Computational Statistics and Data Analysis 56 (2012) 3345–3365

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Bayesian panel data analysis for exploring the impact of subprime financial crisis on the US stock market

Ruey S. Tsay [a], Tomohiro Ando [b,*]

[a] *Booth School of Business, University of Chicago, United States*

[b] *Graduate School of Business Administration, Keio University, 2-1-1 Hiyoshi-Honcho, Kohoku-ku, Yokohama-shi, Kanagawa 223-8523, Japan*

ARTICLE INFO

Article history:
Available online 21 December 2010

Keywords:
Panel data
Markov chain Monte Carlo
Model selection

ABSTRACT

The effects of recent subprime financial crisis on the US stock market are analyzed. To investigate this problem, a Bayesian panel data analysis to identify common factors that explain the movement of stock returns when the dimension is high is developed. For high-dimensional panel data, it is known that previously proposed approaches cannot estimate accurately the variance–covariance matrix. An advantage of the proposed method is that it considers parameter uncertainty in variance–covariance estimation and factor selection. Two new criteria for determining the number of factors in the data are developed and the consistency of the selection criteria as both the number of observations and the cross-section dimension tend to infinity is established. An empirical analysis indicates that the US stock market was subject to 8 common factors before the outbreak of the subprime crisis, but the number of factors reduced substantially after the outbreak. In particular, a small number of common factors govern the fluctuations of the stock market after the collapse of Lehman Brothers. In other words, empirical evidence that the structure of US stock market has changed drastically after the subprime crisis is obtained. It is also shown that the factor models selected by the proposed criteria work well in out-of-sample forecasting of asset returns.

1. Introduction

Common factors that explain the co-movement of asset returns have attracted much interest in mutual fund management and financial econometrics. See, for instance, Fama and French (1993). The turbulence of US financial market occurred in the summer of 2007 has seriously affected the entire US and global banking systems and led to a global economic recession. The goal of this paper is to investigate the effects of the subprime financial crisis on the US stock market. In particular, we seek to detect changes, if any, in the common factors that explain the co-movement of US stock returns. Knowing the common factors is important in investment decision, asset allocation, and risk management. For instance, a quantitative financial model (e.g., an arbitrary pricing model) with too few factors cannot capture the variation of the asset returns whereas a model with too many factors leads to overfitting.

To identify the number of common factors in economic and financial applications, one often employs a large panel data set. This is particularly so in recent years because advances in information technology make it possible to collect and process huge panel data sets. On the other hand, traditional statistical methods such as the vector autoregressive model of multivariate time series analysis often fare poorly in data analysis when the dimension is high, and dimension reduction becomes a necessity. Factor models are perhaps the most commonly used statistical tool to simplify the analysis

* Corresponding author. Tel.: +81 45 564 2039; fax: +81 45 562 3502.
E-mail addresses: ruey.tsay@chicagogsb.edu (R.S. Tsay), andoh@hc.cc.keio.ac.jp, andoh@kbs.keio.ac.jp (T. Ando).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2010.11.028

of huge panel data sets. Indeed, many efforts have been devoted lately in the econometric and statistical literature to factor models for analyzing high-dimensional data. See, for example, Stock and Watson (1998, 2002a, 2004), Forni et al. (2000), Forni and Lippi (2001), Bai and Ng (2002), Bai (2003), and Hallin and Liska (2007) in the econometric literature. In the statistical literature, McLachlan et al. (2003), Lopes and West (2004), Lopes et al. (2008), Carvalho et al. (2008), Ando (2009), Bhattacharya and Dunson (2009), and Frühwirth-Schnatter and Lopes (2010) all consider Bayesian factor analysis. In particular, Lopes and West (2004) treated model uncertainty in Bayesian factor analysis using reversible jump Markov chain Monte Carlo. West (2003) considered Bayesian factor regression models in the "large p, small n" setting.

