DIFFUSION INDEXES

James H. Stock

Kennedy School of Government, Harvard University
and the NBER

and

Mark W. Watson*

Woodrow Wilson School, Princeton University
and the NBER

June 1999

*This is a substantially revised version of NBER Working Paper #6702, August 1998. The authors thank Michael Boldin, Frank Diebold, Gregory Chow, Andrew Harvey, Lucrezia Reichlin, Ken Wallis, and Charles Whiteman for helpful discussions and/or comments, and Lewis Chan and Alexei Onatski for skilled research assistance. This research was supported in part by National Science Foundation grants no. SBR-9409629 and SBR-9730489.

ABSTRACT

This paper considers forecasting a single time series when there are there are many predictors (N) and time series observations (T). When the data follow an approximate dynamic factor model, the predictors can be summarized by a small number of indexes, akin to diffusion indexes. Estimation is discussed for balanced and unbalanced panels. Feasible forecasts are shown to be asymptotically efficient as $N,T\rightarrow\infty$ and $N=O(T^{\rho})$ for any $\rho>0$. The estimated dynamic factors are shown to be (uniformly) consistent, even in the presence of time variation in the population parameters and/or data contamination. The method is used to construct 6, 12, and 24 month ahead forecasts for 8 monthly U.S. macroeconomic time series using 215 predictors in simulated real time from 1970 through 1998. Over this sample period these new forecasts outperform various state-of-the-art benchmark models.

Key Words: Dynamic factor models, forecasting, principal components

JEL Codes: C32, E37

1. Introduction

Recent advances in information technology now make it possible to access in real time, at a reasonable cost, literally thousands of economic time series for major developed economies. This raises the prospect of a new frontier in macroeconomic forecasting, in which a very large number of time series are used to forecast a few key economic quantities such as output or inflation. Time series models currently used for macroeconomic forecasting, however, incorporate only a handful of series: vector autoregressions, for example, typically contain fewer than ten variables. Similarly, standard model selection methods based on information criteria work poorly when the number of candidate predictors is very large. Although thousands of time series are available in real time, some dimension reduction scheme is necessary before such a large number of variables can be used for forecasting.

In this paper, we use dynamic factor models as a framework for dimension reduction in macroeconomic forecasting. The premise is that, for forecasting purposes, a large number of predictor variables can be replaced by a handful of estimated factors. As proposed by Sargent and Sims (1977), a dynamic factor model expresses a N-dimensional time series X_t as the sum of a r-dimensional unobserved factor F_t and an idiosyncractic error vector. The data, X_t, are observed at dates $t=1,\ldots,T$. The dynamic factor model generalizes the familiar static factor model by introducing dynamics into the evolution of the factor and the idiosyncratic errors and by allowing X_t to depend on a distributed lag of F_t. In an exact factor model, the idiosyncratic errors are uncorrelated across series. However, this assumption is implausible for macroeconomic applications, and so instead we consider dynamic factor models that are approximate in the sense of Chamberlain and Rothschild (1983).

This paper makes theoretical, computational, and empirical contributions. As summarized below, there is a large literature on factor models, but for various reasons (discussed in section

2) the existing theoretical results are insufficient to support application of these methods to macroeconomic forecasting. We therefore provide some theoretical econometric results appropriate for this application. Section 2 studies forecasting an individual series, y_{t+1}, using many predictor variables X_t, when (y_{t+1}, X_t) follows a time-invariant approximate dynamic factor model and the factors are estimated by the principal components of X_t. We show that, given the number of factors, the resulting feasible forecast is asymptotically first-order efficient in the sense that the out-of-sample prediction error variance of the feasible forecast equals the prediction error variance of the infeasible conditional expectation $E(y_{T+1} | F_T)$. This result is shown to hold for $N=O(T^{\rho})$ for any $\rho>0$. The implication of this result is that principal components applied to a dynamic factor model can be expected to yield asymptotically efficient forecasts in a variety of applications, including $N>>T$, $N\simeq T$, and $N<<T$, as long as N and T are both large.

In section 3, the assumptions of section 2 are weakened in three main ways: the population factor loadings are allowed to vary over time; the number of estimated factors (k) and true factors (r) are allowed to differ; and the number of estimated factors is allowed increase to infinity as the sample size increases. In addition, stronger results than those in section 2 are presented, including results on uniform consistency of the estimated factors, rates of consistency, and determination of the number of factors needed for forecasting via information criteria. The weaker assumptions and stronger conclusions come at the cost of a stronger condition on the rate of data accumulation, and the results of section 3 apply to the case $N,T\to\infty$ when $N>>T$. The finite sample performance of these methods is examined in a Monte Carlo study, reported in section 4.

The computational contribution of the paper is a modification of principal components analysis using the EM algorithm for the analysis of dynamic factor models with unbalanced panels and mixed sampling frequencies. While conceptually straightforward, this extension is of practical importance in applications to macroeconomic data. For example, series may be available over different sample periods, or the data set might include both monthly and quarterly data. The particulars of the extension to unbalanced panel and other data irregularities are presented in appendix A.

Whether all this is relevant for macroeconomic forecasting is, of course, an empirical question. In section 5, we therefore look at forecasts of eight major monthly economic time series for the United States (four real series and four price inflation series) at the 6, 12, and 24 month horizons, using a simulated out of sample forecast comparison. The full data set spans 1959:1-1998:12 and contains 215 time series. Factors are extracted and forecasts are made using all 215 series and using a balanced panel subset of 149 series. Benchmark models are univariate and vector autoregressions and a multivariate model based on several leading economic indicators. Unemployment-based Phillips curve models are also used as forecast benchmarks for the inflation series. Generally speaking, factor model forecasts based on a small number of factors -- in most cases, one or two -- are found to perform very well, with relative performance improving as the horizon increases. The improvement over the benchmark forecasts can be dramatic, in several cases producing simulated out of sample mean square forecast errors that are one-third less than those of the benchmark models.

This paper is related to several earlier bodies of work. One important motivation, reflected in the title of this paper, is the use of diffusion indexes as developed by business cycle analysts at the National Bureau of Economic Research (NBER). These indexes are averages of contemporaneous values of a large number of time series, where the series and their weights are selected by expert judgment. A classic use of a diffusion index is to measure whether a recession reaches throughout the economy. Diffusion indexes have been used as leading and coincident economic indicators. Because it is an average of many variables, a diffusion index summarizes the information in a large number of economic time series. Our estimated factors are similar in spirit, except that the weights are computed by principal components analysis.

A number of studies have applied dynamic factor models to macroeconomic data. Geweke (1977) and Sargent and Sims (1977) analyzed these models in the frequency domain for a small number of variables. Engle and Watson (1981), Sargent (1989), and Stock and Watson (1991) estimated small-N parametric dynamic factor models by maximum likelihood in the time domain. Quah and Sargent (1993) extended the time domain MLE to more series (N=60). Forni and Reichlin (1996, 1997, 1998) and Forni, Hallin, Lippi and Reichlin (1998) have considered various methods for estimating factors in dynamic factor models with large N. In the first applications of this large cross section approach to macroeconomic data, they apply their methods to large regional and sectoral data sets, for example Forni and Reichlin (1998) extract and analyze common factors from data on productivity and output for 450 U.S. industries. Another closely related body of work uses static approximate factor structures to study asset prices. Contributions include Chamberlain and Rothschild (1983), Connor and Korajczyk (1986, 1988, 1993), Mei (1993), Schneewwiss and Mathes (1995), Bekker et. al. (1996), Geweke and Zhou (1996), and Zhou (1997); also see the survey in Campbell, Lo and McKinley (1996, chapter 6)). The papers most closely related to this one are Connor and Korajczyk (1986, 1988, 1993), Forni and Reichlin (1996, 1997, 1998), and Forni, Hallin, Lippi and Reichlin (1998), and these are discussed in more detail in section 2.

2. The Model and Estimation

2.1. The model

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+1}) admit a dynamic factor model representation with $\bar{r}$ common dynamic factors f_t, (2.1) $X_{it} = \bar{\lambda}_i(L)f_t + e_{it}$

(2.2) $y_{t+1} = \bar{\beta}(L)f_t + \epsilon_{t+1}$

where $e_t = (e_{1t},\ldots,e_{Nt})'$ is the $N\times 1$ idiosyncratic disturbance and $\bar{\lambda}_i(L)$ and $\bar{\beta}(L)$ are lag polynomials in non-negative powers of L. Both f_t and e_t are assumed to follow mean zero stationary stochastic processes (so X_t and y_{t+1} are deviations from their means). The errors in the forecasting equation (2.2) are assumed to be a homoskedastic martingale difference sequence with respect to $F_t=(X_t,f_t,\epsilon_t,X_{t-1},f_{t-1},\epsilon_{t-1},\ldots)$, so $E(\epsilon_{t+1}|F_t)=0$, and $E(\epsilon^2_{t+1}|F_t)=\sigma^2_\epsilon$; it is also assumed that the fourth moment of ϵ_t is finite. Although e_{it} and e_{js} are not required to be uncorrelated for $i\neq j$, the idiosyncratic terms are assumed to have limited dependence across series. Additional technical assumptions are given below.[1]

The assumption that $E(\epsilon_{t+1}|F_t)=0$, and thus $E(y_{t+1}|F_t)=\bar{\beta}(L)f_t$, is implied by the two conditions: (i) $E(y_{t+1}|F_t)$ depends on f_t and its lags but not otherwise on X_t, and (ii) lags of y_t do not enter (2.2). The first is the key condition that produces the desired dimension reduction when N is large. The second condition is not restrictive because y_{t+1} can be interpreted as a quasidifference, thereby implicitly incorporating its lags into (2.2).

It is assumed throughout that $\bar{\lambda}_i(L)$ and $\bar{\beta}(L)$ have finite orders of at most q, with $\bar{\lambda}_i(L)=\Sigma^q_{j=0}\bar{\lambda}_{ij}L^j$ and $\bar{\beta}(L)=\Sigma^q_{j=0}\bar{\beta}_jL^j$. Then (2.1) and (2.2) can be rewritten in a convenient "static" form,

(2.3) $X_t = \Lambda F^0_t + e_t$

(2.4) $y_{t+1} = \beta' F^0_t + \epsilon_{t+1}$

where $F^0_t = (f_t,\ldots,f_{t-q})$ is $r\times 1$, where $r=(q+1)\bar{r}$, the i-th row of Λ in (2.3) is $(\bar{\lambda}_{i0},\ldots,\bar{\lambda}_{iq})$, and $\beta=(\bar{\beta}_0,\ldots,\bar{\beta}_q)'$. The superscript "0" on F^0_t identifies the true values of the unobserved factors. We work with (2.3) and (2.4) throughout and focus on the problem of estimating the r unknown factors $\{F_t^0\}$. Because F_t^0 in general contains lags of the dynamic factors f_t, it may be dynamically singular in the sense that it has a singular spectral density matrix. This could pose problems for the interpretation of the factors, but not for forecasting. Rather, as is evident from (2.2) and (2.4), these lags are needed for forecasting so the dynamic redundancy is irrelevant, at least in population. As discussed in the next section, one can further use model selection methods to eliminate individual factors that do not enter the y_{t+1} equation.

The assumption that $\bar{\lambda}_i(L)$ and $\bar{\beta}(L)$ have finite orders q merits discussion. On the one hand, if the factors affect many variables with arbitrarily long lags, then this assumption would be inappropriate. On the other hand, if the effects of the factors largely occur within a few periods then this assumption might be acceptable for empirical forecasting. Of course, because the dynamics of f_t and e_t are largely unrestricted aside from stationarity and some technical conditions, even with finite order lag polynomials the model allows for a rich dynamic characterization of (X_t, y_{t+1}). Finally, the fact that parametric exact dynamic factors models, which use this finite-order assumption, have met with apparent empirical success (Stock and Watson [1991], Quah and Sargent [1993]) provides some reason to believe this assumption is empirically appropriate. Whether this assumption is acceptable ultimately is an empirical issue.

The model (2.3) and (2.4) contains several important special cases. One is the static factor model in which F_t^0 and $\{e_{jt}\}$ are mutually uncorrelated and i.i.d., and $\{e_{it}\}$ and $\{e_{jt}\}$ are independent for $i \neq j$. If $\{e_{jt}\}$ are serially uncorrelated but are correlated across series, subject to some limit on this correlation, then the model is an approximate static factor model, cf. Chamberlain and Rothschild (1983) and Connor and Korajczyk (1986, 1993). Like Forni, Lippi, Hallin and Reichlin (1998), we permit $\{e_{jt}\}$ to be weakly correlated across series and refer to (2.1) and (2.2) as an approximate dynamic factor model.

2.2. Estimation

The standard estimation method for parametric dynamic factor models is a two-step process: time invariant parameters are first estimated using Gaussian MLE, then these estimated parameters are used in standard signal extraction formulae (like the Kalman smoother) to estimate the unobserved factors. However, when N is very large this is not promising from a computational perspective. We therefore take a different approach and estimate the dynamic factors nonparametrically.

When the panel of data is balanced, we estimate the factors by the method of principal components. Consider the nonlinear least squares objective function,

$$V(F,\Lambda) = (NT)^{-1}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}(X_{it}-\lambda_i'F_t)^2 \tag{2.5}$$

where X_{it} is the observation on variable i at time t, λ_i is the i-th row of Λ, and $F=(F_1, F_2,\ldots,F_T)'$. After concentrating out F, minimizing (2.5) is equivalent to maximizing $tr[\Lambda'(X'X)\Lambda]$, subject to $\Lambda'\Lambda/N = I_r$, where X is the $T\times N$ data matrix with t-th row X_t', and tr(.) denotes the matrix trace. This is the classical principal components problem, which is solved by setting $\hat{\Lambda}$ equal to ($N^{1/2}$ times) the eigenvectors of $X'X$ corresponding to its r largest eigenvalues. Then the principal components estimator of F is,

$$\hat{F}=X\hat{\Lambda}/N. \tag{2.6}$$

Computation of $\hat{F}$ entails computing the eigenvectors of an $N\times N$ matrix, and when $N>T$ a computationally simpler approach is available. By concentrating out Λ, minimizing (2.5) is equivalent to maximizing $tr[F'(XX')F]$, subject to $F'F/T = I$, which yields the estimator $\tilde{F}$ that is ($T^{1/2}$ times) the matrix of the first r eigenvectors of the $T\times T$ matrix XX'. The column spaces of $\hat{F}$ and $\tilde{F}$ are equivalent, so for forecasting purposes they can be used interchangeably, depending on computational convenience.

From equations (2.3) and (2.4) X_t may include any variables known at date t, that are related to the factors that help predict y_{t+1}. Thus any variable dated t or earlier is a potential candidate element of X_t. If we let Z_t denote the set of variables that become available at date t, then X_t can be constructed as $X_t=Z_t$ or alternatively as a "stacked" version of Z_t, $X_t=(Z_t', Z_{t-1}', \ldots, Z_{t-m}')'$. The choice between these two versions of X_t depends whether lags of Z_t contain useful information about y_{t+1} that is not contained in Z_t. We investigate this empirically in section 5.

Computation is more difficult when there are data irregularities. For example, an unbalanced panel arises when there are missing observations and/or series that are available over shorter time spans. Another example is a data set consisting of both monthly and quarterly data. In these cases standard principal components analysis does not apply. Although (2.5), suitably modified for these data irregularities, continues to be a valid objective function, its direct minimization appears to be computationally infeasible when T or N is large. However, the EM algorithm can be used to estimate the factors by solving the minizimation problem iteratively. Details are given in appendix A.

2.3. Efficient forecasting using in (2.3) and (2.4)

Several studies consider the large sample properties of estimated factor models with large N. Connor and Korajczyk (1986, 1988, 1993) argue that factors estimated by principal components are consistent (at a given date) as $N\rightarrow\infty$ with T fixed in a static factor model. Schneeweiss and Mathes (1995) prove consistency of the principal component estimator $\hat{F}$ in an approximate static factor model when EX_tX_t', Λ, and Ee_te_t' are known, but of course these are unknown in empirical applications. Forni and Reichlin (1996, 1998) propose an estimator of the factors in a dynamic factor model based on contemporaneous sample averages under the assumption that the researcher has a-priori information about the space spanned by the factor loading matrix, and prove its consistency when T is fixed and $N \to \infty$. Although Forni and Reichlin (1998) argue that this a-priori information is available for their application to the estimation of sectoral factors using industry-level data, this a-priori knowledge (in their case, grouping industries into sectors and averaging within sectors) is in general lacking in the forecasting applications considered here. Recently, Lippi, Hallin, Forni and Reichlin (1998) considered an approximate dynamic factor model and proved the consistency, as $T,N \to \infty$, of factors estimated by dynamic principal components, but they do not provide joint rates for T and N and their estimator appears to require $N << T$. None of these papers provide explicit sequences linking N and T, so in particular it is unclear whether these results hold in the main cases of interest here, $N = O(T)$ and $N >> T$. Neither do they provide results about forecasting using estimated factors.

We therefore provide a result that establishes the first order asymptotic efficiency of feasible out of sample forecasts based on the estimated factors. To simplify the proofs, the results are presented for a balanced panel.

This result requires additional notation and conditions. Let M_{ij} denote the (i,j) element of a matrix M, adopt the matrix norm $\|M\| = \{tr(M'M)\}^{1/2}$, and let mineval(M) and maxeval(M) denote the minimum and maximum eigenvalue of M. Throughout, c and d denote generic finite positive constants. The first condition concerns the factor loading matrix.

Condition FL (factor loadings)

$|\lambda_{i,m}| \le \bar{\lambda} < \infty$, $i=1,\ldots,N$, $m=1,\ldots,r$; $\text{mineval}(\Lambda'\Lambda/N) \ge d > 0$; $tr(\Lambda'\Lambda/N) \le c < \infty$; and there is a positive definite $r \times r$ matrix D such that $\|\Lambda'\Lambda/N - D\| \to 0$.

The condition that $tr(\Lambda'\Lambda/N) \leq c$ ensures that the expected contribution of the factors to the variance of X_t is finite, while the minimum eigenvalue condition ensures a nontrivial contribution of each factor to the variance of X_t.

Under the next condition, the idiosyncratic errors e_t can have limited temporal and cross-sectional dependence:

Condition M1

$\{e_t\}$ satisfies,

(i) $Ee_{it} = 0$, $E(e_t'e_{t+u}/N) = \gamma(u)$, and $\sum_{u=-\infty}^{\infty} |\gamma(u)| < \infty$,

(ii) $Ee_{it}e_{jt} = \tau_{ij}$, where $\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} |\tau_{ij}| < \infty$,

(iii) $\sup_{i,t} Ee_{it}^4 < \infty$ and $\lim_{N\to\infty} \sup_{s,t} N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} |cov(e_{is}e_{it}, e_{js}e_{jt})| < \infty$.

The true factors F_t^0 are assumed to have covariance matrix Σ_F, and the next condition assures that $F^{0\prime}F^0/T \xrightarrow{p} \Sigma_F$.

Condition M2

(i) $EF_t^0 = 0$, $EF_t^0F_t^{0\prime} = \Sigma_F$, where Σ_F is positive definite.

(ii) $\sup_{\ell,m,t} \sum_{u=-\infty}^{\infty} |cov(F_{\ell t}^0 F_{mt}^0, F_{\ell t+u}^0 F_{mt+u}^0)| < \infty$.

Note that no restriction is made on the dependence between F_t and the errors e_t.

We now turn to the main result of this section. Let $\hat{\beta} = (\sum_{t=1}^{T-1} \hat{F}_t\hat{F}_t')^{-1}(\sum_{t=1}^{T-1} \hat{F}_t y_{t+1})$, so that $\hat{y}_{T+1|T} = \hat{\beta}'\hat{F}_T$ is the feasible forecast of y_{T+1} based on OLS regression using the estimated factors. Let $y_{T+1|T} = E(y_{T+1}|\mathcal{F}_t) = \beta' F_T^0$, the optimal (under quadratic loss) infeasible forecast of y_{T+1} using β and the r true factors F^0. The following theorem shows that the feasible forecast is asymptotically optimal as $(N,T)\to\infty$.

Theorem 1. Let (X_t, y_{t+1}) obey (2.3) and (2.4), and suppose that conditions FL, M1 and M2 hold. Let k=r and $N=O(T^\rho)$ for any $\rho>0$. Then, as $(N,T)\to\infty$, $\hat{y}_{T+1|T} - y_{T+1|T} \xrightarrow{p} 0$ and $E(y_{T+1} - \hat{y}_{T+1|T})^2 \to \sigma_\epsilon^2$.

All proofs are given in appendix B.

It is perhaps not surprising that a forecast efficiency result can be shown, given the results surveyed above that suggest large-N consistency of principal components in a static factor model. The contribution of theorem 1 is to show this holds for essentially any joint sequences $(N,T)\to\infty$. Thus theorem 1 obtains for sequences such that $N>>T$, $N=O(T)$, or $N<<T$.

To keep the proofs and exposition simple, theorem 1 was stated for one-step ahead forecasts when there are no unkown coefficients on lagged dependent variables. However, the empirical work in section 5 considers h-step ahead forecasts and uses lagged dependent variables. The proof of theorem 1 can however be modified to cover these cases.[2]

3. Time-Varying Factor Loadings and Further Results

In this section some of the conditions on the probability model (2.3) and (2.4) are relaxed, and it is no longer assumed that the true number of factors is known. Under these weaker conditions, the theoretical results of the previous section are extended to show the uniform (in t) consistency of the empirical factors and to provide sufficient conditions for information criteria to estimate consistently the number of factors needed to forecast y_{t+1}.

The assumptions of section 2 are relaxed in two ways. First, there is considerable evidence that many economic time series relationships exhibit structural instability. In this section, the factor loadings are therefore allowed to vary over time, specifically, they are modeled as evolving according to a random walk. Accordingly, (2.3) and (2.4) hold, except that Λ is replaced by Λ_t, where,

$$(3.1) \qquad \Lambda_t = \Lambda_{t-1} + h\zeta_t,$$

where h is a diagonal $N \times N$ scaling matrix and ζ_t is a $N \times r$ stochastic disturbances. If h is too large then the factor loadings bear little resemblence to the loadings in the previous period, and analysis of this model does not seem promising. Therefore h is modeled as a sequence of random matrices h_T that satisfy,

Condition TV (time varying factor loadings)

$h_T = \text{diag}(h_{1T},\ldots,h_{NT})$, where h_{iT} is i.i.d., h_T is independent of $(\epsilon_t, e_t, \zeta_t\}$, and $T\kappa_{jT} = O(1)$, $j=1,\ldots,4$, where $\kappa_{jT} \equiv (E|h_{iT}|^j)^{1/j}$.

This condition captures two sources of temporal instability: moderate parameter drift because of structural change for many series, and large jumps because of redefinitions or coding errors for a few. Consider the following example. Suppose a fraction π of the series experience moderate parameter drift of the form $h_{iT}=O_p(1/T)$ (so that $\lambda_{iT}-\lambda_{i0} = O_p(T^{-\frac{1}{2}})$, the same order as conventional sampling uncertainty if F_t were observed)[3], and the remaining series are subject to large parameter drift, for which $\lambda_{iT}-\lambda_{i0} = O_p(1)$ and thus $h_{iT}=O_p(T^{-\frac{1}{2}})$. Then condition TV is satisfied if $\pi=O(1/T^2)$.

Second, in practice the true number of factors is unknown and the number of estimated factors might be large. It therefore is of interest to examine the behavior of the empirical factors when the number of estimated factors, k, differs from the number of true factors, r.

To address the notion that a large number of factors might be estimated, the number of factors is taken to be a sequence, indexed by T, so that $k=k_T$, where it is assumed that $k_T \geq 1$ and $k_T = O(\ln T)$ (including of course the case that k_T is fixed).

These changes require additional moment conditions. Let ζ_{it}' denote the i'th row of ζ_t.

