

Consistent Estimation of the Number of Dynamic Factors in a Large N and T Panel

Dante AMENGUAL

Department of Economics, Princeton University, Princeton, NJ 08544 (amengual@princeton.edu)

Mark W. WATSON

Woodrow Wilson School, Princeton University, Princeton, NJ 08544 (mwatson@princeton.edu)

Bai and Ng proposed a consistent estimator for the number of *static* factors in a large N and T approximate factor model. This article shows how the Bai–Ng estimator can be modified to consistently estimate the number of *dynamic* factors in a restricted dynamic factor model. The modification is straightforward: The standard Bai–Ng estimator is applied to residuals obtained by projecting the observed data onto lagged values of principal-components estimates of the static factors.

KEY WORDS: Approximate factor model; Bai–Ng estimator; Dynamic factor model.

1. INTRODUCTION

Panel datasets with large time series dimension (T) and cross-sectional dimension (N) are being increasingly used in macroeconomics for both forecasting and structural analysis. Often, these data are analyzed in the context of an assumed latent factor structure of the form

$$X_t = \Lambda F_t + e_t \quad (1.1)$$

for $t = 1, \dots, T$, where X_t denotes an $N \times 1$ vector of observed variables, F_t is an $r \times 1$ vector of latent factors, Λ is a matrix of coefficients, and e_t is a vector of errors. When the elements of e_t have weak cross-sectional and serial correlation, the factors F_t summarize the important cross-covariance properties of the variables.

A question of fundamental interest is the number of latent factors, r , that are required in (1.1). Significant progress on addressing this problem was made in Bai and Ng (2002) who proposed consistent estimators of r based on a penalized least squares objective function associated with the classic principal-components estimator. However, in dynamic models, it is important to differentiate between the number of “static” factors (necessary to fit the covariance matrix of X) and the number of “dynamic” factors (necessary to fit the spectral density matrix of X). Whereas the Bai–Ng estimator was developed to estimate the number of static factors, this article shows that it can be easily modified to consistently estimate the number of dynamic factors.

Dynamics can be incorporated in the model by assuming that F_t evolves as a VAR:

$$F_t = \sum_{i=1}^p \Phi_i F_{t-i} + \varepsilon_t, \quad (1.2)$$

with innovations ε_t that can be represented as $\varepsilon_t = G\eta_t$, where G is $r \times q$ with full column rank and η_t is a sequence of shocks with mean 0 and covariance matrix $\Sigma_{\eta\eta} = I_q$; η_t is the vector of dynamic factor shocks. Several articles show how (1.1) and (1.2) can be derived from a restricted version of a general dynamic factor model driven by q dynamic factors; in this case, F_t contains linear combinations of current and lagged values of the dynamic factors and (1.1)–(1.2) is analogous to the companion form representation of the dynamic factor model.

See Bai and Ng (2005, 2007), Forni, Hallin, Lippi, and Reichlin (2005), Giannone, Reichlin, and Sala (2004), and Stock and Watson (2005, 2006).

To see how the Bai–Ng estimator might be used to estimate the number of dynamic factors, q , substitute (1.2) into (1.1) to obtain

$$Y_t = \Gamma \eta_t + e_t, \quad (1.3)$$

where $Y_t = X_t - \sum_{i=1}^p \Lambda \Phi_i F_{t-i}$ and $\Gamma = \Lambda G$. Thus, Y_t can be represented as a factor model with q factors that correspond to the common shocks η_t . Were Y_t observed data, q could be consistently estimated by applying the Bai–Ng estimator to Y_t . This is infeasible because Y_t depends on unknown parameters and lags of the unobserved factors.

This article studies the consistency properties of the Bai–Ng estimator applied to $\hat{Y}_t = X_t - \sum_{i=1}^p \hat{\Pi}_i \hat{F}_{t-i}$, where $\hat{\Pi}_i$ is an estimator of $\Lambda \Phi_i$ and \hat{F}_{t-i} is an estimator of F_{t-i} . The analysis proceeds in two steps. In the first step, the Bai–Ng estimator is shown to remain consistent if the estimation error $\hat{Y}_t - Y_t$ is sufficiently small (specifically $\sum_{t=1}^T \sum_{i=1}^n (\hat{Y}_{it} - Y_{it})^2 = O_p[\max(N, T)]$). The second step shows that the principal-components estimator of F and feasible estimators of Π yield estimators \hat{Y}_{it} that achieve this degree of accuracy. Together these results yield a feasible consistent estimator of the number of dynamic factors.

The estimator studied in this article was proposed in Stock and Watson (2006) and applied to the problem of estimating the number of dynamic factors in a large panel of U.S. macroeconomic time series. Stock and Watson (2006) did not study the consistency properties of the estimator, and that is the purpose of the present article. Other estimators have also been proposed and used in applied work. Notably, Forni et al. (2000) suggested informal methods based on the relative size of eigenvalues from the estimated spectral density matrix for X ; related methods have been proposed and applied in the empirical analysis of Forni, Lippi, and Reichlin (2003), Giannone et al. (2004) and

elsewhere, and Hallin and Liška (2006) showed how a consistent estimator of q can be constructed from the estimated spectrum. Bai and Ng (2007) proposed an estimator for q based on the residual covariance matrix of the VAR in (1.3) estimated using the principal-components estimator of F_t and showed that the estimator is consistent. Section 3 studies the relative performance of various consistent estimators using a simulation study.

More generally, the plan of this article is as follows. Section 2 briefly summarizes the Bai–Ng estimator, shows the estimator remains consistent when applied to data contaminated with a small amount of measurement error, and uses this result to show that the Bai–Ng estimator applied to \hat{Y} is a consistent estimator of the number of dynamic factors. A Monte Carlo study is presented in Section 3 to gauge the performance of the estimator, Section 4 contains some concluding remarks, and the Appendix includes the proofs to the results given in Section 2.

2. ASSUMPTIONS AND ASYMPTOTIC RESULTS

2.1 Review of Existing Work With a Small Extension

We begin by reviewing results for the model (1.1) under a standard set of assumptions. Transposing (1.1) and stacking the T equations yields

$$X = F\Lambda' + e, \quad (2.1)$$

where X is $T \times N$, F is $T \times r$, Λ is $N \times r$, and e is $T \times N$. The t th rows of X , F , and e are X_t' , F_t' , and e_t' ; the i th row of Λ is λ_i' ; the i th element of X_t is X_{it} and similarly for e_{it} , so that $X_{it} = \lambda_i' F_t + e_{it}$.