The usefulness of factor models in economic applications has also been reported in the literature. Stock and Watson (2002b) reported that forecasting errors of many macroeconomic variables are reduced by extracting a small number of common factors from a large panel of economic and financial variables. Bernanke and Boivin (2003) found that the unobserved factors are empirically related to the monetary policy of the US Federal Reserve Banks. Furthermore, factor models are useful tools in forecasting financial variables (Stock and Watson, 2003) and in constructing a core inflation index (Forni and Reichlin, 1998).

For a given panel data set, an important topic in factor modeling is the determination of the optimal (i.e., true) number of factors because the number of factors plays a fundamental role in modeling, interpreting, and forecasting of the data. To select the number of factors, Forni et al. (2000) advocated a heuristic rule based on the number of diverging dynamic eigenvalues of the covariance matrix. Using an information theoretic approach, Bai and Ng (2002) proposed several criteria for the identification of the number of factors. These authors showed that their selection criteria are consistent in the sense that, under certain assumptions, the identified number of factors shrink towards the true number of factors as both the number of observations and the cross-section dimension tend to infinity. Onatski (2005) developed another criterion based on the theory of random matrices.

Our limited experience indicates that the aforementioned methods for selecting the number of factors may fare poorly in finite samples; see also the cases of small N and T in Tables I–VIII of Bai and Ng (2002). The aim of this paper is, therefore, to develop new criteria for factor selection that perform well in finite samples. Our approach is Bayesian and the proposed criteria are referred to as the Panel Data C_p (PDC$_p$) and Panel Data Information Criterion (PDIC), respectively. A special feature of these new criteria is that they consider parameter uncertainty in factor selection. In recent years, many studies reported advantages of treating parameter uncertainty in statistical analysis; See, e.g., Campbell et al. (2003). Since we estimate the factor model by a Bayesian procedure, no criteria are currently available to select the number of factors when both the dimension and sample size go to infinity. The second goal of this paper is to develop criteria that can select a proper factor model when a Bayesian approach is adopted in the analysis.

We establish the consistency of the proposed criteria under certain conditions as both the number of observations and the cross-section dimension tend to infinity. One of the main advantages of our criteria relative to the others available in the literature is that they work well even in the situation when the number of observations and the cross-section dimension are small. Another advantage of the proposed PDC_p criterion is that it is less sensitive than other criteria to the violation of model assumptions. Our simulation study shows that the proposed PDC_p criterion continues to work well, even when there are heteroscedasticity, serial correlation, and fat-tailed features in the data. Similar to that of Amengual and Watson (2007), the proposed criteria can be modified to select the number of dynamic factor models in panel data.

In application, we employ the daily returns of 49 industrial portfolios from the Fama and French database to investigate the impact of subprime crisis on the US stock market. We divided the data span into the following three periods: (1) June 30, 2006 to June 29, 2007 denoting period before the outbreak of the subprime crisis, (2) August 1, 2007 to August 29, 2008 denoting the period after the outbreak of the subprime crisis, but before the Lehman's failure, (3) October 1, 2008 to September 30, 2009 denoting the period after Lehman's failure. We omit one month of returns between the periods because the exact dates of impact that the extreme events have on the market are not certain.

Based on the proposed criteria, we found that the number of common factors reduced substantially after the outbreak of the subprime crisis. We then investigated the correlation structure between the unobserved factors and some well-known factors in the literature, including Fama and French (1993)'s three factors, Momentum factor, Short-Term Reversal factor, and Long-Term Reversal factor. The first latent factor in each period is strongly correlated with the market excess return of the period. More interestingly, we found that some unobserved factors are not correlated with these observable 6 factors. The result indicates that there is room for developing new factors to help explain the US stock returns.

We also evaluate the out-of-sample forecasting performance of the proposed method and show that the proposed method improves the forecasting performance over the model that uses the six commonly used factors mentioned before. Finally, we construct certain portfolios based on the selected factors and demonstrate that the proposed portfolio performs well.