Condition M3

The random variables $\{e_t, \zeta_t, F_t\}$ satisfy:

(a) (i) $E\zeta_{it,m} = 0$, $E\zeta_{it}\zeta_{jt+u}' = \Gamma_{ij}(u)$, and $\Sigma_{u=-\infty}^{\infty} \sup_{i,j,l,m} |\Gamma_{ij,lm}(u)| < \infty$

(ii) $\lim_{N\to\infty} \sup_m N^{-1} \Sigma_{i=1}^{N} \Sigma_{j=1}^{N} \Sigma_{u=-\infty}^{\infty} |\Gamma_{ij,mm}(u)| < \infty$,

(iii) $\sup_{i,s,m} E\zeta_{is,m}^4 < \infty$ and

$\lim_{N\to\infty} \sup_{\ell,m} N^{-1} \Sigma_{i=1}^{N} \Sigma_{j=1}^{N} \sup_{t,u_1,u_2,u_3} |\text{cov}(\zeta_{it,\ell}\zeta_{it+u_1,m}, \zeta_{jt+u_2,\ell}\zeta_{jt+u_3,m})| < \infty$.

(b) (i) $E\zeta_{it}e_{jt+u} = \Psi_{ij}(u)$ and $\sup_i \Sigma_{u=-\infty}^{\infty} \sup_m |\Psi_{ii,m}(u)| < \infty$,

(ii) $\sup_m N^{-1} \Sigma_{i=1}^{N} \Sigma_{j=1}^{N} \sup_{t,u,v} |\text{cov}(e_{it}\zeta_{it+u,m}, e_{jt}\zeta_{jt+v,m})| < \infty$.

(c) (i) $EF_t^0 = 0$, $EF_t^0F_t^{0\prime} = \Sigma_F$, where Σ_F is positive definite.

(ii) $\sup_{\ell,m,t} \Sigma_{u=-\infty}^{\infty} |\text{cov}(F_{\ell t}^0 F_{mt}^0, F_{\ell t+u}^0 F_{mt+u}^0)| < \infty$.

(iii) $\max_{1\leq t\leq T} \|F_t^0\| / (\ln T)^2 \xrightarrow{p} 0$.

Conditions M3(a) and (b) limit the dependence across series and over time of these disturbances. The various disturbances are not assumed to be mutually independent, for example e_t and ζ_t can be dependent, even across series, subject to condition M3(b). Relative to condition M2, condition M3(c) contains the additional assumption that $\max_{1\leq t\leq T} \|F_t^0\| = o_p((\ln T)^2)$. This condition is satisfied by a variety of distributions of $\{F_{it}\}$. For example, if F_t is i.i.d., then if $(\max_{1\leq t\leq T} |F_{it}| - a_T)/b_T$ has a limiting distribution, it will be of the Fréchet, Weibull or Gumbel form; the sequences (a_T, b_T), if they exist, depend on the distribution of $\{F_{it}\}$ (Reiss [1989, p.152]). Three examples of distributions of $\{F_{it}\}$ for which M(c)(iii) is satisfied are the normal, for which $a_T = (2\ln T)^{1/2} + o(1)$ and $b_T = (2\ln T)^{-1/2}$, and the exponential and logistic, for which $a_T = \ln T$ and $b_T = 1$.

3.2. Results

We now turn to the consistency of the esimated factors. This is shown for a rescaled version of the principal components estimator:

$$(3.3) \qquad \hat{\tilde{F}} = \tilde{F}(\tilde{F}'\tilde{F}/T)^{1/2}.$$

The next result states that $\hat{\tilde{F}}_t$ is uniformly consistent for a linear combination of the true factors F_T^0, at the rate δ_{NT}.

Theorem 2. Let X_t and Λ_t obey (2.3) and (3.1), let Λ_0 satisfy condition FL, and suppose that conditions TV, M1, and M3 hold. Let $k_T = O(\ln T)$. If $(N,T)\rightarrow\infty$ and $N=O(T^\rho)$ for some $\rho>2$, then $\delta_{NT}\sup_t\|\hat{F}_t - HF_t^0\| \xrightarrow{p} 0$, where $\delta_{NT} = T^b$ for $b=\min[\frac{1}{2}\rho-1,1]$, where H is a nonrandom $k_T\times r$ matrix with row rank of $\min(k_T,r)$.

The rate condition indicates that this result applies for $N>>T$. Intuitively this arises because time varying factor loadings introduce additional noise in the time dimension, which is compensated for by additional cross sectional averaging.

To interpret theorem 2 it is useful to consider separately the three cases of $k<r$, $k=r$, and $k>r$. When $k<r$, F^0H' are the eigenvectors corresponding to the k largest eigenvalues of $F^0DF^{0\prime}$ and in this sense $\hat{F}_t$ estimates the most important factors. When $k=r$, R is a full rank square matrix so asymptotically $\hat{F}_t$ is a nonsingular transformation of F_t^0. When $k>r$, the row rank of R and thus H is only r, so $\hat{F}_t$ contains k-r redundant estimates of the factors. Intuitively, this arises because the final k-r columns of $\hat{\Lambda}$ are $O_p(1)$ random variables that estimate factor loadings that are zero in population, and cross sectional averages of X_{it}, weighted by these random variables, obey a weak law of large numbers. Equivalently, by the algebra of principal components, the final k-r columns of $\hat{F}_t$ have a variance equal to the corresponding ordered eigenvalues of XX'/NT, which are asymptotically zero.

The final theorem concerns determination of the number of factors useful for forecasting y_{t+1}. Consider the information criterion,

$$\text{(3.4)} \qquad IC_k = \ln(\hat{\sigma}_\epsilon^2(k)) + g(T)k$$

where $\hat{\sigma}_\epsilon^2(k) = SSR(k)/T$, where SSR(k) is the sum of squared residuals from estimation of (2.4) using the k empirical factors $\hat{F}$, and g(T) is a penalty function. Following convention, suppose that $r,k \leq k_{max}$, where k_{max} is finite and known. The information criterion estimate of r, $\hat{r}$, solves $\min_{1 \leq k \leq k_{max}} IC_k$. Let δ_{NT} be as defined in theorem 2.

Theorem 3. Suppose that the conditions of theorem 2 hold except $1 \leq k \leq k_{max} < \infty$.

(a) If $k \geq r$ then $\hat{\sigma}_\epsilon^2(k) \xrightarrow{p} \sigma_\epsilon^2$.

(b) Let $\hat{r}$ be the estimate of r produced by an information criterion with $g(T) \to 0$ and $\delta_{NT} g(T) \to \infty$, where k_{max} is known and $k_{max} \geq r$. Then $\Pr(\hat{r}=r) \to 1$ and $\hat{\sigma}_\epsilon^2(\hat{r}) \xrightarrow{p} \sigma_\epsilon^2$.

Theorem 3(a) states that the efficient forecast of theorem 1 can be achieved even if "too many" factors are estimated and even if there are time varying factor loadings. Still, one worries about estimating more coefficients than needed, so it might be desirable to use an information criterion to reduce the number of factors. Theorem 3(b) therefore provides sufficient conditions under which doing so consistently estimates of the number of factors and produces an efficient forecast.

The conditions on $g(T)$ in theorem 3(b) differ from the standard conditions that justify information criteria, which require $Tg(T)\to\infty$ and $g(T)\to 0$ for consistent order estimation; BIC satisfies $Tg(T)\to\infty$ but not $\delta_{NT}g(T)\to\infty$. A penalty function which does satisfy this condition is,

$$(3.5) \qquad g(T) = \omega \ln T/\delta_{NT}$$

where δ_{NT} is given in theorem 2, that is, $\delta_{NT} = \min(N^{1/2}/T^{1+\epsilon}, T^{1-\epsilon})$, where ϵ is a small positive constant and ω is a positive constant. If for example $N=T^3$, then $\delta_{NT} = T^{1/2-\epsilon}$, so that $g(T)=\omega \ln T/T^{1/2-\epsilon}$, an asymptotically larger penalty than BIC. Thus BIC (and AIC) will produce larger estimates of r than (3.5). It should be emphasized that theorem 3(a) nevertheless implies that AIC, BIC, and (3.5), as well as fixed $k\geq r$, all produce asymptotically first order efficient forecasts. The relative performance of different information criteria, and suitable choices of ω in practice, are investigated in the Monte Carlo study of the next section.

4. Monte Carlo Analysis

The Monte Carlo experiment reported in this section has two objectives. The first is to examine numerically the predictions of theorems 1 and 2. The second is to quantify the finite sample implications of estimating the number of factors by an information criterion, as studied asymptotically in theorem 3.

The experimental design is a parametric dynamic factor model that allows for time varying factor loadings, an autoregressive factor, and idiosyncratic terms that are serially correlated and correlated across series. All the results here are for a balanced panel. The design is, (4.1) $$X_{it} = \Sigma_{j=0}^{q}\lambda'_{ijt}f_{t-j} + e_{it}$$

(4.2) $$f_t = \alpha f_{t-1} + u_t$$

(4.3) $$(1\text{-}aL)e_{it} = (1+b^2)v_{it}+bv_{i+1,t}+bv_{i-1,t}$$

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

where $i=1,\ldots,N$ and $t=1,\ldots,T$, f_t and λ_{ijt} are $\bar{r}\times 1$, $\{e_{it}, v_{it}, \zeta_{ijt}\}$ are i.i.d. N(0,1), u_t is i.i.d. $N(0,I_r)$, and $\{u_t\}$ is independent of $\{e_{it}, v_{it}, \zeta_{ijt}\}$. In static form (2.3) the true number of factors is $r=(q+1)\bar{r}$. The time variation here is a special case of the heterogeneous time variation allowed in section 3.

The initial factor loading matrix Λ_0 is chosen as follows. Let $R_i^2 = \text{var}(\Sigma_{j=0}^{q}\lambda'_{ij0}f_{t-j})/[\text{var}(\Sigma_{j=0}^{q}\lambda'_{ij0}f_{t-j}) + \text{var}(e_{it})]$. Then $\lambda_{ij0} = \lambda_i^*\tilde{\lambda}_{ij0}$, where $\tilde{\lambda}_{ij0}$ is i.i.d. N(0,1) and independent of $\{e_{it}, v_{it}, \zeta_{ijt}, v_t\}$, and λ_i^* is randomly chosen so that R_i^2 has a uniform distribution on [0.1,0.8].[4] The initial values of the factor are drawn from their stationary distribution. Finally the $\{X_{it}\}$ are transformed to have sample mean zero and sample variance one (this transformation is used in the empirical work presented in the next section).

The scalar variable to be forecast is generated as,

(4.5) $$y_{t+1} = \Sigma_{j=0}^{q}\iota' f_{t-j} + \epsilon_{t+1}$$

where ι is a $\bar{r}\times 1$ vector of 1's and ϵ_{t+1} is i.i.d. N(0,1).

The factors were estimated by principal components as discussed in section 2.2 for the balanced panel using X (not stacked). The coefficients β in the forecasting regression (2.4) were estimated by the OLS coefficients $\hat{\beta}$ in the regression of y_{t+1} on $\hat{F}_t$, $t=1,\ldots,T\text{-}1$. The out-of-sample forecast is $\hat{y}_{T+1|T} = \hat{\beta}'\hat{F}_T$. For comparison purposes, the infeasible out-of-sample forecast $\hat{y}^0_{T+1|T} = \hat{\beta}^{0\prime}F^0_t$ was also computed, where $\hat{\beta}^0$ is the OLS estimator from regressing y_{t+1} on F^0_t, $t=1,\ldots,T\text{-}1$.

The free parameters varied in the Monte Carlo experiment are N, T, r, k, α, a, b, and c. The results are summarized by two statistics. The first is a trace R^2 of the multivariate regression of $\hat{F}$ on F_0:

$$(4.6) \qquad R^2_{\hat{F},F^0} = \hat{E}\|P_{F^0}\hat{F}\|^2/\hat{E}\|\hat{F}\|^2 = \hat{E}tr(\hat{F}'P_{F^0}\hat{F})/\hat{E}tr(\hat{F}'\hat{F}),$$

where $\hat{E}$ denotes the expectation estimated by averaging the relevant statistic over the Monte Carlo repetitions and $P_{F^0}=F^0(F^{0\prime}F^0)^{-1}F^{0\prime}$. According to theorem 2, if $k \geq r$ then $R^2_{\hat{F},F^0} \xrightarrow{p} 1$. Values of this statistic considerably less than one indicates a case in which theorem 2 provides a poor approximation to the finite sample performance of $\hat{F}$.

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

$$(4.7) \qquad S^2_{\hat{y},\hat{y}^0} = 1 - \hat{E}(\hat{y}_{T+1|T} - \hat{y}^0_{T+1|T})^2/\hat{E}(\hat{y}^0_{T+1|T})^2$$

Because $E(\hat{y}^0_{T+1|T} - y_{T+1|T})^2 \to 0$, according to theorem 1, $S^2_{\hat{y},\hat{y}^0} \xrightarrow{p} 1$ for k=r and by theorem 3 if $k \geq (q+1)r$ and either k is fixed or it is chosen using an information criterion that satisfies the conditions in the theorem. The results are reported for several information criteria: the AIC, the BIC, and the information criterion with the penalty function (3.5) for various choices of the scaling parameter ω.

The results are summarized in table 1. Panel A presents results for the static factor model with i.i.d. errors and factors. In panel B, this model is extended to idiosyncratic errors that are serially correlated across series. Panel C considers the dynamic factor model with serially correlated factors and lags of the factors entering X_t, and time varying factor loadings are introduced in panel D.

First consider the results for $R^2_{\hat{F},F^0}$. In all cases, $R^2_{\hat{F},F^0}$ exceeds .8, even for T=25 and N=50. As T and N increase, this R^2 increases, for example, for T=100, N=250, r=k=5, $R^2_{\hat{F},F^0}$=.97. Consistent with theorem 2, estimating k>r typically introduces little spurious noise, for example, when T=100, N=250, and r=5, increasing k from 5 to 10 decreases $R^2_{\hat{F},F^0}$ by .02. If the idiosyncratic errors are moderately serially correlated (a=.5), $R^2_{\hat{F},F^0}$ drops only slightly, although it drops further when a=.9 (although this drop is largely eliminated when T is increased). The $R^2_{\hat{F},F^0}$ is also high when the true model is dynamic but the factors are extracted from a static procedure with $k \geq \bar{r}(q+1)$, although some deterioration is noticeable when the factors are highly serially correlated. The greatest deterioration of the estimates of the factors occurs when time variation in the factor weights is introduced. With large time variation (c=10), $R^2_{\hat{F},F^0}$ is between .83 and .87 for the various cases considered. In general, the results improve when T increases, with N, r, and k fixed, and when N increases, with T, r, and k fixed; for fixed T and N, results deteriorate as r increase and k=r, although they deteriorate only slightly as k increases for fixed r.

The forecasting results are consistent with the prediction of theorem 1 that, holding constant N/T and the other design parameters, $S^2_{\hat{y},\hat{y}^0}$ tends to 1 as T and N increase. When T=100 and N=250, $S^2_{\hat{y},\hat{y}^0}$ is generally large across the different design parameters, typically exceeding .95 in the static models. The quality of the forecasts drops in the dynamic models and when there is time variation in the factor loadings. The results for forecasts based on model selection criteria are generally consistent with theorem 3. Generally speaking, for T and N large, forecasts based on the BIC, AIC, or (3.5) with ω=.001 perform similarly, and only slightly worse than those with k=r. However, the forecasts based on (3.5) with larger values of ω such as ω=.01 perform poorly, and even larger values ω perform worse (these results are not shown to save space). These results suggest that the criterion (3.5) is overly conservative.

5. Application to Forecasting U.S. Industrial Production and Inflation

This section reports the results of a simulated real-time forecasting experiment in which forecasts based on the factor model approach are compared to forecasts from a variety of benchmark models. Forecasts were computed for eight major monthly macroeconomic variables for the United States. Four of these are the measures of real economic activity used to construct the Index of Coincident Economic Indicators maintained by The Conference Board (formerly by the U.S. Department of Commerce): total industrial production (ip); real personal income less transfers (gmyxpq); real manufacturing and trade sales (msmtq); and the number of employees on nonagricultural payrolls (lpnag) (additional details are given in appendix C, which lists series by the mnemonics given here in parenthesis). The remaining four series are price indexes: the consumer price index (punew); the personal consumption expenditure implicit price deflator (gmdc); the CPI less food and energy (puxx); and the producer price index for finished goods (pwfsa). One-month ahead forecasts of these variables being of limited practical interest, this experiment compares forecasts at the 6, 12, and 24 month horizons. The complete data set spans 1959:1 - 1998:12.

5.1 Forecasting models

For each series, several forecasting models are compared: "diffusion index" forecasts based on estimated factors; a benchmark univariate autoregression; and benchmark multivariate models. For both the real and price series, one of the benchmark multivariate models is a trivariate vector autoregression, and a second is based on leading economic indicators. As a further comparison, inflation forecasts are also computed using an unemployment-based Phillips curve.

The formal development in sections 2 and 3 considered 1-step ahead prediction. Here the focus is on multistep ahead prediction, and most of the forecasting regressions are projections of a h-step ahead variable y^h_{t+h} onto t-dated predictors, sometimes including lagged transformed values z_t of the variable of interest. Specifically, the real variables are modeled as being I(1) in logarithms. Because all four real variables are treated identically, consider industrial production, for which

$$(5.1) \qquad y^h_{t+h} = (1200/h)\ln(IP_{t+h}/IP_t) \text{ and } z_t = 1200\ln(IP_t/IP_{t-1}).$$

The price indexes are modeled as being I(2) in logarithms.[5] Accordingly, for the CPI (and similarly for the other price series),

$$(5.2) \qquad y^h_{t+h} = (1200/h)\ln(CPI_{t+h}/CPI_t)\text{-}1200\ln(CPI_t/CPI_{t-1}) \text{ and } z_t = 1200\Delta\ln(CPI_t/CPI_{t-1}).$$

Diffusion Index forecasts. The most general diffusion index/factor model forecasting equation that we consider is,

$$(5.3) \qquad y^h_{t+h} = \beta_0 + \Sigma^m_{j=1}\beta_j'\hat{F}_{t-j+1} + \Sigma^p_{j=1}\gamma_j z_{t-j+1} + \epsilon^h_{t+h}$$

where $\{\hat{F}_{it}\}$ are the estimated factors. This modifies (2.4) in four ways. First, the dependent variable is the h-step growth. Second, lags of z_t have been added explicitly, while (as discussed in section 2) in (2.4) they were left implicit and y_{t+1} was interpreted as a quasidifference. Third, an intercept has been explicitly added. Fourth, m-1 lags of the factors have been introduced as predictors. The first three of these modifications are consistent with the foregoing theoretical development (see for example the final paragraph in section 2). The fourth modification, while not indicated by the population model, is made because, if k is small, the lagged dynamic factors that F_t contains in population might not be in the estimate of

F_t, so adding lags of F_t might improve forecasting performance. Given the lag orders and the estimated factors, the coefficients of (5.3) were estimated by OLS.

Results for three variants of (5.3) are reported. The first, denoted in the tables by "DI-AR,Lag", includes lags of the factors and lags of z_t, with k and lag orders m and p estimated by BIC, with $1 \leq k \leq 4$, $1 \leq m \leq 3$, and $0 \leq p \leq 6$. Thus the smallest candidate model that BIC can choose here includes only contemporaneous $\hat{F}_t$ and excludes z_t. The second, denoted "DI-AR", includes contemporaneous $\hat{F}_t$, that is, $m=0$, and k and p are chosen by BIC with $1 \leq k \leq 12$ and $0 \leq p \leq 6$. The third, denoted "DI", includes only contemporaneous $\hat{F}_t$, so $p=0$, $m=0$, and k is chosen by BIC, $1 \leq k \leq 12$.

The data used to construct the factors are 215 monthly monthly time series for the U.S. from 1959:1-1998:12. The series were selected judgmentally to represent 14 main categories of macroeconomic time series: real output and income; employment and hours; real retail, manufacturing and trade sales; consumption; housing starts and sales; real inventories and inventory-sales ratios; orders and unfilled orders; stock prices; exchange rates; interest rates; money and credit quantity aggregates; price indexes; average hourly earnings; and miscellaneous. The list of series is given in Appendix C, and is similar to lists that we have used elsewhere (Stock and Watson [1996, 1998b]). These series were taken from a somewhat longer list, from which we eliminated series with gross problems such as redefinitions. However no further pruning of this list was performed. The series were taken from the May 1999 release of the DRI/McGraw Hill Basic Economics database (formerly Citibase). In general these series represent the fully revised historical series available as of May 1999.

The theory of sections 2 and 3 assumes that X_t is I(0), so these series were subjected to three preliminary steps: possible transformation by taking logarithms, possible first differencing, and screening for outliers. The decision to take logarithms or to first difference the series was made judgmentally. In general, logarithms were taken for all nonnegative series that were not already in rates or percentage units and most series were first differenced. A code summarizing these transformations is given for each series in Appendix C. After these transformations, all series were further standardized to have sample mean zero and unit sample variance. Finally, the transformed data were screened automatically for outliers (generally taken to be coding errors or exceptional events such as sectoral strikes), and observations exceeding ten times the interquartile range from the median were replaced by missing values.

Using this dataset, three sets of empirical factors were considered. The first was computed as discussed in section 2.2 from the subset of 149 variables that are available for the full sample period, thus constituting a balanced panel. The second set of factors was computed using all 215 series; because these constitute an unbalanced panel, the empirical factors were computed as discussed in Appendix A. The third set of factors were computed by stacking the 149 variables in the balanced panel with their first lags, so the augmented data vector has dimension 298. Empirical factors were then estimated by the principal components of the "stacked data" as discussed in section 2.2.

Autoregressive forecast. The autoregressive forecast is a univariate forecast based on (5.3) where the terms involving $\hat{F}$ are excluded. The lag order p is selected recursively by BIC with $0 \leq p \leq 6$, where $p=0$ indicates that z_t and its lags are excluded.

Vector autoregressive forecast. The first multivariate benchmark model is a VAR with four lags each of three variables. The variables in the VAR are a measure of the monthly growth in real activity, the change in monthly inflation, and the change in the 90 day U.S. Treasury bill rate. When used to forecast the real series, the relevant real activity variable was used and the inflation measure was CPI inflation. For forecasting inflation, the relevant price series was used and the real activity measure was industrial production. Multistep forecasts were computed by iterating the VAR forward. This contrasts to the AR forecasts, which were computed by h-step ahead projection rather than iteration.

Multivariate leading indicator forecasts. The leading indicator models are of the form,

$$(5.4) \qquad y^h_{t+h} = \delta_0 + \Sigma^m_{j=1}\delta_i' W_{t-j+1} + \Sigma^p_{j=1}\gamma_j z_{t-j+1} + u^h_{t+h}$$

where W_t is a vector of leading indicators that have been featured in the literature and/or in real-time forecasting applications.

For the real variables, W_t consists of eleven leading indicators that we have used for real-time monthly forecasting in experimental leading and recession indicators; see Stock and Watson (1989)[6]. Five of these leading indicators are also used in the factor estimation step in the diffusion index forecasts. These are: average weekly hours of production workers in manufacturing (lphrm); the capacity utilization rate in manufacturing (ipxmca); housing starts (building permits) (hsbr); the index of help-wanted advertising in newspapers (lhel); and the interest rate on 10-year U.S. Treasury bonds (fygt10). The remaining six leading indicators are: the interest rate spread between 3-month U.S. Treasury bills and 3-month commercial paper; the spread between 10-year and 1-year U.S. Treasury bonds; the number of people working part-time in nonagricultural industries because of slack work; real manufacturers' unfilled orders in durable goods industries; a trade-weighted index of nominal exchange rates between the U.S. and the U.K., West Germany, France, Italy, and Japan; and the National Assiation of Purchasing Managers' index of vendor performance (the percent of companies reporting slower deliveries).

For the inflation forecasts, eight leading indicators are used. These variables were chosen because of their good individual performance in previous inflation forecasting exercises. In particular these variables performed well in at least one of the historical episodes considered in Staiger, Stock and Watson (1997) (also see Stock and Watson [1998b]). Seven of these variables are also used in the factor estimation step in the diffusion index forecasts: the total unemployment rate (lhur); real manufacturing and trade sales (msmtq); housing starts (hsbr); new orders in durable goods industries (mdoq); the nominal M1 money supply (fm1); the federal funds overnight interest rate (fyff); and the interest rate spread between 1-year U.S. Treasury bonds and the federal funds rate (sfygt1). The remaining variable is the trade-weighted exchange rate listed in the previous paragraph.

