



Taylor & Francis
Taylor & Francis Group

Forecasting Using Principal Components from a Large Number of Predictors

Author(s): James H. Stock and Mark W. Watson

Source: *Journal of the American Statistical Association*, Dec., 2002, Vol. 97, No. 460 (Dec., 2002), pp. 1167-1179

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association

Stable URL: <http://www.jstor.com/stable/3085839>

REFERENCES

Linked references are available on JSTOR for this article:

http://www.jstor.com/stable/3085839?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

American Statistical Association and Taylor & Francis, Ltd. are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association*

Forecasting Using Principal Components From a Large Number of Predictors

James H. STOCK and Mark W. WATSON

This article considers forecasting a single time series when there are many predictors (N) and time series observations (T). When the data follow an approximate factor model, the predictors can be summarized by a small number of indexes, which we estimate using principal components. Feasible forecasts are shown to be asymptotically efficient in the sense that the difference between the feasible forecasts and the infeasible forecasts constructed using the actual values of the factors converges in probability to 0 as both N and T grow large. The estimated factors are shown to be consistent, even in the presence of time variation in the factor model.

KEY WORDS: Factor models; Forecasting; Principal components.

1. INTRODUCTION

This article considers forecasting one series using a large number of predictor series. In macroeconomic forecasting, for example, the number of candidate predictor series (N) can be very large, often larger than the number of time series observations (T) available for model fitting. This high-dimensional problem is simplified by modeling the covariability of the series in terms of a relatively few number of unobserved latent factors. Forecasting can then be carried out in a two-step process. First, a time series of the factors is estimated from the predictors; second, the relationship between the variable to be forecast and the factors is estimated by a linear regression. If the number of predictors is large, then precise estimates of the latent factors can be constructed using simple methods even under fairly general assumptions about the cross-sectional and temporal dependence in the variables. We estimate the factors using principal components, and show that these estimates are consistent in an approximate factor model with idiosyncratic errors that are serially and cross-sectionally correlated.

To be specific, let y_t be the scalar time series variable to be forecast and let X_t be a N -dimensional multiple time series of candidate predictors. It is assumed that (X_t, y_{t+h}) admit a factor model representation with r common latent factors F_t ,

$$X_t = \Lambda F_t + e_t, \quad (1)$$

and

$$y_{t+h} = \beta'_F F_t + \beta'_w w_t + \varepsilon_{t+h} \quad (2)$$

where e_t is a $N \times 1$ vector idiosyncratic disturbances, h is the forecast horizon, w_t is a $m \times 1$ vector of observed variables (e.g., lags of y_t), that together with F_t are useful for forecasting y_{t+h} , and ε_{t+h} is the resulting forecast error. Data are available for $\{y_t, X_t, w_t\}_{t=1}^T$, and the goal is to forecast y_{T+h} .

If the idiosyncratic disturbances e_t in (1) were cross-sectionally independent and temporally iid, then (1) is the

classic factor analysis model. In our macroeconomic forecasting application, these assumptions are unlikely to be satisfied, and so we allow the error terms to be both serially correlated and (weakly) cross-sectionally correlated. In this sense, (1) is a serially correlated version of the approximate factor model introduced by Chamberlain and Rothschild (1983) for the study of asset prices. To construct forecasts of y_{T+h} , we form principal components of $\{X_t\}_{t=1}^T$ to serve as estimates of the factors. These estimated factors, together with w_t , are then used in (2) to estimate the regression coefficients. The forecast is constructed as $\hat{y}_{T+h} = \hat{\beta}'_F \hat{F}_T + \hat{\beta}'_w w_T$, where $\hat{\beta}_F$, $\hat{\beta}_w$, and \hat{F}_T are the estimated coefficients and factors.

This article makes three contributions. First, under general conditions on the errors discussed in Section 2, we show that the principal components of X_t are consistent estimators of the true latent factors (subject to a normalization discussed in Sec. 2). Consistency requires that both N and $T \rightarrow \infty$, although there are no restrictions on their relative rates of increase. Second, we show that the feasible forecast, \hat{y}_{T+h} , constructed from the estimated factors together with the estimated coefficients converge to the infeasible forecast that would be obtained if the factors and coefficients were known. Again, this result holds as $N, T \rightarrow \infty$. Thus the feasible forecast is first-order asymptotically efficient. Finally, motivated by the problem of temporal instability in macroeconomic forecasting models, we study the robustness of the consistency results to time variation in the factor model. We show that these results continue to hold when the temporal instability is small (as suggested by empirical work in macroeconomics) and weakly cross-sectionally dependent, in a sense that is made precise in Section 3.

This article is related to a large literature on factor analysis and a much smaller literature on forecasting. The literature on principal components and classical factor models is large and well known (Lawley and Maxwell 1971). Sargent and Sims (1977) and Geweke (1977) extended the classical factor model to dynamic models, and several researchers have applied versions of their dynamic factor model. In most applications of the classic factor model and its dynamic generalization, the dimension of X is small, and so the question of

James H. Stock is Professor, Kennedy School of Government, Harvard University, Cambridge, MA 02138, and the National Bureau of Economic Research (E-mail: james_stock@harvard.edu). Mark W. Watson is Professor, Department of Economics and Woodrow Wilson School, Princeton University, Princeton, NJ 08540, and the National Bureau of Economic Research (E-mail: mwatson@princeton.edu). The results in this article originally appeared in the paper titled "Diffusion Indexes" (NBER Working Paper 6702, August 1998). The authors thank the associate editor and referees, Jushan Bai, Michael Boldin, Frank Diebold, Gregory Chow, Andrew Harvey, Lucrezia Reichlin, Ken Wallis, and Charles Whiteman for helpful discussions and/or comments, and Lewis Chan, Piotr Eliasziw, and Alexei Onatski for skilled research assistance. This research was supported in part by National Science Foundation grants SBR-9409629 and SBR-9730489.

© 2002 American Statistical Association
Journal of the American Statistical Association
December 2002, Vol. 97, No. 460, Theory and Methods
DOI 10.1198/016214502388618960

consistent estimation of the factors is not relevant. However, several authors have noted that with large N , consistent estimation is possible. Connor and Korajczyk (1986, 1988, 1993) discussed the problem in a static model and argue that the factors can be consistently estimated by principle components as $N \rightarrow \infty$ even if the errors terms are weakly cross-sectionally correlated. Forni and Reichlin (1996, 1998) and Forni, Hallin, Lippi, and Reichlin (1999) discussed consistent $(N, T \rightarrow \infty)$ estimation of factors in a dynamic version of the approximate model. Finally, in a prediction problem similar to the one considered here, Ding and Hwang (1999) analyzed the properties of forecasts constructed from principal components in a setting with large N and T . Their analysis is conducted under the assumption that error process $\{e_t, \varepsilon_{t+h}\}$ is cross-sectionally and temporally iid, an assumption that is inappropriate for economic models and when interest focuses on multiperiod forecasts. We highlight the differences between our results and those of others later in the article.

The article is organized as follows. Section 2 presents the model in more detail, discusses the assumptions, and presents the main consistency results. Section 3 generalizes the model to allow temporal instability in the factor model. Section 4 examines the finite-sample performance of these methods in a Monte Carlo study, and Section 5 discusses an application to macroeconomic forecasting.

2. THE MODEL AND ESTIMATION

2.1 Assumptions

As described in Section 1, we focus on a forecasting situation in which N and T are both large. This motivates our asymptotic results requiring that $N, T \rightarrow \infty$ jointly or, equivalently, that $N = N(T)$ with $\lim_{T \rightarrow \infty} N(T) \rightarrow \infty$. No restrictions on the relative rates of N and T are required.

The assumptions about the model are grouped into assumptions about the factors and factor loading, assumptions about the errors in the (1), and assumptions about the regressors and errors in (2).

Assumption F1 (Factors and Factor Loading).

- $(\Lambda' \Lambda / N) \rightarrow I_r$.
- $E(F_t F_t') = \Sigma_{FF}$, where Σ_{FF} is a diagonal matrix with elements $\sigma_{ii} > \sigma_{jj} > 0$ for $i < j$.
- $|\lambda_{i,m}| \leq \bar{\lambda} < \infty$.
- $T^{-1} \sum_t F_t F_t' \xrightarrow{p} \Sigma_{FF}$.

Assumption F1 serves to identify the factors. The nonsingular limiting values of $(\Lambda' \Lambda / N)$ and Σ_{FF} imply that each of the factors provides a nonnegligible contribution to the average variance of x_{it} , where x_{it} is the i th element of X_t and the average is taken over both i and t . Moreover, because $\Lambda F_t = \Lambda R R^{-1} F_t$ for any nonsingular matrix R , a normalization is required to uniquely define the factors. Said differently, the model with factor loadings ΛR and factors $R^{-1} F_t$ is observationally equivalent to the model with factor loadings Λ and factors F_t . Assumption F1(a) restricts R to be orthonormal, and this together with Assumption F1(b) restricts R to be a diagonal matrix with diagonal elements of ± 1 . This identifies the factors up to a change of sign. Equivalently, Assumption F1 provides this normalization (asymptotically)

by associating Λ with the ordered orthonormal eigenvectors of $(NT)^{-1} \sum_{t=1}^T \Lambda F_t F_t' \Lambda'$ and $\{F_t\}_{t=1}^T$ with the principal components of $\{\Lambda F_t\}_{t=1}^T$. The diagonal elements of Σ_{FF} correspond to the limiting eigenvalues of $(NT)^{-1} \sum_{t=1}^T \Lambda F_t F_t' \Lambda'$. For convenience, these eigenvalues are assumed to be distinct. If they were not distinct, then the factors could only be consistently estimated up to an orthonormal transformation.

Assumption F1(b) allows the factors to be serially correlated, although it does rule out stochastic trends and other processes with nonconstant unconditional second moments. The assumption also allows lags of the factors to enter the equations for x_{it} and y_{t+h} . A leading example of this occurs in the dynamic factor model

$$x_{it} = \lambda_i(L) f_t + e_{it} \quad (3)$$

and

$$y_{t+h} = \beta_f(L)' f_t + \beta_w' w_t + \varepsilon_{t+h}, \quad (4)$$

where $\lambda_i(L)$ and $\beta_f(L)$ are lag polynomials is nonnegative powers of the lag operator L . If the lag polynomials have finite order q , then (3)–(4) can be rewritten as (1)–(2) with $F_t = (f_t' f_{t-1}' \cdots f_{t-q}')'$, and Assumption F1(b) will be satisfied if the f_t process is covariance stationary.

In the classical model, the errors or “uniquenesses” are assumed to be iid and normally distributed. This assumption is clearly inappropriate in the macroeconomic forecasting application, because the variables are serially correlated, and many or the variables (e.g., alternative measures of the money supply) may be cross-correlated even after the aggregate factors are controlled for. We therefore modify the classic assumptions to accommodate these complications.

Assumption M1 (Moments of the Errors e_t).

- $E(e_t' e_{t+u} / N) = \gamma_{N,t}(u)$, and $\lim_{N \rightarrow \infty} \sup_t \sum_{u=-\infty}^{\infty} |\gamma_{N,t}(u)| < \infty$.
Let e_{it} denote the i th element of e_t ; then
- $E(e_{it} e_{jt}) = \tau_{ij,t}$, $\lim_{N \rightarrow \infty} \sup_t N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\tau_{ij,t}| < \infty$, and
- $\lim_{N \rightarrow \infty} \sup_{t,s} N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\text{cov}(e_{is} e_{it}, e_{js} e_{jt})| < \infty$.

Assumption M1(a) allows for serial correlation in the e_{it} processes. As in the approximate factor model of Chamberlain and Rothschild (1983) and Connor and Korajczyk (1986, 1993), Assumption M1(b) allows $\{e_{it}\}$ to be weakly correlated across series. Forni et al. (1999) also allowed for serial correlation and cross-correlation with assumptions similar to M1(a)–(b). Normality is not assumed, but M1(c) limits the size of fourth moments.

It is assumed that the forecasting equation (2) is well behaved in the sense that if $\{F_t\}$ were observed, then ordinary least squares (OLS) would provide a consistent estimator of the regression coefficients. The specific assumption is as follows.

Assumption Y1 (Forecasting Equation). Let $z_t = (F_t' w_t')'$ and $\beta = (\beta_f' \beta_w')'$. Then the following hold:

- $E(z_t z_t') = \Sigma_{zz} = \begin{bmatrix} \Sigma_{FF} & \Sigma_{Fw} \\ \Sigma_{wF} & \Sigma_{ww} \end{bmatrix}$ a positive definite matrix.
- $T^{-1} \sum_t z_t z_t' \xrightarrow{p} \Sigma_{zz}$.

- c. $T^{-1} \sum_t z_t \varepsilon_{t+h} \xrightarrow{p} 0$.
- d. $T^{-1} \sum_t \varepsilon_{t+h}^2 \xrightarrow{p} \sigma^2$.
- e. $|\beta| < \infty$.

Assumptions Y1(a)–(c) are a standard set of conditions that imply consistency of OLS from the regression of y_{t+h} onto $(F_t' w_t')$. Here F_t is not observed, and the additional assumptions are useful for showing consistency of the OLS regression coefficients in the regression of y_{t+h} onto $(\hat{F}_t' w_t')$ and the resulting forecast of y_{T+h} .