Asymptotic properties of various statistics generated by this model have been studied in Stock and Watson (2002), Bai and Ng (2002), Bai (2003), and Bai and Ng (2005, 2007) under a similar set of assumptions. The focus is on datasets in which both N and T are large, so that the asymptotics assume that $N, T \rightarrow \infty$ jointly [equivalently that $N = N(T)$ with $\lim_{T \rightarrow \infty} N(T) = \infty$]. The minimum value of N and T plays an important role in the analysis and this value is denoted by $s_{NT} = \min(N, T)$. The remaining assumptions concern moments and dependence properties of the variables; for the purposes of this article, the following assumptions suffice:

$$(A1) \ E(F_t F_t') = I_r.$$

(A2) $E(\lambda_i \lambda_i') = \Sigma_{\Lambda\Lambda}$, where $\Sigma_{\Lambda\Lambda}$ is a diagonal matrix with elements $\sigma_{ii} > \sigma_{jj} > 0$ for $i < j$. (When Λ is deterministic, $\Sigma_{\Lambda\Lambda}$ is interpreted as the limiting empirical average.)

$$(A3) \ T^{-1} \sum_{t=1}^T F_t F_t' \xrightarrow{p} I_r.$$

$$(A4) \ N^{-1} \sum_{i=1}^N \lambda_i \lambda_i' \xrightarrow{p} \Sigma_{\Lambda\Lambda}.$$

$$(A5) \ (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T e_{it}^2 \xrightarrow{p} \sigma_e^2 > 0.$$

(A6) For some integer $m \geq 2$ and for all integers $j \leq m$, $E \text{trace}[(ee')^j] = O(NT \times \max[N, T]^{j-1})$.

$$(A7) \ E \sum_{t=1}^T \sum_{s=1}^T (\sum_{i=1}^N \lambda_i' F_t e_{is})^2 = O(NT^2).$$

$$(A8) \ E \sum_{t=1}^T \sum_{i=1}^N \lambda_i' \lambda_i e_{it}^2 = O(NT).$$

$$(A9) \ E \sum_{i=1}^N \|\sum_{t=1}^T F_t e_{it}\|^2 = O(NT).$$

Assumptions (A1)–(A5) rule out explosive or trending behavior in both the time series and cross-sectional dimensions; the particular values of $E(F_t F_t')$ and $E(\lambda_i \lambda_i')$ listed in

(A1) and (A2) are normalizations (because $\Lambda F_t = \Lambda H H^{-1} F_t$ for arbitrary H), and assumption (A5) rules out degenerate cases in which the factors explain all of the variance of the X_{it} 's. Assumption (A6) limits the variability and dependence in the errors e_{it} . For $j = 1$, it implies that $\sum_{i=1}^N \sum_{t=1}^T E(e_{it}^2) = O(NT)$; for $j = 2$, it implies that $\sum_{i=1}^N \sum_{j=1}^N (\sum_{t=1}^T e_{it} e_{jt})^2 = \sum_{t=1}^T \sum_{\tau=1}^T (\sum_{i=1}^N e_{it} e_{i\tau})^2 = O_p(NT \times \max[N, T])$, and so forth for larger values of j . Assumptions (A7)–(A9) limit the dependence across elements of Λ , F , and e . Importantly, all of these assumptions hold for sequences of iid random variables with the appropriate number of moments, and assumptions (A6)–(A9) can be interpreted as relaxing the iid assumption to allow weak dependence.

The Bai–Ng estimators of r are based on penalized least squares objective functions. The penalty function depends on a deterministic function $g(N, T)$ that satisfies $g(N, T) \rightarrow 0$ and $s_{NT}^\delta g(N, T) \rightarrow \infty$ for $\delta = (m - 1)/m$, where m is given in assumption (A6). The least squares objective function is conveniently written in terms of the eigenvalues of the XX' moment matrix. Let ω_i denote the i th largest eigenvalue of $(NT)^{-1} XX'$ and consider the least squares problem: $\min_{\{\lambda_i^k\}, \{F_t^k\}} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T (X_{it} - \lambda_i^k F_t^k)^2$, where λ_i^k and F_t^k are arbitrary $k \times 1$ vectors. The usual principal-components calculations imply that the average predicted sum of squares associated with the least squares solution is given by $R(k, X) = \sum_{i=1}^k \omega_i$. Letting $\hat{\sigma}_X^2 = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T X_{it}^2$ denote the average total sum of squares, the penalized average sum of squared residuals is $\text{PC}(k, X) = \hat{\sigma}_X^2 - R(k, X) + kg(N, T)$, and the Bai–Ng “PC” estimator is

$$\widehat{\text{BN}}^{\text{PC}}(X) = \arg \min_{0 \leq k \leq r^{\max}} \text{PC}(k, X). \quad (2.2)$$

Letting $\text{IPC}(k, X) = \ln[\hat{\sigma}_X^2 - R(k, X)] + kg(N, T)$, the Bai–Ng “IPC” estimator is

$$\widehat{\text{BN}}^{\text{IPC}}(X) = \arg \min_{0 \leq k \leq r^{\max}} \text{IPC}(k, X), \quad (2.3)$$

where r^{\max} is a finite constant that satisfies $r \leq r^{\max}$.

Consistency of the Bai–Ng estimator is given in the following lemma.

Lemma 1 (Bai–Ng). Under assumptions (A1)–(A9), $\widehat{\text{BN}}^{\text{PC}}(X) \xrightarrow{p} r$ and $\widehat{\text{BN}}^{\text{IPC}}(X) \xrightarrow{p} r$.

As discussed in the last section, we will study consistency of the Bai–Ng estimators applied to variables measured with error (\tilde{Y}_t in the notation of the last section). The following result shows that the Bai–Ng estimators remain consistent in the presence of sufficiently small measurement error.