In all cases, the leading indicators were transformed so that W_t is I(0). This entailed taking logarithms of variables not already in rates, and differencing all variables except the interest rate spreads, housing starts, the index of vendor performance, and the help wanted index.

For each variable to be forecast, p and m in (5.4) were determined by recursive BIC with $1 \leq m \leq 4$ and $0 \leq p \leq 6$, so 28 possible models were compared in each time period.

Phillips curve forecasts. The unemployment-based Phillips curve is a key tool in applied macroeconomic forecasting and is considered by many to have been a reliable method for forecasting inflation over this period, cf. Gordon (1982) and, more recently, the Congressional Budget Office (1996), Fuhrer (1995), Gordon (1997), Staiger, Stock and Watson (1997), and Tootel (1994). The Phillips curve inflation forecasts considered here have the form (5.4), where W_t consists of: the unemployment rate (LHUR) and m-1 of its lags; the relative price of food and energy (current and one lagged value only); and Gordon's (1982) variable that controls for the imposition and removal of the Nixon wage and price controls.[7] The lag lengths m and p was chosen by recursive BIC, where $1 \leq m \leq 6$ and $0 \leq p \leq 6$.

5.2. Simulated real-time experimental design

Estimation and forecasting was conducted to simulate real-time forecasting. This entailed fully recursive parameter estimation, factor estimation, model selection, etc. The first simulated out of sample forecast was made in 1970:1. To construct this forecast, the data were standardized, the parameters and factors were estimated, and the models were selected, using data available from 1959:1 through 1970:1 (the first date for the regressions was 1960:1, with earlier observations used for initial conditions as needed). Thus regressions (5.3) and (5.4) were run for t=1960:1,...,1970:1-h, then the values of the regressors at t=1970:1 were used to forecast $y^h_{1970:1+h}$. All parameters, factors, etc. were then reestimated, information criteria were recomputed, and models were selected using data from 1959:1 through 1970:2, and forecasts from these models were then computed for $y^h_{1970:2+h}$. The final simulated out of sample forecast was made in 1998:12-h for $y^h_{1998:12}$.

5.3 Forecasting results

The results for the real variables are reported in detail in table 2 for 12-month ahead forecasts, and summaries for 6- and 24-month ahead forecasts are reported in table 3. Two sets of statistics are reported. The first is the mean squared error (MSE) of the candidate forecasting model, computed relative to the MSE of the univariate autoregressive forecast (so the AR forecast has a relative MSE of 1.00). For example, the simulated out of sample MSE of the leading indicator (LI) forecast of industrial production is 86% that of the AR forecast at the 12 month horizon. Autocorrelation consistent standard errors for these relative MSEs, calculated following West (1996), are reported in parentheses. The second set of statistics are the coefficient on the candidate forecast from the forecast combining regression,

$$(5.5) \qquad y_{t+h} = \alpha\hat{y}^h_{t+h|t} + (1-\alpha)\hat{y}^{h,AR}_{t+h|t} + u_{t+h}$$

where $\hat{y}^h_{t+h|t}$ is the candidate h-step ahead forecast and $\hat{y}^{h,AR}_{t+h|t}$ is the benchmark h-step ahead AR forecast. HAC standard errors for α are reported in parentheses. For example, α is estimated to be .57 when the candidate forecast is the leading indicator forecast at the 12 month horizon, with a standard error of .13, so the hypothesis that the weight on the leading indicator forecast is zero ($\alpha=0$) is rejected at the 5% level, but so is the hypothesis that the leading indicator forecast receives unit weight.

We now turn to the results for the real variables. First consider the diffusion index forecasts with factors estimated using the full data set (the unbalanced panel). These forecasts with BIC factor selection generally improve substantially over the benchmark univariate and multivariate forecasts. The DI-AR,Lag model, which allows recursive BIC selection across own lags and lags of the factors, outperforms all three benchmark models in 10 of the 12 variable/horizon combinations, the exceptions being 6- and 12-month ahead forecasts of employment. In most cases the performance of the simpler DI forecasts, which exclude lags of $\hat{F}_t$ and z_t, is comparable to or even better than that of the DI-AR,Lag forecasts. This is rather surprising, because it implies that essentially all the predictable dynamics of these series are accounted for by the estimated factors. In some cases, the improvement over the benchmark forecasts are quite substantial, for example, for industrial production at the 12 month horizon the DI-AR,Lag forecast has a forecast error variance 57% that of the AR model and two-thirds that of the leading indicator model. The relative improvements are more modest at the 6 month horizon. At the 24 month horizon, the multivariate benchmark forecasts break down and perform worse than the univariate forecast, however the DI-AR,Lag, DI-AR, and DI forecasts continue to outperform the AR benchmark very substantially.

The performance of comparable models is usually better when the empirical factors from the full data set are used, relative to those from the balanced panel subset. Performance is not improved by using empirical factors from augmenting the balanced panel with its first lag; for these real series, doing so does comparably, or somewhat worse, than using the empirical factors from the unstacked balanced panel.

Inspection of the final panels of tables 2 and 3 reveals a striking finding: simply using DI or DI-AR forecasts with two factors captures most of the forecasting improvement. In most cases, incorporating BIC factor and lag order selection provides little or no improvement over just using two factors, with no lags of the factors and no lagged dependent variables.

The results for the price series are given in tables 4 and 5. There are three notable differences in these results, relative to those for the real variables. First, the DI-AR,Lag forecasts outperform all the benchmark forecasts less often, in only 6 of the 12 variable/horizon combinations. Second, including lagged inflation dramatically improves the forecasts, and without this the DI forecasts are actually worse than the autoregressive forecasts. Third, other factor forecasts generally outperform the DI-AR,Lag forecasts. Notably, the full data set DI-AR forecast with k=1 (and no lagged factors) outperforms all the benchmarks in 11 of 12 cases, and typically improves upon the DI-AR lag. Thus most of the forecasting gains seem to come from using a single factor.

As with the real variables, forecasts based on the stacked data perform less well than those based on the unstacked data. While the full data set forecasts are typically better than the balanced panel subset forecasts for the 6 and 12 month horizons, at the 24 month horizon the balanced panel forecasts (not shown in the tables) outperform the full data set forecasts.

Additional analysis of factor-based forecasts of CPI and consumption deflator inflation, and additional comparisons of these forecasts to other Phillips-curve forecasts and to forecasts based on other leading indicators, are contained in Stock and Watson (1998b). Three findings from that study are worth noting here. First, the DI-AR and DI-AR,Lag forecasts are found to perform well relative to a large number of additional multivariate benchmarks. Second, the forecasts reported here can be further improved upon using a single-factor forecast, where the factor is computed from a set of variables that all measure real economic activity. Forecasts based on this real economic activity factor have MSEs approximately 10% less than the best forecasts reported in table 4. Finally, similar rankings of methods are obtained using I(1) forecasting models, rather than the I(2) models used here, that is, when first rather than second differences of log prices are used for the forecasting equation and factor estimation.

In interpreting these results, it should be stressed that the multivariate leading indicator models are sophisticated forecasting tools that provide a stiff benchmark against which to judge the diffusion index forecasts. In our judgment, the performance of the leading indicator models reported here overstates their true potential out of sample performance, because the lists of leading indicators used to construct the forecasts were chosen by model selection methods based on their forecasting performance over the past two decades, as discussed in section 5.2. In this light, we consider the performance of the various diffusion index models to be encouraging.

5.4. Empirical factors

Because the factors are identified only up to a $k \times k$ matrix, detailed discussion of the individual factors is unwarranted. Nevertheless the finding that good forecasts can be made with only one or two factors suggests briefly characterizing the first few factors.

Figure 1 therefore displays the R^2s of the regressions of the 215 individual time series against each of the first six empirical factors from the balanced panel subset, estimated over the full sample period. These R^2s are plotted as bar charts with one chart for each factor. (The series are grouped by category and ordered numerically using the ordering in the appendix.) Broadly speaking, the first factor loads primarily on output and employment; the second factor on interest rate spreads, unemployment rates, and capacity utilization rates; the third, on interest rates; the fourth, on stock returns; the fifth, on inflation; and the sixth, on housing starts. Taken together, these six factors account for 39% of the variance of the 215 monthly time series in the full data set, as measured by the trace-R^2; the first twelve factors together account for 53% of the variance of these series.[8]

6. Discussion and Conclusions

We find several features of the empirical results surprising and intriguing. Few theoretical macroeconomic models suggest a linear factor structure for the overall macroeconomy, yet six factors account for much of the variance of our 215 time series. Even if a factor structure describes the joint behavior of these series, this does not imply that a forecast based on dynamic factors estimated by principal components should outperform forecasts based on leading indicators or other specialized models that have been fine tuned through years of experience. Yet, forecasts based on just the first few factors perform well for *both* real activity measures and inflation series, series that measure quite different economic concepts and have quite different univariate time series properties. Evidently, these results raise numerous issues for future empirical and theoretical research.

Several methodological issues remain. One is to explore estimation methods that might be more efficient in the presence of heteroskedastic 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 paper 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.

Another important extension is to real time forecasting with mixed frequency data (daily, weekly, monthly and quarterly). The EM algorithm presented in appendix A addresses this case, but it has not yet been implemented empirically. Other issues that arise in real time include data revisions and the nonsynchronous timing of data releases. Work on these and related issues is ongoing.

Footnotes

1. The differing dates on y and X in (2.1) and (2.2) reflect the use of X_t to forecast y_{t+1}. This representation can be taken as primitive or alternatively can be derived from a dynamic factor model for (X_t, y_t). Specifically, suppose that $y_t = \bar{\beta}^0(L)f_t + \epsilon_t^0$, where ϵ_t^0 is a martingale difference with respect to F_{t-1}, and further suppose that $E(f_t | F_{t-1}) = \bar{\beta}^1(L)f_{t-1}$ (a standard assumption in parametric dynamic factor models). Then $y_t = \bar{\beta}_0^0 \bar{\beta}^1(L)f_{t-1} + L^{-1}[\bar{\beta}^0(L) - \bar{\beta}_0^0]f_{t-1} + [\bar{\beta}_0^0 f_t - E(f_t | F_{t-1})] + \epsilon_t^0$. This yields (2.2) by setting $\bar{\beta}(L) = \bar{\beta}_0^0 \bar{\beta}^1(L) + L^{-1}[\bar{\beta}^0(L) - \bar{\beta}_0^0]$ and $\epsilon_t = [\bar{\beta}_0^0 f_t - E(f_t | F_{t-1})] + \epsilon_t^0$, by noting that $E(\epsilon_t | F_{t-1}) = 0$, and by shifting the time subscript.

2. The extension to lagged dependent variables is handled by repeated application of appendix lemma B1(c). The extension to h-step ahead forecasts, and thus overlapping data, is handled by modifying the argument used to show $g_{2T} \xrightarrow{p} 0$, where g_{2T} is defined in the proof of theorem 1.

3. The 1/T nesting is also the local neighborbood within which break tests such as the Quandt likelihood ratio test would have nondegenerate asymptotic power were F_t observed. Stock and Watson (1996, 1998a) argue that the 1/T nesting is empirically plausible for many macroeconomic time series.

4. That is, $\lambda_i^* = [R_i^2(R_i^2-1)^{-1}(\mathrm{var}(\Sigma_{j=0}^q \tilde{\lambda}_{ij0}' F_{t-j}))^{-1}]^{1/2}$ (this uses $\mathrm{var}(e_{it}) = 1$), where R_i^2 is i.i.d. U(0.1,0.8).

5. The I(2) specification is consistent with standard non-accelerationist Phillips curve equations and is a good description of the series over much of the sample period. However, I(1) specifications also provide adequate descriptions of the data, particularly in the early and late parts of the sample. Stock and Watson (1998b) find little difference in I(1) and I(2) factor model forecasts for these prices over the sample period studied here, so for the sake of brevity we limit our analysis here to the I(2) specification.

6. The list used here consists of the leading indicators used to produce the XRI and the XRI-2, which are released monthly (and documented) at the web site http://www.nber.org.

7. The wage and price control dummy is introduced for forecasts made in 1971:7+h, before which it produces singular regressions.

8. The contributions to the trace-R^2 by the first six factors are, respectively: 0.137, 0.085, 0.048, 0.040, 0.034, and 0.041, for a total of 0.385.

Appendix A: EM Estimation with an Unbalanced Panel and Data Irregularities

In practice, when N is large one encounters various data irregularities, including occasionally missing observations, unbalanced panels, and mixed frequency (e.g., monthly and quarterly) data. In this case, a modification of the least squares objective function (2.5) is appropriate. For example, consider the unbalanced panel and let $I_{it}=1$ if X_{it} is observed and $=0$ otherwise; then (2.5) becomes,

$$V^{\dagger}(F,\Lambda) = \Sigma_{i=1}^{N}\Sigma_{t=1}^{T} I_{it}(X_{it}-\lambda_i'F_t)^2. \quad \text{(A.1)}$$

Eigenvalue calculations cannot be used directly to minimize (A.1), and iterative methods must be used instead. This appendix summarizes a method method based on the EM algorithm that has proven to be easy and effective.

To motivate this EM algorithm notice that $V(F,\Lambda)$ in (2.5) is proportional to the log-likelihood under the assumption that X_{it} are iid $N(\lambda_i'F_t,1)$, in which case the least squares estimators are the Gaussian MLEs; this is also true for the least squares estimators that minimize (A.1). Because $V^{\dagger}$ is just a "missing data" version of V and because minimization of V is computationally simple, a simple EM algorithm can be constructed to minimize $V^{\dagger}$.

The j-th iteration of the algorithm is defined as follows. Let $\hat{\Lambda}$ and $\hat{F}$ denote estimates of Λ and F constructed from the (j-1)'st iteration, and let

$$Q(X^{\dagger},\hat{F},\hat{\Lambda},F,\Lambda) = E_{\hat{F},\hat{\Lambda}}[V(F,\Lambda)\,|\,X^{\dagger}] \quad \text{(A.2)}$$

where $X^{\dagger}$ denotes the full set of observed data and $E_{\hat{F},\hat{\Lambda}}[V(F,\Lambda)\,|\,X^{\dagger}]$ is the expected value of the "complete data" log-likelihood $V(F,\Lambda)$, evaluated using the conditional density of $X\,|\,X^{\dagger}$ evaluated at $\hat{F}$ and $\hat{\Lambda}$. The estimates of F and Λ at iteration j solve $\text{Min}_{F,\Lambda}Q(X^{\dagger},\hat{F},\hat{\Lambda},F,\Lambda)$.

To carry out the calculations, note that

$$\text{(A.3)}\qquad Q(X^{\dagger},\hat{F},\hat{\Lambda},F,\Lambda)=\Sigma_i\Sigma_t\{E_{\hat{F},\hat{\Lambda}}(X_{it}^2|X^{\dagger})+(\lambda_i'F_t)^2-2\hat{X}_{it}(\lambda_i'F_t)\}$$

where $\hat{X}_{it}\doteq E_{\hat{F},\hat{\Lambda}}(X_{it}|X^{\dagger})$. The first term on the right hand side of (A.3) does not depend on F or Λ, and so for purposes of minimization it can be replaced by $\Sigma_i\Sigma_t\hat{X}_{it}{}^2$. This implies that the values of F and Λ that minimize (A.3) can be calculated as the minimizers of $\hat{V}(F,\Lambda)$ = $\Sigma_i\Sigma_t(\hat{X}_{it}-\lambda_i'F_t)^2$. At the jth step this reduces to the principal component eigenvalue calculations discussed in section 2.2, where the missing data are replaced by their expectation conditional on the observed data and using the parameter values from the previous iteration. If the full data set contains a subset that constitutes a balanced panel, then starting values for $\hat{F}$ in the EM iteration can be obtained using $\tilde{F}$ from the balanced panel subset.

We now provide some additional details on the calculation of $\hat{X}_{it}$ for some important special cases. Let $\underline{X}_i$ = $(X_{i1},\ldots,X_{iT})'$, and let $\underline{X}_i^{\dagger}$ be the vector of observations on the i-th variable. Suppose that $\underline{X}_i^{\dagger}=A_i\underline{X}_i$; this can be done in the cases of missing values and temporal aggregation, for example. Then $E(\underline{X}_i|X^{\dagger})$ = $E(\underline{X}_i|\underline{X}_i^{\dagger})$ = $F\lambda_i+A_i'(A_iA_i')^-(\underline{X}_i^{\dagger}-A_iF\lambda_i)$, where $(A_iA_i')^-$ is the generalized inverse of A_iA_i'. The particulars of these calculations are now presented for some important special cases. In the first four special cases discussed below, this level of generality is unecessary and the formula for $\hat{X}_{it}$ follows quite simply from the nature of the data irregularity.

A. Missing observations. Suppose some observations on X_{it} are missing. Then, during iteration j, the elements of the estimated balanced panel are constructed as $\hat{X}_{it}=X_{it}$ if X_{it} observed, and $\hat{X}_{it}=\hat{\lambda}_i'\hat{F}_t$ otherwise. The estimate of F is then updated by computing the eigenvectors corresponding to the largest r eigenvalues of $N^{-1}\Sigma_i\hat{\underline{X}}_i\hat{\underline{X}}_i$ where $\hat{\underline{X}}_i=(\hat{X}_{i1},$

$\hat{X}_{i2},\ldots,\hat{X}_{iT})'$. The estimate of Λ is updated by the OLS regression of $\hat{X}$ onto this updated estimate of F.

B. Mixed monthly/quarterly data - I(0) stock variables. A series that is observed quarterly and is a stock variable would be the point-in-time level of a variable at the end of the quarter, say the level of inventories at the end of the quarter. If this series is I(0) then it is handled as in case A, that is, it is treated as a monthly series with missing observations in the first and second months of the quarter.

C. Mixed monthly/quarterly data - I(0) flow variables. A quarterly flow variable is the average (or sum) of unobserved monthly values. If this series is I(0), it can be treated as follows. The unobserved monthly series, X_{it}, is measured only as the time aggregate X^q_{it} where $X^q_{it}=(1/3)(X_{i,t-2}+X_{i,t-1}+X_{it})$ for t=3,6,9,12,..., and X^q_{it} is missing for all other values of t. In this case estimation procedes as in case A, but with $\hat{X}_{it}=\hat{\lambda}_i'\hat{F}_t + \hat{e}_{it}$, where $\hat{e}_{it}=X^q_{i\tau}-\hat{\lambda}_i'(\hat{F}_{\tau-2}+\hat{F}_{\tau-1}+\hat{F}_{\tau})/3$, where $\tau=3$ when t=1,2,3, $\tau=6$, when t=4,5,6, etc.

D. Mixed monthly/quarterly data - I(1) stock variables. Suppose that underlying monthly data are I(1) and let X^q_{it} denote the quarterly first difference stock variable, assumed to be measured in the third month of every quarter, and let X_{it} denote the monthly first difference of the variable. Then $X^q_{it}=(X_{i,t-2}+X_{i,t-1}+X_{it})$ for t=3,6,9,12,..., and X^q_{it} is missing for all other values of t. In this case estimation procedes as in case A, but with $\hat{X}_{it}=\hat{\lambda}_i'\hat{F}_t+(1/3)\hat{e}_{it}$, where $\hat{e}_{it}=X^q_{i\tau}-\hat{\lambda}_i'(\hat{F}_{\tau-2}+\hat{F}_{\tau-1}+\hat{F}_{\tau})$, where $\tau=3$ when t=1,2,3, $\tau=6$, when t=4,5,6, etc.

E. Mixed monthly/quarterly data - I(1) flow variables.

Construction of $\hat{X}_{it}$ is more difficult here than in the earlier cases. Here the general regression formula given above can be implemented after specifying $\underline{X}_i^\dagger$ and A_i. Let the quarterly first differences be denoted by X^q_{it}, which is assumed to observed at the end of every quarter. The vector of observations is then $\underline{X}^\dagger_i=(X^q_{i3}, X^q_{i6},\ldots,X^q_{i\tau})'$, where τ denotes the month of the last quarterly observation. If the underlying quarterly data are averages of monthly series, and if the monthly first differences are denoted by X_{it}, then $X^q_{it}=(1/3)(X_{i,t}+2X_{i,t-1}+3X_{it-2}+2X_{it-3}+X_{it-4})$ for $t=3,6,9,12,\ldots$, and this implicitly defines the rows of A_i. Then the estimate of $\underline{X}_i$ is given by $\hat{\underline{X}}_i=F\lambda_i+A_i'(A_iA_i')^{-1}(\underline{X}^\dagger_i-A_iF\lambda_i)$.

Appendix B: Proofs of Theorems

The proof of theorem 1 makes use of the following results. First, adopt some additional notation. Let $P_F = F(F'F)^{-1}F'$, let $M_F = I - P_F$, and let $\underline{e}_i = (e_{i1}, \ldots e_{iT})'$.

Lemma B1. Under the conditions of theorem 1,

(a) $tr[T^{-1}F^{0\prime}(P_{\tilde{F}} - P_{F^0})F^0] \xrightarrow{p} 0$;

(b) $tr[T^{-1}\tilde{F}'(P_{\tilde{F}} - P_{F^0})\tilde{F}] \xrightarrow{p} 0$; and

(c) if the $T \times 1$ vector z is such that $z'z/T \xrightarrow{p} c < \infty$, $T^{-1}z'(P_{\tilde{F}} - P_{F^0})z \xrightarrow{p} 0$.

Proof of Lemma B1

(a) As discussed in section 2.2, $\tilde{F} = \operatorname{argmax}_{\{F:\ F'F/T = I\}}\tilde{Q}(F)$, where $\tilde{Q}(F) = (NT)^{-1}tr(F'XX'F) = (NT)^{-1}\sum_{i=1}^{N}\underline{X}_i'P_F\underline{X}_i$. Define $\tilde{Q}^*(F) = (NT)^{-1}\sum_{i=1}^{N}\lambda_i'F^{0\prime}P_F F^0\lambda_i$. By direct calculation, $\tilde{Q}(F) - \tilde{Q}^*(F) = A_{4T}(F) + A_{5T}(F)$, where $A_{4T}(F) = 2(NT)^{-1}\sum_{i=1}^{N}\underline{e}_i'P_F F^0\lambda_i$ and $A_{5T}(F) = (NT)^{-1}\sum_{i=1}^{N}\underline{e}_i'P_F\underline{e}_i$. It is shown below that

(B.1) $$\sup_F|A_{4T}(F)| \xrightarrow{p} 0 \text{ and } \sup_F|A_{5T}(F)| \xrightarrow{p} 0$$

for F: $F'F/T = I_r$ (this normalization is maintained for the rest of this proof). Thus

(B.2) $$\sup_F|\tilde{Q}(F) - \tilde{Q}^*(F)| \xrightarrow{p} 0.$$

Let $\tilde{F}^*$ solve $\max_F\tilde{Q}^*(F)$, and write, $\tilde{Q}^*(\tilde{F}) - \tilde{Q}^*(F^0) = [\tilde{Q}^*(\tilde{F}) - \tilde{Q}(\tilde{F})] + [\tilde{Q}(\tilde{F}) - \tilde{Q}^*(\tilde{F}^*)] + [\tilde{Q}^*(\tilde{F}^*) - \tilde{Q}^*(F^0)]$. By (B.2), $\tilde{Q}^*(\tilde{F}) - \tilde{Q}(\tilde{F}) \xrightarrow{p} 0$ and $\tilde{Q}(\tilde{F}) - \tilde{Q}^*(\tilde{F}^*) = \sup_F\tilde{Q}(F) - \sup_F\tilde{Q}^*(F) \xrightarrow{p} 0$, so $\tilde{Q}^*(\tilde{F}) - \tilde{Q}^*(F^0) = [\tilde{Q}^*(\tilde{F}) - \tilde{Q}(\tilde{F})] + o_p(1)$. Now $\tilde{Q}^*(F) =$

$\text{tr}[T^{-1}F^{0\prime}P_F F^0(\Lambda'\Lambda/N)] = \text{tr}[(F'F^0/T)(\Lambda'\Lambda/N)(F^{0\prime}F/T)]$ because $F'F/T = I$. Because $(\Lambda'\Lambda/N)$ is $r \times r$ and full rank by condition FL, evidently $\tilde{Q}(F)$ is maximized by $\tilde{F}^* = F^0\tilde{R}$, where $\tilde{R}$ is such that $F'F/T = I$, so $\tilde{R}'(F^{0\prime}F^0/T)\tilde{R} = I$ and $(F^{0\prime}F^0/T) = (\tilde{R}\tilde{R}')^{-1}$. Thus, $P_{\tilde{F}^*} = T^{-1}F^0\tilde{R}[\tilde{R}'(F^{0\prime}F^0/T)\tilde{R}]^{-1}\tilde{R}'F^0 = P_{F^0}$, so $\tilde{Q}^*(\tilde{F}^*) = \text{tr}[T^{-1}F^{0\prime}P_{F^0}F^0(\Lambda'\Lambda/N)] = \tilde{Q}^*(F^0)$. Thus $\tilde{Q}^*(\tilde{F}) - \tilde{Q}^*(F^0) \xrightarrow{p} 0$. But $\tilde{Q}^*(F^0) - \tilde{Q}^*(\tilde{F}) = \text{tr}[T^{-1}F^{0\prime}(P_{F^0} - P_{\tilde{F}})F^0(\Lambda'\Lambda/N)] \geq \text{tr}[T^{-1}F^{0\prime}(P_{F^0} - P_{\tilde{F}})F^0]\text{mineval}(\Lambda'\Lambda/N)$. By FL, $\text{mineval}(\Lambda'\Lambda/N) \to c > 0$, so $\text{tr}[T^{-1}F^{0\prime}(P_{F^0} - P_{\tilde{F}})F^0] \xrightarrow{p} 0$.