2.2 Estimation

In “small- N ” dynamic factor models, forecasts are generally constructed using a three-step process (see, e.g., Stock and Watson 1989). First, parametric models are postulated for the joint stochastic process $\{y_{t+h}, X_t, w_t, e_t\}$, and the sample data $\{y_{t+h}, X_t, w_t\}_{t=1}^{T-h}$ are used to estimate the parameters of this process, typically using a Gaussian Maximum likelihood estimator (MLE). Next, these estimated parameters are used in signal extraction algorithms to estimate the unknown value of F_T . Finally, the forecast of y_{T+h} is constructed using this estimated value of the factor and the estimated parameters. When N is large, this process requires estimating many parameters using iterative nonlinear methods, which can be computationally prohibitive. We therefore take a different approach and estimate the dynamic factors nonparametrically using the method of principal components.

Consider the nonlinear least squares objective function,

$$V(\tilde{F}, \tilde{\Lambda}) = (NT)^{-1} \sum_i \sum_t (x_{it} - \tilde{\lambda}_i \tilde{F}_t)^2, \quad (5)$$

written as a function of hypothetical values of the factors (\tilde{F}) and factor loadings ($\tilde{\Lambda}$), where $\tilde{F} = (\tilde{F}_1 \tilde{F}_2 \dots \tilde{F}_T)'$ and $\tilde{\lambda}_i$ is the i th row of $\tilde{\Lambda}$. Let \hat{F} and $\hat{\Lambda}$ denote the minimizers of $V(\tilde{F}, \tilde{\Lambda})$. After concentrating out \hat{F} , minimizing (5) is equivalent to maximizing $\text{tr}[\tilde{\Lambda}' X' X \tilde{\Lambda}]$ subject to $\tilde{\Lambda}' \tilde{\Lambda} / N = I_r$, where X is the $T \times N$ data matrix with t th row X_t' and $\text{tr}(\cdot)$ denotes the matrix trace. This is the classical principal components problem, which is solved by setting $\hat{\Lambda}$ equal to the eigenvectors of $X'X$ corresponding to its r largest eigenvalues. The resulting principal components estimator of F is then

$$\hat{F} = X' \hat{\Lambda} / N. \quad (6)$$

Computation of \hat{F} requires the eigenvectors of the $N \times N$ matrix $X'X$; when $N > T$, a computationally simpler approach uses the $T \times T$ matrix XX' . By concentrating out $\hat{\Lambda}$, minimizing (5) is equivalent to maximizing $\text{tr}[\hat{F}'(XX')\hat{F}]$, subject to $\hat{F}'\hat{F}/T = I_r$, which yields the estimator, say \tilde{F} , which is the matrix of the first r eigenvectors of XX' . The column spaces of \hat{F} and \tilde{F} are equivalent, and so for forecasting purposes they can be used interchangeably, depending on computational convenience.

2.3 Consistent Estimation of Factors and Forecasts (1) and (2)

The first result presented in this section shows that the principal component estimator is pointwise (for any date t) consistent and has limiting mean squared error (MSE) over all t

that converges in probability to 0. Because Assumption F1 does not identify the sign of the factors, the theorem is stated in terms of sign-adjusted estimators.

Theorem 1. Let S_i denote a variable with value of ± 1 , let $N, T \rightarrow \infty$, and suppose that F1 and M1 hold. Suppose that k factors are estimated, where k may be \leq or $>$ r , the true number of factors. Then S_i can be chosen so that the following hold:

- a. For $i = 1, 2, \dots, r$, $T^{-1} \sum_{t=1}^T (S_i \hat{F}_{it} - F_{it})^2 \xrightarrow{p} 0$.
- b. For $i = 1, 2, \dots, r$, $S_i \hat{F}_{it} \xrightarrow{p} F_{it}$.
- c. For $i = r+1, \dots, k$, $T^{-1} \sum_{t=1}^T \hat{F}_{it}^2 \xrightarrow{p} 0$.

The details of the proof are provided in the Appendix; here we offer only a few remarks to provide some insight into problem and the need for the assumptions given in the preceding section. The estimation problem would be considerably simplified if it happened that Λ were known, because then F_t could be estimated by the least squares regression of $\{x_{it}\}_{i=1}^N$ onto $\{\lambda_i\}_{i=1}^N$. Consistency of the resulting estimator would then be studied by analyzing $\hat{F}_t - F_t = (\Lambda' \Lambda / N)^{-1} (N^{-1} \sum_i \lambda_i e_{it})$. Because $N \rightarrow \infty$, $(\Lambda' \Lambda / N) \rightarrow I_r$ [by F1(a)], and $N^{-1} \sum_i \lambda_i e_{it} \xrightarrow{p} 0$ [by M1(a) and F1(c)], the consistency of \hat{F}_t would follow directly. Alternatively, if F were known, then λ_i could be estimated by regression $\{x_{it}\}_{t=1}^T$ onto $\{F_t\}_{t=1}^T$, and consistency would be studied analyzing $(T^{-1} \sum_t F_t F_t')^{-1} T^{-1} \sum_t F_t e_{it}$, as $T \rightarrow \infty$ in a similar fashion. Because both F and Λ are unknown, both N and $T \rightarrow \infty$ are needed, and as it turns out, the proof is more complicated than these two simple steps suggest. The strategy that we have used is to show that the first r eigenvectors of $(NT)^{-1} X'X$ behave like the first r eigenvectors of $(NT)^{-1} \Lambda' F' F \Lambda$ (Assumption M1 is critical in this regard), and then show that these eigenvectors can be used to construct a consistent estimator of F (Assumption F1 is critical in this regard).

The next result shows that the feasible forecast (constructed using the estimated factors and estimated parameters) converges to the optimal infeasible forecast and thus is asymptotically efficient. In addition, it shows that the feasible regression coefficient estimators are consistent.

The result assumes that the forecasting equation is estimated using the $k = r$ factors. This is with little loss of generality, because there are several methods for consistently estimating the number of factors. For example, using analysis similar to that in Theorem 1, Bai and Ng (2001) constructed estimators of r based on penalized versions of the minimized value of (5), and in an earlier version of this article (Stock and Watson 1998a), we developed a consistent estimator of r based on the fit of the forecasting equation (2).

Theorem 2. Suppose that Y1 and the conditions of Theorem 1 hold. Let $\hat{\beta}_F$ and $\hat{\beta}_w$ denote the OLS estimates of β_F and β_w from the regression of $\{y_{t+h}\}_{t=1}^{T-h}$ onto $\{\hat{F}_t, w_t\}_{t=1}^{T-h}$. Then the following conditions hold:

- a. $(\hat{\beta}_F' \hat{F}_T + \hat{\beta}_w' w_T) - (\beta_F' F_T + \beta_w' w_T) \xrightarrow{p} 0$.
- b. $\hat{\beta}_w - \beta_w \xrightarrow{p} 0$ and S_i (defined in Theorem 1) can be chosen so that $S_i \hat{\beta}_{iF} - \beta_{iF} \xrightarrow{p} 0$ for $i = 1, \dots, r$.

The theorem follows directly from Theorem 1 together with Assumption Y1. The details of the proof are given in the Appendix.

3. TIME-VARYING FACTOR LOADINGS

In practice, when macroeconomic forecasts are constructed using many variables over a long period, some degree of temporal instability is inevitable. In this section we model this instability as stochastic drift in the factor loadings, and show that if this drift is not too large and not too dependent across series (in a sense made precise later), then the results of Theorems 1 and 2 continue to hold. Thus the principal components estimator and forecast are robust to small and idiosyncratic shifts in the factor loadings.

Specifically, replace the time-invariant factor model (1) with

$$x_{it} = \lambda_{it} F_t + e_{it} \quad (7)$$

and

$$\lambda_{it} = \lambda_{it-1} + g_{it} \zeta_{it} \quad (8)$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$, where g_{iT} is a scalar and ζ_{it} is an $r \times 1$ vector of random variables. This formulation implies that factor loadings for the i th variable shift by an amount, $g_{it} \zeta_{it}$, in time period t . The assumptions given in this section limit this time variation in two ways. First, the scalar g_{iT} is assumed to be small [$g_{iT} \sim O_p(T^{-1})$] which is consistent with the empirical literature in macroeconomics that estimates the degree of temporal instability in macroeconomic relations (Stock and Watson 1996, 1998b). This nesting implies that means that $\lambda_{iT} - \lambda_{i0} \sim O_p(T^{-1/2})$. Second, ζ_{it} is assumed to have weak cross-sectional dependence. That is, whereas some of the x variables may undergo related shifts in a given period, wholesale shifts involving a large number of the x 's are ruled out. Presumably such wholesale shifts could be better represented by shifts in the factors rather than in the factor loadings. In any event, this section shows that when these assumptions hold (along with technical assumptions given later), then the instability does not affect the consistency of the principal components estimator of F_t .

To motivate the additional assumptions used in this section, rewrite (7) as

$$x_{it} = \lambda_{i0} F_t + a_{it} \quad (9)$$

where $a_{it} = e_{it} + (\lambda_{it} - \lambda_{i0}) F_t = e_{it} + g_{it} \sum_{s=1}^t \zeta'_{is} F_s$. This equation has the same general form of the time-invariant factor model studied in the last section, with λ_{i0} and a_{it} in (9) replacing λ_i and e_{it} in the time-invariant model. This section introduces two new sets of assumptions that imply that λ_{i0} and a_{it} in (9) satisfy the assumptions concerning λ_i and e_{it} from the preceding section. This means that the conclusions of Theorems 1 and 2 will continue to hold for the time-varying factor model of this section.

The first new assumption is as follows.

Assumption F2.

- g_{iT} is independent of F_t , e_{jt} , and ζ_{jt} for all i, j , and t and $\sup_{i,j,k,m} T[E(|g_{iT} g_{jT} g_{kT} g_{mT}|)^{1/4}] < \kappa < \infty$ for all i, j, k , and m .

- The initial values of the values loadings satisfy $N^{-1} \sum_i \lambda'_{i0} \lambda_{i0} = \Lambda'_0 \Lambda_0 / N \xrightarrow{p} I_r$ and $\sup_{i,j} |\lambda_{ij,0}| < \bar{\lambda}$, where $\lambda_{ij,0}$ is the j th element of λ_{i0} .

As discussed earlier, Assumption F2(a) makes the amount of time variation small. Assumption F2(b) means that the initial value of the factor loadings satisfy the same assumptions as the time-invariant factor loading of the preceding section.

The next assumption limits the dependence in ζ_{it} . This assumption is written in fairly general form, allowing for some dependence in the random variables in the model.

Assumption M2. Let $\zeta_{it,m}$ denote the m th element of ζ_{it} . Then the following hold:

- $\lim_{T \rightarrow \infty} \sup_{i,s} \sum_{v=1-s}^{T-s} \sup_{l,q,l,m} |E(F_{lt} F_{mq} \zeta_{is,l} \zeta_{is+v,m})| < \infty$.
- $\lim_{T \rightarrow \infty} \sup_{i,s} \sum_{v=1-s}^{T-s} \sup_{m,q} |E(e_{is} F_{mq} \zeta_{is+v,m})| < \infty$.
- $\lim_{N \rightarrow \infty} N^{-1} \sum_i \sum_j \sup_{l,t,m,q,s,v} |E(F_{lt} F_{mq} \zeta_{is,l} \zeta_{jv,m})| < \infty$.
- $\lim_{N \rightarrow \infty} N^{-1} \sum_i \sum_j \sup_{l,s,q,m} |E(e_{it} F_{ms} \zeta_{jq,m})| < \infty$.
- $\lim_{N \rightarrow \infty} N^{-1} \sum_i \sum_j \sup_{\{l_k\}_{k=1}^2, \{l_k\}_{k=1}^6} |\text{cov}(e_{it_1} e_{it_2}, F_{l_{1t_3}} \zeta_{j_{t_4}, l_1} F_{l_{2t_5}} \zeta_{j_{t_6}, l_2})| < \infty$.
 - $\lim_{N \rightarrow \infty} N^{-1} \sum_i \sum_j \sup_{l, \{l_k\}_{k=1}^5} |\text{cov}(e_{it_1} e_{it_2}, e_{j_{t_3}} F_{l_{t_4}} \times \zeta_{j_{t_5}, l_1})| < \infty$.
 - $\lim_{N \rightarrow \infty} N^{-1} \sum_i \sum_j \sup_{\{l_k\}_{k=1}^4, \{l_k\}_{k=1}^8} |\text{cov}(F_{l_{1t_1}} \zeta_{it_2, l_1} \times F_{l_{2t_3}} \zeta_{it_4, l_2}, F_{l_{3t_5}} \zeta_{j_{t_6}, l_3} F_{l_{4t_7}} \zeta_{j_{t_8}, l_4})| < \infty$.
 - $\lim_{N \rightarrow \infty} N^{-1} \sum_i \sum_j \sup_{\{l_k\}_{k=1}^3, \{l_k\}_{k=1}^7} |\text{cov}(F_{l_{1t_1}} \zeta_{it_2, l_1} \times F_{l_{2t_3}} \zeta_{it_4, l_2}, e_{j_{t_5}} F_{l_{3t_6}} \zeta_{j_{t_7}, l_3})| < \infty$.
 - $\lim_{N \rightarrow \infty} N^{-1} \sum_i \sum_j \sup_{\{l_k\}_{k=1}^2, \{l_k\}_{k=1}^6} |\text{cov}(e_{it_1} F_{l_{1t_2}} \zeta_{it_3, l_1}, e_{j_{t_4}} F_{l_{2t_5}} \zeta_{j_{t_6}, l_2})| < \infty$.