Lemma 2. Suppose (A1)–(A9) are satisfied and $\tilde{X} = X + b$, where $T^{-1} N^{-1} \sum_{i=1}^N \sum_{t=1}^T b_{it}^2 = O_p(s_{NT}^{-1})$. Then $\widehat{\text{BN}}^{\text{PC}}(\tilde{X}) \xrightarrow{p} r$ and $\widehat{\text{BN}}^{\text{IPC}}(\tilde{X}) \xrightarrow{p} r$.

Bai and Ng (2002) showed consistency of $\widehat{\text{BN}}^{\text{PC}}(X)$ and $\widehat{\text{BN}}^{\text{IPC}}(X)$ for $\delta = 1$ using assumptions like those in (A1)–(A9), but without (A6). However, there was an error in their proof. As shown in their errata, their proof is valid using a stronger condition on e . In particular, for $e = R\xi H$, where R and H are $N \times N$ and $T \times T$ matrices with bounded eigenvalues, ξ is required to be a $T \times N$ matrix of independently distributed random variables with mean 0 and bounded seventh moments.

2.2 Consistent Estimation of the Number of Dynamic Factors

The results from Lemma 2 suggest that the estimators $\widehat{\text{BN}}^{\text{PC}}(\hat{Y})$ and $\widehat{\text{BN}}^{\text{IPC}}(\hat{Y})$ will be consistent for the number of dynamic factors if the error $\hat{Y} - Y$ is small. We consider two versions of \hat{Y} that are sufficiently accurate for this purpose. Both rely on a first-stage estimate of F . Thus, let \hat{F} and $\hat{\Lambda}$ denote the principal-components estimators of F and Λ constructed from (2.1) using a consistent estimator of r . Let $(\hat{\Phi}_1, \hat{\Phi}_2, \dots, \hat{\Phi}_p)$ denote the ordinary least squares (OLS) estimators from the regression of \hat{F}_t onto $(\hat{F}_{t-1}, \dots, \hat{F}_{t-p})$. The first version of \hat{Y} is

$$\hat{Y}_t^A = X_t - \sum_{i=1}^p \hat{\Lambda} \hat{\Phi}_i \hat{F}_{t-i}. \quad (2.4)$$

The second version of \hat{Y} uses direct estimates of the regression of X_t onto lags of \hat{F}_t . Let $(\hat{\Pi}_1^{\text{OLS}}, \hat{\Pi}_2^{\text{OLS}}, \dots, \hat{\Pi}_p^{\text{OLS}})$ denote the OLS estimators from the regression of X_t onto $(\hat{F}_{t-1}, \dots, \hat{F}_{t-p})$. The second version of \hat{Y} is

$$\hat{Y}_t^B = X_t - \sum_{i=1}^p \hat{\Pi}_i^{\text{OLS}} \hat{F}_{t-i}, \quad (2.5)$$

which does not impose the cross-equation constraint $\Pi_i = \Lambda \Phi_i$.

Consistency of the Bai–Ng estimator for q is then readily shown if (1) the factor model (1.3) for Y satisfies the analogs of conditions (A1)–(A9), and (2) the estimators \hat{F} , $\hat{\Lambda}$, $\hat{\Phi}$, and $\hat{\Pi}^{\text{OLS}}$ are sufficiently accurate. Thus, to begin, assume that (A1)–(A9) hold with η replacing F and Γ replacing Λ . [Note that the normalization in (A1)–(A2) can be achieved by appropriate choice of G .] Stock and Watson (2002) and Bai (2003) discussed the accuracy of the estimators \hat{F} and $\hat{\Lambda}$ under assumptions like those listed as (A1)–(A9). As in Bai and Ng (2005, 2007), $\hat{\Phi}$ will be $T^{1/2}$ consistent under a standard set of assumptions for the VAR for F_t :

(A10) Let $\mathbf{F}_t = (F'_{t-1}, \dots, F'_{t-p})'$. Then

1. The stochastic process $\{F_t\}$ is stationary and ergodic.
2. $E(\mathbf{F}_t \mathbf{F}_t')$ is nonsingular.
3. $\text{vec}(\mathbf{F}_t \eta'_t)$ is a martingale difference sequence with finite second moments.

Finally, accuracy of $\hat{\Pi}^{\text{OLS}}$ requires the additional assumption:

(A11) $E \sum_{i=1}^N \|\sum_{t=1}^T \mathbf{F}_t e_{it}\|^2 = O(NT)$.

We then have the following result.

Theorem 1. Consider the model (1.1)–(1.3). Suppose that (1.1) satisfies (A1)–(A9), that the analogous assumptions are satisfied for (1.3), and that (A10) is satisfied. Then

- (a) $\widehat{\text{BN}}^{\text{PC}}(\hat{Y}^A) \xrightarrow{P} q$ and $\widehat{\text{BN}}^{\text{IPC}}(\hat{Y}^A) \xrightarrow{P} q$.
- (b) In addition, suppose that (A11) is satisfied. Then $\widehat{\text{BN}}^{\text{PC}}(\hat{Y}^B) \xrightarrow{P} q$ and $\widehat{\text{BN}}^{\text{IPC}}(\hat{Y}^B) \xrightarrow{P} q$.

The remaining ingredient in the testing problem is p , the number of lags in the VAR. It is straightforward to show that, under the usual VAR assumptions, p can be estimated consistently by the Bayesian information criterion (BIC).

In some models, innovations in a subset of the X_t variables may depend on only a subset of the dynamic shocks η_t . For example in Bernanke, Boivin, and Elias (2005) and Stock and Watson (2006), X_t is partitioned into a set of “slow moving” variables and other variables, $X_t = (X_t^{\text{slow}}, X_t^{\text{other}})'$ where innovations in X_t^{slow} depend on only a subset of the η_t . It is straightforward to show that the size of this subset can be consistently estimated ($N_{\text{slow}}, T \rightarrow \infty$) using $\widehat{\text{BN}}^{\text{IPC}}$ applied to the relevant subset of elements of \hat{Y} .

3. COMPARING THE ESTIMATORS USING SIMULATED DATA

3.1 Experimental Design

The experimental design is taken from Bai and Ng (2007) where four data-generating processes (DGPs) are considered.