(b) Using the normalization $\tilde{F}'\tilde{F}/T = I_r$, part (a), and condition M2, we have,

$$\begin{aligned} \text{tr}[T^{-1}\tilde{F}'(P_{\tilde{F}} - P_{F^0})\tilde{F}] &= r - \text{tr}[(\tilde{F}'F^0/T)(F^{0\prime}F^0/T)^{-1}(F^{0\prime}\tilde{F}/T)] \\ &= r - \text{tr}[(F^{0\prime}F^0/T)^{-1}T^{-1}F^{0\prime}(P_{\tilde{F}} - P_{F^0})F^0] - r \\ &\leq \text{tr}[T^{-1}F^{0\prime}(P_{F^0} - P_{\tilde{F}})F^0]\text{maxeval}[(F^{0\prime}F^0/T)^{-1}] \xrightarrow{p} 0. \end{aligned}$$

(c) Without loss of generality write $z = M_{\tilde{F},F^0}z + \tilde{F}a + F^0b$, where $M_{\tilde{F},F^0}$ is the orthogonal projection operator for the combined column space of $(\tilde{F}, F^0)$. Thus $T^{-1}z'(P_{\tilde{F}} - P_{F^0})z = (a'\tilde{F}' + b'F^{0\prime})(P_{\tilde{F}} - P_{F^0})(\tilde{F}a + F^0b) \leq \{\|a\|[\text{tr}(T^{-1}\tilde{F}'(P_{\tilde{F}} - P_{F^0})\tilde{F})]^{1/2} + \|b\|[\text{tr}(T^{-1}F^{0\prime}(P_{\tilde{F}} - P_{F^0})F^0)]^{1/2}$. The assumption $z'z/T \xrightarrow{p} c$ implies that $\|a\|$ and $\|b\|$ are $O_p(1)$. It follows from parts (a) and (b) that $T^{-1}z'(P_{\tilde{F}} - P_{F^0})z \xrightarrow{p} 0$.

It remains to show (B.1). Looking ahead to theorem 2, (B.1) is shown here for the case that the number of estimated factors, k_T, can differ from r, where $k_T = O(\ln T)$. First consider $A_{4T}(F)$. Use $F'F/T = I_{k_T}$ to write $\frac{1}{2}A_{4T}(F) = T^{-1}\sum_{t=1}^{T}\phi_t'\upsilon_t$, where $\phi_t = (F^{0\prime}F/T)F_t'$ and $\upsilon_t = N^{-1}\sum_{i=1}^{N}\lambda_i e_{it}$. Because υ_t does not depend on F, $\sup_F|\frac{1}{2}A_{4T}(F)| \leq (\sup_F T^{-1}\sum_{t=1}^{T}\phi_t'\phi_t)^{1/2}(T^{-1}\sum_{t=1}^{T}\upsilon_t'\upsilon_t)^{1/2}$. Now, $T^{-1}\sum_{t=1}^{T}\phi_t'\phi_t = T^{-1}\sum_{t=1}^{T}F_t(F'F^0/T)(F^{0\prime}F/T)F_t' = \text{tr}[T^{-1}F^{0\prime}P_F F^0]$. Thus $\sup_F T^{-1}\sum_{t=1}^{T}\phi_t'\phi_t \leq \text{tr}(F^{0\prime}F^0/T) \xrightarrow{p} \text{tr}(\Sigma_F)$. Also, as in condition M1 let $\tau_{ij} = E(e_{it}e_{jt})$, so

$$ET^{-1}\sum_{t=1}^{T}v_t'v_t = N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i'\lambda_j\tau_{ij}$$

(B.3) $$\leq N^{-1}(N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}|\tau_{ij}|)\sup_i\|\lambda_i\|^2 \to 0$$

by assumptions M1 and FL. Thus $\sup_F|A_{4T}(F)| \xrightarrow{p} 0$.

Next turn to $A_{5T}(F)$ and use $F'F/T = I_{k_T}$ to write,

(B.4) $$A_{5T}(F) = (NT)^{-1}\sum_{i=1}^{N}\underline{e}_i'P_F\underline{e}_i = T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}F_t'F_s\omega_{ts} + T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}F_t'F_s\gamma(t\text{-}s)$$

where $\omega_{ts} = N^{-1}e_t'e_s - \gamma(t\text{-}s)$. Because $T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(F_t'F_s)^2 = tr[(F'F/T)^2] = k_T$, $|T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}F_t'F_s\omega_{ts}| \leq [k_T T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\omega_{ts}^2]^{1/2}$. Also,

$$Ek_T T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\omega_{ts}^2 \leq k_T\sup_{t,s}E\omega_{ts}^2$$

(B.5) $$\leq (k_T/N)\sup_{t,s}N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}|cov(e_{it}e_{is},e_{jt}e_{js})| \to 0$$

by condition M1. Thus $\sup_F|T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}F_t'F_s\omega_{ts}| \xrightarrow{p} 0$. Finally, turn to the second term in (B.4): $\sup_F|T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}F_t'F_s\gamma(t\text{-}s)| \leq [k_T/T]^{1/2}\sum_{u=-\infty}^{\infty}|\gamma(u)| \to 0$ by condition M1. Thus $\sup_F|A_{5T}(F)| \xrightarrow{p} 0$.

Inspection of the proof of this lemma reveals that limits are taken with respect to N only twice, in (B.3) and (B.5). In both cases, T does not appear in these expressions. In all the limits taken as $T\to\infty$, N does not enter the expressions. Thus these limits hold jointly if $N\to\infty$ and $T\to\infty$, where there is no restriction on the joint rate. In particular, these limits hold if $N=O(T^\rho)$ for any constant $\rho>0$. □

Proof of theorem 1

Define the full sample estimator $\hat{\beta}^\dagger = (\sum_{t=1}^{T}\hat{F}_t\hat{F}_t')^{-1}(\sum_{t=1}^{T}\hat{F}_t y_{t+1})$ and let $\hat{y}_{T+1|T}^\dagger = \hat{\beta}^{\dagger\prime}\hat{F}_T$. Because $\hat{F}_T$ is $O_p(1)$, $\hat{\beta}^\dagger - \hat{\beta} \xrightarrow{p} 0$ and $\hat{y}_{T+1|T}^\dagger - \hat{y}_{T+1|T} \xrightarrow{p} 0$, so it suffices to prove the result for $\hat{\beta}^\dagger$ and $\hat{y}^\dagger_{T+1|T}$. Also, although the theorem is stated in terms of $\hat{F}$, it is convenient instead to prove it for statistics based on $\tilde{F}$ because $\tilde{F}'\tilde{F}/T=I$. This change is without loss of generality because the column spaces of $\tilde{F}$ and $\hat{F}$ are identical. Accordingly, define $\tilde{\beta} = (\Sigma_{t=1}^T \tilde{F}_t\tilde{F}_t')^{-1}(\Sigma_{t=1}^T \tilde{F}_t y_{t+1})$ and let $\tilde{y}_{T+1|T} = \tilde{\beta}'\tilde{F}_T$; then $\tilde{y}_{T+1|T} = \hat{y}^\dagger_{T+1|T}$. To prove this theorem, it therefore suffices to show that $\tilde{y}_{T+1|T} - y_{T+1|T} \xrightarrow{p} 0$.

Let $y = (y_2,\ldots,y_{T+1})'$. Now,

$$\tilde{y}_{T+1|T} - y_{T+1|T} = \tilde{F}_T'\tilde{\beta} - F_T^{0\prime}\beta = \tilde{F}_T'(\tilde{F}'\tilde{F})^{-1}\tilde{F}'y - F_T^{0\prime}\beta$$
$$= g_{1T}'\beta + g_{2T}'\tilde{F}_T$$

where $g_{1T} = (F^{0\prime}\tilde{F}/T)\tilde{F}_T - F_T^0$ and $g_{2T} = \tilde{F}'\epsilon/T$. It will be shown that $g_{1T} \xrightarrow{p} 0$ and $g_{2T} \xrightarrow{p} 0$, from which it follows that $\tilde{y}_{T+1|T} - y_{T+1|T} \xrightarrow{p} 0$.

First consider g_{1T}. It will be shown that $Eg_{1T}'(F^{0\prime}F^0/T)^{-1}g_{1T} \to 0$, from which it follows that $g_{1T} \xrightarrow{p} 0$ (because $F^{0\prime}F^0/T \xrightarrow{p} \Sigma_F$, which is positive definite). Now,

$$g_{1T}'(F^{0\prime}F^0/T)^{-1}g_{1T} = [\tilde{F}_T'(\tilde{F}'F^0/T) - F_T^{0\prime}](F^{0\prime}F^0/T)^{-1}[(F_T^{0\prime}\tilde{F}/T)\tilde{F}_T - F_T^0]$$
$$= G_{1T}^a + G_{1T}^b - 2G_{1T}^c$$

where $G_{1T}^a = \tilde{F}_T'(T^{-1}\tilde{F}'P_{F^0}\tilde{F})\tilde{F}_T$, $G_{1T}^b = F_T^{0\prime}(F^{0\prime}F^0/T)^{-1}F_T^0$, and $G_{1t}^c = F_T^{0\prime}(F^{0\prime}F^0/T)^{-1}(F^{0\prime}\tilde{F}/T)\tilde{F}_T$.

Consider the terms G_{1T}^a, G_{1T}^b, and G_{1T}^c. First, $G_{1T}^a = tr(\tilde{F}_T\tilde{F}_T') + tr[T^{-1}\tilde{F}'(P_{F^0}-P_{\tilde{F}})\tilde{F}(\tilde{F}_T\tilde{F}_T')]$. Now $|tr[T^{-1}\tilde{F}'(P_{F^0}-P_{\tilde{F}})\tilde{F}(\tilde{F}_T\tilde{F}_T')]| \le tr[T^{-1}\tilde{F}'(P_{\tilde{F}}-P_{F^0})\tilde{F}]\,\|\tilde{F}_T\|^2 \xrightarrow{p} 0$ by lemma B1(b) because $\|\tilde{F}_T\| = O_p(1)$. By the normalization of $\tilde{F}$, $Etr(\tilde{F}_T\tilde{F}_T') \to r$; thus $EG_{1T}^a \to r$. Second, by condition M2, $F^{0\prime}F^0/T \xrightarrow{p} \Sigma_F$, so $EG_{1T}^b \to Etr(\Sigma_F F_T^0 F_T^{0\prime}) = r$. Third, because lemma B1(b) implies $\|T^{-1}\tilde{F}'(P_{\tilde{F}}-P_{F^0})\tilde{F}\| \xrightarrow{p} 0$ and

$\tilde{F}'P_{\tilde{F}}\tilde{F}/T = I$, $T^{-1}\tilde{F}'P_{F^0}\tilde{F} \xrightarrow{p} \Sigma_{\tilde{F},F^0}\Sigma_F^{-1}\Sigma_{F^0,\tilde{F}} = I_r$. Thus $(\tilde{F}'F^0/T) \xrightarrow{p} \Sigma_{\tilde{F},F^0} = \Sigma_F^{1/2}$, which is full rank because Σ_F is positive definite. Now use this observation to write,

$$G_{1T}^c = F_T^{0\prime}\Sigma^{-1}_{\tilde{F},F^0}(T^{-1}\tilde{F}'P_{F^0}\tilde{F})\tilde{F}_T + o_p(1)$$
$$= tr\{[T^{-1}\tilde{F}'(P_{F^0}-P_{\tilde{F}})\tilde{F}](\tilde{F}_T F_T^{0\prime}\Sigma^{-1}_{\tilde{F},F^0})\} + F_T^{0\prime}\Sigma^{-1}_{\tilde{F},F^0}\tilde{F}_T.$$

Now $EF_T^{0\prime}\Sigma^{-1}_{\tilde{F},F^0}\tilde{F}_T = r$ and, by lemma B1(b), $|tr[T^{-1}\tilde{F}'(P_{F^0}-P_{\tilde{F}})\tilde{F}]| \xrightarrow{p} 0$. Thus $G_{1T}^c \to r$, so $Eg_{1T}'(F^{0\prime}F^0/T)^{-1}g_{1T} \to 0$ so $g_{1T} \xrightarrow{p} 0$.

Next consider g_{2T}. Because $\epsilon'\epsilon/T \xrightarrow{p} \sigma_\epsilon^2$, $g_{2T}'g_{2T} = (\epsilon'\tilde{F}/T)(\tilde{F}'\epsilon/T) = T^{-1}\epsilon'P_{\tilde{F}}\epsilon = T^{-1}\epsilon'P_{F^0}\epsilon + T^{-1}\epsilon'(P_{\tilde{F}}-P_{F^0})\epsilon = T^{-1}\epsilon'P_{F^0}\epsilon + o_p(1)$ by lemma B1(c). Thus $Eg_{2T}'g_{2T} = ET^{-1}\epsilon'P_{F^0}\epsilon + o(1) = T^{-1}ET^{-1}\Sigma_{t=1}^T\epsilon_{t+1}^2F_t^{0\prime}(F^{0\prime}F^0/T)^{-1}F_t^0 + o(1) = \sigma_\epsilon^2 r/T + o(1) \to 0$. Thus $g_{2T} \xrightarrow{p} 0$.

Note that all the limits in this proof, other than those that rely on lemma B1, are taken as $T\to\infty$ without reference to N. Thus the only restriction on the joint limits of (N,T) arise from lemma B1 which, as was discussed at the end of its proof, holds for $N,T \to \infty$ and $N=O(T^\rho)$ for any $\rho>0$. □

The proofs of theorems 2 and 3 make use of inequalities similar to those used in the proof of theorem 1, but because the conditions for theorems 2 and 3 are different and because theorem 2 provides a stronger result on uniform consistency at the rate δ_{NT}, these (and additional) inequalities are shown to hold under the different conditions of theorems 2 and 3. These are collected in lemma B2.

Lemma B2. Let $F'F/T = I$. Under the conditions of theorem 2,

(a) $\delta_{NT}k_T\sup_{s,t}|F_s^{0\prime}\Lambda_0'e_t/N| \xrightarrow{p} 0$;

(b) $\delta_{NT}k_T\sup_{s,t}|F_s^{0\prime}(\Lambda_s-\Lambda_0)'e_t/N| \xrightarrow{p} 0$;

(c) $\delta_{NT}k_T\sup_{s,t}|e_s'e_t/N - \gamma(s-t)| \xrightarrow{p} 0$;

(d) $\delta_{NT}k_T\sup_{s,t}|F_s^{0\prime}(\Lambda_s'\Lambda_t/N - \Lambda_0'\Lambda_0/N)F_t^0| \xrightarrow{p} 0$;

(e) $\delta_{NT}k_T\sup_{s,t}|X_s'X_t/N - F_s^{0\prime}(\Lambda_0'\Lambda_0/N)F_t^0 - \gamma(s-t)| \xrightarrow{p} 0$;

(f) $\sup_F|N^{-1}\Sigma_{i=1}^N\|\Delta_i\|^2 \xrightarrow{p} 0$.

(g) $\sup_F|N^{-1}\Sigma_{i=1}^N(e_i'F/T)\Delta_i| \xrightarrow{p} 0$; and

(h) $\sup_F|N^{-1}\Sigma_{i=1}^N\lambda_{i0}'(F^{0\prime}F/T)\Delta_i| \xrightarrow{p} 0$.

Proof of Lemma B2

The proof uses the following facts. Let μ_t and $\nu_{t,s}$ be random matrices indexed by $t=1,\ldots,T$, $s=1,\ldots,T$. Then:

(B.6a) If $T\sup_t E(\|\mu_t\|^q) \to 0$ for $q\geq 1$, then $\sup_t\|\mu_t\| \xrightarrow{p} 0$;

(B.6b) If $T^2\sup_{s,t}E(\|\nu_{s,t}\|^q) \to 0$ for $q\geq 1$, then $\sup_{s,t}\|\nu_{s,t}\| \xrightarrow{p} 0$; and

(B.6c) If both $\sup_{s,t}\|E\nu_{s,t}\| \to 0$ and $T^2\sup_{s,t}E(\|\nu_{s,t}-E\nu_{s,t}\|^2) \to 0$, then $\sup_{s,t}\|\nu_{s,t}\| \xrightarrow{p} 0$.

Also note that the rates given in theorem 2 imply the limits,

(B.7) $k_T^2/\delta_{NT} \to 0$, $\delta_{NT}k_T^6/T \to 0$, and $\delta_{NT}^2T^2k_T^2(\ln T)^8/N \to 0$.

(a) Let $\nu_{s,t}=\delta_{NT}k_TF_s^{0\prime}\Lambda_0'e_t/N$ and use (B.6b) with $q=2$. Then $T^2\sup_{s,t}E\nu_{s,t}^2 = T^2\delta_{NT}^2k_T^2E[F_s^{0\prime}\Lambda_0'e_t/N]^2 \leq \delta_{NT}^2k_T^2E(F_s^{0\prime}F_s^0)E(e_t'\Lambda_0\Lambda_0'e_t/N^2)$. Now $E(F_s^{0\prime}F_s^0) = tr(\Sigma_F)$ and $E(e_t'\Lambda_0\Lambda_0'e_t/N^2) \leq (r\bar{\lambda}^2/N)N^{-1}\Sigma_{i=1}^N\Sigma_{j=1}^N|\tau_{ij}|$, so $T^2\sup_{s,t}E\nu_{t,s}^2 \leq (\delta_{NT}^2T^2k_T^2/N)tr(\Sigma_F)r\bar{\lambda}^2N^{-1}\Sigma_{i=1}^N\Sigma_{j=1}^N|\tau_{ij}| \to 0$ by condition M3 and (B.7).

(b) Let $\nu_{s,t} = \delta_{NT}k_T(\Lambda_s-\Lambda_0)'e_t/N$ and note that, with probability one, $|\delta_{NT}k_TF_s^{0\prime}(\Lambda_s-\Lambda_0)'e_t/N| = |F_s^{0\prime}\nu_{s,t}| \leq [\|F_s^0\|/(\ln T)^2\|][(\ln T)^2\|\nu_{s,t}\|]$. Thus the result follows from M3(c)(iii) if $(\ln T)^2\sup_{s,t}\|\nu_{s,t}\| \xrightarrow{p} 0$. This is now shown using (B.6c).

Now $Ee_t'(\Lambda_s-\Lambda_0)/N = N^{-1}\Sigma_{i=1}^{N}\Sigma_{r=1}^{s}Eh_{iT}\zeta_{ir}e_{it} = \kappa_{1T}N^{-1}\Sigma_{i=1}^{N}\Sigma_{r=1}^{s}\Psi_{ii}(t-r)$, so

$$(\ln T)^2\sup_{s,t}\|E\nu_{s,t}\| = (\ln T)^2\sup_{s,t}\delta_{NT}k_T\|\kappa_{1T}N^{-1}\Sigma_{i=1}^{N}\Sigma_{r=1}^{s}\Psi_{ii}(t-r)\|$$
$$\leq [(\ln T)^2\delta_{NT}k_T/T](T\kappa_{1T})\sup_i\Sigma_{u=-\infty}^{\infty}\|\Psi_{ii}(u)\|$$

which tends to zero by (B.7) and M3(b)(i). In addition,

$$(\ln T)^2T^2\sup_{s,t}E\|\nu_{s,t}-E\nu_{s,t}\|^2 = (\ln T)^2T^2\sup_{s,t}\delta_{NT}^2k_T^2\Sigma_{m=1}^{r}\mathrm{var}[N^{-1}\Sigma_{i=1}^{N}e_{it}(\lambda_{is,m}-\lambda_{i0,m})]$$
$$\leq [(\ln T)^2\delta_{NT}^2k_T^2T^2/N](T\kappa_{2T})^2\{r(\sup_{i,t}Ee_{it}^4)^{1/2}(\sup_{i,m,s,}E\zeta_{is,m}^4)^{1/2}$$
$$+ N^{-1}\Sigma_{i=1}^{N}\Sigma_{j=1}^{N}\sup_{t,s,s'}|tr[cov(e_{it}\zeta_{is},e_{jt}\zeta_{js'})]|\}$$

which tends to zero by (B.7) and conditions M1 and M3.

(c) Let $\nu_{s,t}=\delta_{NT}k_T[e_s'e_t/N - \gamma(t-s)]$ and use (B.6b) with q=2. Now $T^2\sup_{s,t}E\|\nu_{s,t}\|^2 = (\delta_{NT}^2k_T^2T^2/N)\sup_{s,t}|N^{-1}\Sigma_{i=1}^{N}\Sigma_{j=1}^{N}cov(e_{is}e_{it},e_{js}e_{jt})| \to 0$ by (B.7) and M1.

(d) $$\sup_{s,t}\|\delta_{NT}k_TF_s^{0\prime}[(\Lambda_s'\Lambda_t-\Lambda_0'\Lambda_0)/N]F_t^0\| \leq \sup_{s,t}\delta_{NT}k_T\|F_s^0\|^2\|(\Lambda_s'\Lambda_t-\Lambda_0'\Lambda_0)/N]\|$$
$$\leq [\sup_t\|F_s^0\|/(\ln T)^2]^2\{(\ln T)^4\sup_{s,t}\|\nu_{s,t}\| + 2(\ln T)^4\sup_t\|\mu_t\|\}$$

where $\nu_{s,t} = \delta_{NT}k_T(\Lambda_s-\Lambda_0)'(\Lambda_t-\Lambda_0)/N$ and $\mu_t = \delta_{NT}k_T(\Lambda_t-\Lambda_0)'\Lambda_0/N$.