This assumption essentially repeats Assumption M1 for the components of the composite error term a_{it} in (9). To interpret the assumption, consider the leading case in which the various components $\{\varepsilon_t\}$, $\{F_t\}$, $\{e_{it}\}$, and $\{\zeta_{it}\}$ are independent and have mean 0. Then, assuming that the F_t have finite fourth moments, and given the assumptions made in the last section, Assumption M2 is satisfied if (a) $\lim_{T \rightarrow \infty} \sup_{i,s} \sum_{v=1-s}^{T-s} \sup_{l,m} |E(\zeta_{is,l} \zeta_{is+v,m})| < \infty$; (b) $\lim_{N \rightarrow \infty} N^{-1} \sum_i \sum_j \sup_{l,s,v} |E(\zeta_{is,l} \zeta_{jv,m})| < \infty$; and (c) $\lim_{N \rightarrow \infty} N^{-1} \sum_i \sum_j \sup_{\{l_k\}_{k=1}^4, m, \{l_k\}_{k=1}^4} |\text{cov}(\zeta_{it_1, l_1} \zeta_{it_2, l_2}, \zeta_{j_{t_3}, l_3} \times \zeta_{j_{t_4}, l_4})| < \infty$, which are the analogs of the assumptions in M1 applied to the ζ error terms.

These two new assumptions yield the main result of this section, which follows.

Theorem 3. Given F1(b), F1(d), F2, M1, and M2, the results of Theorems 1 and 2 continue to hold.

The proof is given in the Appendix.

4. MONTE CARLO ANALYSIS

In this section we study some of the finite-sample properties of the principal components estimator and forecast using a small Monte Carlo experiment. The framework used in the preceding two sections was quite rich, allowing for distributed lags of potentially serially correlated factors to enter the x and y equations, error terms that were conditionally heteroscedastic and serially and cross-correlated, and factor loadings that

evolved through time. The design used here incorporates all of these features, and the data are generated according to

$$x_{it} = \sum_{j=0}^q \lambda_{ijt} f_{t-j} + e_{it}, \quad (10)$$

$$\lambda_{ijt} = \lambda_{ijt-1} + (c/T) \zeta_{ijt}, \quad (11)$$

$$f_t = \alpha f_{t-1} + u_t, \quad (12)$$

$$(1 - aL)e_{it} = (1 + b^2)v_{it} + bv_{i+1,t} + bv_{i-1,t}, \quad (13)$$

$$v_{it} = \sigma_{it} \eta_{it}, \quad (14)$$

and

$$\sigma_{it}^2 = \delta_0 + \delta_1 \sigma_{it-1}^2 + \delta_1 v_{it-1}^2, \quad (15)$$

where $i = 1, \dots, N$, $t = 1, \dots, T$, f_t and λ'_{ijt} are $\bar{r} \times 1$, and the variables $\{\zeta_{ijt}\}$, $\{u_{jt}\}$, and $\{\eta_{it}\}$ are mutually independent iid $N(0, 1)$ random variables. Equation (10) is dynamic factor model with q lags of \bar{r} factors that, as shown in Section 2, can be represented as the static factor model (1) with $r = \bar{r}(1 + q)$ factors. From (12), the factors evolve as a vector autoregressive [VAR(1)] model with common scalar autoregressive (AR) parameter α . From (13), the error terms in the factor equation are serially correlated, with an AR(1) coefficient of a , and cross-correlated, [with spatial moving average [MA(1)] coefficient b]. The innovations v_{it} are conditionally heteroscedastic and follow a GARCH(1, 1) process with parameters δ_0 , δ_1 , and δ_2 [see (14) and (15)]. Finally, from (11), the factor loadings evolve as random walk, with innovation standard deviation proportional to c .

The scalar variable to be forecast is generated as

$$y_{t+1} = \sum_{j=0}^q \iota' f_{t-j} + \varepsilon_{t+1}, \quad (16)$$

where ι is an $\bar{r} \times 1$ vector of 1s and $\varepsilon_{t+1} \sim \text{iid } N(0, 1)$ and is independent of the other errors in (10)–(15).

The other design details are as follows. The initial factor loading matrix, Λ_0 , was chosen as a function of R_i^2 , the fraction of the variance of x_{i0} explained by F_0 . The value of R_i^2 was chosen as an iid random variable equal to 0 with probability π and drawn from a uniform distribution on $[.1, .8]$ with probability $1 - \pi$. A nonzero value of π allows for the inclusion of x 's unrelated to the factors. Given this value of R_i^2 , the initial factor loading was computed as $\lambda_{ij0} = \lambda^*(R_i^2) \bar{\lambda}_{ij0}$, where $\lambda^*(R_i^2)$ is a scalar and $\bar{\lambda}_{ij0} \sim \text{iid } N(0, 1)$ and independent of $\{\eta_{it}, \zeta_{ijt}, u_{jt}\}$. The initial values of the factors were drawn from their stationary distribution. The parameter δ_0 was chosen so that the unconditional variance of v_{it} was unity.

Principal components were used to estimate k factors, as discussed in Section 2.2. These k estimated factors were used to estimate r (the true number of factors) using methods described later, and the coefficients β in the forecasting regression (2) were estimated by the OLS coefficients $\hat{\beta}$ in the regression of y_{t+1} onto \hat{F}_{jt} , $j = 1, \dots, \hat{r}$, $t = 1, \dots, T - 1$, where \hat{r} is the estimated number of factors. The out-of-sample forecast is $\hat{y}_{T+1/T} = \sum_{j=1}^{\hat{r}} \hat{\beta}_j \hat{F}_{jT}$. For comparison purposes, the infeasible out-of-sample forecast $\tilde{y}_{T+1/T} = \tilde{\beta}' F_T$ was also computed, where $\tilde{\beta}$ is the OLS estimator obtained from regressing y_{t+1} onto F_t , $t = 1, \dots, T - 1$. The free parameters in

the Monte Carlo experiment are $N, T, \bar{r}, q, k, \pi, a, b, c, \delta_1$, and δ_2 .

The results are summarized by two statistics. The first statistic is a trace R^2 of the multivariate regression of \hat{F} onto F ,

$$R_{\hat{F}, F}^2 = \hat{E} \|P_F \hat{F}\|^2 / \hat{E} \|\hat{F}\|^2 = \hat{E} \text{tr}(\hat{F}' P_F \hat{F}) / \hat{E} \text{tr}(\hat{F}' \hat{F}), \quad (17)$$

where \hat{E} denotes the expectation estimated by averaging the relevant statistic over the Monte Carlo repetitions and $P_F = F(F'F)^{-1}F'$. According to Theorem 1, $R_{\hat{F}, F}^2 \xrightarrow{P} 1$.

The second statistic measures how close the forecast based on the estimated factors is to the infeasible forecast based on the true factors,

$$S_{\hat{y}, \tilde{y}}^2 = 1 - \hat{E}(\hat{y}_{T+1/T} - \tilde{y}_{T+1/T})^2 / \hat{E}(\hat{y}_{T+1/T}^2). \quad (18)$$

Because $\hat{y}_{T+1/T} - \tilde{y}_{T+1/T} \xrightarrow{P} 0$ when $k = r$ from Theorem 1, $S_{\hat{y}, \tilde{y}}^2$ should be close to 1 for appropriately large N and T . $S_{\hat{y}, \tilde{y}}^2$ is computed for several choices of \hat{r} . First, as a benchmark, results are shown for $\hat{r} = r$. Second, \hat{r} is formed using three of the information criteria suggested by Bai and Ng (2001). These criteria have the form $IC_p(k) = \ln(\hat{V}_k) + kg(T, N)$, where \hat{V}_k is the minimized value of the objective function (5) for a model with k factors and $g_j(T, N)$ is a penalty function. Three of the penalty functions suggested by Bai and Ng are used:

$$g_1(N, T) = \left(\frac{N+T}{NT} \right) \ln \left(\frac{NT}{N+T} \right),$$

$$g_2(N, T) = \left(\frac{N+T}{NT} \right) \ln C_{NT}^2,$$

$$g_3(N, T) = \left(\frac{\ln(C_{NT}^2)}{C_{NT}^2} \right),$$

where $C_{NT}^2 = \min(N, T)$, resulting in criteria labeled IC_{p1} , IC_{p2} , and IC_{p3} . The minimizers of these criteria yield a consistent estimator of r , and interest here focuses on their relative small-sample accuracy. Finally, results are shown with \hat{r} computed using the conventional Akaike information criterion (AIC) and Bayes information criterion (BIC) applied to the forecasting equation (2).

The results are summarized in Table 1. Panel A presents results for the static factor model with iid errors and factors and with large N and T ($N, T \geq 100$). Panel B gives corresponding results for small values of N and T ($N, T \leq 50$). Panel C adds irrelevant x_{it} 's to the model ($\pi > 0$). Panel D extends the model to idiosyncratic errors that are serially correlated, cross-correlated, conditionally heteroscedastic, or some combination of these. Panel E considers the dynamic factor model with serially correlated factors and/or lags of the factors entering X_t . Finally, panel F gives time-varying factor loadings.

First, consider the results for the static factor model shown in panel A. The values of $R_{\hat{F}, F}^2$ exceed .85 except when many redundant factors are estimated. The smallest value of $R_{\hat{F}, F}^2$ is .69, which obtains when N and T are relatively small ($N = T = 100$) and there are 10 redundant factors ($r = 5$ and

Table 1. Monte Carlo Results

T	N	\bar{r}	k	q	α	a	b	c	π	δ_0	δ_1	δ_2	$R^2_{\bar{F},F}$	$S^2_{\bar{y},\bar{y}}$					AIC	BIC
														k = r	IC _{p1}	IC _{p2}	IC _{p3}			
A. Static factor models, large N and T																				
100	100	5	5	0	0	0	0	0	0	1.00	0	0	.94	.93	.93	.93	.93	.93	.91	
100	250	5	5	0	0	0	0	0	0	1.00	0	0	.97	.96	.96	.96	.96	.96	.95	
100	500	5	5	0	0	0	0	0	0	1.00	0	0	.98	.98	.98	.98	.98	.97	.96	
100	100	5	10	0	0	0	0	0	0	1.00	0	0	.78	.93	.93	.93	.93	.91	.91	
100	250	5	10	0	0	0	0	0	0	1.00	0	0	.85	.96	.96	.96	.96	.94	.95	
100	500	5	10	0	0	0	0	0	0	1.00	0	0	.88	.98	.98	.98	.98	.96	.96	
100	100	5	15	0	0	0	0	0	0	1.00	0	0	.69	.93	.93	.93	.93	.89	.91	
100	250	5	15	0	0	0	0	0	0	1.00	0	0	.77	.96	.96	.96	.96	.93	.95	
100	500	5	15	0	0	0	0	0	0	1.00	0	0	.81	.98	.98	.98	.98	.95	.96	
100	100	10	10	0	0	0	0	0	0	1.00	0	0	.88	.83	.56	.16	.83	.82	.74	
100	250	10	10	0	0	0	0	0	0	1.00	0	0	.95	.92	.88	.82	.92	.92	.85	
100	500	10	10	0	0	0	0	0	0	1.00	0	0	.97	.95	.94	.93	.95	.95	.89	
B. Static factor models, small N and T																				
25	50	2	2	0	0	0	0	0	0	1.00	0	0	.92	.90	.90	.90	.90	.89	.87	
50	50	2	2	0	0	0	0	0	0	1.00	0	0	.94	.93	.93	.93	.93	.93	.92	
25	50	5	5	0	0	0	0	0	0	1.00	0	0	.84	.65	.57	.35	.66	.58	.49	
50	50	5	5	0	0	0	0	0	0	1.00	0	0	.87	.82	.75	.52	.82	.81	.74	
25	50	2	10	0	0	0	0	0	0	1.00	0	0	.56	.90	.90	.90	.28	.62	.79	
50	50	2	10	0	0	0	0	0	0	1.00	0	0	.62	.93	.93	.93	.70	.86	.91	
C. Irrelevant X's																				
100	333	5	5	0	0	0	0	0	.25	1.00	0	0	.97	.96	.96	.96	.96	.96	.95	
100	333	5	10	0	0	0	0	0	.25	1.00	0	0	.81	.96	.96	.96	.96	.94	.95	
100	333	10	10	0	0	0	0	0	.25	1.00	0	0	.94	.92	.43	.30	.80	.91	.83	
D. Dependent errors																				
100	250	5	5	0	0	.5	0	0	0	1.00	0	0	.97	.96	.96	.96	.96	.96	.95	
100	250	5	10	0	0	.5	0	0	0	1.00	0	0	.80	.96	.96	.96	.96	.95	.95	
100	250	5	5	0	0	.9	0	0	0	1.00	0	0	.85	.84	.84	.84	.84	.84	.82	
100	250	5	5	0	0	.0	0	0	0	.05	.90	.05	.97	.97	.97	.97	.97	.96	.95	
100	250	5	5	0	0	.9	0	0	0	.05	.90	.05	.85	.85	.85	.85	.85	.84	.82	
100	250	5	5	0	0	0	1.0	0	0	1.00	.00	0	.97	.96	.96	.96	.96	.96	.95	
100	250	5	10	0	0	0	1.0	0	0	1.00	.00	0	.82	.96	.96	.96	.96	.94	.95	
100	250	5	10	0	0	0	1.0	0	0	.05	.90	.05	.82	.96	.96	.96	.96	.94	.95	
E. Dynamic factors																				
100	250	5	10	1	0	0	0	0	0	1.00	0	0	.95	.92	.87	.80	.92	.91	.84	
100	250	5	15	1	0	0	0	0	0	1.00	0	0	.84	.92	.87	.80	.92	.89	.84	
100	250	5	5	0	.9	0	0	0	0	1.00	0	0	.89	.80	.78	.76	.79	.79	.75	
100	250	5	10	0	.9	0	0	0	0	1.00	0	0	.76	.80	.78	.76	.79	.77	.75	
100	250	5	5	0	.9	.5	1.0	0	0	1.00	0	0	.88	.79	.79	.79	.79	.78	.75	
100	250	5	5	0	.9	.5	1.0	0	0	.05	.90	.05	.88	.78	.78	.78	.78	.77	.73	
100	250	5	10	0	.9	.5	1.0	0	0	1.00	0	0	.71	.79	.79	.79	.74	.77	.75	
F. Time-varying factor loading																				
100	250	5	5	0	0	0	0	5	0	1.00	0	0	.93	.93	.93	.93	.93	.92	.91	
100	250	5	5	0	0	0	0	10	0	1.00	0	0	.86	.89	.89	.89	.89	.88	.87	
100	250	5	5	0	0	0	0	5	0	.05	.90	.05	.93	.93	.93	.93	.93	.93	.92	
100	250	5	10	0	0	0	0	5	0	1.00	0	0	.81	.93	.93	.93	.88	.85	.89	
100	250	5	10	0	0	0	0	10	0	1.00	0	0	.76	.89	.77	.79	.74	.78	.82	
100	250	5	5	0	.9	.5	1.0	5	0	1.00	0	0	.86	.78	.78	.78	.78	.77	.74	
100	250	5	5	0	.9	.5	1.0	10	0	1.00	0	0	.82	.75	.75	.75	.75	.74	.71	
100	250	5	10	0	.9	.5	1.0	5	0	1.00	0	0	.72	.78	.72	.74	.67	.75	.74	
100	250	5	10	0	.9	.5	1.0	10	0	1.00	0	0	.71	.75	.58	.59	.58	.66	.70	
100	250	5	5	0	.9	.5	1.0	10	.25	.05	.90	.05	.81	.71	.71	.71	.71	.71	.69	
100	250	5	15	1	.9	.5	1.0	10	.25	.05	.90	.05	.73	.57	.44	.45	.44	.59	.60	