DGP1. In the first design (DGP1), $X_{it} = \lambda'_i F_t + e_{it}$ and $F_t = \Phi F_{t-1} + G \eta_t$, where F_t is 5×1 and η_t is 3×1 , so that $r = 5$ and $q = 3$; $\{\lambda_i\}$, $\{e_{it}\}$, and $\{\eta_t\}$ are mutually independent, with $\{\lambda_i\}$ and $\{\eta_t\}$ iid standard normal random variables/vectors; Φ is a diagonal matrix with elements $(.2, .375, .55, .725, .90)$, and the columns of G are randomly chosen from the unit sphere and are independent of the other random variables. To allow cross-sectional dependence in the idiosyncratic errors, e_t is $N(0, \Omega)$, where $\Omega_{ij} = \rho^{|i-j|}$. Results are presented for $\rho = 0$ and $\rho = .5$. DGP2 is the same as DGP1, but with $r = 3$ and $\Phi = .5 \times I_3$.

In the final two designs, X_t is a moving average of factors f_t that follow an autoregressive (AR) (DGP3) or moving average (MA) (DGP4) process. In DGP3, $X_{it} = (\lambda_{i0} + \lambda_{i1} L) f_t + e_{it}$ and $f_t = \phi f_{t-1} + \eta_t$, where f_t is 2×1 , so that $r = 4$ and $q = 2$. This model can be written as (1.1) and (1.2) with $F_t = (f'_t, f'_{t-1})'$, $\lambda'_i = (\lambda'_{i0}, \lambda'_{i1})'$, $\Phi = \begin{bmatrix} \phi & 0 \\ 0 & 1 \end{bmatrix}$, and $G = [I_2 \ 0_{2 \times 2}]'$. The factor loadings and errors are generated as in DGP1, and $\phi = .5 \times I_2$. In DGP4, $X_{it} = (\lambda_{i0} + \lambda_{i1} L + \lambda_{i2} L^2) f_t + e_{it}$ and $f_t = (I_2 + \Theta L) \eta_t$, where f_t is 2×1 , so that $r = 6$ and $q = 2$. In this design, $F_t = (f'_t, f'_{t-1}, f'_{t-2})'$, but now F_t follows a MA process, so that the VAR in (1.2) serves as an approximation. The MA coefficient matrix is diagonal with elements .2 and .9.

Estimators. The $\widehat{\text{BN}}^{\text{IPC}}$ estimators are implemented using the penalty factor $g(N, T) = \ln(s_{NT})/A$, where $A = NT/(N + T)$. (This is the “IPC2” penalty factor in Bai and Ng 2002.) r is estimated using $\widehat{\text{BN}}^{\text{IPC}}(X)$, where X is the standardized version of the data generated by DGP1–DGP4, and where $r^{\text{max}} = 10$. q is estimated using $\widehat{\text{BN}}^{\text{IPC}}(\hat{Y}^A)$ and $\widehat{\text{BN}}^{\text{IPC}}(\hat{Y}^B)$ constructed using two lags of \hat{F}_t for both specifications.

Two alternative estimators, \hat{q}_3 and \hat{q}_4 from Bai and Ng (2007), were also constructed. These estimators use the eigenvalues of the residual covariance matrix of the VAR for \hat{F}_t to estimate q . Specifically, let $\hat{\varepsilon}_t = \hat{F}_t - \sum_{i=1}^p \hat{\Phi}_i \hat{F}_{t-i}$, where \hat{F}_t is the $\hat{r} \times 1$ vector of factors estimated by principal components using the normalization $N^{-1} \hat{\Lambda}' \hat{\Lambda} = I_{\hat{r}}$ and $T^{-1} \hat{F}' \hat{F} = \text{diag}(\hat{\sigma}_{ii})$, $\hat{\Sigma}_{\hat{\varepsilon}\hat{\varepsilon}} = T^{-1} \sum_{t=p+1}^T \hat{\varepsilon}_t \hat{\varepsilon}'_t$ denote the estimated covariance matrix, and the ordered eigenvalues of $\hat{\Sigma}_{\hat{\varepsilon}\hat{\varepsilon}}$ are denoted by $c_1 \geq c_2 \geq \dots \geq c_{\hat{r}}$. Let $D_{1,k} = [c_{k+1}^2 / \sum_{i=1}^{\hat{r}} c_i^2]^{1/2}$ and $D_{2,k} = [\sum_{i=k+1}^{\hat{r}} c_i^2 / \sum_{i=1}^{\hat{r}} c_i^2]^{1/2}$; the estimators are $\hat{q}_3 = \min_k [k : D_{1,k} < m/s_{NT}^{2/5}]$, and $\hat{q}_4 = \min_k [k : D_{2,k} < m/s_{NT}^{2/5}]$, where m is a positive constant. Following Bai and Ng (2007), we implement these estimators using $m = 1.0$.

3.2 Results

Results are shown in Table 1 for each DGP and various values of N and $T = 100$. Panel A shows results with $\rho = 0$ (so that the e_{it} errors are mutually uncorrelated), and panel B shows results with $\rho = .5$ (so that e_{it} are correlated in the cross section).

Looking first at panel A, five results stand out. First, the estimators are quite accurate for N as small as 50, at least for the simple designs considered. All of the estimators produce the correct answer in more than 98% of the simulations

when $N = 50$ and perform nearly as well when $N = 40$. Second, the constraint $\Pi_i = \Lambda \Phi_i$ used by $\widehat{BN}^{IPC}(\hat{Y}^A)$ but ignored by $\widehat{BN}^{IPC}(\hat{Y}^B)$ is useful: $\widehat{BN}^{IPC}(\hat{Y}^A)$ has a smaller root mean squared error than $\widehat{BN}^{IPC}(\hat{Y}^B)$ in all of the cases considered in the table. Third, for DGP1 and DGP2, $\widehat{BN}^{IPC}(\hat{Y}^A)$ achieves a higher proportion of correct values of q than the other estimators; for DGP3 and DGP4, \hat{q}_3 achieves the highest proportion of correct values. Fourth, in DGP1, while Σ_{FF} has rank 5, two of its eigenvalues are small and $\widehat{BN}^{IPC}(X)$ tends to underestimate the number of static factors when N is large. In spite of