First, show that $(\ln T)^4\sup_{s,t}\|\nu_{s,t}\| \xrightarrow{p} 0$ using (B.6c):

$$(\ln T)^4\sup_{s,t}\|E\nu_{s,t}\| = \sup_{s,t}(\ln T)^4\delta_{NT}k_T\|N^{-1}\Sigma_{i=1}^{N}\kappa_{2T}^2\Sigma_{r=1}^{s}\Sigma_{r'=1}^{t}E\zeta_{ir}\zeta_{ir'}'\|$$
$$\leq [(\ln T)^4\delta_{NT}k_T/T](T\kappa_{2T})^2\sup_i\Sigma_{u=-\infty}^{\infty}\|\Gamma_{ii}(u)\|$$

which tends to zero by (B.7) and M3. Also, by (B.7) and M3,

$$
\begin{aligned}
&(\ln T)^4 T^2 \sup_{t,s} E\|\nu_{t,s} - E\nu_{t,s}\|^2 \\
&\quad = (\ln T)^4 \sup_{t,s} \delta_{NT}^2 k_T^2 T^2 E\|N^{-1}\Sigma_{i=1}^N \Sigma_{r=1}^s \Sigma_{r'=1}^t [h_{iT}^2 \zeta_{ir}\zeta_{ir'}' - E(h_{iT}^2 \zeta_{ir}\zeta_{ir'}')]\|^2 \\
&\quad = (\ln T)^4 \sup_{t,s} \delta_{NT}^2 k_T^2 T^2 \\
&\qquad \times \Sigma_{\ell=1}^r \Sigma_{m=1}^r E\{N^{-1}\Sigma_{i=1}^N \Sigma_{r=1}^s \Sigma_{r'=1}^t [h_{iT}^2 \zeta_{ir,\ell}\zeta_{ir',m} - E(h_{iT}^2 \zeta_{ir,\ell}\zeta_{ir',m})]\}^2 \\
&\quad \le [(\ln T)^4 \delta_{NT}^2 k_T^2 T^2/N] r^2 (T\kappa_{4T})^4 \{\sup_{i,s,m} E\zeta_{is,m}^4 \\
&\qquad + \sup_{\ell,m} N^{-1}\Sigma_{i=1}^N \Sigma_{j=1}^N \sup_{t,u_1,u_2,u_3} |\mathrm{cov}(\zeta_{it,\ell}\zeta_{it+u_1,m}, \zeta_{jt+u_2,\ell}\zeta_{jt+u_3,m})| < \infty \\
&\quad \to 0
\end{aligned}
$$

by M3(a) so that $(\ln T)^4 \sup_{t,s}\|\nu_{t,s}\| \xrightarrow{p} 0$. Next, show that $(\ln T)^4 \sup_t \|\mu_t\| \xrightarrow{p} 0$ using (B.6a) with q=2:

$$
\begin{aligned}
T(\ln T)^8 \sup_t E\|\mu_t\|^2 &= (\ln T)^8 T \sup_t \delta_{NT}^2 k_T^2 E\|(\Lambda_t - \Lambda_0)'\Lambda_0/N\|^2 \\
&\le [(\ln T)^8 \delta_{NT}^2 k_T^2/N] r^2 \bar{\lambda}^2 (T\kappa_{2T})^2 \sup_m N^{-1}\Sigma_{i=1}^N \Sigma_{j=1}^N \Sigma_{u=-\infty}^{\infty} |\Gamma_{ij,mm}(u)|
\end{aligned}
$$

which tends to zero by (B.7) and M3, so $\sup_t \|\mu_t\| \xrightarrow{p} 0$.

(e) Write $\delta_{NT}k_T[X_s'X_t/N - F_s^{0\prime}(\Lambda_0'\Lambda_0/N)F_t^0 - \gamma(s-t)] = \Sigma_{i=1}^4 \nu_{i,st}$, where

$$
\begin{aligned}
\nu_{1,st} &= \delta_{NT}k_T F_s^{0\prime}[(\Lambda_s'\Lambda_t - \Lambda_0'\Lambda_0)/N]F_t^0 \\
\nu_{2,st} &= \delta_{NT}k_T[e_s'e_t/N - \gamma(s-t)] \\
\nu_{3,st} &= 2\delta_{NT}k_T F_s^{0\prime}(\Lambda_s - \Lambda_0)'e_t/N \\
\nu_{4,st} &= 2\delta_{NT}k_T F_s^{0\prime}\Lambda_0'e_t/N.
\end{aligned}
$$

It was shown in parts (a)-(d) that $\sup_{s,t}\|\nu_{i,st}\| \xrightarrow{p} 0$, i=1,...,4, and the desired result follows.

(f) Define $A_{ts} = N^{-1}\Sigma_{i=1}^{N}(\lambda_{it}-\lambda_{i0})(\lambda_{is}-\lambda_{i0})'$. Then

$$\begin{aligned}
\sup_F N^{-1}\Sigma_{i=1}^{N}\|\Delta_i\|^2 &= \sup_F T^{-2}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T}F_t'F_sF_t^{0\prime}A_{ts}F^0 \\
&\le [\sup_F T^{-2}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T}(F_t'F_s)^2]^{1/2}[T^{-2}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T}(F_t^{0\prime}A_{ts}F^0)^2]^{1/2} \\
&\le [k_T T^{-2}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T}\|F_t^0\|^2\|F_s^0\|^2\|A_{ts}\|^2]^{1/2} \\
&\le \{[\sup_t\|F_t^0\|/(\ln T)^2]^4 k_T(\ln T)^8T^{-2}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T}\|A_{ts}\|^2\}^{1/2}.
\end{aligned}$$

By M3(c)(iii), $\sup_t\|F_t^0\|/(\ln T)^2 \xrightarrow{p} 0$. It will now be shown that $Ek_T(\ln T)^8T^{-2}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T}\|A_{ts}\|^2 \to 0$, from which it follows that $\sup_F N^{-1}\Sigma_{i=1}^{N}\|\Delta_i\|^2 \xrightarrow{p} 0$. To simplify the proof, this is shown for $r=1$; the proof for general r is similar. First, write $A_{ts} = N^{-1}\Sigma_{i=1}^{N}h_{iT}^2\Sigma_{r=1}^{t}\Sigma_{r'=1}^{s}\zeta_{ir}\zeta_{ir'}$. Thus $\sup_{t,s}(EA_{ts})^2 \le T^{-2}[(T\kappa_{2T})^2N^{-1}\Sigma_{i=1}^{N}\Sigma_{u=-\infty}^{\infty}|\Gamma_{ii}(u)|]^2 = c_1/T^2$, say, where c_1 is a finite constant. Also,

$$\begin{aligned}
E[A_{ts}-EA_{ts}]^2 &= E[N^{-1}\Sigma_{i=1}^{N}\Sigma_{r=1}^{t}\Sigma_{r'=1}^{s}h_{iT}^2(\zeta_{ir}\zeta_{ir'}-\Gamma_{ii}(r-r'))]^2 \\
&\quad + E[N^{-1}\Sigma_{i=1}^{N}(h_{iT}^2-\kappa_{2T}^2)\Sigma_{r=1}^{t}\Sigma_{r'=1}^{s}\Gamma_{ii}(r-r')]^2 \\
&\le N^{-1}(T\kappa_{4T})^4N^{-1}\Sigma_{i=1}^{N}\Sigma_{j=1}^{N}\sup_{t,u_1,u_2,u_3}|\mathrm{cov}(\zeta_{it}\zeta_{it+u_1},\zeta_{jt+u_2}\zeta_{jt+u_3})| \\
&\quad + (NT^2)^{-1}(T\kappa_{4T})^4(N^{-1}\Sigma_{i=1}^{N}\Sigma_{u=-\infty}^{\infty}|\Gamma_{ii}(u)|)^2
\end{aligned}$$

so, by M3(a), $\sup_{t,s}E[A_{ts}-EA_{ts}]^2 \le c_2/N + c_3/NT^2$, say, where c_2 and c_3 are finite constants. Thus $Ek_T(\ln T)^8T^{-2}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T}\|A_{ts}\|^2 \le k_T(\ln T)^8\sup_{t,s}E\|A_{ts}\|^2 \le k_T(\ln T)^8[c_1/T^2+c_2/N+c_3/NT^2] \to 0$.

(g) Now $\sup_F|N^{-1}\Sigma_{i=1}^{N}(\underline{e}_i'F/T)\Delta_i| \le [\sup_F A_{5T}(F)]^{1/2}[\sup_F N^{-1}\Sigma_{i=1}^{N}\|\Delta_i\|^2]^{1/2}$, where $A_{5T}(F) = N^{-1}\Sigma_{i=1}^{N}\underline{e}_i'P_F\underline{e}_i$. It was shown in the proof of lemma B1 that $\sup_F|A_{5T}(F)| \xrightarrow{p} 0$ (that proof is valid under the conditions of theorem 2). The result follows from this and result (f).

(h) $\sup_F|N^{-1}\Sigma_{i=1}^N\lambda_{i0}'(F^{0\prime}F/T)\Delta_i|^2 \le [\sup_F N^{-1}\Sigma_{i=1}^N\lambda_{i0}'(F^{0\prime}F/T)(F'F^0/T)\lambda_{i0}][\sup_F N^{-1}\Sigma_{i=1}^N\|\Delta_i\|^2]$

$= \sup_F tr[(\Lambda_0'\Lambda_0/N)(T^{-1}F^{0\prime}P_F F^0)][\sup_F N^{-1}\Sigma_{i=1}^N\|\Delta_i\|^2]$

Now $\sup_F tr[(\Lambda_0'\Lambda_0/N)(T^{-1}F^{0\prime}P_F F^0)] \le tr(\Lambda_0'\Lambda_0/N)tr(F^{0\prime}F^0/T) \xrightarrow{p} tr(D\Sigma_F)$, so the result follows by applying result (f).

Proof of theorem 2

Note that $\hat{\hat{F}} = \hat{F}(\hat{F}'\hat{F}/T)^{1/2} = N^{-1}X\tilde{\Lambda}$, where $\tilde{\Lambda} = T^{-1}X'\tilde{F}$, where $\tilde{F}$ is (as in section 2) $T^{1/2}$ times the ordered eigenvectors of XX'. Thus we have the identity, $\hat{\hat{F}}_t-H_{NT}F_t^0 = T^{-1}\Sigma_{s=1}^T\tilde{F}_s\gamma(s\text{-}t) + T^{-1}\Sigma_{s=1}^T\tilde{F}_s\xi_{st}$, where $H_{NT} = (\tilde{F}'F^0/T)(\Lambda_0'\Lambda_0/N)$ and $\xi_{st} = X_s'X_t/N - F_s^{0\prime}(\Lambda_0'\Lambda_0/N)F_t^0 - \gamma(s\text{-}t)$. Thus

$$\delta_{NT}\sup_t\|\hat{\hat{F}}_t-H_{NT}F_t^0\| \le \delta_{NT}\sup_t\|T^{-1}\Sigma_{s=1}^T\tilde{F}_s\gamma(s\text{-}t)\| + \delta_{NT}\sup_t\|T^{-1}\Sigma_{s=1}^T\tilde{F}_s\xi_{st}\|.$$

Because $E\tilde{F}_s'\tilde{F}_r \le k_T$, $E\delta_{NT}^2\sup_t\|T^{-1}\Sigma_{s=1}^T\tilde{F}_s\gamma(s\text{-}t)\|^2 = \delta_{NT}^2\sup_t T^{-2}\Sigma_{r=1}^T\Sigma_{s=1}^T E\tilde{F}_s'\tilde{F}_r\gamma(s\text{-}t)\gamma(r\text{-}t) \le (k_T\delta_{NT}^2/T^2)[\Sigma_{u=-\infty}^{\infty}|\gamma(u)|]^2 \to 0$ by (B.7). Also,

$$\delta_{NT}\sup_t\|T^{-1}\Sigma_{s=1}^T\tilde{F}_s\xi_{st}\| \le \delta_{NT}\sup_T\{[T^{-2}\Sigma_{r=1}^T\Sigma_{s=1}^T(\tilde{F}_s'\tilde{F}_r)^2]^{1/2}[T^{-2}\Sigma_{r=1}^T\Sigma_{s=1}^T\xi_{st}^2\xi_{rt}^2]^{1/2}\}^{1/2}$$
$$\le \delta_{NT}k_T^{1/4}\sup_{s,t}|\xi_{st}| \xrightarrow{p} 0$$

by lemma B2(e). Thus $\delta_{NT}\sup_t\|\hat{\hat{F}}_t-H_{NT}F_t^0\| \xrightarrow{p} 0$.

It remains to characterize H_{NT}. Let $\tilde{Q}(F)$ and $\tilde{Q}^*(F)$ be defined as in the proof of lemma B1 with λ_{i0} replacing λ_i, so $\tilde{Q}^*(F) = (NT)^{-1}\Sigma_{i=1}^N\lambda_{i0}'F^{0\prime}P_F F^0\lambda_{i0}$. Algebra reveals that, $\hat{Q}_{NT}(F) - \tilde{Q}_{NT}(F) = \Sigma_{i=1}^5 A_{iT}(F)$, where for $F'F/T = I$, $A_{1T}(F) = N^{-1}\Sigma_{i=1}^N\Delta_i'\Delta_i$, $A_{2T}(F) =$

$2N^{-1}\Sigma_{i=1}^{N}\Delta_i'(F'F^0/T)\lambda_{i0}$, $A_{3T}(F) = 2N^{-1}\Sigma_{i=1}^{N}(\underline{e}_i'F/T)\Delta_i$, $A_{4T}(F) =$ $2N^{-1}\Sigma_{i=1}^{N}(\underline{e}_i'F/T)(F'F^0/T)\lambda_{i0}$, and $A_{5T}(F) = N^{-1}\Sigma_{i=1}^{N}(\underline{e}_i'F/T)(F'\underline{e}_i/T)$.

By lemma B2(f)-(h), $\sup_F|A_{iT}(F)| \xrightarrow{p} 0$, i=1,2,3. The results $\sup_F|A_{iT}(F)| \xrightarrow{p} 0$, i=4,5 follow from (B.1) because (B.1) was proven under the conditions of theorem 2 (note that the terms in ζ_t and λ_{it} do not enter the expressions for $A_{4T}(F)$ and $A_{5T}(F)$). Thus $\sup_F|\hat{Q}_{NT}(F) - \tilde{Q}_{NT}(F)| \xrightarrow{p} 0$. Let $\tilde{F}^*$ maximize $\tilde{Q}^*(F)$. The argument following (B.2) applies here; thus,

$$\text{(B.8)} \qquad \tilde{Q}^*(\tilde{F}^*) - \tilde{Q}^*(\tilde{F}) = tr\{[T^{-1}F^{0\prime}(P_{\tilde{F}^*}-P_{\tilde{F}})F^0](\Lambda_0'\Lambda_0/N)\} \xrightarrow{p} 0.$$

Result (B.8) is now used to characterize H_{NT} for (i) $k_T \le r$ and (ii) $k_T > r$.

(i) $k_T \le r$. Use $F'F/T = I$ to write $\tilde{Q}^*(F) = tr[T^{-1}(\Lambda_0'\Lambda_0/N)^{1/2\prime}F^{0\prime}P_F F^0(\Lambda_0'\Lambda_0/N)^{1/2}] =$ $tr[T^{-1}F'F^1F^{1\prime}F] + o_p(1)$, where $F^1 = F^0D^{1/2}$. For $k \le r$, $\tilde{Q}^*(F)$ evidently is maximized for $\tilde{F}^* =$ $T^{1/2}F^1S$, where S is an O(1) $r \times k_T$ matrix such that the columns of $\tilde{F}^*$ are $T^{1/2}$ times the eigenvectors corresponding to the k_T largest eigenvalues of $F^1F^{1\prime}$. Now (B.8) implies $T^{-1}F^{1\prime}P_{\tilde{F}^*}F^1$ - $T^{-1}F^{1\prime}P_{\tilde{F}}F^1 \xrightarrow{p} 0$. But $T^{-1}F^{1\prime}P_{\tilde{F}^*}F^1 = (F^{1\prime}F^1/T)S[S'(F^{1\prime}F^1/T)S]^{-1}S'(F^{1\prime}F^1/T) \xrightarrow{p} \Sigma^1S[S'\Sigma^1S]^{-1}S'\Sigma^1$ where $\Sigma^1 = D^{1/2\prime}\Sigma_F D^{1/2}$. Thus $T^{-1}F^{1\prime}P_{\tilde{F}}F^1 \xrightarrow{p} D^{1/2\prime}\Sigma_{F^0\tilde{F}}\Sigma_{F^0\tilde{F}}'D^{1/2} = \Sigma^1S(S'\Sigma^1S)^{-1}S'\Sigma^1$, so $\tilde{F}'F^0/T \xrightarrow{p}$ $\Sigma_{\tilde{F}F^0} = (S'\Sigma^1S)^{-1/2\prime}S'\Sigma^1D^{-1/2}$. Thus, for $k \le r$, $\|H_{NT}-H\| \xrightarrow{p} 0$, where $H = \Sigma_{\tilde{F}F^0}D$. Because S has full rank and D and Σ_F are positive definite, $\Sigma_{\tilde{F}F^0}$ has full row rank so H has row rank $k_T = \min(r,k_T)$.

(ii) $k_T > r$. Now $\tilde{F}^* = F^0$ and $\tilde{Q}^*(\tilde{F}^*) - \tilde{Q}^*(\tilde{F}) = tr\{[T^{-1}F^{0\prime}(P_{F^0}-P_{\tilde{F}})F^0](\Lambda_0'\Lambda_0/N)\} \ge$ $\text{mineval}(\Lambda_0'\Lambda_0/N)tr[T^{-1}F^{0\prime}(P_{F^0}-P_{\tilde{F}})F^0]$, so by FL and (B.8), $tr[T^{-1}F^{0\prime}(P_{F^0}-P_{\tilde{F}})F^0] \xrightarrow{p} 0$ so $\|T^{-1}F^{0\prime}(P_{F^0}-P_{\tilde{F}})F^0\| \xrightarrow{p} 0$. Thus $T^{-1}F^{0\prime}P_{\tilde{F}}F^0 \xrightarrow{p} \Sigma_{F^0\tilde{F}}\Sigma_{F^0\tilde{F}}' = \Sigma_F$, where $\Sigma_{F^0\tilde{F}}$ is $r \times k_T$. Because Σ_F is positive definite, $\text{rank}(\Sigma_{F^0\tilde{F}}) = r$, so $\|H_{NT}-H\| \xrightarrow{p} 0$ where $H =$ $\Sigma_{F^0\tilde{F}}'D$ is $k_T \times r$ and has row rank $r = \min(r,k_T)$. □

Next turn to theorem 3. It is useful first to set out some additional notation and preliminary results. When $k > r$, it is convenient to consider forecasts based on $\tilde{F}$ rather than $\hat{F}$ (this is done without loss of generality because they have identical column spaces in finite samples). Partition $\tilde{F}$ as $\tilde{F} = [\tilde{F}^a \ \ \tilde{F}^b]$, where $\tilde{F}^a$ is $T \times r$ and $\tilde{F}^b$ is $T \times (k\text{-}r)$, where the column space of $\tilde{F}^a$ equals the column space of the first r ordered eigenvectors of XX', and the columns of $\tilde{F}^b$ are (in order) $T^{1/2}$ times the next k-r eigenvectors of this matrix. The uniqueness of the eigenvectors implies that, given X, the column space of $\tilde{F}^a$ does not depend on k when $k \geq r$. Define the $r \times r$ matrix $H^a = \Sigma_{\tilde{F}^a F^0}\Sigma_F^{-1}$, which by the argument in the proof of theorem 2 is full rank. Then theorem 2 implies that $\delta_{NT}\sup_t \| \tilde{F}_t^a - H^a F_t^0 \| \xrightarrow{p} 0$.

The following lemma is used in the proof of theorem 3.

Lemma B3.

Under the assumptions of theorem 3,

(a) Let z be a $T \times 1$ random vector with $z'z/T \xrightarrow{p} c < \infty$. Then $\delta_{NT}|z'(P_{\tilde{F}^a} - P_{F^0})z/T| \xrightarrow{p} 0$.

(b) $\delta_{NT}\beta' F^{0\prime} P_{\tilde{F}^b} F^0 \beta/T \xrightarrow{p} 0$

(c) $\delta_{NT}\epsilon' P_{\tilde{F}^b}\epsilon/T \xrightarrow{p} 0$

(d) $\delta_{NT} y' P_{\tilde{F}^b} y/T \xrightarrow{p} 0$

(e) $\delta_{NT}|y'(P_{\tilde{F}^a} - P_{F^0})y/T| \xrightarrow{p} 0$

(f) $\delta_{NT}|y'(P_{\tilde{F}} - P_{F^0})y/T| \xrightarrow{p} 0$.

<u>Proof</u>

Because H_a is invertible, $P_{F^0} = P_{H^a F^0}$, so without loss of generality set $H^a = I_r$.

(a) Let $V_{00} = F^{0\prime}F^0/T$, $V_{z0} = z'F^0/T$, $V_{za} = z'\tilde{F}^a/T$, $V_{aa} = \tilde{F}^{a\prime}\tilde{F}^a/T$, etc. Then,

$$\delta_{NT}|z'(P_{\tilde{F}^a}-P_{F^0})z/T| \le \delta_{NT}\|V_{za}-V_{z0}\|\,\|V_{aa}^{-1}\|\,\|V_{za}\|$$
$$+ \delta_{NT}\|V_{z0}\|\,\|V_{aa}^{-1}-V_{00}^{-1}\|\,\|V_{za}\|$$
$$+ \delta_{NT}\|V_{z0}\|\,\|V_{00}^{-1}\|\,\|V_{za}-V_{z0}\|$$

By condition M3(c), $\|V_{00}\| \le r^{1/2}d$ and $\|V_{00}^{-1}\| \le r^{1/2}c$. Also, $\delta_{NT}\|V_{za}-V_{z0}\| \le \|T^{-1/2}z\|\,\delta_{NT}\sup_t\|\tilde{F}_t^a-F_t^0\| \xrightarrow{p} 0$, and $\delta_{NT}\|V_{00}-V_{aa}\| \le \delta_{NT}\|(\tilde{F}^a-F^0)'(\tilde{F}^a-F^0)/T\| + 2\delta_{NT}\|(\tilde{F}^a-F^0)'F^0/T\| \xrightarrow{p} 0$. By the continuity of the inverse, $\delta_{NT}\|V_{00}^{-1}-V_{aa}^{-1}\| \xrightarrow{p} 0$. By assumption, $\|T^{-1/2}z\|$ and $\|V_{0z}\|$ are $O_p(1)$. The result follows.

(b) Use $\tilde{F}^{a\prime}\tilde{F}^b=0$ to write, $M_{\tilde{F}} = I-P_{\tilde{F}^a}-P_{\tilde{F}^b} = M_{\tilde{F}^a}-P_{\tilde{F}^b}$, so $z'M_{\tilde{F}^a}z = z'P_{\tilde{F}^b}z + z'M_{\tilde{F}}z \ge z'P_{\tilde{F}^b}z$. Also, $z'M_{\tilde{F}^a}z = z'M_{F^0}z + z'(P_{F^0}-P_{\tilde{F}^a})z$. Thus, $\delta_{NT}z'P_{\tilde{F}^b}z/T \le \delta_{NT}z'M_{F^0}z/T + \delta_{NT}|z'(P_{\tilde{F}^a}-P_{F^0})z|/T$. Now let $z = F_0\beta$. Then $\beta'F^{0\prime}M_{F^0}F^0\beta = 0$ and $\delta_{NT}|\beta'F^{0\prime}(P_{\tilde{F}^a}-P_{F^0})F^0\beta/T| \xrightarrow{p} 0$ by part a of this lemma, because $\|T^{-1/2}z\| = \|T^{-1/2}F^0\beta\|^2 \xrightarrow{p} \beta'\Sigma_F\beta < \infty$ and $\|V_{0z}\| = \|T^{-1}F^{0\prime}F^0\beta\| \xrightarrow{p} \|\Sigma_F\beta\| \le \infty$; thus $\delta_{NT}\beta'F^{0\prime}P_{\tilde{F}^b}F^0\beta/T \xrightarrow{p} 0$.

(c) This follows because ϵ_{t+1} is a martingale difference sequence with respect to $\{X_t, y_t, F_t, X_{t-1}, y_{t-1}, F_{t-1},\ldots\}$ and because $P_{\tilde{F}^b}$ is idempotent with rank q-r.

(d) $\delta_{NT}y'P_{\tilde{F}^b}y/T \le [(\delta_{NT}\beta'F^{0\prime}P_{\tilde{F}^b}F^0\beta/T)^{1/2} + (\delta_{NT}\epsilon'P_{\tilde{F}^b}\epsilon/T)^{1/2}]^2$, which converges in probability to zero by parts (b) and (c).

(e) This follows from (a) with x=y because $\|T^{-1/2}y\|^2 \xrightarrow{p} Ey^2$ and $\|T^{-1}F^{0\prime}y\| \xrightarrow{p} \|\Sigma_{F,T}\beta\|$.