NOTE: Based on 2,000 Monte Carlo draws using the design described in Section 4 of the text. The statistic $R^2_{\bar{F},F}$ is the trace of a regression of the estimated factors on the true factor, and $S^2_{\bar{y},\bar{y}}$ is a measure of the closeness of the forecast based on the estimated factors to the forecast based on the true factor; see the text for definitions. The values are reported using the correct number of factors and for \hat{r} estimated using an information criterion, where $0 \leq \hat{r} \leq k$.

$k = 15$). The values of $S_{\hat{y}, \hat{y}}^2$ generally exceed .9, and this is true for all methods used to estimate the number of factors. The only important exception is when $k = r = 10$ and $N = T = 100$; in this case, IC_{p1} and IC_{p2} perform poorly. The penalty factors for IC_{p1} and IC_{p2} are larger than for IC_{p3} [e.g., when $T = N$, $g_2(N, T) = 2g_3(N, T)$], and apparently these large penalties lead to serious underfitting.

Performance deteriorates somewhat in small samples, as shown in panel B. With only two factors, $S_{\hat{y}, \hat{y}}^2$ is near .9 for $k = r$, so that the forecasts perform nearly as well as the infeasible forecasts. When k is much larger than r ($k = 10$ and $r = 2$), IC_{p3} performs poorly because of overfitting, particularly when T is very small ($T = 25$). All of the methods deteriorate when there are five factors; for example, when $T = 25$, the values of $S_{\hat{y}, \hat{y}}^2$ are closer to .6.

Panel C suggests that including irrelevant series has little effect on the estimators and forecasts when N and T are large. Results are shown for $\pi = .25$ (so that 25% of the series are unrelated to the factors) and $N = 333$. The results for the models with five factors are nearly identical to the results in panel A with the same number of relevant series ($N = 250$). The results for 10 factors are also similar to those in panel A, although panel C shows some deterioration of the forecasts using IC_{p1} and IC_{p2} .

From panel D, moderate amounts of serial or spatial correlation in the errors have little effect on the estimators and forecasts. For example, on the one hand, when moderate serial correlation is introduced by setting $a = .5$, the results in the table are very similar to the results with $a = 0$; similarly, there is little change when spatial correlation is introduced by setting $b = 1.0$. On the other hand, some deterioration in performance occurs when the degree of serial correlation is large (compare the entries with $a = .9$ to those with $a = .5$). Conditional heteroscedasticity has no apparent effect on the performance of the estimator and forecasts.

From panel E, introducing lags of the factors has little effect on the quality of the estimators and forecasts: the results with $\bar{r} = 5$ and $q = 1$ (so that $r = 10$) are essentially identical to the static factor model with 10 factors. However, a high degree of serial correlation in the factor process ($\alpha = .9$) does result in some deterioration of performance. For example, when $T = 100$, $N = 250$, and $r = 5$, $R_{\hat{F}, F}^2 = .97$ in the static factor model ($\alpha = 0$), and this falls to .89 when $\alpha = .9$.

Finally, panel F shows the effect of time variation on the factor loadings in isolation and together with other complications. There appears to be only moderate deterioration of the forecast performance even for reasonably large amount of temporal instability (e.g., $S_{\hat{y}, \hat{y}}^2$ remains high even as the parameter governing time-variation increases from 0 to 10). However, when all of the complications are present (i.e., serially correlated dynamic factors, $k > r$, serial and cross-correlated heteroscedastic errors, time-varying factor loading, and a large number of unrelated x 's), forecast performance deteriorates significantly, as shown in the last few entries in panel F.

5. AN EMPIRICAL EXAMPLE

In related empirical work we have applied factor models and principal components to forecast several macroeconomic variables (see Stock and Watson 1999, 2002). In this

section we describe a forecasting experiment for the Federal Reserve Board's Index of Industrial Production, an important monthly indicator of macroeconomic activity. The variables making up X_t are 149 monthly macroeconomic variables representing several different facets of the macroeconomy (e.g., production, consumption, employment, price inflation, interest rates). We have described the variables in detail in earlier work (Stock and Watson 2002). The sample period is January 1959–December 1998. Principal components of X_t were used to construct forecasts of $y_{t+12} = \ln(IP_{t+12}/IP_t)$, where IP_t is the index of industrial production for date t . These 12-month-ahead forecasts were constructed in each month starting in 1970:1 and extending through 1997:12, using previously available data to estimate unknown parameters and factors.

To simulate real-time forecasting, we used data dated T and earlier in all calculations for constructing forecasts at time T . Thus for example, to compute the forecast in $T = 1970:1$, the variables making up X_t were standardized using data from $t = 1959:1$ – $1970:1$, and principal components were computed. These estimated values of F_t were used together with y_{t+12} for $t = 1959:1$ – $1969:1$ to estimate β in (2). Model selection with $k = 10$ based on IC_{p1} , IC_{p2} , IC_{p3} , AIC, and BIC were used to determine the number of factors to include in the regression. Finally, the forecasts constructed in $T = 1970:1$ were formed as $\hat{y}_{t+12/t} = \hat{\beta}\hat{F}_t$. This process was repeated for 1970:2–1997:12.

We also computed forecasts using four other methods: a univariate autoregression in which y_{t+12} was regressed on lags of $\ln(IP_t/IP_{t-1})$, a vector autoregression that included the rate of price inflation and short-term interest rates in addition to the rate of growth of the industrial production index, a leading-indicator model in which y_{t+12} was regressed on 11 leading indicators chosen by Stock and Watson (1989) as good predictors of aggregate macroeconomic activity; and an autoregressive-augmented principal components model in which y_{t+12} was regressed on the estimated factors and lags of $\ln(IP_t/IP_{t-1})$. We gave details of the specification, including lag length choice and exact description of the variables, in earlier work (Stock and Watson 2002).

Table 2 shows the MSE of the resulting forecasts, where we have shown each MSE relative to the MSE for the univariate autoregression. The first three rows show results from the benchmark AR, VAR, and leading indicator models. The next row shows the results for the principal components forecasts, with the number of factors determined by IC_{p3} . (The results for the other selection procedures are similar and thus are not reported.) This is followed by principal components forecasts using a fixed number of factors ($k = 1$ – $k = 4$). Finally, the last row shows the principal components forecasting model (with r estimated by IC_{p3}) augmented with BIC-selected lags of the growth rate of industrial production. (Again, results for other selection procedures are very similar and are not reported.)

Both the leading indicator and VAR models perform slightly better than the univariate AR in this simulated out-of-sample experiment. However, the gains are not large. The factor models offer substantial improvement. The results suggest that nearly all of the forecasting gain comes from the first two or three factors and that once these factors are included, no additional gain is realized from including lagged values of IP

Table 2. Simulated Out-of-Sample Forecasting Results Industrial Production, 12-Month Horizon

Forecast method	Relative MSE
Univariate autoregression	1.00
Vector autoregression	.97
Leading indicators	.86
Principal components	.58
Principal components, $k = 1$.94
Principal components, $k = 2$.62
Principal components, $k = 3$.55
Principal components, $k = 4$.56
Principal components, AR	.69
Root MSE, AR model	.049

NOTE: For each forecast method, this table shows the ratio of the MSE of the forecast made by the method for that row to the MSE of a univariate autoregressive forecast with lag length selected by the BIC. The final line presents the root MSE for the autoregressive model in native (decimal growth rate) units at an annual rate.

growth. We have already reported similar results for other real macroeconomic variables (Stock and Watson 2002).

6. DISCUSSION

This article has shown that forecasts of a single series based on principal components of a large number of predictors are first-order asymptotically efficient as $N, T \rightarrow \infty$ for general relationships between N and T in the context of an approximate factor model with dynamics. The Monte Carlo results suggest that these theoretical implications provide a useful guide to empirical work in datasets of the size typically encountered in macroeconomic forecasting. The empirical results summarized here and reported in more detail elsewhere suggest that these methods can contribute to substantial improvement in forecasts beyond conventional models using a small number of variables.

Several methodologic issues remain. One issue is to explore estimation methods that might be more efficient in the presence of heteroscedastic and serially correlated uniquenesses. Another is to develop a distribution theory for the estimated factors that goes beyond the consistency results shown here and provides measures of the sampling uncertainty of the estimated factors. A third theoretical extension is to move beyond the $I(0)$ framework of this article and to introduce strong persistence into the series; for example, by letting some of the factors have a unit autoregressive root, which would permit some of the observed series to contain a common stochastic trend.

APPENDIX: PROOFS OF THEOREMS

We begin with some notation.

Define $\sum_i = \sum_{i=1}^N$ and $\sum_t = \sum_{t=1}^T$.

Let γ denote an $N \times 1$ vector and let $\Gamma = \{\gamma | \gamma' \gamma / N = 1\}$, $R(\gamma) = N^{-2} T^{-1} \gamma' \sum_t x_t x_t' \gamma$, and $R^*(\gamma) = N^{-2} T^{-1} \gamma' \sum_t \Lambda F_t F_t' \Lambda' \gamma$.

We begin by collecting a set of results used in the proof.

Results (R1)–(R19) Hold Under Assumptions F1 and M1

(R1) $N^{-1} \sum_i e_{it}^2 \sim O_p(1)$.

Proof. $N^{-1} \sum_i e_{it}^2 = N^{-1} \sum_i \tau_{ii,t} + N^{-1} \sum_i (e_{it}^2 - \tau_{ii,t})$.

The term $N^{-1} \sum_i \tau_{ii,t}$ is $O(1)$ from Assumption M1(b) (because $N^{-1} \sum_i \tau_{ii,t} \leq N^{-1} \sum_i \sum_j |\tau_{ij,t}| \sim O(1)$). So it suffices that the second term converges to 0 in probability. Now

$$\begin{aligned} E \left[N^{-1} \sum_i (e_{it}^2 - \tau_{ii,t}) \right]^2 &= N^{-2} \sum_i \sum_j E(e_{it}^2 - \tau_{ii,t})(e_{jt}^2 - \tau_{jj,t}) \\ &= N^{-2} \sum_i \sum_j \text{cov}(e_{it}^2, e_{jt}^2) \\ &\leq N^{-2} \sum_i \sum_j |\text{cov}(e_{it}^2, e_{jt}^2)| \\ &\leq \sup_{t,s} N^{-2} \sum_{i=1}^N \sum_{j=1}^N |\text{cov}(e_{is} e_{it}, e_{js} e_{jt})| \rightarrow 0, \end{aligned}$$

by Assumption M1(c). Thus $N^{-1} \sum_i (e_{it}^2 - \tau_{ii,t}) \xrightarrow{p} 0$.

(R2) $\sup_{\gamma \in \Gamma} (N^2 T)^{-1} \gamma' e' e \gamma \xrightarrow{p} 0$.