Table 1. Simulation Results: $cov(e_{it}, e_{jt}) = \rho^{li-jl}$

N	T	$\widehat{BN}^{IPC}(\hat{Y}^A)$				$\widehat{BN}^{IPC}(\hat{Y}^B)$				\hat{q}_3				\hat{q}_4				$\widehat{BN}^{IPC}(X)$		
		<q	=q	>q	RMSE	<q	=q	>q	RMSE	<q	=q	>q	RMSE	<q	=q	>q	RMSE	<r	=r	>r
Panel A: $\rho = 0$																				
DGP1: $X_{it} = \lambda_i' F_t + e_{it}; F_t = \Phi F_{t-1} + G\eta_t; r = 5, q = 3$																				
20	100	.00	.84	.16	.84	.00	.63	.37	2.50	.31	.69	.00	.56	.15	.83	.01	.41	.27	.17	.56
30	100	.00	1.00	.00	.07	.00	.98	.02	.17	.05	.95	.00	.23	.04	.96	.00	.20	.64	.24	.12
40	100	.00	1.00	.00	.01	.00	1.00	.00	.03	.01	.99	.00	.09	.01	.99	.00	.08	.79	.18	.03
50	100	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00	.05	.00	1.00	.00	.05	.85	.14	.01
100	100	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	.86	.14	.00
DGP2: $X_{it} = \lambda_i' F_t + e_{it}; F_t = \Phi F_{t-1} + G\eta_t; r = 3, q = 3$																				
20	100	.00	.85	.15	.76	.00	.67	.33	2.31	.23	.77	.00	.48	.16	.84	.00	.40	.00	.52	.48
30	100	.00	1.00	.00	.05	.00	.98	.02	.16	.03	.97	.00	.16	.03	.97	.00	.16	.00	.92	.08
40	100	.00	1.00	.00	.00	.00	1.00	.00	.03	.00	1.00	.00	.04	.00	1.00	.00	.04	.00	.99	.01
50	100	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00
100	100	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00
DGP3: $X_{it} = (\lambda_{i0} + \lambda_{i1}L)' f_t + e_{it}; f_t = \Phi f_{t-1} + \eta_t; r = 4, q = 2$																				
20	100	.00	.74	.26	.79	.00	.51	.49	2.20	.03	.95	.01	.22	.00	.73	.26	.66	.00	.09	.91
30	100	.00	.92	.08	.28	.00	.85	.15	.40	.00	.99	.00	.08	.00	.90	.10	.41	.00	.40	.60
40	100	.00	.98	.02	.15	.00	.95	.05	.23	.00	1.00	.00	.04	.00	.96	.04	.24	.00	.61	.39
50	100	.00	1.00	.00	.06	.00	.99	.01	.11	.00	1.00	.00	.02	.00	.99	.01	.12	.00	.76	.24
100	100	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	.99	.01
DGP4: $X_{it} = (\lambda_{i0} + \lambda_{i1}L + \lambda_{i2}L^2)' f_t + e_{it}; f_t = (I + \Theta L)\eta_t; r = 6, q = 2$																				
20	100	.00	.73	.26	.81	.00	.51	.49	2.09	.02	.97	.00	.16	.01	.92	.07	.32	.00	.25	.75
30	100	.00	.95	.05	.22	.00	.88	.12	.35	.00	1.00	.00	.04	.00	.99	.01	.12	.00	.69	.31
40	100	.00	.99	.01	.08	.00	.98	.02	.13	.00	1.00	.00	.00	.00	1.00	.00	.04	.00	.84	.16
50	100	.00	1.00	.00	.03	.00	1.00	.00	.06	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	.93	.07
100	100	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00
Panel B: $\rho = .5$																				
DGP1: $X_{it} = \lambda_i' F_t + e_{it}; F_t = \Phi F_{t-1} + G\eta_t; r = 5, q = 3$																				
20	100	.00	.00	1.00	6.49	.00	.00	1.00	6.97	.32	.68	.00	.58	.07	.87	.06	.37	.00	.00	1.00
30	100	.00	.30	.70	2.15	.00	.02	.98	5.30	.06	.94	.00	.25	.01	.93	.07	.28	.01	.03	.96
40	100	.00	.73	.27	.67	.00	.32	.68	2.32	.01	.99	.00	.09	.00	.97	.03	.17	.14	.19	.67
50	100	.00	.92	.08	.30	.00	.70	.30	.86	.00	1.00	.00	.05	.00	.99	.01	.08	.35	.31	.34
100	100	.00	1.00	.00	.00	.00	1.00	.00	.01	.00	1.00	.00	.00	.00	1.00	.00	.00	.82	.18	.00
DGP2: $X_{it} = \lambda_i' F_t + e_{it}; F_t = \Phi F_{t-1} + G\eta_t; r = 3, q = 3$																				
20	100	.00	.00	1.00	6.28	.00	.00	1.00	6.96	.27	.73	.00	.52	.04	.90	.06	.33	.00	.00	1.00
30	100	.00	.25	.75	2.14	.00	.03	.97	5.18	.03	.97	.00	.18	.00	.95	.05	.24	.00	.02	.98
40	100	.00	.71	.29	.73	.00	.36	.64	2.15	.00	1.00	.00	.05	.00	.99	.01	.09	.00	.25	.75
50	100	.00	.92	.08	.30	.00	.73	.27	.75	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	.60	.40
100	100	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00
DGP3: $X_{it} = (\lambda_{i0} + \lambda_{i1}L)' f_t + e_{it}; f_t = \Phi f_{t-1} + \eta_t; r = 4, q = 2$																				
20	100	.00	.00	1.00	7.48	.00	.00	1.00	7.96	.03	.92	.04	.28	.00	.58	.42	.80	.00	.00	1.00
30	100	.00	.47	.53	1.76	.00	.08	.92	5.32	.00	.98	.02	.15	.00	.55	.45	.88	.00	.00	1.00
40	100	.00	.83	.17	.50	.00	.48	.52	1.63	.00	.99	.01	.10	.00	.63	.37	.81	.00	.03	.97
50	100	.00	.95	.05	.22	.00	.80	.20	.58	.00	1.00	.00	.06	.00	.79	.21	.61	.00	.15	.85
100	100	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00	.08	.00	.90	.10
DGP4: $X_{it} = (\lambda_{i0} + \lambda_{i1}L + \lambda_{i2}L^2)' f_t + e_{it}; f_t = (I + \Theta L)\eta_t; r = 6, q = 2$																				
20	100	.00	.01	.99	7.51	.00	.00	1.00	7.96	.03	.97	.01	.18	.00	.81	.19	.49	.00	.00	1.00
30	100	.00	.47	.53	1.75	.00	.10	.90	5.03	.00	1.00	.00	.05	.00	.87	.13	.41	.00	.01	.99
40	100	.00	.84	.16	.48	.00	.51	.49	1.46	.00	1.00	.00	.01	.00	.94	.06	.28	.00	.13	.87
50	100	.00	.96	.04	.21	.00	.82	.18	.53	.00	1.00	.00	.00	.00	.98	.02	.16	.00	.39	.61
100	100	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	.00	.98	.01