The paper is organized as follows. Section 2 describes the factor model and the associated assumptions used in the paper. It also briefly reviews the asymptotic principal component analysis of Connor and Korajczyk (1986, 1988); see also Forni et al. (2000) and Stock and Watson (1998). Section 3 introduces a Bayesian estimation procedure of principal components under the additional assumption of elliptical distributions. It also presents a Markov chain Monte Carlo (MCMC) algorithm for estimating the posterior distribution in a panel data. We demonstrate details of the MCMC procedure using multivariate normal, including the use of singular multivariate normal distribution. In Section 4 we propose the new model selection criteria, Panel Data C_p and Panel Data Information Criterion, and establish their consistency property. Section 5 conducts Monte Carlo simulations. In the simulation, we compare the performance of the proposed criteria with others available in the literature for several data generating processes. We find that the proposed Panel Data C_p criterion outperforms the criteria of Bai and Ng (2002) when the time series and cross-sectional dimensions are small or when the common factors have serial dependence. When the sample size or the cross-sectional dimension is large, all of the criteria considered perform well for most data generating processes. Section 6 contains the empirical application to US stock market. Finally, Section 7 concludes.

2. A factor model

Factor models are a useful econometric tool for describing the variations in a large number of time series by a small number of common factors. Suppose a set of T observations $\mathbf{Y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T\}$ are generated from the following r-factor model:

$$\boldsymbol{y}_t = \boldsymbol{\Lambda}_0 \boldsymbol{f}_t^0 + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T, \tag{1}$$

where $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{Nt})'$ is a N-dimensional random vector, $\boldsymbol{f}_t^0 = (f_{1t}^0, \ldots, f_{rt}^0)'$ is the r-dimensional random vector of factors, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$ is the N-dimensional random noise vector with mean $\mathbf{0}$ and variance that satisfies the Assumption C below, and $\boldsymbol{\Lambda}_0 = (\boldsymbol{\lambda}_1^0, \ldots, \boldsymbol{\lambda}_N^0)'$ is the $N \times r$ matrix of factor loadings. We exclude the singular situation that $\boldsymbol{\Lambda}_0 = \mathbf{O}$ and $\boldsymbol{f}_t^0 = \mathbf{0}$ for $t = 1, \ldots, T$.

Let $\|\mathbf{A}\| = [\mathrm{tr}(\mathbf{A}'\mathbf{A})]^{1/2}$ be the usual norm of the matrix $\mathbf{A}$, where tr denotes the trace of a square matrix. In this paper, we adopt the same assumptions as those of Bai and Ng (2002) for the factor model.

Assumption A. The factors satisfy $E(\boldsymbol{f}_t^0) = \mathbf{0}$, $E\|\boldsymbol{f}_t^0\|^4 < \infty$, and $T^{-1}\sum_{t=1}^T \boldsymbol{f}_t^0 \boldsymbol{f}_t^{0\prime} \to \boldsymbol{\Sigma}_{F_0}$ as $T \to \infty$, where $\boldsymbol{\Sigma}_{F_0}$ is a $r \times r$ positive definite matrix.

Assumption B. The factor loading matrix $\boldsymbol{\Lambda}_0$ satisfies $\|\boldsymbol{\lambda}_i^0\| < \infty$ and

$$\|N^{-1}\boldsymbol{\Lambda}_0{}'\boldsymbol{\Lambda}_0 - \boldsymbol{\Sigma}_{\Lambda_0}\| \to \mathbf{0}, \quad \text{as } N \to \infty,$$

where $\boldsymbol{\Sigma}_{\Lambda_0}$ is a $r \times r$ positive definite matrix.

Assumption C. The noise term $\boldsymbol{\varepsilon}_t$ in Eq. (1) follows a class of elliptical distributions with zero mean and may have cross-section dependence.