Proof.

$$\begin{aligned} (N^2 T)^{-1} \gamma' e' e \gamma &= (N^2 T)^{-1} \sum_t \sum_i \sum_j \gamma_i \gamma_j e_{it} e_{jt} \\ &= N^{-2} \sum_i \sum_j \gamma_i \gamma_j \left(T^{-1} \sum_t e_{it} e_{jt} \right) \\ &\leq \left(N^{-2} \sum_i \sum_j \gamma_i^2 \gamma_j^2 \right)^{1/2} \\ &\quad \times \left(N^{-2} \sum_i \sum_j \left(T^{-1} \sum_t e_{it} e_{jt} \right)^2 \right)^{1/2} \end{aligned}$$

but $N^{-2} \sum_i \sum_j \gamma_i^2 \gamma_j^2 = (\gamma' \gamma / N)^2$, and for all $\gamma \in \Gamma$, $(\gamma' \gamma / N) = 1$. Thus

$$\sup_{\gamma \in \Gamma} (N^2 T)^{-1} \gamma' e' e \gamma \leq \left(N^{-2} \sum_i \sum_j \left(T^{-1} \sum_t e_{it} e_{jt} \right)^2 \right)^{1/2}.$$

Now

$$N^{-2} \sum_i \sum_j \left(T^{-1} \sum_t e_{it} e_{jt} \right)^2 = N^{-2} T^{-2} \sum_i \sum_j \sum_t \sum_s e_{it} e_{is} e_{jt} e_{js}$$

and

$$\begin{aligned} E \left[N^{-2} T^{-2} \sum_i \sum_j \sum_t \sum_s e_{it} e_{is} e_{jt} e_{js} \right] \\ &= N^{-2} T^{-2} \sum_i \sum_j \sum_t \sum_s \gamma_{i,t,s} \gamma_{j,t,s} \\ &\quad + N^{-2} T^{-2} \sum_i \sum_j \sum_t \sum_s E[(e_{it} e_{is} - \gamma_{i,t,s})(e_{jt} e_{js} - \gamma_{j,t,s})], \end{aligned}$$

where $\gamma_{i,t,s} = E(e_{it} e_{is})$.

The first term is

$$T^{-2} \sum_t \sum_u \left(N^{-1} \sum_i \gamma_{i,t,t+u} \right) \left(N^{-1} \sum_j \gamma_{j,t,t+u} \right) = T^{-2} \sum_t \sum_u \gamma_{N,t}(u)^2,$$

because $(N^{-1} \sum_i \gamma_{i,t,t+u}) = N^{-1} E \sum_i e_{it} e_{i,t+u} = \gamma_{N,t}(u)$ defined in M1(a). Now the absolute summability of $|\gamma_{N,t}(u)|$ in M1(a) implies square summability, so that $\lim_{N \rightarrow \infty} \sup_t \sum_u \gamma_{N,t}(u)^2 < \infty$. This implies that $N^{-2} T^{-2} \sum_i \sum_j \sum_t \sum_s \gamma_{i,t,s} \gamma_{j,t,s} \rightarrow 0$.

The second term is

$$N^{-2}T^{-2} \sum_i \sum_j \sum_t \sum_s \text{cov}(e_{it}e_{is}, e_{jt}e_{js}) \\ \leq N^{-2} \sup_{t,s} \sum_i \sum_j |\text{cov}(e_{it}e_{is}, e_{jt}e_{js})| \rightarrow 0$$

by Assumption M1(c).

(R3) Let q_t denote a sequence of random variables with $T^{-1} \sum_t q_t^2 \sim O_p(1)$. Then

$$\sup_{\gamma \in \Gamma} |T^{-1} \sum_t q_t (N^{-1} \sum_i \gamma_i e_{it})| \xrightarrow{p} 0.$$

Proof. $\sup_{\gamma \in \Gamma} |T^{-1} \sum_t q_t (N^{-1} \sum_i \gamma_i e_{it})| \leq (T^{-1} \sum_t q_t^2)^{1/2} \times (\sup_{\gamma \in \Gamma} T^{-1} \sum_i (N^{-1} \sum_i \gamma_i e_{it})^2)^{1/2}$.

The first term is $O_p(1)$ by assumption, and so the result follows from

$$\sup_{\gamma \in \Gamma} T^{-1} \sum_i \left(N^{-1} \sum_i \gamma_i e_{it} \right)^2 = \sup_{\gamma \in \Gamma} (TN^2)^{-1} \sum_i \sum_j \gamma_i \gamma_j e_{it} e_{jt} \\ = \sup_{\gamma \in \Gamma} (N^2 T)^{-1} \gamma' e' e \gamma \xrightarrow{p} 0,$$

where the limit follows from (R2).

(R4) $\sup_{\gamma \in \Gamma} |T^{-1} \sum_t F_{jt} (N^{-1} \sum_i \gamma_i e_{it})| \xrightarrow{p} 0$ for $j = 1, 2, \dots, r$.

Proof. Because $T^{-1} \sum_t F_{jt}^2 \xrightarrow{p} \sigma_{jj}$ [from Assumption F1(d)], the result follows from (R3).

(R5) $\sup_{\gamma \in \Gamma} (N^2 T)^{-1} |\gamma' \Lambda F' e \gamma| \xrightarrow{p} 0$.

Proof. $(N^2 T)^{-1} \gamma' \Lambda F' e \gamma = \sum_j (\gamma' \underline{\lambda}_j / N) T^{-1} \sum_t F_{jt} (N^{-1} \sum_i \gamma_i e_{it})$, so that

$$|(N^2 T)^{-1} \gamma' \Lambda F' e \gamma| \leq \sum_j |(\gamma' \underline{\lambda}_j / N)| \left| T^{-1} \sum_t F_{jt} \left(N^{-1} \sum_i \gamma_i e_{it} \right) \right|.$$

Thus,

$$\sup_{\gamma \in \Gamma} |(N^2 T)^{-1} \gamma' \Lambda F' e \gamma| \\ \leq \left(\max_j \sup_{\gamma \in \Gamma} |\gamma' \underline{\lambda}_j / N| \right) \sum_{j=1}^r \sup_{\gamma \in \Gamma} \left| T^{-1} \sum_t F_{jt} \left(N^{-1} \sum_i \gamma_i e_{it} \right) \right| \\ \leq \left\{ \sup_{\gamma \in \Gamma} (\gamma' \gamma / N)^{1/2} \right\} \left\{ \max_j (\underline{\lambda}'_j \underline{\lambda}_j / N)^{1/2} \right\} \\ \times \sum_{j=1}^r \sup_{\gamma \in \Gamma} \left| T^{-1} \sum_t F_{jt} \left(N^{-1} \sum_i \gamma_i e_{it} \right) \right| \xrightarrow{p} 0,$$

where the last line follows from $(\gamma' \gamma / N) = 1$, F1(a), and (R4).

(R6) $\sup_{\gamma \in \Gamma} |R(\gamma) - R^*(\gamma)| \xrightarrow{p} 0$.

Proof. $R(\gamma) - R^*(\gamma) = (N^2 T)^{-1} \gamma' e' e \gamma + 2(N^2 T)^{-1} \gamma' \Lambda F' e \gamma$ and

$$\sup_{\gamma \in \Gamma} |R(\gamma) - R^*(\gamma)| \leq \sup_{\gamma \in \Gamma} (N^2 T)^{-1} |\gamma' e' e \gamma| + \sup_{\gamma \in \Gamma} (N^2 T)^{-1} |\gamma' \Lambda F' e \gamma|,$$

where the two terms on the rhs of the inequality converge to 0 in probability by (R2) and (R5).

(R7) $|\sup_{\gamma \in \Gamma} R(\gamma) - \sup_{\gamma \in \Gamma} R^*(\gamma)| \xrightarrow{p} 0$.

Proof. $|\sup_{\gamma \in \Gamma} R(\gamma) - \sup_{\gamma \in \Gamma} R^*(\gamma)| \leq \sup_{\gamma \in \Gamma} |R(\gamma) - R^*(\gamma)| \xrightarrow{p} 0$, where the first inequality follows by the definition of the sup and the convergence follows from (R6).

(R8) $\sup_{\gamma \in \Gamma} R^*(\gamma) \xrightarrow{p} \sigma_{11}$.

Proof. Write $\Lambda' \Lambda / N = (\Lambda' \Lambda / N)^{1/2} (\Lambda' \Lambda / N)^{1/2'}$ to denote the Choleski factorization of $\Lambda' \Lambda / N$. Let γ be represented as $\gamma = \Lambda (\Lambda' \Lambda / N)^{-1/2} \delta + V$, where $V' \Lambda = 0$. Note that $\gamma' \gamma / N = \delta' \delta + V' V / N$, so that for all $\gamma \in \Gamma$, $\delta' \delta \leq 1$. Thus we can write

$$\sup_{\gamma \in \Gamma} R^*(\gamma) = \sup_{\delta, \delta' \delta \leq 1} \delta' (\Lambda' \Lambda / N)^{1/2'} (F' F / T) (\Lambda' \Lambda / N)^{1/2} \delta = \hat{\sigma}_{11},$$

where $\hat{\sigma}_{11}$ is the largest eigenvalue of $(\Lambda' \Lambda / N)^{1/2'} (F' F / T) (\Lambda' \Lambda / N)^{1/2}$. But $(\Lambda' \Lambda / N)^{1/2} \rightarrow I$ by F1(i) and $F' F / T \xrightarrow{p} \Sigma_{FF}$ by F1(d), so that $(\Lambda' \Lambda / N)^{1/2'} (F' F / T) (\Lambda' \Lambda / N)^{1/2} \xrightarrow{p} \Sigma_{FF}$ and (by continuity of eigenvalues) $\hat{\sigma}_{11} \xrightarrow{p} \sigma_{11}$.

(R9) $\sup_{\gamma \in \Gamma} R(\gamma) \xrightarrow{p} \sigma_{11}$.

Proof. This follows from (R7) and (R8).

(R10) Let $\hat{\lambda}_1 = \arg \sup_{\gamma \in \Gamma} R(\gamma)$; then $R^*(\hat{\lambda}_1) \xrightarrow{p} \sigma_{11}$.

Proof. This follows from (R6) and (R9).

(R11) Let $\hat{\lambda}_1$ denote the first column of $\hat{\Lambda}$ and let $S_1 = \text{sign}(\hat{\lambda}_1' \underline{\lambda}_1)$, meaning $S_1 = 1$ if $\hat{\lambda}_1' \underline{\lambda}_1 \geq 0$ and $S_1 = -1$ if $\hat{\lambda}_1' \underline{\lambda}_1 < 0$.

Then $(S_1 \hat{\lambda}_1' \Lambda / N) \xrightarrow{p} \ell'_1$, where $\ell_1 = (100 \dots 0)'$.

Proof. For particular values of $\hat{\delta}$ and \hat{V} , we can write $\hat{\lambda}_1 = \Lambda (\Lambda' \Lambda / N)^{-1/2} \hat{\delta} + \hat{V}$, where $\hat{V}' \Lambda = 0$ and $\hat{\delta}' \hat{\delta} \leq 1$. (Note that $\hat{\delta}$ is $r \times 1$.) Let $C_{NT} = (\Lambda' \Lambda / N)^{1/2'} (F' F / T) (\Lambda' \Lambda / N)^{1/2}$ and note that $R^*(\hat{\lambda}_1) = \hat{\delta}' C_{NT} \hat{\delta}$. Thus

$$R^*(\hat{\lambda}_1) - \sigma_{11} = \hat{\delta}' (C_{NT} - \Sigma_{FF}) \hat{\delta} + \hat{\delta}' \Sigma_{FF} \hat{\delta} - \sigma_{11} \\ = \hat{\delta}' (C_{NT} - \Sigma_{FF}) \hat{\delta} + (\hat{\delta}'_1 - 1) \sigma_{11} + \sum_{i=2}^r \hat{\delta}'_i \sigma_{ii}.$$

Because $C_{NT} \xrightarrow{p} \Sigma_{FF}$ and $\hat{\delta}$ is bounded, the first term on the right side of this expression is $o_p(1)$. This result together with (R10) implies that $(\hat{\delta}'_1 - 1) \sigma_{11} + \sum_{i=2}^r \hat{\delta}'_i \sigma_{ii} \xrightarrow{p} 0$. Because $\sigma_{ii} > 0$, $i = 1, \dots, r$ [Assumption F1(b)], this implies that $\hat{\delta}'_1 \xrightarrow{p} 1$ and $\hat{\delta}'_i \xrightarrow{p} 0$ for $i > 1$. Note that this result, together with $\hat{\lambda}'_1 \hat{\lambda}_1 / N = 1$, implies that $\hat{V}' \hat{V} / N \xrightarrow{p} 0$.

The result then follows from the assumption that $\Lambda' \Lambda / N \rightarrow I_r$ [Assumption F1(a)].

(R12) Suppose that the $N \times r$ matrix $\hat{\Lambda}$ is formed as the r ordered eigenvectors of $X'X$ normalized as $\hat{\Lambda}' \hat{\Lambda} / N = I$ (with the first column corresponding to the largest eigenvalue, etc.) Let S denote $S = \text{diag}(\text{sign}(\hat{\Lambda}' \Lambda))$. Then $S \hat{\Lambda}' \Lambda / N \xrightarrow{p} I$.

Proof. The result for the first column of $S \hat{\Lambda}' \Lambda / N$ is given in (R11). The results for the other columns mimic the steps in (R8)–(R11), for the other principal components, that is, by maximizing $R(\cdot)$ and $R^*(\cdot)$ sequentially using orthonormal subspaces of Γ . For later reference, we note that this process yields a representation of the j th column of $\hat{\Lambda}$ as

$$\hat{\lambda}_j = \Lambda (\Lambda' \Lambda / N)^{-1/2} \hat{\delta}_j + \hat{V}_j,$$

where $\hat{V}'_j \Lambda = 0$, $\hat{V}'_j \hat{V}_j / N \xrightarrow{p} 0$, and $\hat{\delta}'_{ij} \xrightarrow{p} 0$ for $i \neq j$ and $\hat{\delta}'_{jj} \xrightarrow{p} 1$.