NOTE: The first two columns show the values of N and T used in the simulations. The next four columns summarize the results for the estimator $\widehat{BN}^{IPC}(\hat{Y}^A)$; the columns labeled $<q$, $=q$, and $>q$ show the fraction of estimates that were less than, equal to, and greater than q ; the column labeled RMSE is the root mean squared error of the estimates. The same entries are provided for the other estimators of q . The final three columns summarize the results for the estimates of r . Results are based on 5,000 simulations.

Downloaded by [Princeton University] at 12:52 25 June 2012

this, $\widehat{\text{BN}}^{\text{IPC}}(\hat{Y}^a)$ and $\widehat{\text{BN}}^{\text{IPC}}(\hat{Y}^b)$ accurately estimate the number of dynamic factors. Finally, comparing the results from DGP3 and DGP4, the AR approximation for DGP4 does not appear to lead to a serious deterioration of performance in any of the estimators.

Panel B shows that the performance of the $\widehat{\text{BN}}^{\text{IPC}}$ deteriorates when there is cross-sectional correlation in the errors: $\widehat{\text{BN}}^{\text{IPC}}(X)$ tends to overestimate r , the number of static factors, and, while not as severe, this upward bias is also evident in $\widehat{\text{BN}}^{\text{IPC}}(\hat{Y}^a)$ and $\widehat{\text{BN}}^{\text{IPC}}(\hat{Y}^b)$. \hat{q}_3 and \hat{q}_4 suffer only a small deterioration in accuracy. Both $\widehat{\text{BN}}^{\text{IPC}}(\hat{Y})$ and \hat{q} provide accurate estimates of the number of dynamic factors when $N = 100$.

4. SUMMARY AND CONCLUDING REMARKS

This article has proposed a modification of the Bai–Ng (2002) estimator and shown that the modification provides a consistent estimator for the number of dynamic factors in an approximate dynamic factor model. The modification uses a result (Lemma 2) that shows that the Bai–Ng estimator remains consistent even when the data are contaminated with a suitably small amount of error. This result may prove useful in other settings, for example, in models in which the equation for X_{it} has the form $X_{it} = \lambda'_i F_t + \beta'_i Z_{it} + e_{it}$, where Z_{it} are observed regressors and β_i must be estimated. We leave these calculations for future work.

ACKNOWLEDGMENTS

This research is an outgrowth of joint work with Jim Stock, who we thank for his comments and suggestions. Thanks also to Jushan Bai, Serena Ng, two referees, and the associate editor for their useful comments. This work was funded in part by NSF grant SBR-0214131.

APPENDIX: PROOFS

This appendix summarizes key details of proofs to the results given in the text. A complete set of proofs is given in the detailed appendix (D-Appendix hereafter) available at <http://www.princeton.edu/~mwatson>.

Proof of Lemma 1

This is a version of theorem 1 and corollary 1 in Bai and Ng (2002) under slightly different assumptions. See D-Appendix for a detailed proof using the assumptions listed previously.

Proof of Lemma 2

Let $\tilde{\omega}_k$ denote the k th ordered eigenvalue of $(NT)^{-1} \tilde{X} \tilde{X}'$. As shown in D-Appendix, Lemma 2 is implied by (1) $\tilde{\omega}_k - \omega_k = o_p(1)$ for $k \leq r$ and (2) $\tilde{\omega}_k - \omega_k = O_p(s_{NT}^{-\delta})$ for $k > r$. To verify (1) and (2), let μ denote the largest eigenvalue of $(NT)^{-1} bb'$. Then

$$\omega_k + \mu - 2(\omega_k \mu)^{1/2} \leq \tilde{\omega}_k \leq \omega_k + \mu + 2(\omega_k \mu)^{1/2} \quad (\text{A.1})$$

follows from Horn and Johnson (1991, thm. 3.3.16). By the assumption of the lemma, $\text{trace}(bb') = O_p(s_{NT}^{-1})$, so that $\mu =$

$O_p(s_{NT}^{-1})$. For $k \leq r$, $\omega_k \xrightarrow{p} \sigma_{kk}$ (D-Appendix R11), so that $\tilde{\omega}_k - \omega_k = o_p(1)$ for $k = 1, \dots, r$ follows from (A.1), and this shows (1). For $k > r$, $\omega_k = O_p(s_{NT}^{-\delta})$ (D-Appendix R28); thus, (A.1) implies $\tilde{\omega}_k - \omega_k = O_p(s_{NT}^{-1}) + O_p(s_{NT}^{-(1+\delta)/2})$, and this shows (2).

Proof of Theorem 1

Let $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_p)$, $\Pi = \Lambda \Phi$, and $\mathbf{F}_t = (F'_{t-1}, \dots, F'_{t-p})'$, so that $F_t = \Phi \mathbf{F}_t + G \eta_t$ and $Y_t = X_t - \Pi \mathbf{F}_t$. Let π'_i denote the i th row of Π and γ'_i denote the i th row of Γ . Then $X_{it} = \eta'_i \gamma_i + \mathbf{F}'_t \pi_i + e_{it}$. The following results are versions of theorem 1 in Bai and Ng (2002) (see D-Appendix):

$$T^{-1} \sum_{t=1}^T \|\hat{\mathbf{F}}_t - J_{NT} \mathbf{F}_t\|^2 = O_p(s_{NT}^{-1}), \quad (\text{A.2})$$

where J_{NT} is an $r \times r$ matrix that satisfies $J_{NT} \xrightarrow{p} J$ a nonsingular matrix,

$$T^{-1} \sum_{t=1}^T \|\hat{\mathbf{F}}_t - \mathbf{J}_{NT} \mathbf{F}_t\|^2 = O_p(s_{NT}^{-1}), \quad (\text{A.3})$$

where \mathbf{J}_{NT} is a $(pr) \times (pr)$ matrix that satisfies $\mathbf{J}_{NT} \xrightarrow{p} \mathbf{J}$ a nonsingular matrix,

$$N^{-1} \sum_{i=1}^N \|\hat{\lambda}_i - J_{NT}^{-1} \lambda_i\|^2 = O_p(s_{NT}^{-1}). \quad (\text{A.4})$$

The following lemma is useful.

