tilde{F}_t$, we have $\tilde{e}_{it} = e_{it} - (c_{it} - \tilde{c}_{it})$. Thus, $\tilde{e}_{it} \tilde{e}_{jt} = e_{it} e_{jt} - e_{it} (c_{jt} - \tilde{c}_{jt}) - e_{jt} (c_{it} - \tilde{c}_{it}) + (c_{it} - \tilde{c}_{it})(c_{jt} - \tilde{c}_{jt})$ and

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

As shown in the working version of this paper, the first term is $O_p(T^{-1/2})$, and the second and third terms are zero if $\sqrt{n}/T \rightarrow 0$. These three terms are dominated by IV , whose norm can be rewritten as

$$\begin{aligned} &\left\| \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{T} \sum_{t=1}^T (c_{it} - \tilde{c}_{it})(c_{jt} - \tilde{c}_{jt}) \lambda_i \lambda'_j \right\| \\ &= \frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n (c_{it} - \tilde{c}_{it}) \lambda_i \right\|^2. \end{aligned}$$

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

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{i=1}^n (c_{it} - \tilde{c}_{it}) \lambda_i \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (H^{-1'} \lambda_i - \tilde{\lambda}_i)' \tilde{F}_t \lambda_i + \frac{1}{\sqrt{n}} \sum_{i=1}^n \lambda'_i H^{-1} (H F_t - \tilde{F}_t) \lambda_i \end{aligned}$$

and

$$\begin{aligned} & \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n (c_{it} - \tilde{c}_{it}) \lambda_i \right\|^2 \\ & \leq 2 \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \lambda_i (H^{-1'} \lambda_i - \tilde{\lambda}_i)' \right\|^2 \|\tilde{F}_t\|^2 \\ & \quad + 2 \|H^{-1}\|^2 \left(\frac{1}{n} \sum_{i=1}^n \|\lambda_i\|^2 \right)^2 \cdot n \cdot \|F_t - HF_t\|^2, \end{aligned}$$

where the inequality follows from $(a + b)^2 \leq 2a^2 + 2b^2$. Summing over t and dividing by T ,

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

By Lemma A.2(i), $a \rightarrow 0$ if $\sqrt{n}/T \rightarrow 0$, and $b = O_p(n)O_p(\min[N, T]^{-1}) \rightarrow 0$ if $\frac{n}{T} \rightarrow 0$ and $\frac{n}{N} \rightarrow 0$.

The proof of part (ii) is similar and can be found in the working version of this paper. Q.E.D.

REFERENCES

- ANDREWS, D. W. K. (1991): "Heteroskedastic and Autocorrelation Consistent Matrix Estimation," *Econometrica*, 59, 817–854. [1141]
- ANGELINI, E., J. HENRY, AND R. MESTRE (2001): "Diffusion Index-Based Inflation Forecasts for the Euro Area," Working Paper 61, European Central Bank. [1133]
- ARTIS, M., A. BANERJEE, AND M. MARCELLINO (2001): "Factor Forecasts for the U.K.," Discussion Paper 3119, CEPR. [1133]
- BAI, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–172. [1138]
- BAI, J., AND S. NG (2004): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," <http://www-personal.umich.edu/ngse/research.html>. [1135]
- (2006): "Evaluating Latent and Observed Factors in Macroeconomics and Finance," *Journal of Econometrics*, 113, 507–537. [1141]
- BANERJEE, A., M. MARCELLINO, AND I. MASTEN (2004): "Forecasting Macroeconomic Variables for the Acceding Countries," Working Paper 260, IGER. [1133]
- BERNANKE, B., AND J. BOIVIN (2003): "Monetary Policy in a Data Rich Environment," *Journal of Monetary Economics*, 50, 525–546. [1137]

- BERNANKE, B., J. BOIVIN, AND P. ELIASZ (2005): "Factor Augmented Vector Autoregressions (FVARs) and the Analysis of Monetary Policy," *Quarterly Journal of Economics*, 120, 387–422. [1133,1134,1137]
- CHAMBERLAIN, G., AND M. ROTHSCILD (1983): "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets," *Econometrica*, 51, 1281–1304. [1136]
- CONLEY, T. (1999): "GMM Estimation with Cross-Section Dependence," *Journal of Econometrics*, 92, 1–45. [1141]
- CRISTADORO, R., M. FORNI, L. REICHLIN, AND V. GIOVANNI (2001): "A Core Inflation Index for the Euro Area," Manuscript. [1133]
- DRISCOLL, H., AND A. KRAAY (1998): "Consistent Covariance Matrix Estimation with Spatially-Dependent Panel Data," *Review of Economics and Statistics*, 80, 549–560. [1141]
- FAVERO, C., M. MARCELLINO, AND F. NEGLIA (2005): "Principal Components at Work: The Empirical Analysis of Monetary Policy with Large Data Sets," *Journal of Applied Econometrics*, 20, 603–620. [1137]
- FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2001): "Do Financial Variables Help in Forecasting Inflation and Real Activity in the Euro Area," Manuscript. [1133]
- GIANNONE, D., L. REICHLIN, AND L. SALA (2002): "Tracking Greenspan: Systematic and Unsystematic Monetary Policy Revisited," Manuscript. [1137]
- GREENE, W. (2003): *Econometric Analysis* (Fifth Ed.). Englewood Cliffs, NJ: Prentice-Hall. [1139]
- NEWWEY, W., AND K. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708. [1141]
- PAGAN, A. (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25, 221–247. [1137]
- SHINTANI, M. (2002): "Nonlinear Analysis of Business Cycles Using Diffusion Indexes: Applications to Japan and the U.S.," Mimeo, Vanderbilt University. [1133]
- STOCK, J. H., AND M. W. WATSON (2001): "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, 47, 1–48. [1133]
- (2002a): "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179. [1134,1136,1142]
- (2002b): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business & Economic Statistics*, 20, 147–162. [1133]