(R13) For $j = 1, \dots, r$, $T^{-1} \sum_t \hat{F}_{jt}^2 = R(\hat{\lambda}_j) \xrightarrow{p} \sigma_{jj}$.

Proof. $T^{-1} \sum_i \hat{F}_{it}^2 = R(\hat{\lambda}_j)$ by the definitions of \hat{F} and $R(\cdot)$. The convergence result follows from (R9) for $j = 1$ and from the steps outlined in (R12) for $j = 2, \dots, r$.

(R14) For $i > r$, $T^{-1} \sum_i \hat{F}_{it}^2 \xrightarrow{p} 0$.

Proof. Let $\hat{\gamma}$ denote the i th ordered eigenvector of $X'X$, $i > r$, normalized so that $\hat{\gamma}'\hat{\gamma}/N = 1$. Then $T^{-1} \sum_i \hat{F}_{it}^2 = R(\hat{\gamma})$ follows from the definition of $R(\cdot)$ and \hat{F}_{it} .

Now, for particular values of $\hat{\alpha}$ and \hat{V} , we can write

$$\hat{\gamma} = \Lambda(\Lambda'\Lambda/N)^{-1/2}\hat{\alpha} + \hat{V},$$

where $\hat{V}'\Lambda = 0$, $\hat{\alpha}'\hat{\alpha} \leq 1$, and $\hat{V}'\hat{V}/N \leq 1$. Now, by construction, $\hat{\gamma}'\hat{\lambda}_j = 0$ for $j = 1, \dots, r$. Using the representation for $\hat{\lambda}_j$ given in (R12), we can write $N^{-1}\hat{\gamma}'\hat{\lambda}_j = \hat{\alpha}'\hat{\delta}_j + \hat{V}'\hat{V}_j/N = 0$. Because from (R12), $\hat{V}_j'\hat{V}_j \xrightarrow{p} 0$ and $\hat{V}'\hat{V}/N \leq 1$, $\hat{V}'\hat{V}_j \xrightarrow{p} 0$. Thus $\hat{\alpha}'\hat{\delta}_j \xrightarrow{p} 0$, so $\hat{\alpha}'[\hat{\delta}_1 \dots \hat{\delta}_r] \xrightarrow{p} 0$. But $[\hat{\delta}_1 \dots \hat{\delta}_r] \xrightarrow{p} I_r$, so $\hat{\alpha} \xrightarrow{p} 0$.

Thus $R^*(\hat{\gamma}) = \hat{\alpha}'(\Lambda'\Lambda/N)^{1/2}(F'F/T)(\Lambda'\Lambda/N)^{1/2}\hat{\alpha} \xrightarrow{p} 0$. The result then follows from (R6).

(R15) $S_j\hat{F}_{jt} - F_{jt} \xrightarrow{p} 0$ for $j = 1, \dots, r$.

Proof. $S_j\hat{F}_{jt} - F_{jt} = S_j\hat{\lambda}_j'X_t/N - F_{jt} = (S_j\hat{\lambda}_j'\hat{\lambda}_j/N - 1)F_{jt} + \sum_{i \neq j}(S_j\hat{\lambda}_j'\hat{\lambda}_i/N)F_{it} + S_j\hat{\lambda}_j'e_{it}/N$.

Because $S(\hat{\Lambda}'\Lambda/N) \xrightarrow{p} I$ from (R12), $(S_j\hat{\lambda}_j'\hat{\lambda}_j/N - 1) \xrightarrow{p} 0$ and $S_j\hat{\lambda}_j'\hat{\lambda}_i/N \xrightarrow{p} 0$ for $i \neq j$. Because $|F_T|$ is $O_p(1)$ [which follows from $E(F_T F_T') = \Sigma_{FF}$], $(S_j\hat{\lambda}_j'\hat{\lambda}_j/N - 1)F_{jt} \xrightarrow{p} 0$ and $\sum_{i \neq j}(S_j\hat{\lambda}_j'\hat{\lambda}_i/N) \times F_{it} \xrightarrow{p} 0$. The result then follows by showing that $S_j\hat{\lambda}_j'e_{it}/N \xrightarrow{p} 0$.

Now

$$N^{-1}S_j\hat{\lambda}_j'e_{it} = N^{-1} \sum_i S_j\hat{\lambda}_{ij}e_{it} = N^{-1} \sum_i (S_j\hat{\lambda}_{ij} - \lambda_{ij})e_{it} + N^{-1} \sum_i \lambda_{ij}e_{it}.$$

Also,

$$\left| N^{-1} \sum_i (S_j\hat{\lambda}_{ij} - \lambda_{ij})e_{it} \right| \leq \left(N^{-1} \sum_i (S_j\hat{\lambda}_{ij} - \lambda_{ij})^2 \right)^{1/2} \left(N^{-1} \sum_i e_{it}^2 \right)^{1/2}$$

and

$$N^{-1} \sum_i (S_j\hat{\lambda}_{ij} - \lambda_{ij})^2 = (\hat{\lambda}_j'\hat{\lambda}_j/N) + (\hat{\lambda}_j'\hat{\lambda}_j/N) - 2(S_j\hat{\lambda}_j'\hat{\lambda}_j/N) \xrightarrow{p} 0$$

by (R12). Finally, $N^{-1} \sum_i e_{it}^2 \sim O_p(1)$ by (R1). Thus $|N^{-1} \sum_i (S_j\hat{\lambda}_{ij} - \lambda_{ij})e_{it}| \xrightarrow{p} 0$ by Slutsky's theorem. This means that $S_j\hat{\lambda}_j'e_{it}/N = N^{-1} \sum_i \lambda_{ij}e_{it} + o_p(1)$. Now

$$\begin{aligned} E \left[N^{-1} \sum_i \lambda_{ij}e_{it} \right]^2 &= N^{-2} \sum_i \sum_k \lambda_{ij}\lambda_{kj} E(e_{it}e_{kt}) \\ &= N^{-2} \sum_i \sum_k \lambda_{ij}\lambda_{kj}\tau_{ik,t} \leq \bar{\lambda}^2 N^{-2} \sum_i \sum_k |\tau_{ik,t}| \rightarrow 0, \end{aligned}$$

where the final inequality uses the bound on λ_{ij} given in F1(c) and the limit uses M1(b). Thus $S_j\hat{\lambda}_j'e_{it}/N \xrightarrow{p} 0$.

(R16) Let q_t denote a sequence of random variables with $T^{-1} \sum_i q_t^2 \xrightarrow{p} \sigma_q^2$ and $T^{-1} \sum_i F_t q_t \xrightarrow{p} \Sigma_{Fq}$. Then $T^{-1} \sum_i S \hat{F}_t q_t \xrightarrow{p} \Sigma_{Fq}$.

Proof.

$$\begin{aligned} T^{-1} \sum_i S \hat{F}_t q_t &= N^{-1} T^{-1} \sum_i S \hat{\Lambda}' X_t q_t \\ &= T^{-1} \sum_i (S \hat{\Lambda}' \Lambda / N) F_t q_t + T^{-1} N^{-1} \sum_i S \hat{\Lambda}' e_t q_t. \end{aligned}$$

Because $(S \hat{\Lambda}' \Lambda / N) \xrightarrow{p} I$ by (R12) and $T^{-1} \sum_i F_t q_t \xrightarrow{p} \Sigma_{Fq}$ by assumption,

$$T^{-1} \sum_i (S \hat{\Lambda}' \Lambda / N) F_t q_t \xrightarrow{p} \Sigma_{Fq}.$$

Now the j th element of $T^{-1} N^{-1} \sum_i S \hat{\Lambda}' e_t q_t$ satisfies

$$\begin{aligned} \left| T^{-1} N^{-1} \sum_i S_j \hat{\lambda}_j' e_t q_t \right| &= \left| T^{-1} \sum_i q_t \left(N^{-1} \sum_i \hat{\lambda}_{ij} e_{it} \right) \right| \\ &\leq \sup_{\gamma \in \Gamma} \left| T^{-1} \sum_i q_t \left(N^{-1} \sum_i \gamma_i e_{it} \right) \right| \xrightarrow{p} 0, \end{aligned}$$

where the final inequality follows because $\hat{\lambda}_j \in \Gamma$ and the limit follows from (R3).

(R17) $T^{-1} \sum_i S \hat{F}_t F_t \xrightarrow{p} \Sigma_{FF}$.

Proof. This follows from (R16) with $q_t = F_{jt}$, $j = 1, \dots, r$. Assumption F1(d) shows that this choice of q_t satisfies the restriction in (R16).

(R18) $T^{-1} \sum_i \hat{F}_t \hat{F}_t' \xrightarrow{p} \Sigma_{FF}$.

Proof. This follows from (R16) and (R17). Set $q_t = S_j \hat{F}_{jt}$. (R13) shows that q_t satisfies the conditions of (R16).

(R19) For $i = 1, 2, \dots, r$, $T^{-1} \sum_i (S_i \hat{F}_{it} - F_{it})^2 \xrightarrow{p} 0$.

Proof. $T^{-1} \sum_i (S_i \hat{F}_{it} - F_{it})^2 = T^{-1} \sum_i \hat{F}_{it}^2 + T^{-1} \sum_i F_{it}^2 - 2T^{-1} \sum_i S_i \hat{F}_{it} F_{it} \xrightarrow{p} \sigma_{ii} + \sigma_{ii} - 2\sigma_{ii} = 0$ by (R18), F1(d), and (R17).

(R20)–(R23) Hold Given F1, M1, and Y1

(R20) $T^{-1} \sum_i S \hat{F}_t w_t' \xrightarrow{p} \Sigma_{Fw}$.

Proof. This follows from (R16), with q_t equal to w_{jt} . Y1(b) shows that this definition of q_t satisfies the conditions of (R16).

(R21) $T^{-1} \sum_i S \hat{F}_t \varepsilon_{t+h} \xrightarrow{p} 0$.

Proof. This follows from (R16) and Y1(c) with q_t equal to ε_{t+h} . Y1(d) shows that this definition of q_t satisfies the conditions of (R16).

(R22) With $\hat{\beta}$ partitioned as $\hat{\beta} = (\hat{\beta}_z' \hat{\beta}_w')'$, $\hat{\beta}_w - \beta_w \xrightarrow{p} 0$ and $S_i \hat{\beta}_{iz} - \beta_{iz} \xrightarrow{p} 0$ for $i = 1, \dots, r$.

Proof. Write

$$\begin{aligned} &\begin{pmatrix} S \hat{\beta}_z \\ \hat{\beta}_w \end{pmatrix} - \begin{pmatrix} \beta_z \\ \beta_w \end{pmatrix} \\ &= \begin{pmatrix} T^{-1} \sum_i \hat{F}_t \hat{F}_t' & T^{-1} S \sum_i \hat{F}_t w_t' \\ T^{-1} \sum_i w_t \hat{F}_t' S & T^{-1} \sum_i w_t w_t' \end{pmatrix}^{-1} \begin{pmatrix} T^{-1} S \sum_i \hat{F}_t \varepsilon_{t+h} \\ T^{-1} \sum_i w_t \varepsilon_{t+h} \end{pmatrix}. \end{aligned}$$

Because $T^{-1} \sum_i \hat{F}_t \hat{F}_t' \xrightarrow{p} \Sigma_{FF}$ (R18), $T^{-1} S \sum_i \hat{F}_t w_t' \xrightarrow{p} \Sigma_{Fw}$ (R20), $T^{-1} \sum_i w_t w_t' \xrightarrow{p} \Sigma_{ww}$ [Y1(b)], $T^{-1} S \sum_i \hat{F}_t \varepsilon_{t+h} \xrightarrow{p} 0$ (R21), and $T^{-1} \sum_i w_t \varepsilon_{t+h} \xrightarrow{p} 0$ [Y1(c)]; and because Σ_{zz} is nonsingular [Y1(a)], the result follows by Slutsky's theorem.

(R23) Let $\hat{z}_t = (\hat{F}_{1t} \hat{F}_{2t} \dots \hat{F}_{rt} w_t')'$ and $\hat{\beta} = (\sum_{t=1}^{T-h} \hat{z}_t \hat{z}_t')^{-1} (\sum_{t=1}^{T-h} \hat{z}_t y_{t+h})$; then $\hat{\beta}' \hat{z}_T - \beta' z_T \xrightarrow{p} 0$.

Proof. Let $R = \begin{bmatrix} S & 0 \\ 0 & I_{n_w} \end{bmatrix}$, where n_w denotes the number of elements in w_t .

$$\begin{aligned} \hat{\beta}' \hat{z}_T - \beta' z_T &= (R \hat{\beta})' R \hat{z}_T - \beta' z_T \\ &= (R \hat{\beta} - \beta)' z_T + (R \hat{\beta})' (R \hat{z}_T - z_T). \end{aligned}$$

Because z_T is $O_p(1)$ [because $E(z_T z_T') = \Sigma_{zz}$ by Y1(a)], and $R\hat{\beta} - \beta \xrightarrow{p} 0$ (R22), the first term vanishes in probability by Slutsky's theorem. Similarly, because β is finite [Assumption Y1(e)] and $R\hat{z}_T - z_T \xrightarrow{p} 0$ (R15), the second term vanishes in probability by Slutsky's theorem.