(f) This follows from (d) and (e). □

Proof of theorem 3

(a) Let $\hat{\epsilon}_0 = y_t - \hat{\beta}^{0\prime}F_t^0$, where $\hat{\beta}^0 = (F^{0\prime}F^0)^{-1}(F^{0\prime}y)$, and write

$$\hat{\sigma}_\epsilon^2(q) - \sigma_\epsilon^2 = [\hat{\epsilon}'\hat{\epsilon}/T - \hat{\epsilon}^{0\prime}\hat{\epsilon}^0/T] + [\hat{\epsilon}^{0\prime}\hat{\epsilon}^0/T - \epsilon'\epsilon/T] + [\epsilon'\epsilon/T - \sigma_\epsilon^2].$$

Consider the three bracketed terms separately.

(i) $\hat{\epsilon}'\hat{\epsilon}/T - \hat{\epsilon}^{0\prime}\hat{\epsilon}^0/T = y'(P_{\tilde{F}}-P_{F^0})y/T \xrightarrow{p} 0$ by lemma B3(f).

(ii) The moment conditions imply that $\hat{\beta}^0 \xrightarrow{p} \beta$, from which it follows that $\hat{\epsilon}^{0\prime}\hat{\epsilon}^0/T - \epsilon'\epsilon/T \xrightarrow{p} 0$.

(iii) This follows from the moment assumptions on ϵ_t.

(b) The proof proceeds by showing (i) $\Pr[\hat{r}>r] \to 0$ and (ii) $\Pr[\hat{r}<r] \to 0$, hence $\Pr[\hat{r}=r] \to 1$. The results $\hat{\sigma}_\epsilon^2(\hat{r}) \xrightarrow{p} \sigma_\epsilon^2$ follows from the consistency of r and from part (a) of this theorem.

(i) This holds trivially if $k_{max}=r$ so suppose that $k_{max}>r$. Now

$$\begin{aligned}\Pr[\hat{r}>r] &\le \Pr[\min_{k=r+1,\ldots,k_{max}} IC_k < IC_r] \\ &\le \Sigma_{k=r+1}^{k_{max}} \Pr[\delta_{NT}(IC_k-IC_r) < 0] \\ &= \Sigma_{k=r+1}^{k_{max}} \Pr[\delta_{NT}\ln(\hat{\sigma}_\epsilon^2(k)/\hat{\sigma}_\epsilon^2(r)) + (k-r)\delta_{NT}g(T) < 0].\end{aligned}$$

By assumption, $\delta_{NT}g(T)\to\infty$. Thus, because $k>r$, to prove $\Pr[\hat{r}>r]\to 0$ it suffices to show that $\delta_{NT}\ln[\hat{\sigma}_\epsilon^2(k)/\hat{\sigma}_\epsilon^2(r)] \xrightarrow{p} 0$ for $k=r+1,\ldots,k_{max}$. But because $\hat{\sigma}_\epsilon^2(k) \le \hat{\sigma}_\epsilon^2(k-1) \le \cdots \le \hat{\sigma}_\epsilon^2(r)$, it suffices to show this for $k=k_{max}$. Now,

$$\begin{aligned}\delta_{NT}\ln[\hat{\sigma}_\epsilon^2(k_{max})/\hat{\sigma}_\epsilon^2(r)] &= \delta_{NT}\ln[1+\{\delta_{NT}(\hat{\sigma}_\epsilon^2(k_{max})-\hat{\sigma}_\epsilon^2(r))/\hat{\sigma}_\epsilon^2(r)\}/\delta_{NT}] \\ &= \delta_{NT}\ln[1+\{\delta_{NT}(y'P_{\tilde{F}^b}y/T)/\hat{\sigma}_\epsilon^2(r)\}/\delta_{NT}].\end{aligned}$$

From part (a), $\hat{\sigma}^2_\epsilon(r) \xrightarrow{p} \sigma^2_\epsilon$, and from Lemma B3(d), $\delta_{NT}(y'P_{\tilde{F}^b}y/T) \xrightarrow{p} 0$, so the final expression above converges to zero in probability and the desired result follows.

(ii) This holds trivially if $r=1$ so suppose that $r>1$. Following the reasoning in (i), it suffices to show that $\Pr[IC_k \leq IC_r] \to 0$, $1 \leq k < r$. Now $IC_k - IC_r = \ln[\hat{\sigma}^2_\epsilon(k)/\hat{\sigma}^2_\epsilon(r)] + (k-r)g(T)$. Because $g(T) \to 0$ and $\hat{\sigma}^2_\epsilon(r) \xrightarrow{p} \sigma^2_\epsilon$, $\Pr[IC_k \leq IC_r] \to 0$ if $\text{plim}[\hat{\sigma}^2_\epsilon(k) - \hat{\sigma}^2_\epsilon(r)] > 0$. For a given $k<r$, from theorem 2 and the equivalence of the column spaces of $\hat{F}$ and $\tilde{F}$, there is a $k \times r$ matrix $\tilde{H}$ with full row rank such that $\sup_t \| \tilde{F}_T - \tilde{H}F^0_t \| \xrightarrow{p} 0$. Thus

$$
\begin{aligned}
\hat{\sigma}^2_\epsilon(k) - \hat{\sigma}^2_\epsilon(r) &= T^{-1}\beta'F^{0\prime}F^0\beta - T^{-1}\beta'F^{0\prime}P_{\tilde{F}}F^0\beta + o_p(1) \\
&= T^{-1}\beta'F^{0\prime}F^0\beta - T^{-1}\beta'F^{0\prime}P_{F^0\tilde{H}'}F^0\beta + o_p(1) \\
&= \beta'[\Sigma_F - \Sigma_F\tilde{H}'(\tilde{H}\Sigma_F\tilde{H}')^{-1}\tilde{H}\Sigma_F]\beta + o_p(1).
\end{aligned}
$$

Because the term in brackets in the final line is positive definite, $\text{plim}[\hat{\sigma}^2_\epsilon(r) - \hat{\sigma}^2_\epsilon(k)] > 0$, which yields the desired result. □

Appendix C: Data Description

This appendix lists the time series used to construct the diffusion index forecasts discussed in section 5. The format is: series number; series mnemonic; data span used; transformation code; and brief series description. The transformation codes are: 1 = no transformation; 2 = first difference; 4 = logarithm; 5 = first difference of logarithms; 6 = second difference of logarithms. An asterisk after the date denotes a series that was included in the unbalanced panel but not the balanced panel, either because of missing data or because of gross outliers which were treated as missing data. The series were either taken directly from the DRI-McGraw Hill Basic Economics database, in which case the original mnemonics are used, or they were produced by authors' calculations based on data from that database, in which case the authors calculations and original DRI/McGraw series mnemonics are summarized in the data description field. The following abbreviations appear in the data definitions: SA = seasonally adjusted; NSA = not seasonally adjusted; SAAR = seasonally adjusted at an annual rate; FRB = Federal Reserve Board; AC = Authors calculations

Real output and income (Out)

1.	ip	1959:01-1998:12	5	industrial production: total index (1992=100,sa)
2.	ipp	1959:01-1998:12	5	industrial production: products, total (1992=100,sa)
3.	ipf	1959:01-1998:12	5	industrial production: final products (1992=100,sa)
4.	ipc	1959:01-1998:12	5	industrial production: consumer goods (1992=100,sa)
5.	ipcd	1959:01-1998:12	5	industrial production: durable consumer goods (1992=100,sa)
6.	ipcn	1959:01-1998:12	5	industrial production: nondurable condsumer goods (1992=100,sa)
7.	ipe	1959:01-1998:12	5	industrial production: business equipment (1992=100,sa)
8.	ipi	1959:01-1998:12	5	industrial production: intermediate products (1992=100,sa)
9.	ipm	1959:01-1998:12	5	industrial production: materials (1992=100,sa)
10.	ipmd	1959:01-1998:12*	5	industrial production: durable goods materials (1992=100,sa)
11.	ipmnd	1959:01-1998:12	5	industrial production: nondurable goods materials (1992=100,sa)
12.	ipmfg	1959:01-1998:12	5	industrial production: manufacturing (1992=100,sa)
13.	ipd	1959:01-1998:12	5	industrial production: durable manufacturing (1992=100,sa)
14.	ipn	1959:01-1998:12	5	industrial production: nondurable manufacturing (1992=100,sa)
15.	ipmin	1959:01-1998:12	5	industrial production: mining (1992=100,sa)
16.	iput	1959:01-1998:12	5	industrial production: utilities (1992-=100,sa)
17.	ipx	1967:01-1998:12*	1	capacity util rate: total industry (% of capacity,sa)(frb)
18.	ipxmca	1959:01-1998:12	1	capacity util rate: manufacturing,total(% of capacity,sa)(frb)
19.	ipxdca	1967:01-1998:12*	1	capacity util rate: durable mfg (% of capacity,sa)(frb)
20.	ipxnca	1967:01-1998:12*	1	capacity util rate: nondurable mfg (% of capacity,sa)(frb)
21.	ipxmin	1967:01-1998:12*	1	capacity util rate: mining (% of capacity,sa)(frb)
22.	ipxut	1967:01-1998:12*	1	capacity util rate: utilities (% of capacity,sa)(frb)
23.	pmi	1959:01-1998:12	1	purchasing managers' index (sa)
24.	pmp	1959:01-1998:12	1	NAPM production index (percent)

25.	gmpyq	1959:01-1998:12*	5	personal income (chained) (series #52) (bil 92$,saar)
26.	gmyxpq	1959:01-1998:12	5	personal income less transfer payments (chained) (#51) (bil 92$,saar)

Employment and hours (EMP)

27.	lhel	1959:01-1998:12	5	index of help-wanted advertising in newspapers (1967=100;sa)
28.	lhelx	1959:01-1998:12	4	employment: ratio; help-wanted ads:no. unemployed clf
29.	lhem	1959:01-1998:12	5	civilian labor force: employed, total (thous.,sa)
30.	lhnag	1959:01-1998:12	5	civilian labor force: employed, nonagric.industries (thous.,sa)
31.	lhur	1959:01-1998:12	1	unemployment rate: all workers, 16 years & over (%,sa)
32.	lhu680	1959:01-1998:12	1	unemploy.by duration: average(mean)duration in weeks (sa)
33.	lhu5	1959:01-1998:12	1	unemploy.by duration: persons unempl.less than 5 wks (thous.,sa)
34.	lhu14	1959:01-1998:12	1	unemploy.by duration: persons unempl.5 to 14 wks (thous.,sa)
35.	lhu15	1959:01-1998:12	1	unemploy.by duration: persons unempl.15 wks + (thous.,sa)
36.	lhu26	1959:01-1998:12	1	unemploy.by duration: persons unempl.15 to 26 wks (thous.,sa)
37.	lpnag	1959:01-1998:12	5	employees on nonag. payrolls: total (thous.,sa)
38.	lp	1959:01-1998:12	5	employees on nonag payrolls: total, private (thous,sa)
39.	lpgd	1959:01-1998:12	5	employees on nonag. payrolls: goods-producing (thous.,sa)
40.	lpmi	1959:01-1998:12*	5	employees on nonag. payrolls: mining (thous.,sa)
41.	lpcc	1959:01-1998:12	5	employees on nonag. payrolls: contract construction (thous.,sa)
42.	lpem	1959:01-1998:12	5	employees on nonag. payrolls: manufacturing (thous.,sa)
43.	lped	1959:01-1998:12	5	employees on nonag. payrolls: durable goods (thous.,sa)
44.	lpen	1959:01-1998:12	5	employees on nonag. payrolls: nondurable goods (thous.,sa)
45.	lpsp	1959:01-1998:12	5	employees on nonag. payrolls: service-producing (thous.,sa)
46.	lptu	1959:01-1998:12*	5	employees on nonag. payrolls: trans. & public utilities (thous.,sa)
47.	lpt	1959:01-1998:12	5	employees on nonag. payrolls: wholesale & retail trade (thous.,sa)
48.	lpfr	1959:01-1998:12	5	employees on nonag. payrolls: finance,insur.&real estate (thous.,sa
49.	lps	1959:01-1998:12	5	employees on nonag. payrolls: services (thous.,sa)
50.	lpgov	1959:01-1998:12	5	employees on nonag. payrolls: government (thous.,sa)
51.	lw	1964:01-1998:12*	2	avg. weekly hrs. of prod. wkrs.: total private (sa)
52.	lphrm	1959:01-1998:12	1	avg. weekly hrs. of production wkrs.: manufacturing (sa)
53.	lpmosa	1959:01-1998:12	1	avg. weekly hrs. of prod. wkrs.: mfg.,overtime hrs. (sa)
54.	pmemp	1959:01-1998:12	1	NAPM employment index (percent)

Real retail, manufacturing and trade sales (RTS)

55.	msmtq	1959:01-1998:12	5	manufacturing & trade: total (mil of chained 1992 dollars)(sa)
56.	msmq	1959:01-1998:12	5	manufacturing & trade:manufacturing;total(mil of chained 1992 dollars)(sa)
57.	msdq	1959:01-1998:12	5	manufacturing & trade:mfg; durable goods (mil of chained 1992 dollars)(sa)
58.	msnq	1959:01-1998:12	5	manufact. & trade:mfg;nondurable goods (mil of chained 1992 dollars)(sa)
59.	wtq	1959:01-1998:12	5	merchant wholesalers: total (mil of chained 1992 dollars)(sa)
60.	wtdq	1959:01-1998:12	5	merchant wholesalers:durable goods total (mil of chained 1992 dollars)(sa)
61.	wtnq	1959:01-1998:12	5	merchant wholesalers:nondurable goods (mil of chained 1992 dollars)(sa)
62.	rtq	1959:01-1998:12	5	retail trade: total (mil of chained 1992 dollars)(sa)
63.	rtnq	1959:01-1998:12	5	retail trade:nondurable goods (mil of 1992 dollars)(sa)

Consumption (PCE)

64.	gmcq	1959:01-1998:12	5	personal consumption expend (chained)-total (bil 92$,saar)
65.	gmcdq	1959:01-1998:12	5	personal consumption expend (chained)-total durables (bil 92$,saar)
66.	gmcnq	1959:01-1998:12	5	personal consumption expend (chained)-nondurables (bil 92$,saar)
67.	gmcsq	1959:01-1998:12	5	personal consumption expend (chained)-services (bil 92$,saar)
68.	gmcanq	1959:01-1998:12	5	personal cons expend (chained)-new cars (bil 92$,saar)

Housing starts and sales (HSS)

69.	hsfr	1959:01-1998:12	4	housing starts:nonfarm(1947-58);total farm&nonfarm(1959-)(thous.,sa
70.	hsne	1959:01-1998:12	4	housing starts:northeast (thous.u.)s.a.
71.	hsmw	1959:01-1998:12	4	housing starts:midwest(thous.u.)s.a.
72.	hssou	1959:01-1998:12	4	housing starts:south (thous.u.)s.a.
73.	hswst	1959:01-1998:12	4	housing starts:west (thous.u.)s.a.
74.	hsbr	1959:01-1998:12	4	housing authorized: total new priv housing units (thous.,saar)
75.	hsbne	1960:01-1998:12*	4	houses authorized by build. permits:northeast(thou.u.)s.a

76.	hsbmw	1960:01-1998:12*	4	houses authorized by build. permits:midwest(thou.u.)s.a.
77.	hsbsou	1960:01-1998:12*	4	houses authorized by build. permits:south(thou.u.)s.a.
78.	hsbwst	1960:01-1998:12*	4	houses authorized by build. permits:west(thou.u.)s.a.
79.	hns	1963:01-1998:12*	4	new 1-family houses sold during month (thous,saar)
80.	hnsne	1973:01-1998:12*	4	one-family houses sold:northeast(thou.u.,s.a.)
81.	hnsmw	1973:01-1998:12*	4	one-family houses sold:midwest(thou.u.,s.a.)
82.	hnssou	1973:01-1998:12*	4	one-family houses sold:south(thou.u.,s.a.)
83.	hnswst	1973:01-1998:12*	4	one-family houses sold:west(thou.u.,s.a.)
84.	hnr	1963:01-1998:12*	4	new 1-family houses, month's supply @ current sales rate(ratio)
85.	hniv	1963:01-1998:12*	4	new 1-family houses for sale at end of month (thous,sa)
86.	hmob	1959:01-1998:12	4	mobile homes: manufacturers' shipments (thous.of units,saar)
87.	contc	1964:01-1998:12*	4	construct.put in place:total priv & public 1987$(mil$,saar)
88.	conpc	1964:01-1998:12*	4	construct.put in place:total private 1987$(mil$,saar)
89.	conqc	1964:01-1998:12*	4	construct.put in place:public construction 87$(mil$,saar)
90.	condo9	1959:01-1998:10*	4	construct.contracts: comm'l & indus.bldgs(mil.sq.ft.floor sp.;sa)

Real inventories and inventory-sales ratios (Inv)

91.	ivmtq	1959:01-1998:12	5	manufacturing & trade inventories:total (mil of chained 1992)(sa)
92.	ivmfgq	1959:01-1998:12	5	inventories, business, mfg (mil of chained 1992 dollars, sa)
93.	ivmfdq	1959:01-1998:12	5	inventories, business durables (mil of chained 1992 dollars, sa)
94.	ivmfnq	1959:01-1998:12	5	inventories, business, nondurables (mil of chained 1992 dollars, sa)
95.	ivwrq	1959:01-1998:12	5	manufacturing & trade inv:merchant wholesalers (mil of chained 1992 dollars)(s
96.	ivrrq	1959:01-1998:12	5	manufacturing & trade inv:retail trade (mil of chained 1992 dollars)(sa)
97.	ivsrq	1959:01-1998:12	2	ratio for mfg & trade: inventory/sales (chained 1992 dollars, sa)
98.	ivsrmq	1959:01-1998:12	2	ratio for mfg & trade:mfg;inventory/sales (87$)(s.a.)
99.	ivsrwq	1959:01-1998:12	2	ratio for mfg & trade:wholesaler;inventory/sales(87$)(s.a.)
100.	ivsrrq	1959:01-1998:12	2	ratio for mfg & trade:retail trade;inventory/sales(87$)(s.a.)
101.	pmnv	1959:01-1998:12	1	napm inventories index (percent)

Orders and unfilled orders (Ord)

102.	pmno	1959:01-1998:12	1	napm new orders index (percent)
103.	pmdel	1959:01-1998:12	1	napm vendor deliveries index (percent)
104.	mocmq	1959:01-1998:12	5	new orders (net)-consumer goods & materials, 1992 dollars (bci)
105.	mdoq	1959:01-1998:12	5	new orders, durable goods industries, 1992 dollars (bci)
106.	msondq	1959:01-1998:12	5	new orders, nondefense capital goods, in 1992 dollars (bci)
107.	mo	1959:01-1998:12	5	mfg new orders: all manufacturing industries, total (mil$,sa)
108.	mowu	1959:01-1998:12	5	mfg new orders: mfg industries with unfilled orders(mil$,sa)
109.	mdo	1959:01-1998:12	5	mfg new orders: durable goods industries, total (mil$,sa)
110.	mduwu	1959:01-1998:12	5	mfg new orders:durable goods indust with unfilled orders(mil$,sa)
111.	mno	1959:01-1998:12	5	mfg new orders: nondurable goods industries, total (mil$,sa)
112.	mnou	1959:01-1998:12	5	mfg new orders: nondurable gds ind.with unfilled orders(mil$,sa)
113.	mu	1959:01-1998:12	5	mfg unfilled orders: all manufacturing industries, total (mil$,sa)
114.	mdu	1959:01-1998:12	5	mfg unfilled orders: durable goods industries, total (mil$,sa)
115.	mnu	1959:01-1998:12	5	mfg unfilled orders: nondurable goods industries, total (mil$,sa)
116.	mpcon	1959:01-1998:12	5	contracts & orders for plant & equipment (bil$,sa)
117.	mpconq	1959:01-1998:12	5	contracts & orders for plant & equipment in 1992 dollars (bci)

Stock prices (SPr)

118.	fsncom	1959:01-1998:12	5	NYSE common stock price index: composite (12/31/65=50)
119.	fsnin	1966:01-1998:12*	5	NYSE common stock price index: industrial (12/31/65=50)
120.	fsntr	1966:01-1998:12*	5	NYSE common stock price index: transportation (12/31/65=50)
121.	fsnut	1966:01-1998:12*	5	NYSE common stock price index: utility (12/31/65=50)
122.	fsnfi	1966:01-1998:12*	5	NYSE common stock price index: finance (12/31/65=50)
123.	fspcom	1959:01-1998:12	5	S&P's common stock price index: composite (1941-43=10)
124.	fspin	1959:01-1998:12	5	S&P's common stock price index: industrials (1941-43=10)
125.	fspcap	1959:01-1998:12	5	S&P's common stock price index: capital goods (1941-43=10)
126.	fsptr	1970:01-1998:12*	5	S&P's common stock price index: transportation (1970=10)
127.	fsput	1959:01-1998:12	5	S&P's common stock price index: utilities (1941-43=10)
128.	fspfi	1970:01-1998:12*	5	S&P's common stock price index: financial (1970=10)

129.	fsdxp	1959:01-1998:12	1	S&P's composite common stock: dividend yield (% per annum)
130.	fspxe	1959:01-1998:12	1	S&P's composite common stock: price-earnings ratio (%,nsa)
131.	fsnvv3	1974:01-1997:07*	5	NYSE mkt composition:reptd share vol by size,5000+ shrs,%
Exchange rates (EXR)				
132.	exrus	1959:01-1998:12	5	United States effective exchange rate (merm)(index no.)
133.	exrger	1959:01-1998:12	5	foreign exchange rate: Germany (deutsche mark per U.S.$)
134.	exrsw	1959:01-1998:12	5	foreign exchange rate: Switzerland (swiss franc per U.S.$)
135.	exrjan	1959:01-1998:12	5	foreign exchange rate: Japan (yen per U.S.$)
136.	exruk	1959:01-1998:12*	5	foreign exchange rate: United Kingdom (cents per pound)
137.	exrcan	1959:01-1998:12	5	foreign exchange rate: Canada (canadian $ per U.S.$)
Interest rates (Int)				
138.	fyff	1959:01-1998:12*	2	interest rate: federal funds (effective) (% per annum,nsa)
139.	fycp90	1959:01-1998:12*	2	interest rate: 90 day commercial paper, (ac) (% per ann,nsa)
140.	fygm3	1959:01-1998:12*	2	interest rate: U.S.treasury bills,sec mkt,3-mo.(% per ann,nsa)
141.	fygm6	1959:01-1998:12*	2	interest rate: U.S.treasury bills,sec mkt,6-mo.(% per ann,nsa)
142.	fygtl	1959:01-1998:12*	2	interest rate: U.S.treasury const maturities,1-yr.(% per ann,nsa)
143.	fygt5	1959:01-1998:12	2	interest rate: U.S.treasury const maturities,5-yr.(% per ann,nsa)
144.	fygt10	1959:01-1998:12	2	interest rate: U.S.treasury const maturities,10-yr.(% per ann,nsa)
145.	fyaaac	1959:01-1998:12	2	bond yield: moody's aaa corporate (% per annum)
146.	fybaac	1959:01-1998:12	2	bond yield: moody's baa corporate (% per annum)
147.	fwafit	1973:01-1994:04*	1	weighted avg foreign interest rate(%,sa)
148.	fyfha	1959:01-1998:12	2	secondary market yields on fha mortgages (% per annum)
149.	sfycp	1959:01-1998:12	1	spread fycp - fyff
150.	sfygm3	1959:01-1998:12	1	spread fygm3 - fyff
151.	sfygm6	1959:01-1998:12	1	spread fygm6 - fyff
152.	sfygtl	1959:01-1998:12	1	spread fygtl - fyff
153.	sfygt5	1959:01-1998:12	1	spread fygt5 - fyff
154.	sfygt10	1959:01-1998:12	1	spread fygt10 - fyff
155.	sfyaaac	1959:01-1998:12	1	spread fyaaac - fyff
156.	sfybaac	1959:01-1998:12	1	spread fybaac - fyff
157.	sfyfha	1959:01-1998:12	1	spread fyfha - fyff
Money and credit quantity aggregates (Mon)				
158.	fm1	1959:01-1998:12	6	money stock: m1(curr,trav.cks,dem dep,other ck'able dep)(bil$,sa)
159.	fm2	1959:01-1998:12	6	money stock:m2(m1+o'nite rps,euro$,g/p&b/d mmmfs&sav&sm time dep(bil$,
160.	fm3	1959:01-1998:12	6	money stock: m3(m2+lg time dep,term rp's&inst only mmmfs)(bil$,sa)
161.	fml	1959:01-1998:09*	6	money stock:l(m3 + other liquid assets) (bil$,sa)
162.	fm2dq	1959:01-1998:12	5	money supply-m2 in 1992 dollars (bci)
163.	fmfba	1959:01-1998:12	6	monetary base, adj for reserve requirement changes(mil$,sa)
164.	fmrra	1959:01-1998:12	6	depository inst reserves:total,adj for reserve req chgs(mil$,sa)
165.	fmrnbc	1959:01-1998:12	6	depository inst reserves:nonborrow+ext cr,adj res req cgs(mil$,sa)
166.	fcls	1973:01-1998:12*	5	loans & sec @ all coml banks: total (bils,sa)
167.	fcsgv	1973:01-1998:12*	5	loans & sec @ all coml banks: U.S.govt securities (bil$,sa)
168.	fclre	1973:01-1998:12*	5	loans & sec @ all coml banks: real estate loans (bil$,sa)
169.	fclin	1973:01-1998:12*	5	loans & sec @ all coml banks: loans to individuals (bil$,sa)
170.	fclnbf	1973:01-1994:01*	5	loans & sec @ all coml banks: loans to nonbank fin inst(bil$,sa)
171.	fclnq	1959:01-1998:12*	5	commercial & industrial loans oustanding in 1992 dollars (bci)
172.	fclbmc	1959:01-1998:12*	1	wkly rp lg com'l banks:net change com'l & indus loans(bil$,saar)
173.	cci30m	1959:01-1995:09*	1	consumer instal.loans: delinquency rate,30 days & over, (%,sa)
174.	ccint	1975:01-1995:09*	1	net change in consumer instal cr: total (mil$,sa)
175.	ccinv	1975:01-1995:09*	1	net change in consumer instal cr: automobile (mil$,sa)
176.	ccinrv	1980:01-1995:09*	1	net change in consumer instal cr: revolving(mil$,sa)
Price indexes (Pri)				
177.	pmcp	1959:01-1998:12	1	napm commodity prices index (percent)
178.	pwfsa	1959:01-1998:12	6	producer price index: finished goods (82=100,sa)
179.	pwfcsa	1959:01-1998:12	6	producer price index:finished consumer goods (82=100,sa)