Lemma 3. Let $\hat{\pi}_i$ denote an estimator of π_i and $b_{ii} = \hat{\mathbf{F}}'_i \hat{\pi}_i - \mathbf{F}'_i \pi_i$. If $N^{-1} \sum_{i=1}^N \|\hat{\pi}_i - \mathbf{J}_{NT}^{-1} \pi_i\|^2 = O_p(s_{NT}^{-1})$, then $T^{-1} N^{-1} \times \sum_{i=1}^T \sum_{i=1}^N b_{ii}^2 = O_p(s_{NT}^{-1})$.

Proof. Write $\hat{\mathbf{F}}_t = \mathbf{J}_{NT} \mathbf{F}_t + (\hat{\mathbf{F}}_t - \mathbf{J}_{NT} \mathbf{F}_t)$ and $\hat{\pi}_i = \mathbf{J}_{NT}^{-1} \pi_i + (\hat{\pi}_i - \mathbf{J}_{NT}^{-1} \pi_i)$, so that $b_{ii} = \mathbf{F}'_i \mathbf{J}_{NT} (\hat{\pi}_i - \mathbf{J}_{NT}^{-1} \pi_i) + (\hat{\mathbf{F}}_t - \mathbf{J}_{NT} \mathbf{F}_t)' \times \mathbf{J}_{NT}^{-1} \pi_i + (\hat{\mathbf{F}}_t - \mathbf{J}_{NT} \mathbf{F}_t)' (\hat{\pi}_i - \mathbf{J}_{NT}^{-1} \pi_i)$. Thus,

$$\begin{aligned} T^{-1} N^{-1} \sum_{t=1}^T \sum_{i=1}^N b_{ii}^2 &\leq \left[T^{-1} \sum_{t=1}^T \|\mathbf{F}_t\|^2 \right] \|\mathbf{J}_{NT}\|^2 \left[N^{-1} \sum_{i=1}^N \|\hat{\pi}_i - \mathbf{J}_{NT}^{-1} \pi_i\|^2 \right] \\ &+ \left[T^{-1} \sum_{t=1}^T \|\hat{\mathbf{F}}_t - \mathbf{J}_{NT} \mathbf{F}_t\|^2 \right] \|\mathbf{J}_{NT}^{-1}\|^2 \left[N^{-1} \sum_{i=1}^N \|\pi_i\|^2 \right] \\ &+ \left[T^{-1} \sum_{t=1}^T \|\hat{\mathbf{F}}_t - \mathbf{J}_{NT} \mathbf{F}_t\|^2 \right] \left[N^{-1} \sum_{i=1}^N \|\hat{\pi}_i - \mathbf{J}_{NT}^{-1} \pi_i\|^2 \right], \end{aligned}$$

and the result follows from $T^{-1} \sum_{t=1}^T \|\mathbf{F}_t\|^2 = O_p(1)$, (A10), $\|\mathbf{J}_{NT}\|^2 \xrightarrow{p} \|\mathbf{J}\|^2 < \infty$, $N^{-1} \sum_{i=1}^N \|\hat{\pi}_i - \mathbf{J}_{NT}^{-1} \pi_i\|^2 = O_p(s_{NT}^{-1})$ (assumption of the lemma), and $T^{-1} \sum_{t=1}^T \|\hat{\mathbf{F}}_t - \mathbf{J}_{NT} \mathbf{F}_t\|^2 = O_p(s_{NT}^{-1})$ [from (A.3)].

Part (a) of Theorem 1. The feasible OLS estimator of Φ is

$$\hat{\Phi} = \left[T^{-1} \sum_{t=p+1}^T \hat{F}_t \hat{F}_t' \right] \left[T^{-1} \sum_{t=p+1}^T \hat{F}_t \hat{F}_t' \right]^{-1}.$$

Using $F_t = \Phi F_t + G\eta_t$, \hat{F}_t can be written as

$$\hat{F}_t = J_{NT} \Phi J_{NT}^{-1} \hat{F}_t + J_{NT} G \eta_t + (\hat{F}_t - J_{NT} F_t) - J_{NT} \Phi J_{NT}^{-1} (\hat{F}_t - J_{NT} F_t).$$

Thus,

$$\begin{aligned} \hat{\Phi} - J_{NT} \Phi J_{NT}^{-1} &= \left[J_{NT} G T^{-1} \sum_{t=p+1}^T \eta_t \hat{F}_t' + T^{-1} \sum_{t=p+1}^T (\hat{F}_t - J_{NT} F_t) \hat{F}_t' \right. \\ &\quad \left. - J_{NT} \Phi J_{NT}^{-1} T^{-1} \sum_{t=p+1}^T (\hat{F}_t - J_{NT} F_t) \hat{F}_t' \right] \\ &\quad \times \left[T^{-1} \sum_{t=p+1}^T \hat{F}_t \hat{F}_t' \right]^{-1}. \end{aligned}$$

Straightforward calculations (see D-Appendix) show that each of the terms $T^{-1} \sum_{t=p+1}^T \eta_t \hat{F}_t'$, $T^{-1} \sum_{t=p+1}^T (\hat{F}_t - J_{NT} F_t) \hat{F}_t'$, and $T^{-1} \sum_{t=p+1}^T (\hat{F}_t - J_{NT} F_t) \hat{F}_t'$ are $O_p(s_{NT}^{-1/2})$ and that $T^{-1} \times \sum_{t=p+1}^T \hat{F}_t \hat{F}_t' \xrightarrow{p} \mathbf{J}E(\mathbf{F}_t \mathbf{F}_t') \mathbf{J}$, which is nonsingular. Thus, $\hat{\Phi} - J_{NT} \Phi J_{NT}^{-1} = O_p(s_{NT}^{-1/2})$.