Proof of Theorem 1

Part a is proved by (R19); part b, by (R15); and part c, by (R14).

Proof of Theorem 2

Part a is proved by (R23); part b, by (R22).

Proof of Theorem 3

The model can be written as

$$X_t = \Lambda_0 F_t + a_t,$$

where $a_{it} = e_{it} + J_{iT} \xi'_{it} F_t$, $J_{iT} = T g_{iT}$, $\xi_{it} = T^{-1} \sum_{s=1}^t \xi_{is}$, and where (from Assumption F2) Λ_0 satisfies the same conditions as Λ in Assumption F1. Thus the results follows if the error terms a_t satisfy the assumptions in M1.

We prove a set of set of results (S0–S5) that yields the results. S0 is a preliminary result. S1–S3 show that a_t satisfies the assumption in M1.

S0.

Let $J_{iT} = T g_{iT}$. Then constants $\kappa_1 - \kappa_4$ can be chosen so that

$$\begin{aligned} \sup_i E|J_{iT}| &< \kappa_1, \\ \sup_{i,j} E|J_{iT} J_{jT}| &< \kappa_2, \\ \sup_{i,j,k} E|J_{iT} J_{jT} J_{kT}| &< \kappa_3, \\ \text{and} \\ \sup_{i,j,k,l} E|J_{iT} J_{jT} J_{kT} J_{lT}| &< \kappa_4. \end{aligned}$$

This follows from F2(a).

S1.

The error term a satisfies Assumption M1(a). That is, $\lim_{N,T \rightarrow \infty} \sup_t \sum_{u=1-t}^{T-t} |E(N^{-1} \sum_i a_{it} a_{it+u})| < \infty$.

Proof. $a_{it} a_{it+u} = e_{it} e_{it+u} + J_{iT}^2 (F'_t \xi_{it})(F'_{t+u} \xi_{it+u}) + J_{iT} e_{it} F'_{t+u} \xi_{it+u} + J_{iT} e_{it+u} F'_t \xi_{it}$.

We consider each part in turn.

S1(a). $\lim_{N \rightarrow \infty} \sup_t \sum_u |E(N^{-1} \sum_i e_{it} e_{it+u})| < \infty$.

Proof. This is Assumption M1(a).

S1(b). $\lim_{N \rightarrow \infty} \sup_t \sum_u |E(N^{-1} \sum_i J_{iT}^2 (F'_t \xi_{it})(F'_{t+u} \xi_{it+u}))| < \infty$.

Proof. $J_{iT}^2 (F'_t \xi_{it})(F'_{t+u} \xi_{it+u}) = J_{iT}^2 \sum_l \sum_m F_{lt} F_{mt+u} \xi_{it,l} \xi_{it+u,m}$, and we show the result for each term in the sum:

$$\begin{aligned} &\sup_t \sum_{u=1-t}^{T-t} \left| E \left(N^{-1} \sum_i J_{iT}^2 F_{lt} F_{mt+u} \xi_{it,l} \xi_{it+u,m} \right) \right| \\ &= \sup_t T^{-2} \sum_{u=1-t}^{T-t} \left| E \left(N^{-1} \sum_{i=1}^N J_{iT}^2 \sum_{s=1}^t \sum_{q=1}^{t+u} F_{lt} F_{mt+u} \xi_{is,l} \xi_{iq,m} \right) \right| \\ &\leq \kappa_2 \sup_t T^{-2} \sum_{u=1-t}^{T-t} N^{-1} \sum_{i=1}^N \sum_{s=1}^t \sum_{q=1}^{t+u} |E(F_{lt} F_{mt+u} \xi_{is,l} \xi_{iq,m})| \\ &\leq \kappa_2 \sup_t T^{-2} \sum_{u=1-t}^{T-t} N^{-1} \sum_{i=1}^N \sum_{s=1}^t \sum_{q=1}^{t+u-s} |E(F_{lt} F_{mt+u} \xi_{is,l} \xi_{is+v,m})| \end{aligned}$$

$$\begin{aligned} &\leq \kappa_2 \sup_t T^{-2} \sum_{u=1-t}^{T-t} N^{-1} \sum_{i=1}^N \sum_{s=1}^t \sum_{v=1-p}^{T-p} |E(F_{lt} F_{mt+u} \xi_{ip,l} \xi_{ip+v,m})| \\ &\quad \times \sup_{a,b,p,l,m} |E(F_{la} F_{mb} \xi_{ip,l} \xi_{ip+v,m})| \\ &\leq \kappa_2 \sup_t T^{-1} \sum_{u=1-t}^{T-t} N^{-1} \sum_{i=1}^N \sum_{v=1-p}^{T-p} |E(F_{la} F_{mb} \xi_{ip,l} \xi_{ip+v,m})| \\ &\quad \times \sup_{a,b,p,l,m} |E(F_{la} F_{mb} \xi_{ip,l} \xi_{ip+v,m})| \\ &\leq \kappa_2 \sup_t T^{-1} \sum_{u=1-t}^{T-t} \sup_i \sum_{v=1-p}^{T-p} \sup_{a,b,p,l,m} |E(F_{la} F_{mb} \xi_{ip,l} \xi_{ip+v,m})| \\ &\leq \kappa_2 \sup_i \sum_{v=1-p}^{T-p} \sup_{a,b,p,l,m} |E(F_{la} F_{mb} \xi_{ip,l} \xi_{ip+v,m})| \\ &< \infty, \end{aligned}$$

where the first line follows from the definition of ξ , the next line follows from independence of J_{iT} and the bound given in S0, the next line redefines the index of summation, the next line follows from definition of the sup, the next line follows because the summand in the summation over s does not depend on s , the next line follows from the definition of the sup, the next line follows because the summand in the summation over u does not depend on u , and the final line follows from Assumption M2(a).

S1(c). $\lim_{N \rightarrow \infty} \sup_t \sum_{u=1-t}^{T-t} |E[N^{-1} \sum_i J_{iT} e_{it} (F'_{t+u} \xi_{it+u})]| < \infty$.

Proof. $J_{iT} e_{it} F'_{t+u} \xi_{it+u} = J_{iT} \sum_m e_{it} F_{mt+u} \xi_{it+u,m}$ and we show the result for each term in the sum:

$$\begin{aligned} &\sup_t \sum_{u=1-t}^{T-t} \left| E \left[N^{-1} \sum_i J_{iT} e_{it} F_{mt+u} \xi_{it+u,m} \right] \right| \\ &= \sup_t T^{-1} \sum_{u=1-t}^{T-t} \left| E \left[N^{-1} \sum_{i=1}^N J_{iT} e_{it} F_{mt+u} \xi_{is,m} \right] \right| \\ &\leq \kappa_1 \sup_t T^{-1} \sum_{u=1-t}^{T-t} N^{-1} \sum_{i=1}^N \sum_{s=1}^{t+u} |E(e_{it} F_{mt+u} \xi_{is,m})| \\ &\leq \kappa_1 \sup_t T^{-1} \sum_{u=1-t}^{T-t} N^{-1} \sum_{i=1}^N \sum_{v=1-p}^{T-p} \sup_{s,p,m} |E(e_{ip} F_{ms} \xi_{ip+v,m})| \\ &\leq \kappa_1 \sup_t T^{-1} \sum_{u=1-t}^{T-t} \sup_i \sum_{v=1-p}^{T-p} \sup_{s,p,m} |E(e_{ip} F_{ms} \xi_{ip+v,m})| \\ &\leq \kappa_1 \sup_i \sum_{v=1-p}^{T-p} \sup_{s,p,m} |E(e_{ip} F_{ms} \xi_{ip+v,m})| \\ &< \infty, \end{aligned}$$

where the first line follows from the definition of ξ , the next line follows from independence of J_{iT} and the bound given in S0, the next line follows definition of the sup, the next line follows from the definition of the sup, the next line follows because the summand in the summation over u does not depend on u , and the final line follows from Assumption M2(b).

S1(d). $\lim_{N \rightarrow \infty} \sup_t \sum_{u=1-t}^{T-t} |E[N^{-1} \sum_i J_{iT} e_{it+u} (F'_t \xi_{it})]| < \infty$.

Proof. The proof parallels S1(c).

S2.

The error terms a , satisfy M1(b); that is, $\lim_{N \rightarrow \infty} \sup_t N^{-1} \sum_{i=1}^N \sum_{j=1}^N |E(a_{it} a_{jt})| < \infty$.

Proof. $a_{it} a_{jt} = e_{it} e_{jt} + J_{iT} J_{jT} (F'_t \xi_{it})(F'_j \xi_{jt}) + J_{jT} e_{it} F'_t \xi_{it} + J_{iT} e_{jt} F'_j \xi_{jt}$.

We consider each part in turn.

$$\mathbf{S2(a)}. \lim_{N \rightarrow \infty} \sup_t N^{-1} \sum_{i=1}^N \sum_{j=1}^N |E(e_{it}e_{jt})| < \infty.$$

Proof. This is Assumption M1(a).

$$\mathbf{S2(b)}. \lim_{N \rightarrow \infty} \sup_t N^{-1} \sum_{i=1}^N \sum_{j=1}^N |E[J_{iT}J_{jT}(F'_t\xi_{it})(F'_t\xi_{jt})]| < \infty.$$

Proof. $J_{iT}J_{jT}(F'_t\xi_{it})(F'_t\xi_{jt}) = \sum_l \sum_m J_{iT}J_{jT}F_{lt}\xi_{it,l}F_{mt}\xi_{jt,m}$, and we show the required result for each term in the sum:

$$\begin{aligned} & \sup_t N^{-1} \sum_{i=1}^N \sum_{j=1}^N |E[J_{iT}J_{jT}F_{lt}\xi_{it,l}F_{mt}\xi_{jt,m}]| \\ &= \sup_t N^{-1} T^{-2} \sum_{i=1}^N \sum_{j=1}^N \left| E \left[J_{iT}J_{jT} \sum_{s=1}^t \sum_{q=1}^t F_{ls}\xi_{is,l}F_{mq}\xi_{jq,m} \right] \right| \\ &\leq \kappa_2 \sup_t N^{-1} T^{-2} \sum_{i=1}^N \sum_{j=1}^N \sum_{s=1}^t \sum_{q=1}^t |E(F_{ls}\xi_{is,l}F_{mq}\xi_{jq,m})| \\ &\leq \kappa_2 N^{-1} \sum_{i=1}^N \sum_{j=1}^N \sup_{l,t,s,v,q,m} |E(F_{lt}\xi_{is,l}F_{mv}\xi_{jq,m})| \\ &< \infty, \end{aligned}$$

where the first line follows from the definition of ξ , the next line follows from the independence of J and the bound in S0, the next line follows from the definition of the sup and the fact that there are t terms in the summations involving s and q , and the final line follows from M2c.

$$\mathbf{S2(c)}. \lim_{N \rightarrow \infty} \sup_t N^{-1} \sum_{i=1}^N \sum_{j=1}^N |E(J_{iT}e_{jt}F'_t\xi_{it})| < \infty.$$

Proof. $J_{iT}e_{jt}F'_t\xi_{it} = J_{iT}e_{jt} \sum_m F_{mt}\xi_{it,m}$, and we show the required result for each term in the sum:

$$\begin{aligned} & \sup_t N^{-1} \sum_{i=1}^N \sum_{j=1}^N |E(J_{iT}e_{jt}F_{mt}\xi_{it,m})| \\ &= \sup_t T^{-1} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \left| E \sum_{s=1}^t (J_{iT}e_{jt}F_{ms}\xi_{is,m}) \right| \\ &\leq \kappa_1 \sup_t T^{-1} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \sum_{s=1}^t |E(e_{jt}F_{ms}\xi_{is,m})| \\ &\leq \kappa_1 N^{-1} \sum_{i=1}^N \sum_{j=1}^N \sup_{t,s,q,m} |E(e_{jt}F_{ms}\xi_{iq,m})| \\ &< \infty, \end{aligned}$$

where the first line follows from the definition of ξ , the next line follows from the independence of J and the bound in S0, the next line follows from the definition of the sup and the fact that there are t terms in the summations involving s , and the final line follows from M2(d).

$$\mathbf{S2(d)}. \lim_{N \rightarrow \infty} \sup_t N^{-1} \sum_{i=1}^N \sum_{j=1}^N |E(J_{jT}e_{it}F'_t\xi_{jt})| < \infty.$$

Proof. The results is implied by S2(c).

S3.

The error terms a satisfy M1(c); that is

$$\lim_{N \rightarrow \infty} \sup_{t,s} N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\text{cov}(a_{is}a_{it}, a_{js}a_{jt})| < \infty.$$

Proof. $a_{it}a_{is} = e_{it}e_{is} + J_{iT}^2(F'_t\xi_{it})(F'_s\xi_{is}) + J_{iT}e_{it}F'_s\xi_{is} + J_{iT}e_{is}F'_t\xi_{it}$, and so $\text{cov}(a_{is}a_{it}, a_{js}a_{jt})$ is made of 16 terms, which have 6 possible forms:

- (a) $\text{cov}(e_{is}e_{it}, e_{js}e_{jt})$
- (b) $\text{cov}(e_{is}e_{it}, J_{jT}^2(F'_t\xi_{jt})(F'_s\xi_{js}))$

- (c) $\text{cov}(e_{is}e_{it}, J_{jT}e_{jt}F'_s\xi_{js})$
- (d) $\text{cov}(J_{iT}^2(F'_t\xi_{it})(F'_s\xi_{is}), J_{jT}^2(F'_t\xi_{jt})(F'_s\xi_{js}))$
- (e) $\text{cov}(J_{iT}^2(F'_t\xi_{it})(F'_s\xi_{is}), J_{jT}e_{jt}F'_s\xi_{js})$
- (f) $\text{cov}(J_{iT}e_{it}F'_s\xi_{is}, J_{jT}e_{jt}F'_s\xi_{js})$.