180.	pwimsa	1959:01-1998:12*	6	producer price index:intermed mat.supplies & components(82=100,sa)
181.	pwcmsa	1959:01-1998:12*	6	producer price index:crude materials (82=100,sa)
182.	pwfxsa	1967:01-1998:12*	6	producer price index: finished goods,excl. foods (82=100,sa)
183.	pw160a	1974:01-1998:12*	6	producer price index: crude materials less energy (82=100,sa)
184.	pw150a	1974:01-1998:12*	6	producer price index: crude nonfood mat less energy (82=100,sa)
185.	psm99q	1959:01-1998:12	6	index of sensitive materials prices (1990=100)(bci-99a)
186.	punew	1959:01-1998:12	6	cpi-u: all items (82-84=100,sa)
187.	pu81	1967:01-1998:12*	6	cpi-u: food & beverages (82-84=100,sa)
188.	puh	1967:01-1998:12*	6	cpi-u: housing (82-84=100,sa)
189.	pu83	1959:01-1998:12	6	cpi-u: apparel & upkeep (82-84=100,sa)
190.	pu84	1959:01-1998:12	6	cpi-u: transportation (82-84=100,sa)
191.	pu85	1959:01-1998:12	6	cpi-u: medical care (82-84=100,sa)
192.	puc	1959:01-1998:12	6	cpi-u: commodities (82-84=100,sa)
193.	pucd	1959:01-1998:12	6	cpi-u: durables (82-84=100,sa)
194.	pus	1959:01-1998:12	6	cpi-u: services (82-84=100,sa)
195.	puxf	1959:01-1998:12	6	cpi-u: all items less food (82-84=100,sa)
196.	puxhs	1959:01-1998:12	6	cpi-u: all items less shelter (82-84=100,sa)
197.	puxm	1959:01-1998:12	6	cpi-u: all items less midical care (82-84=100,sa)
198.	pcgold	1975:01-1998:12*	6	commodities price:gold,london noon fix,avg of daily rate,$ per oz
199.	gmdc	1959:01-1998:12	6	pce,impl pr defl:pce (1987=100)
200.	gmdcd	1959:01-1998:12	6	pce,impl pr defl:pce; durables (1987=100)
201.	gmdcn	1959:01-1998:12	6	pce,impl pr defl:pce; nondurables (1987=100)
202.	gmdcs	1959:01-1998:12	6	pce,impl pr defl:pce; services (1987=100)

Average hourly earnings (AHE)

203.	leh	1964:01-1998:12*	6	avg hr earnings of prod wkrs: total private nonagric ($,sa)
204.	lehcc	1959:01-1998:12	6	avg hr earnings of constr wkrs: construction ($,sa)
205.	lehm	1959:01-1998:12	6	avg hr earnings of prod wkrs: manufacturing ($,sa)
206.	lehtu	1964:01-1998:12*	6	avg hr earnings of nonsupv wkrs: trans & public util($,sa)
207.	lehtt	1964:01-1998:12*	6	avg hr earnings of prod wkrs:wholesale & retail trade(sa)
208.	lehfr	1964:01-1998:12*	6	avg hr earnings of nonsupv wkrs: finance,insur,real est($,sa)
209.	lehs	1964:01-1998:12*	6	avg hr earnings of nonsupv wkrs: services ($,sa)

Miscellaneous (Oth)

210.	fste	1986:01-1998:12*	5	U.S.mdse exports: total exports(f.a.s. value)(mil.$,s.a.)
211.	fstm	1986:01-1998:12*	5	U.S.mdse imports: general imports(c.i.f. value)(mil.$,s.a.)
212.	ftmd	1986:01-1998:12*	5	U.S.mdse imports: general imports (customs value)(mil$,s.a.)
213.	fstb	1986:01-1998:12*	2	U.S.mdse trade balance:exports less imports(fas/cif)(mil$,s.a.)
214.	ftb	1986:01-1998:12*	2	U.S.mdse trade balance:exp.(fas) less imp.(custom)(mil$,s.a.)
215.	hhsntn	1959:01-1998:12	1	u. of mich. index of consumer expectations(bcd-83)

References

Bekker, P., P. Dobbelstein, and T. Wansbeek (1996), "The APT Model as Reduced-Rank Regression," *Journal of Business and Economic Statistics*, 14, 199-202.

Campbell, J.Y., A.W. Lo and A. C. MacKinley (1996), *The Econometrics and Financial Markets*, Princeton University Press, Princeton, New Jersey.

Chamberlain, G. and M. Rothschild (1983), "Arbitrage Factor Stucture, and Mean-Variance Analysis of Large Asset Markets," *Econometrica*, Vol. 51, No. 5.

Congressional Budget Office (1996), *The Economic and Budget Outlook: Fiscal Years 1997-2006*, May.

Connor, G. and R.A. Korajczyk (1986), "Performance Measurement with the Arbitrage Pricing Theory," *Journal of Financial Economics*, 15, 373-394

Connor, G. and R.A. Korajczyk (1988), "Risk and Return in an Equilibrium APT: Application of a New Test Methodology," *Journal of Financial Economics* 21, 255-289.

Connor, G. and R.A. Korajczyk (1993), "A Test for the Number of Factors in an Approximate Factor Model," *Journal of Finance*, Vol. XLVIII, No. 4.

Engle, R.F., and M.W. Watson (1981), "A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates," *Journal of the American Statistical Association 76*, no. 376, 774-781.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (1998), "The Generalized Dynamnic Factor Model: Identification and Estimation," manuscript, Université Libre de Bruxelles.

Forni, M. and L. Reichlin (1996), "Dynamic Common Factors in Large Cross-Sections," *Empirical Economics*, 21, 27-42.

Forni, M. and L. Reichlin (1997), "National Policies and Local Economies: Europe and the United States," CEPR Discussion paper no. 1632.

Forni, M. and L. Reichlin (1998), "Lets Get Real: A Dynamic Factor Analytical Approach to Disaggregated Business Cycle," *Review of Economic Studies*, 65, 453-474.

Fuhrer, Jeffrey C. (1995), "The Phillips Curve is Alive and Well," *New England Economic Review of the Federal Reserve Bank of Boston*, March/April 1995, 41-56.

Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series," in D.J. Aigner and A.S. Goldberger (eds.), *Latent Variables in Socio-Economic Models*, North-Holland, Amsterdam, Ch. 19.

Geweke, J. and G. Zhou (1996), "Measuring the Pricing Error of the Arbitrage Pricing Theory," *Review of Financial Studies*, 9, 557-587.

Mei, J. (1993), "A Semiautoregression Approach to the Arbitrage Pricing Theory," *Journal of Finance*48, 599-620.

Gordon, Robert J. (1982), "Price Inertia and Ineffectiveness in the United States," *Journal of Political Economy 90*, 1087-1117.

Gordon, Robert J. (1997), "The Time-Varying NAIRU and its Implications for Economic Policy," *Journal of Economic Perspectives*, 11-32.

Quah, D. and T.J. Sargent (1993), "A Dynamic Index Model for Large Cross Sections," ch. 7 in J.H. Stock and M.W. Watson (eds.), *Business Cycles, Indicators, and Forecasting*, University of Chicago Press for the NBER, 1993, 285-306.

Reiss, R.D. (1989), *Approximate Distributions of Order Statistics*. Springer-Verlag: New York.

Sargent, T. (1989), "Two Models of Measurements and the Investment Accelerator," *Journal of Political Economy*, 97, 251-287.

Sargent, T.J, and C.A. Sims (1977), "Business Cycle Modeling without Pretending to have Too Much a-priori Economic Theory," in C. Sims et al. (eds.), *New Methods in Business Cycle Research*, Minneapolis: Federal Reserve Bank of Minneapolis.

Schneeweiss, H. and H. Mathes (1995), "Factor Analysis and Principal Components," *Journal of Multivariate Analysis* 55, 105-124.

Staiger, D., J.H. Stock and M.W. Watson, "The NAIRU, Unemployment, and Monetary Policy," *Journal of Economic Perspectives* 11 (Winter 1997), 33-51.

Stock, J.H. and M.W. Watson (1989), "New Indexes of Coincident and Leading Economic Indicators," *NBER Macroeconomics Annual*, 351-393.

Stock, J.H. and M.W. Watson (1991), "A Probability Model of the Coincident Economic Indicators," in K. Lahiri and G.H. Moore (eds.), *Leading Economic Indicatorss: New Approaches and Forecasting Records*, ch. 4. New York: Cambridge University Press, 63-85.

Stock, J.H. and M.W. Watson (1996), "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics*, 14, 11-30.

Stock, J.H. and M.W. Watson (1998a), "Median Unbiased Estimation of Coefficient Variance in a Time Varying Parameter Model," *Journal of the American Statistical Association*, 93, 349-358.

Stock, J.H. and M.W. Watson (1998b), "Forecasting Inflation," NBER Working Paper no. 7023; forthcoming, *Journal of Monetary Economics*.

Tootell, Geoffrey M.B. (1994), "Restructuring, the NAIRU, and the Phillips Curve," *New England Economic Review of the Federal Reserve Bank of Boston*, September/October, 31-44.

West, K.D. (1996), "Asymptotic Inference about Predictive Ability," *Econometrica*, 64, 1067-1084.

Zhou, G. (1997), "Security Return Factors as Linear Combinations of Economic Variables," manuscript, Washington University in St. Louis.

Table 1

Monte Carlo Results

T	N	$\bar{r}$	k	q	α	a	b	c	$R^2_{\hat{F},F^0}$	$S^2_{\hat{Y},Y^0}$: k=r	BIC	AIC	(3.5), ω = 0.001	0.005	0.010
									A. Static factor models						
25	50	5	5	0	0.0	0.0	0.0	0	0.88	0.70	0.57	0.67	0.71	0.70	0.61
25	100	5	5	0	0.0	0.0	0.0	0	0.92	0.81	0.68	0.76	0.81	0.80	0.76
25	250	5	5	0	0.0	0.0	0.0	0	0.95	0.88	0.74	0.83	0.88	0.88	0.86
25	500	5	5	0	0.0	0.0	0.0	0	0.96	0.90	0.74	0.85	0.90	0.90	0.89
50	50	5	5	0	0.0	0.0	0.0	0	0.89	0.80	0.76	0.80	0.80	0.55	0.15
50	100	5	5	0	0.0	0.0	0.0	0	0.94	0.89	0.84	0.89	0.89	0.78	0.40
50	250	5	5	0	0.0	0.0	0.0	0	0.96	0.94	0.90	0.93	0.94	0.92	0.76
50	500	5	5	0	0.0	0.0	0.0	0	0.97	0.96	0.91	0.95	0.96	0.95	0.89
50	50	10	10	0	0.0	0.0	0.0	0	0.82	0.54	0.39	0.53	0.55	0.17	0.04
50	100	10	10	0	0.0	0.0	0.0	0	0.89	0.72	0.52	0.70	0.72	0.36	0.12
50	250	10	10	0	0.0	0.0	0.0	0	0.95	0.87	0.66	0.84	0.87	0.75	0.34
50	500	10	10	0	0.0	0.0	0.0	0	0.97	0.92	0.71	0.88	0.92	0.86	0.62
100	250	5	5	0	0.0	0.0	0.0	0	0.97	0.96	0.95	0.96	0.96	0.63	0.17
100	500	5	5	0	0.0	0.0	0.0	0	0.98	0.97	0.96	0.97	0.97	0.87	0.37
100	250	5	10	0	0.0	0.0	0.0	0	0.95	0.96	0.95	0.94	0.95	0.56	0.17
100	500	5	10	0	0.0	0.0	0.0	0	0.97	0.98	0.96	0.95	0.96	0.86	0.37
100	100	10	10	0	0.0	0.0	0.0	0	0.90	0.83	0.75	0.83	0.73	0.06	0.01
100	250	10	10	0	0.0	0.0	0.0	0	0.95	0.93	0.87	0.92	0.91	0.16	0.05
100	500	10	10	0	0.0	0.0	0.0	0	0.97	0.95	0.89	0.95	0.94	0.34	0.09
100	100	10	15	0	0.0	0.0	0.0	0	0.86	0.83	0.76	0.81	0.75	0.05	0.00
100	250	10	15	0	0.0	0.0	0.0	0	0.93	0.92	0.85	0.90	0.90	0.15	0.03
100	500	10	15	0	0.0	0.0	0.0	0	0.95	0.96	0.89	0.93	0.93	0.37	0.09

Table 1, continued

T	N	$\bar{r}$	k	q	α	a	b	c	$R^2_{\hat{F},F^0}$	$S^2_{\hat{Y},Y^0}$: k=r	BIC	AIC	(3.5), ω = 0.001	0.005	0.010
B. Correlated errors															
100	250	5	5	0	0.0	0.5	0.0	0	0.97	0.96	0.94	0.96	0.95	0.59	0.17
100	500	5	5	0	0.0	0.5	0.0	0	0.98	0.97	0.96	0.97	0.97	0.84	0.32
100	250	5	10	0	0.0	0.5	0.0	0	0.93	0.96	0.95	0.95	0.96	0.61	0.16
100	500	5	10	0	0.0	0.5	0.0	0	0.95	0.97	0.96	0.96	0.96	0.87	0.35
100	250	5	5	0	0.0	0.9	0.0	0	0.86	0.84	0.81	0.83	0.83	0.48	0.14
200	250	5	5	0	0.0	0.9	0.0	0	0.95	0.96	0.96	0.96	0.93	0.08	0.00
100	250	5	5	0	0.0	0.0	1.0	0	0.97	0.96	0.95	0.96	0.96	0.64	0.18
100	500	5	5	0	0.0	0.0	1.0	0	0.98	0.98	0.96	0.97	0.97	0.87	0.36
100	250	5	10	0	0.0	0.0	1.0	0	0.94	0.96	0.94	0.94	0.95	0.60	0.17
100	500	5	10	0	0.0	0.0	1.0	0	0.96	0.98	0.96	0.95	0.96	0.86	0.38
C. Dynamic factor models															
100	250	5	10	1	0.0	0.0	0.0	0	0.95	0.92	0.85	0.91	0.90	0.18	0.04
100	500	5	10	1	0.0	0.0	0.0	0	0.97	0.95	0.88	0.94	0.94	0.36	0.09
100	250	5	15	1	0.0	0.0	0.0	0	0.93	0.91	0.84	0.89	0.89	0.17	0.05
100	500	5	15	1	0.0	0.0	0.0	0	0.95	0.95	0.88	0.92	0.92	0.35	0.09
100	250	5	5	0	0.9	0.0	0.0	0	0.92	0.78	0.75	0.78	0.77	0.44	0.22
100	500	5	5	0	0.9	0.0	0.0	0	0.93	0.79	0.76	0.79	0.79	0.60	0.39
100	250	5	10	0	0.9	0.0	0.0	0	0.90	0.77	0.73	0.75	0.75	0.45	0.24
100	500	5	10	0	0.9	0.0	0.0	0	0.91	0.79	0.76	0.78	0.78	0.60	0.36
100	250	5	5	0	0.9	0.5	1.0	0	0.91	0.76	0.72	0.76	0.75	0.43	0.21
100	500	5	5	0	0.9	0.5	1.0	0	0.92	0.78	0.76	0.78	0.78	0.58	0.34
100	250	5	10	0	0.9	0.5	1.0	0	0.87	0.79	0.74	0.77	0.77	0.48	0.24
100	500	5	10	0	0.9	0.5	1.0	0	0.89	0.80	0.76	0.79	0.79	0.61	0.35

Table 1, continued

T	N	$\bar{r}$	k	q	α	a	b	c	$R^2_{\hat{F},F^0}$	$S^2_{\hat{y},\hat{y}^0}$ k=r	BIC	AIC	(3.5), ω = 0.001	0.005	0.010
									D. Time-varying factor loadings						
100	250	5	5	0	0.0	0.0	0.0	5	0.93	0.93	0.92	0.93	0.92	0.55	0.17
100	500	5	5	0	0.0	0.0	0.0	5	0.94	0.94	0.92	0.93	0.93	0.81	0.34
100	250	5	5	0	0.0	0.0	0.0	10	0.87	0.89	0.88	0.89	0.89	0.47	0.11
100	500	5	5	0	0.0	0.0	0.0	10	0.87	0.89	0.88	0.89	0.89	0.73	0.27
100	250	5	10	0	0.0	0.0	0.0	5	0.91	0.93	0.91	0.87	0.90	0.57	0.16
100	500	5	10	0	0.0	0.0	0.0	5	0.92	0.94	0.91	0.86	0.87	0.81	0.33
100	250	5	10	0	0.0	0.0	0.0	10	0.85	0.89	0.85	0.80	0.82	0.44	0.12
100	500	5	10	0	0.0	0.0	0.0	10	0.85	0.90	0.85	0.81	0.81	0.70	0.25
100	250	5	5	0	0.9	0.5	1.0	5	0.89	0.76	0.73	0.76	0.75	0.43	0.21
100	500	5	5	0	0.9	0.5	1.0	5	0.89	0.76	0.72	0.76	0.76	0.58	0.35
100	250	5	5	0	0.9	0.5	1.0	10	0.85	0.74	0.70	0.74	0.72	0.43	0.22
100	500	5	5	0	0.9	0.5	1.0	10	0.85	0.73	0.70	0.73	0.73	0.54	0.32
100	250	5	10	0	0.9	0.5	1.0	5	0.86	0.76	0.73	0.74	0.75	0.45	0.24
100	500	5	10	0	0.9	0.5	1.0	5	0.87	0.79	0.76	0.76	0.77	0.59	0.36
100	250	5	10	0	0.9	0.5	1.0	10	0.83	0.75	0.72	0.69	0.72	0.43	0.25
100	500	5	10	0	0.9	0.5	1.0	10	0.84	0.73	0.70	0.66	0.67	0.55	0.33

Notes: Based on 2000 Monte Carlo draws using the design described in section 4 of the text. The statistic $R^2_{\hat{F},F^0}$ is the trace R^2 of a regression of the estimated factors on the true factor, and $S^2_{\hat{y},\hat{y}^0}$ is measure of the closeness of the forecast of y_{T+1} based on the estimated factors to factor to the forecast based on the true factor; see the text for definitions. The $S^2_{\hat{y},\hat{y}^0}$ values are reported using the correct number of factors (when k>r, the first r factors are used) and for $\hat{r}$ estimated using an information criterion, where $1 \le \hat{r} \le k$.

Table 2

Simulated out-of-sample forecasting results: Real variables, 12 month horizon

Forecast Method	industrial production Rel. MSE	industrial production $\hat{\alpha}$	personal income Rel. MSE	personal income $\hat{\alpha}$	mfg & trade sales Rel. MSE	mfg & trade sales $\hat{\alpha}$	nonag. employment Rel. MSE	nonag. employment $\hat{\alpha}$
Benchmark models								
AR	1.00		1.00		1.00		1.00	
LI	0.86 (0.27)	0.57 (0.13)	0.97 (0.21)	0.52 (0.15)	0.82 (0.25)	0.63 (0.17)	0.89 (0.23)	0.56 (0.14)
VAR	0.97 (0.07)	0.75 (0.68)	0.98 (0.05)	0.68 (0.34)	0.98 (0.04)	0.73 (0.58)	1.05 (0.09)	0.22 (0.41)
Full data set (N=215)								
DI-AR, Lag	0.57 (0.27)	0.76 (0.13)	0.77 (0.14)	0.76 (0.13)	0.48 (0.25)	0.99 (0.15)	0.91 (0.13)	0.63 (0.18)
DI-AR	0.63 (0.25)	0.71 (0.12)	0.86 (0.16)	0.61 (0.12)	0.57 (0.24)	0.84 (0.18)	0.99 (0.31)	0.51 (0.20)
DI	0.52 (0.26)	0.88 (0.17)	0.86 (0.16)	0.61 (0.12)	0.56 (0.23)	0.94 (0.20)	0.92 (0.26)	0.55 (0.20)
Balanced panel (N=149)								
DI-AR, Lag	0.67 (0.25)	0.70 (0.13)	0.82 (0.15)	0.70 (0.13)	0.56 (0.23)	0.91 (0.16)	0.88 (0.14)	0.68 (0.18)
DI-AR	0.67 (0.25)	0.70 (0.12)	0.92 (0.14)	0.57 (0.12)	0.61 (0.23)	0.80 (0.17)	0.88 (0.22)	0.58 (0.17)
DI	0.59 (0.25)	0.81 (0.17)	0.92 (0.14)	0.57 (0.12)	0.57 (0.23)	0.91 (0.18)	0.84 (0.21)	0.62 (0.16)
Stacked balance panel								
DI-AR	0.65 (0.25)	0.70 (0.12)	0.93 (0.15)	0.56 (0.12)	0.61 (0.22)	0.89 (0.19)	1.02 (0.30)	0.49 (0.14)
DI	0.62 (0.25)	0.81 (0.18)	0.93 (0.15)	0.56 (0.12)	0.66 (0.21)	0.85 (0.20)	0.95 (0.24)	0.53 (0.14)
Full data set; m=1, p=BIC, k fixed								
DI-AR, k=1	1.06 (0.11)	0.27 (0.34)	1.03 (0.08)	0.34 (0.41)	0.98 (0.06)	0.63 (0.46)	1.01 (0.09)	0.49 (0.24)
DI-AR, k=2	0.63 (0.25)	0.76 (0.14)	0.78 (0.14)	0.77 (0.14)	0.53 (0.24)	0.93 (0.15)	0.77 (0.13)	0.82 (0.15)
DI-AR, k=3	0.56 (0.26)	0.84 (0.14)	0.77 (0.15)	0.77 (0.13)	0.52 (0.23)	0.99 (0.16)	0.84 (0.14)	0.75 (0.20)
DI-AR, k=4	0.54 (0.26)	0.85 (0.14)	0.76 (0.15)	0.78 (0.14)	0.51 (0.23)	1.01 (0.16)	0.83 (0.15)	0.73 (0.19)
Full data set; m=1, p=0, k fixed								
DI, k=1	1.03 (0.07)	0.30 (0.49)	1.01 (0.09)	0.46 (0.34)	0.98 (0.05)	0.67 (0.49)	1.01 (0.09)	0.48 (0.24)
DI, k=2	0.55 (0.25)	0.89 (0.15)	0.78 (0.14)	0.76 (0.13)	0.57 (0.24)	0.95 (0.17)	0.78 (0.13)	0.83 (0.16)
DI, k=3	0.51 (0.25)	1.00 (0.16)	0.77 (0.15)	0.77 (0.13)	0.60 (0.21)	1.02 (0.19)	0.84 (0.14)	0.76 (0.19)
DI, k=4	0.49 (0.25)	1.00 (0.16)	0.76 (0.15)	0.78 (0.14)	0.59 (0.22)	1.03 (0.20)	0.82 (0.15)	0.75 (0.18)
RMSE, AR Model	0.049		0.027		0.045		0.017	

Notes to table 2: For each variable/forecast method combination, the first entry is 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 BIC ("AR" in the table), and the HAC standard error of that ratio appears next in parentheses. The second pair of entries are the estimated forecast combining coefficient $\hat{\alpha}$ from regression (5.5) and its HAC standard error. All forecasts are simulated out of sample. The LI (leading indicator), Phillips curve (for inflation series), DI, DI-AR, and DI-AR,Lag forecasts were computed with BIC lag and/or variable selection, see the text for details. The method for computing the factors (full data set, balanced panel, stacked balanced panel) are indicated in italics above the associated panel of results. The final line presents the root MSE for the AR model in native (decimal growth rate) units at an annual rate.