To complete the proof, let $\hat{\pi}_i = \hat{\Phi}' \hat{\lambda}_i$ and write $\hat{\lambda}_i = J_{NT}^{-1} \lambda_i + (\hat{\lambda}_i - J_{NT}^{-1} \lambda_i)$ and $\hat{\Phi} = J_{NT} \Phi J_{NT}^{-1} + (\hat{\Phi} - J_{NT} \Phi J_{NT}^{-1})$, so that $\hat{\pi}_i - J_{NT}^{-1} \pi_i = J_{NT}^{-1} \Phi J_{NT}^{-1} (\hat{\lambda}_i - J_{NT}^{-1} \lambda_i) + (\hat{\Phi} - J_{NT} \Phi J_{NT}^{-1}) J_{NT}^{-1} \times \lambda_i + (\hat{\Phi} - J_{NT} \Phi J_{NT}^{-1}) (\hat{\lambda}_i - J_{NT}^{-1} \lambda_i)$. Thus, $N^{-1} \sum_{i=1}^N \|\hat{\pi}_i - J_{NT}^{-1} \pi_i\|^2 = O_p(s_{NT}^{-1})$ follows from $\hat{\Phi} - J_{NT} \Phi J_{NT}^{-1} = O_p(s_{NT}^{-1/2})$ and $N^{-1} \sum_{i=1}^N \|\hat{\lambda}_i - J_{NT}^{-1} \lambda_i\|^2 = O_p(s_{NT}^{-1})$. Part (a) then follows from Lemma 3.

Part (b) of Theorem 1. The feasible OLS estimator of π_i is

$$\hat{\pi}_i^{\text{OLS}} = \left[T^{-1} \sum_{t=p+1}^T \hat{F}_t \hat{F}_t' \right]^{-1} \left[T^{-1} \sum_{t=p+1}^T \hat{F}_t X_{it} \right].$$

Using $X_{it} = F_t' \pi_i + \eta_t' \gamma_i + e_{it} = \hat{F}_t' J_{NT}^{-1} \pi_i - (\hat{F}_t - J_{NT} F_t)' J_{NT}^{-1} \times \pi_i + \eta_t' \gamma_i + e_{it}$,

$$\hat{\pi}_i^{\text{OLS}} - J_{NT}^{-1} \pi_i = \left[T^{-1} \sum_{t=p+1}^T \hat{F}_t \hat{F}_t' \right]^{-1}$$

$$\begin{aligned} &\times \left[T^{-1} \sum_{t=p+1}^T \hat{F}_t (\hat{F}_t - J_{NT} F_t)' J_{NT}^{-1} \pi_i \right. \\ &\quad \left. + T^{-1} \sum_{t=p+1}^T \hat{F}_t \eta_t' \gamma_i + T^{-1} \sum_{t=p+1}^T \hat{F}_t e_{it} \right]. \end{aligned}$$

Straightforward calculations (see D-Appendix) show that each of the terms $N^{-1} \sum_{i=1}^N \|T^{-1} \sum_{t=p+1}^T \hat{F}_t (\hat{F}_t - J_{NT} F_t)' J_{NT}^{-1} \times \pi_i\|^2$, $N^{-1} \sum_{i=1}^N \|T^{-1} \sum_{t=p+1}^T \hat{F}_t \eta_t' \gamma_i\|^2$, and $N^{-1} \sum_{i=1}^N \|T^{-1} \times \sum_{t=p+1}^T \hat{F}_t \eta_t' \gamma_i\|^2$ are $O_p(s_{NT}^{-1})$ and that $T^{-1} \sum_{t=p+1}^T \hat{F}_t \hat{F}_t' \xrightarrow{p} \mathbf{J}E(\mathbf{F}_t \mathbf{F}_t') \mathbf{J}$, which is nonsingular. Thus, $N^{-1} \sum_{i=1}^N \|\hat{\pi}_i^{\text{OLS}} - J_{NT}^{-1} \pi_i\|^2 = O_p(s_{NT}^{-1})$ and the result follows from Lemma 3.

[Received October 2005. Revised July 2006.]

REFERENCES

- Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171.
- Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.
- (2005), "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," unpublished manuscript, University of Michigan.
- (2006), "Determining the Number of Factors in Approximate Factor Models, Errata," available at <http://www.umich.edu/~ngse/papers/correctionEcta2.pdf>.
- (2007), "Determining the Number of Primitive Shocks in Factor Models," *Journal of Business & Economic Statistics*, 25, 52–60.
- Bernanke, B. S., Boivin, J., and Elias, P. (2005), "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *Quarterly Journal of Economics*, 120, 387–422.
- Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000), "The Generalized Factor Model: Identification and Estimation," *The Review of Economics and Statistics*, 82, 540–554.
- (2005), "The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting," *Journal of the American Statistical Association*, 100, 830–840.
- Forni, M., Lippi, M., and Reichlin, L. (2003), "Opening the Black Box: Structural Factor Models versus Structural VARs," CEPR Discussion Paper 4133.
- Giannone, D., Reichlin, L., and Sala, L. (2004), "Monetary Policy in Real Time," in *NBER Macroeconomics Annual 2004*, No. 19, eds. M. Gertler and K. Rogoff, MIT Press, pp. 161–224.
- Hallin, M., and Liška, R. (2006), "The Generalized Dynamic Factor Model: Determining the Number of Factors," *Journal of the American Statistical Association*, to appear.
- Horn, R. A., and Johnson, C. R. (1991), *Topics in Matrix Analysis*, Cambridge, U.K.: Cambridge University Press.
- Stock, J. H., and Watson, M. W. (2002), "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.
- (2005), "Implications of Dynamic Factor Models for VAR Analysis," NBER Working Paper w11467.
- (2005), "Macroeconomic Forecasting Using Many Predictors," in *Handbook of Economic Forecasting*, eds. G. Elliott, C. Granger, and A. Timmerman, North Holland, pp. 515–554.