We consider each of these terms in turn.

$$\mathbf{S3(a)}. \lim_{N \rightarrow \infty} \sup_{t,s} N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\text{cov}(e_{is}e_{it}, e_{js}e_{jt})| < \infty.$$

Proof. This is Assumption M1(c).

$$\mathbf{S3(b)}. \lim_{N \rightarrow \infty} \sup_{t,s} N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\text{cov}(e_{is}e_{it}, J_{jT}^2(F'_t\xi_{jt})(F'_s\xi_{js}))| < \infty.$$

Proof. $\text{cov}(e_{is}e_{it}, J_{jT}^2(F'_t\xi_{jt})(F'_s\xi_{js})) = \sum_l \sum_m \text{cov}(e_{is}e_{it}, J_{jT}^2F_{lt}\xi_{jt,l}F_{ms}\xi_{js,m})$, and it suffices to show the result for each term in the sum:

$$\begin{aligned} & \sup_{t,s} N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\text{cov}(e_{is}e_{it}, J_{jT}^2F_{lt}\xi_{jt,l}F_{ms}\xi_{js,m})| \\ &= \sup_{t,s} T^{-2} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \left| \sum_{q=1}^s \sum_{h=1}^s \text{cov}(e_{is}e_{it}, J_{jT}^2F_{lt}\xi_{jq,l}F_{ms}\xi_{jh,m}) \right| \\ &\leq \kappa_2 \sup_{t,s} T^{-2} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \sum_{q=1}^s \sum_{h=1}^s |\text{cov}(e_{is}e_{it}, F_{lt}\xi_{jq,l}F_{ms}\xi_{jh,m})| \\ &\leq \kappa_2 N^{-1} \sum_{i=1}^N \sum_{j=1}^N \sup_{t,s,q,h,m,l} |\text{cov}(e_{is}e_{it}, F_{lt}\xi_{jq,l}F_{ms}\xi_{jh,m})| \\ &< \infty, \end{aligned}$$

where the last line follows from M2(e)(1).

$$\mathbf{S3(c)}. \lim_{N \rightarrow \infty} \sup_{t,s} N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\text{cov}(e_{is}e_{it}, J_{jT}e_{jt}F'_s\xi_{js})| < \infty.$$

Proof. $\text{cov}(e_{is}e_{it}, J_{jT}e_{jt}F'_s\xi_{js}) = \sum_m \text{cov}(e_{is}e_{it}, J_{jT}e_{jt}F_{ms}\xi_{js,m})$, and it suffices to show the result for each term in the sum:

$$\begin{aligned} & \sup_{t,s} N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\text{cov}(e_{is}e_{it}, J_{jT}e_{jt}F_{ms}\xi_{js,m})| \\ &= \sup_{t,s} T^{-1} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \left| \sum_{h=1}^s \text{cov}(e_{is}e_{it}, J_{jT}e_{jt}F_{ms}\xi_{jh,m}) \right| \\ &\leq \kappa_1 \sup_{t,s} T^{-1} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \sum_{h=1}^s |\text{cov}(e_{is}e_{it}, e_{jt}F_{ms}\xi_{jh,m})| \\ &\leq \kappa_1 N^{-1} \sum_{i=1}^N \sum_{j=1}^N \sup_{t,s,h,m} |\text{cov}(e_{is}e_{it}, e_{jt}F_{ms}\xi_{jh,m})| \\ &< \infty, \end{aligned}$$

where the last line follows from M2(e)(2).

$$\mathbf{S3(d)}. \lim_{N \rightarrow \infty} \sup_{t,s} N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\text{cov}(J_{iT}^2(F'_t\xi_{it})(F'_s\xi_{is}), J_{jT}^2(F'_t\xi_{jt})(F'_s\xi_{js}))| < \infty.$$

Proof. $\text{cov}(J_{iT}^2(F'_t\xi_{it})(F'_s\xi_{is}), J_{jT}^2(F'_t\xi_{jt})(F'_s\xi_{js})) = \sum_{l_1} \sum_{l_2} \times \sum_{l_3} \sum_{l_4} \text{cov}(J_{iT}^2F_{l_1t}\xi_{it,l_1}F_{l_2s}\xi_{is,l_2}, J_{jT}^2F_{l_3t}\xi_{jt,l_3}F_{l_4s}\xi_{js,l_4})$, and it suffices to show the result for each term in the sum:

$$\begin{aligned} & \sup_{t,s} N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\text{cov}(J_{iT}^2F_{l_1t}\xi_{it,l_1}F_{l_2s}\xi_{is,l_2}, J_{jT}^2F_{l_3t}\xi_{jt,l_3}F_{l_4s}\xi_{js,l_4})| \\ &= \sup_{t,s} T^{-4} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \left| \left(\sum_{q_1}^t \sum_{q_2}^s \sum_{q_3}^t \sum_{q_4}^s \text{cov}(J_{iT}^2F_{l_1t}\xi_{iq_1,l_1}F_{l_2s}\xi_{iq_2,l_2}, \right. \right. \\ & \quad \left. \left. J_{jT}^2F_{l_3t}\xi_{jq_3,l_3}F_{l_4s}\xi_{jq_4,l_4}) \right) \right| \end{aligned}$$

$$\begin{aligned} &\leq \kappa_4 \sup_{t,s} T^{-4} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \sum_{q_1=1}^t \sum_{q_2=1}^s \sum_{q_3=1}^t \sum_{q_4=1}^s \\ &\quad \times |\text{cov}(F_{l_1 t} \zeta_{i q_1, l_1}, F_{l_2 s} \zeta_{i q_2, l_2}, F_{l_3 t} \zeta_{j q_3, l_3}, F_{l_4 s} \zeta_{j q_4, l_4})| \\ &\leq \kappa_4 N^{-1} \sum_{i=1}^N \sum_{j=1}^N \sup_{t,s, \{l_k, q_k\}_{k=1}^4} |\text{cov}(F_{l_1 t} \zeta_{i q_1, l_1}, F_{l_2 s} \zeta_{i q_2, l_2}, \\ &\quad F_{l_3 t} \zeta_{j q_3, l_3}, F_{l_4 s} \zeta_{j q_4, l_4})| < \infty, \end{aligned}$$

where the last line follows from M2(e)(3).

$$\mathbf{S3(e)}: \lim_{N \rightarrow \infty} \sup_{t,s} N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\text{cov}(J_{iT}^2 (F'_t \xi_{it})(F'_s \xi_{is}), J_{jT} e_{jt} F'_s \xi_{js})| < \infty.$$

Proof. $\text{cov}(J_{iT}^2 (F'_t \xi_{it})(F'_s \xi_{is}), J_{jT} e_{jt} F'_s \xi_{js}) = \sum_{l_1} \sum_{l_2} \sum_{l_3} \times \text{cov}(J_{iT}^2 F_{l_1 t} \xi_{it, l_1} F_{l_2 s} \xi_{is, l_2}, J_{jT} e_{jt} F_{l_3 s} \xi_{js, l_3})$, and it suffices to show the result for each term in the sum:

$$\begin{aligned} &\sup_{t,s} N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\text{cov}(J_{iT}^2 F_{l_1 t} \xi_{it, l_1} F_{l_2 s} \xi_{is, l_2}, J_{jT} e_{jt} F_{l_3 s} \xi_{js, l_3})| \\ &= \sup_{t,s} T^{-3} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \left| \left(\sum_{q_1=1}^t \sum_{q_2=1}^s \sum_{q_3=1}^s \text{cov}(J_{iT}^2 F_{l_1 t} \zeta_{i q_1, l_1} F_{l_2 s} \zeta_{i q_2, l_2}, \right. \right. \\ &\quad \left. \left. J_{jT} e_{jt} F_{l_3 s} \zeta_{j q_3, l_3}) \right) \right| \\ &\leq \kappa_3 \sup_{t,s} T^{-3} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \left| \left(\sum_{q_1=1}^t \sum_{q_2=1}^s \sum_{q_3=1}^s |\text{cov}(F_{l_1 t} \zeta_{i q_1, l_1} F_{l_2 s} \zeta_{i q_2, l_2}, \right. \right. \\ &\quad \left. \left. e_{jt} F_{l_3 s} \zeta_{j q_3, l_3}) \right) \right| \\ &\leq \kappa_3 N^{-1} \sum_{i=1}^N \sum_{j=1}^N \sup_{t,s, \{l_k, q_k\}_{k=1}^3} |\text{cov}(F_{l_1 t} \zeta_{i q_1, l_1} F_{l_2 s} \zeta_{i q_2, l_2}, e_{jt} F_{l_3 s} \zeta_{j q_3, l_3})| \\ &< \infty, \end{aligned}$$

where the last line follows from M2(e)(4).

$$\mathbf{S3(f)}: \lim_{N \rightarrow \infty} \sup_{t,s} N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\text{cov}(J_{iT} e_{it} F'_s \xi_{is}, J_{jT} e_{jt} F'_s \xi_{js})| < \infty.$$

Proof. $\text{cov}(J_{iT} e_{it} (F'_s \xi_{is}), J_{jT} e_{jt} F'_s \xi_{js}) = \sum_l \sum_m \text{cov}(J_{iT} e_{it} F_{ls} \xi_{is, l}, J_{jT} e_{jt} F_{ms} \xi_{js, m})$, and it suffices to show the result for each term in the sum:

$$\begin{aligned} &\sup_{t,s} N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\text{cov}(J_{iT} e_{it} F_{ls} \xi_{is, l}, J_{jT} e_{jt} F_{ms} \xi_{js, m})| \\ &= \sup_{t,s} T^{-2} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \left| \sum_{q_1=1}^s \sum_{q_2=1}^s \text{cov}(J_{iT} e_{it} F_{ls} \zeta_{i q_1, l}, J_{jT} e_{jt} F_{ms} \zeta_{j q_2, m}) \right| \end{aligned}$$

$$\begin{aligned} &\leq \kappa_2 \sup_{t,s} T^{-2} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \sum_{q_1=1}^s \sum_{q_2=1}^s |\text{cov}(e_{it} F_{ls} \zeta_{i q_1, l}, e_{jt} F_{ms} \zeta_{j q_2, m})| \\ &\leq \kappa_2 N^{-1} \sum_{i=1}^N \sum_{j=1}^N \sup_{t,s, q_1, q_2, l, m} |\text{cov}(e_{it} F_{ls} \zeta_{i q_1, l}, e_{jt} F_{ms} \zeta_{j q_2, m})| \\ &< \infty, \end{aligned}$$

where the last line follows from M2(e)(5).

[Received April 2000, Revised September 2001]

REFERENCES

- Bai, J., and Ng, S. (2001), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.
- Chamberlain, G., and Rothschild, M. (1983), "Arbitrage Factor Structure, and Mean-Variance Analysis of Large Asset Markets," *Econometrica*, 51, 1281–1304.
- Connor, G., and Korajczyk, R. A. (1986), "Performance Measurement With the Arbitrage Pricing Theory," *Journal of Financial Economics*, 15, 373–394.
- (1988), "Risk and Return in an Equilibrium APT: Application of a New Test Methodology," *Journal of Financial Economics*, 21, 255–289.
- (1993), "A Test for the Number of Factors in an Approximate Factor Model," *Journal of Finance*, XLVIII, 1263–1291.
- Ding, A. A., and Hwang, J. T. (1999), "Prediction Intervals, Factor Analysis Models, and High-Dimensional Empirical Linear Prediction," *Journal of the American Statistical Association*, 94, 446–455.
- Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000), "The Generalized Dynamic Factor Model: Identification and Estimation," *The Review of Economics and Statistics*, 82, 540–552.
- Forni, M., and Reichlin, L. (1996), "Dynamic Common Factors in Large Cross-Sections," *Empirical Economics*, 21, 27–42.
- (1998), "Let's Get Real: A Dynamic Factor Analytical Approach to Disaggregated Business Cycle," *Review of Economic Studies*, 65, 453–474.
- Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series," in *Latent Variables in Socio-Economic Models*, eds. D. J. Aigner and A. S. Goldberger, Amsterdam: North-Holland, Ch. 19.
- Lawley, D. N., and Maxwell, A. E. (1971), *Factor Analysis as a Statistical Method*, New York: American Elsevier Publishing.
- Sargent, T. J., and Sims, C. A. (1977), "Business Cycle Modeling Without Pretending to Have Too Much A Priori Economic Theory," in *New Methods in Business Cycle Research*, eds. C. Sims et al., Minneapolis: Federal Reserve Bank of Minneapolis.
- Stock, J. H., and Watson, M. W. (1989), "New Indexes of Coincident and Leading Economic Indicators," *NBER Macroeconomics Annual*, 351–393.
- (1996), "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics*, 14, 11–30.
- (1998a), "Diffusion Indexes," Working Paper 6702, NBER.
- (1998b), "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model," *Journal of the American Statistical Association*, 93, 349–358.
- (1999), "Forecasting Inflation," *Journal of Monetary Economics*, 44, 293–335.
- (2002), "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20, 147–162.