Table 3
Simulated out-of-sample forecasting results: Real variables, 6 and 24 month horizons

Forecast Method	industrial production Rel. MSE	$\hat{\alpha}$	personal income Rel. MSE	$\hat{\alpha}$	mfg & trade sales Rel. MSE	$\hat{\alpha}$	nonag. employment Rel. MSE	$\hat{\alpha}$
A. Horizon = 6 months								
Benchmark models								
AR	1.00		1.00		1.00		1.00	
LI	0.70 (0.25)	0.68 (0.13)	0.83 (0.15)	0.64 (0.11)	0.77 (0.19)	0.68 (0.14)	0.75 (0.19)	0.67 (0.12)
VAR	1.01 (0.05)	0.43 (0.39)	0.99 (0.03)	0.63 (0.43)	0.99 (0.04)	0.64 (0.45)	1.06 (0.07)	0.12 (0.34)
Full data set (N=215)								
DI-AR, Lag	0.69 (0.25)	0.69 (0.14)	0.77 (0.12)	0.86 (0.15)	0.63 (0.18)	0.89 (0.17)	0.94 (0.16)	0.56 (0.18)
DI-AR	0.77 (0.30)	0.62 (0.16)	0.81 (0.16)	0.66 (0.13)	0.70 (0.20)	0.76 (0.17)	1.02 (0.32)	0.49 (0.19)
DI	0.74 (0.25)	0.68 (0.17)	0.81 (0.16)	0.65 (0.13)	0.67 (0.20)	0.79 (0.18)	0.96 (0.28)	0.52 (0.19)
Balanced panel (N=149)								
DI-AR, Lag	0.73 (0.25)	0.68 (0.16)	0.79 (0.13)	0.78 (0.13)	0.66 (0.17)	0.87 (0.17)	0.93 (0.17)	0.58 (0.21)
DI-AR	0.78 (0.28)	0.62 (0.16)	0.81 (0.15)	0.66 (0.11)	0.76 (0.19)	0.70 (0.17)	0.97 (0.28)	0.52 (0.19)
DI	0.73 (0.24)	0.69 (0.15)	0.81 (0.15)	0.66 (0.11)	0.68 (0.19)	0.81 (0.17)	0.95 (0.26)	0.53 (0.18)
Full data set; m=1, p=BIC, k fixed								
DI-AR, k=1	0.97 (0.15)	0.58 (0.33)	0.91 (0.07)	0.80 (0.23)	0.99 (0.11)	0.52 (0.29)	0.94 (0.12)	0.60 (0.19)
DI-AR, k=2	0.67 (0.22)	0.77 (0.15)	0.76 (0.11)	0.90 (0.14)	0.64 (0.18)	0.86 (0.16)	0.84 (0.13)	0.71 (0.16)
DI-AR, k=3	0.64 (0.23)	0.81 (0.15)	0.75 (0.12)	0.89 (0.14)	0.64 (0.18)	0.88 (0.17)	0.88 (0.14)	0.66 (0.17)
DI-AR, k=4	0.64 (0.23)	0.80 (0.15)	0.74 (0.13)	0.87 (0.14)	0.63 (0.18)	0.87 (0.15)	0.91 (0.16)	0.60 (0.18)
RMSE, AR Model	0.030		0.016		0.028		0.008	

Table 3, continued

Forecast Method	industrial production Rel. MSE	industrial production $\hat{\alpha}$	personal income Rel. MSE	personal income $\hat{\alpha}$	mfg & trade sales Rel. MSE	mfg & trade sales $\hat{\alpha}$	nonag. employment Rel. MSE	nonag. employment $\hat{\alpha}$
B. Horizon = 24 months								
Benchmark models								
AR	1.00		1.00		1.00		1.00	
LI	1.09 (0.28)	0.45 (0.14)	1.29 (0.31)	0.30 (0.20)	1.08 (0.21)	0.45 (0.14)	1.07 (0.31)	0.47 (0.15)
VAR	1.01 (0.10)	0.44 (0.48)	0.98 (0.06)	0.63 (0.34)	1.03 (0.06)	0.13 (0.85)	1.06 (0.13)	0.35 (0.31)
Full data set (N=215)								
DI-AR, Lag	0.57 (0.24)	0.88 (0.13)	0.70 (0.20)	0.94 (0.23)	0.66 (0.18)	0.95 (0.18)	0.82 (0.15)	0.88 (0.26)
DI-AR	0.59 (0.25)	0.88 (0.15)	0.76 (0.22)	0.80 (0.26)	0.70 (0.20)	0.89 (0.19)	0.74 (0.19)	0.97 (0.24)
DI	0.55 (0.26)	0.91 (0.14)	0.76 (0.22)	0.80 (0.25)	0.70 (0.20)	0.89 (0.19)	0.74 (0.19)	0.97 (0.24)
Balanced panel (N=149)								
DI-AR, Lag	0.57 (0.25)	0.87 (0.14)	0.76 (0.19)	0.86 (0.23)	0.64 (0.20)	0.94 (0.18)	0.74 (0.17)	1.06 (0.25)
DI-AR	0.58 (0.25)	0.87 (0.14)	0.83 (0.20)	0.74 (0.24)	0.67 (0.19)	0.93 (0.18)	0.76 (0.18)	0.94 (0.25)
DI	0.58 (0.25)	0.87 (0.14)	0.83 (0.20)	0.74 (0.24)	0.67 (0.20)	0.94 (0.19)	0.75 (0.18)	0.94 (0.24)
Full data set; m=1, p=BIC, k fixed								
DI-AR, k=1	1.12 (0.19)	0.10 (0.46)	1.07 (0.09)	-0.81 (1.00)	0.97 (0.04)	0.90 (0.62)	1.03 (0.07)	0.33 (0.46)
DI-AR, k=2	0.76 (0.19)	0.68 (0.11)	0.88 (0.13)	0.68 (0.17)	0.65 (0.20)	0.87 (0.14)	0.72 (0.16)	0.99 (0.17)
DI-AR, k=3	0.58 (0.24)	0.89 (0.13)	0.72 (0.19)	0.90 (0.18)	0.70 (0.17)	0.89 (0.14)	0.79 (0.16)	0.95 (0.24)
DI-AR, k=4	0.56 (0.24)	0.90 (0.14)	0.70 (0.20)	0.93 (0.23)	0.67 (0.18)	0.95 (0.18)	0.78 (0.16)	0.96 (0.24)
RMSE, AR Model	0.075		0.046		0.070		0.031	

Notes: See the notes to table 2.

Table 4
Simulated out-of-sample forecasting results: Price inflation, 12 month horizon

Forecast Method	CPI Rel. MSE	CPI $\hat{\alpha}$	consumption deflator Rel. MSE	consumption deflator $\hat{\alpha}$	CPI exc. food&energy Rel. MSE	CPI exc. food&energy $\hat{\alpha}$	producer price index Rel. MSE	producer price index $\hat{\alpha}$
Benchmark models								
AR	1.00		1.00		1.00		1.00	
LI	0.79 (0.15)	0.76 (0.15)	0.95 (0.12)	0.58 (0.17)	1.00 (0.16)	0.50 (0.21)	0.82 (0.15)	0.75 (0.19)
Phillips Curve	0.82 (0.13)	0.95 (0.20)	0.92 (0.10)	0.72 (0.23)	0.79 (0.18)	0.80 (0.22)	0.87 (0.14)	0.96 (0.30)
VAR	0.91 (0.09)	0.74 (0.20)	1.02 (0.06)	0.45 (0.20)	0.99 (0.05)	0.56 (0.21)	1.29 (0.14)	0.25 (0.12)
Full data set (N=215)								
DI-AR, Lag	0.72 (0.14)	0.91 (0.14)	0.90 (0.09)	0.65 (0.13)	0.84 (0.15)	0.76 (0.20)	0.83 (0.13)	0.78 (0.21)
DI-AR	0.71 (0.16)	0.83 (0.13)	0.90 (0.10)	0.62 (0.13)	0.85 (0.15)	0.74 (0.20)	0.82 (0.14)	0.75 (0.20)
DI	1.30 (0.16)	0.34 (0.08)	1.40 (0.16)	0.25 (0.08)	1.55 (0.31)	0.24 (0.06)	2.40 (0.88)	0.13 (0.07)
Balanced panel (N=149)								
DI-AR, Lag	0.70 (0.14)	0.94 (0.12)	0.90 (0.08)	0.67 (0.15)	0.84 (0.15)	0.77 (0.21)	0.86 (0.11)	0.77 (0.21)
DI-AR	0.69 (0.15)	0.88 (0.13)	0.87 (0.10)	0.66 (0.12)	0.85 (0.15)	0.73 (0.20)	0.85 (0.14)	0.71 (0.19)
DI	1.30 (0.16)	0.32 (0.08)	1.34 (0.13)	0.26 (0.09)	1.57 (0.33)	0.20 (0.07)	2.44 (0.87)	0.14 (0.06)
Stacked balance panel								
DI-AR	0.73 (0.15)	0.82 (0.12)	0.87 (0.09)	0.65 (0.12)	0.85 (0.15)	0.77 (0.21)	0.81 (0.14)	0.75 (0.20)
DI	1.54 (0.31)	0.28 (0.08)	1.51 (0.18)	0.25 (0.08)	1.55 (0.32)	0.23 (0.06)	3.06 (1.89)	0.11 (0.06)
Full data set; m=1, p=BIC, k fixed								
DI-AR, k=1	0.64 (0.15)	1.14 (0.14)	0.77 (0.12)	0.96 (0.16)	0.71 (0.17)	1.25 (0.23)	0.76 (0.16)	0.95 (0.24)
DI-AR, k=2	0.67 (0.14)	1.07 (0.13)	0.83 (0.09)	0.83 (0.14)	0.72 (0.17)	0.97 (0.19)	0.77 (0.15)	0.93 (0.23)
DI-AR, k=3	0.76 (0.13)	0.91 (0.15)	0.94 (0.07)	0.61 (0.14)	0.86 (0.14)	0.73 (0.20)	0.86 (0.11)	0.78 (0.21)
DI-AR, k=4	0.74 (0.14)	0.89 (0.15)	0.91 (0.09)	0.64 (0.14)	0.87 (0.15)	0.72 (0.21)	0.82 (0.13)	0.79 (0.21)
Full data set; m=1, p=0, k fixed								
DI, k=1	1.60 (0.34)	0.25 (0.07)	1.56 (0.20)	0.22 (0.09)	1.55 (0.31)	0.23 (0.06)	2.76 (1.61)	0.12 (0.07)
DI, k=2	1.56 (0.31)	0.26 (0.07)	1.58 (0.20)	0.21 (0.08)	1.62 (0.39)	0.22 (0.07)	2.72 (1.56)	0.13 (0.07)
DI, k=3	1.57 (0.32)	0.24 (0.08)	1.60 (0.20)	0.17 (0.08)	1.69 (0.43)	0.18 (0.07)	2.68 (1.49)	0.13 (0.07)
DI, k=4	1.56 (0.25)	0.25 (0.07)	1.56 (0.19)	0.21 (0.08)	1.67 (0.40)	0.19 (0.07)	2.55 (0.99)	0.16 (0.06)
RMSE, AR Model	0.021		0.015		0.019		0.033	

Notes: see the notes to table 2.

Table 5, continued

Forecast Method	CPI Rel. MSE	CPI $\hat{\alpha}$	consumption deflator Rel. MSE	consumption deflator $\hat{\alpha}$	CPI exc. food&energy Rel. MSE	CPI exc. food&energy $\hat{\alpha}$	producer price index Rel. MSE	producer price index $\hat{\alpha}$
B. Horizon = 24 months								
Benchmark models								
AR	1.00		1.00		1.00		1.00	
LI	0.70 (0.21)	0.76 (0.12)	0.70 (0.20)	0.78 (0.11)	0.99 (0.29)	0.51 (0.25)	0.65 (0.22)	0.84 (0.19)
Phillips Curve	0.84 (0.12)	0.77 (0.08)	0.81 (0.15)	0.80 (0.09)	0.72 (0.21)	0.93 (0.19)	0.77 (0.19)	1.00 (0.06)
VAR	0.92 (0.08)	0.80 (0.22)	0.98 (0.06)	0.57 (0.18)	1.00 (0.06)	0.49 (0.34)	1.18 (0.12)	0.29 (0.10)
Full data set (N=215)								
DI-AR, Lag	0.74 (0.23)	0.74 (0.18)	0.75 (0.16)	0.79 (0.13)	0.92 (0.26)	0.58 (0.28)	0.82 (0.14)	0.68 (0.12)
DI-AR	0.75 (0.25)	0.67 (0.16)	0.71 (0.21)	0.73 (0.12)	0.96 (0.33)	0.53 (0.27)	0.77 (0.17)	0.68 (0.13)
DI	1.18 (0.22)	0.40 (0.12)	1.21 (0.18)	0.38 (0.12)	1.40 (0.22)	0.30 (0.07)	2.09 (0.72)	0.19 (0.09)
Balanced panel (N=149)								
DI-AR, Lag	0.59 (0.22)	0.95 (0.12)	0.67 (0.18)	0.84 (0.10)	0.84 (0.22)	0.69 (0.24)	0.76 (0.14)	0.78 (0.13)
DI-AR	0.70 (0.24)	0.72 (0.13)	0.70 (0.20)	0.75 (0.12)	0.87 (0.29)	0.61 (0.25)	0.86 (0.15)	0.62 (0.11)
DI	1.07 (0.20)	0.46 (0.12)	1.08 (0.18)	0.45 (0.12)	1.43 (0.22)	0.27 (0.07)	2.10 (0.70)	0.19 (0.08)
Full data set; m=1, p=BIC, k fixed								
DI-AR, k=1	0.63 (0.20)	1.04 (0.18)	0.68 (0.17)	0.97 (0.15)	0.60 (0.25)	1.12 (0.20)	0.73 (0.17)	0.93 (0.22)
DI-AR, k=2	0.61 (0.21)	1.07 (0.17)	0.72 (0.16)	0.92 (0.13)	0.64 (0.24)	0.96 (0.17)	0.68 (0.19)	0.97 (0.20)
DI-AR, k=3	0.80 (0.17)	0.82 (0.23)	0.80 (0.12)	0.83 (0.13)	0.94 (0.25)	0.56 (0.29)	0.81 (0.11)	0.80 (0.14)
DI-AR, k=4	0.76 (0.20)	0.81 (0.21)	0.74 (0.15)	0.83 (0.14)	0.92 (0.26)	0.59 (0.29)	0.78 (0.14)	0.78 (0.14)
RMSE, AR Model	0.052		0.038		0.046		0.077	

Notes: See the notes to table 2.

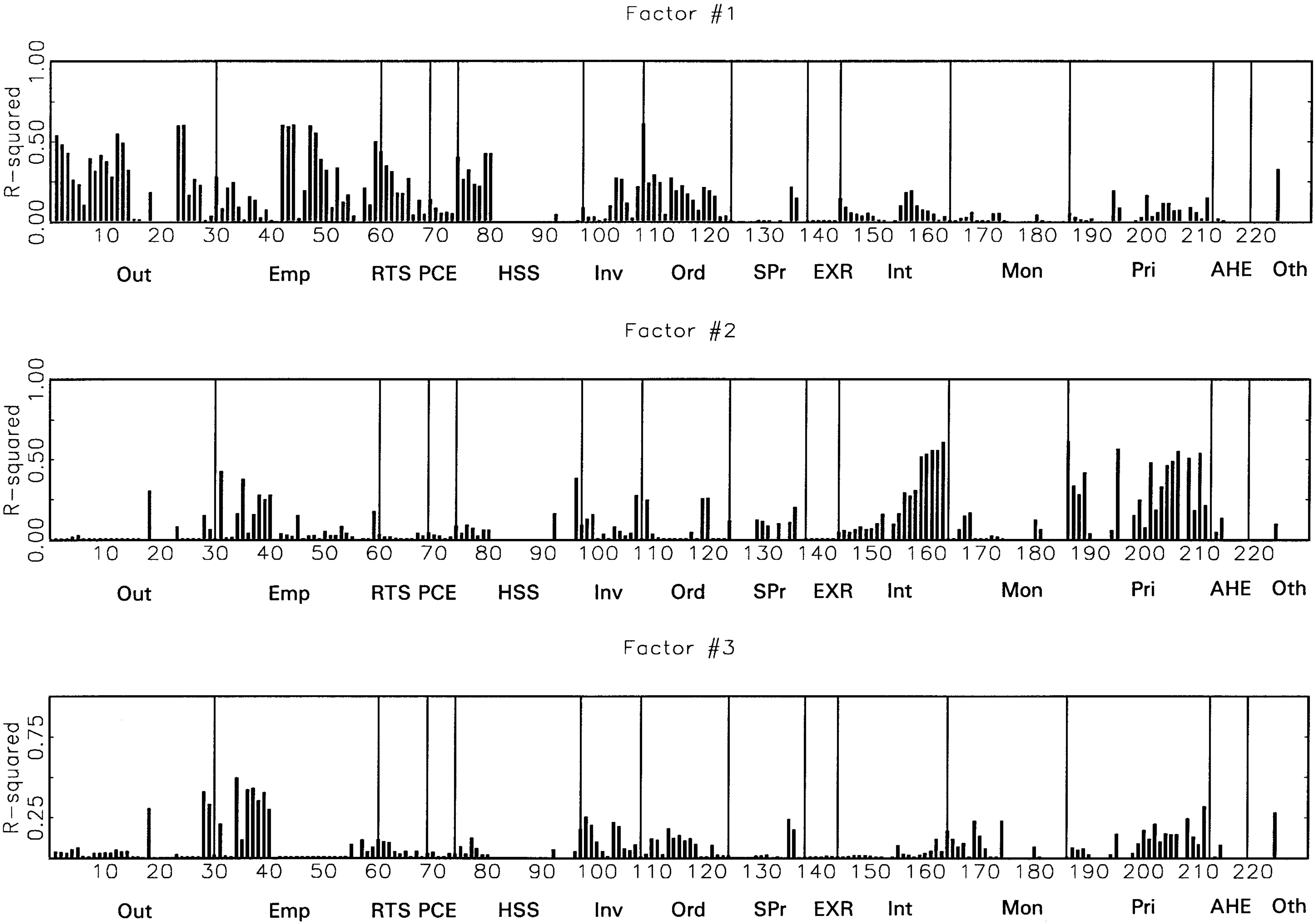

Figure 1. R^2s between factors and individual time series, grouped by category (see Appendix B)

Categories: Real output and income (Out); Employment and hours (Emp); Real retail, manufacturing and trade sales (RTS); Consumption (PCE); Housing starts and sales (HSS); Real inventories and inventory-sales ratios (Inv); Orders and unfilled orders (Ord); Stock prices (SPr); Exchange rates (EXR); Interest rates (Int); Money and credit quantity aggregates (Mon); Price indexes (Pri); Average hourly earnings (AHE); Miscellaneous (Oth)

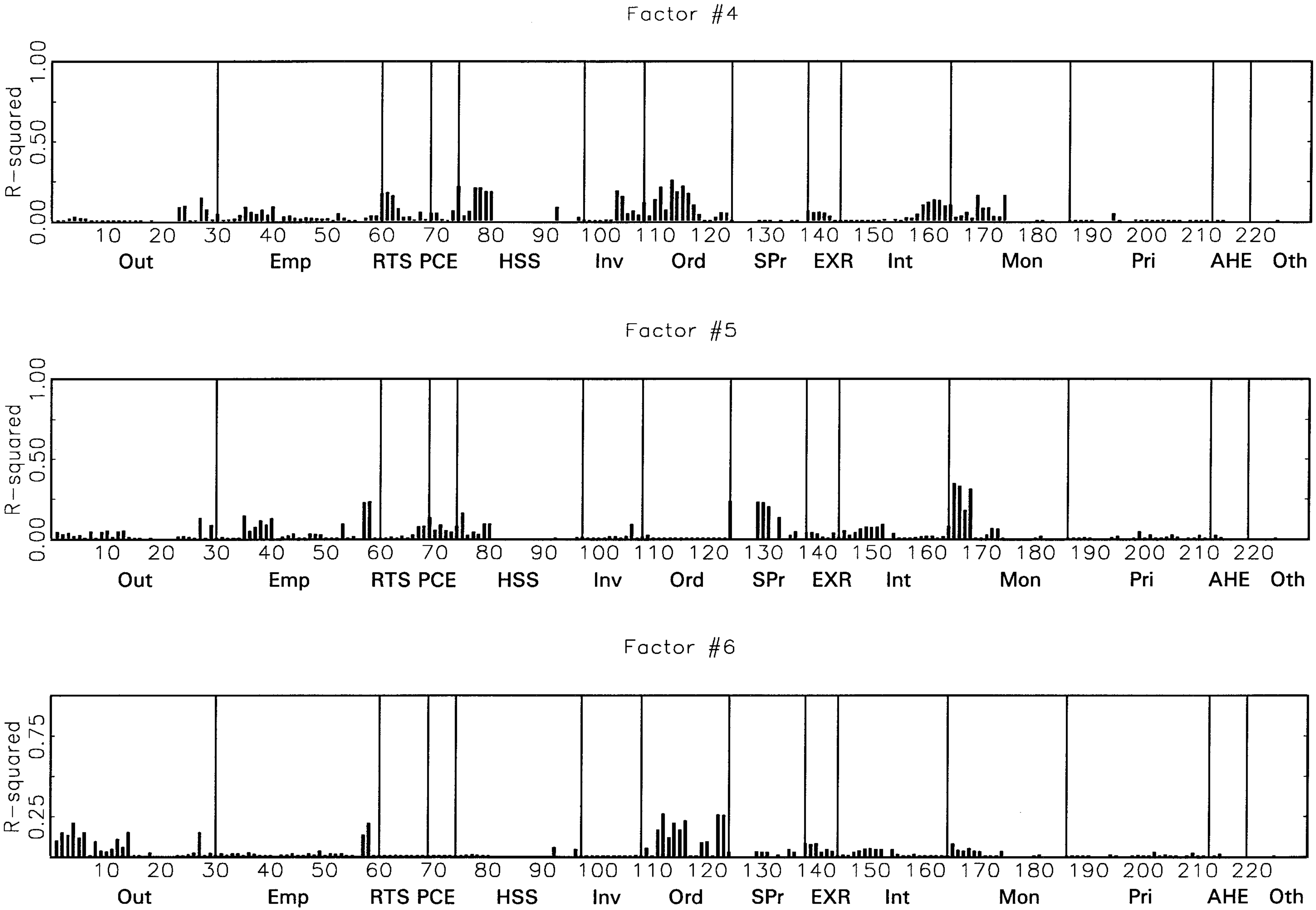

Figure 1 (continued)