Assumption D. We allow certain weak dependence between factors and idiosyncratic errors such that

$$E\left(\frac{1}{N}\sum_{i=1}^N \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T \boldsymbol{f}_t^0 \varepsilon_{it} \right\|^2\right) < C$$

with constant $C < \infty$.

The factor model in Eq. (1) can be expressed in the matrix form:

$$\mathbf{Y} = \mathbf{F}_0 \boldsymbol{\Lambda}_0{}' + \mathbf{E}, \tag{2}$$

where $\mathbf{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)'$, $\mathbf{F}_0 = (\boldsymbol{f}_1^0, \ldots, \boldsymbol{f}_T^0)'$, $\mathbf{E} = (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_T)'$ with the error term $\boldsymbol{\varepsilon}_t$ being a multivariate random vector with mean $\mathbf{0}$ and positive definite covariance matrix that satisfies the Assumption C.

To estimate the unknown parameters of the factor model in Eq. (2), one can adopt the method of asymptotic principal components studied before by Connor and Korajczyk (1986, 1988), Forni et al. (2000) and Stock and Watson (1998). Consider the objective function

$$\ell(\mathbf{F}_r, \boldsymbol{\Lambda}_r) = \frac{1}{NT}\mathrm{tr}\{(\mathbf{Y} - \mathbf{F}_r\boldsymbol{\Lambda}_r')'(\mathbf{Y} - \mathbf{F}_r\boldsymbol{\Lambda}_r')\}, \tag{3}$$

where the $T \times r$ matrix $\mathbf{F}_r$ and the $N \times r$ matrix $\boldsymbol{\Lambda}_r$ are subject to the normalization condition $\mathbf{F}_r'\mathbf{F}_r/T = \mathbf{I}_r$ or $\boldsymbol{\Lambda}_r'\boldsymbol{\Lambda}_r/N = \mathbf{I}_r$, where $\mathbf{I}_r$ is the $r \times r$ identity matrix. The estimates $\widehat{\mathbf{F}}_r$ and $\widehat{\boldsymbol{\Lambda}}_r$ of $\mathbf{F}_0$ and $\boldsymbol{\Lambda}_0$ can be obtained by minimizing the objective function $\ell(\mathbf{F}_r, \boldsymbol{\Lambda}_r)$. Specifically, the asymptotic principal component estimates are

$$\widehat{\boldsymbol{\Lambda}}_r = \sqrt{N}\widehat{\mathbf{A}}_r, \qquad \widehat{\mathbf{F}}_r = \widehat{\mathbf{X}}_r\widehat{\mathbf{L}}_r/\sqrt{N}, \tag{4}$$

where the $T \times r$ and $N \times r$ matrices $\widehat{\mathbf{X}}_r$ and $\widehat{\mathbf{A}}_r$ are the first r columns of the matrices $\mathbf{X}$ and $\mathbf{A}$ in the singular value decomposition of $\mathbf{Y}$ such that $\mathbf{Y} = \mathbf{XLA}'$, where $\mathbf{L}$ is a diagonal matrix consisting of the ordered singular values of $\mathbf{Y}$, and columns of $\mathbf{X}$ and $\mathbf{A}$ satisfy the orthogonality restrictions $\mathbf{A}_r'\mathbf{A}_r = \mathbf{I}_r$ and $\mathbf{X}_r'\mathbf{X}_r = \mathbf{I}_r$. In (4), $\widehat{\mathbf{L}}_r = \mathrm{diag}\{l_1, l_2, \ldots, l_r\}$ consists of the first r singular values in $\mathbf{L}$ satisfying $l_1 > \cdots > l_r > 0$.

Note that properties of the asymptotic principal components estimates $\widehat{\boldsymbol{\Lambda}}_r$ and $\widehat{\mathbf{F}}_r$ in (4) are obtained in the literature under the uncorrelated Gaussian noise assumption; see, for instance, Smidl and Quinn (2007). However, the objective function in (3) does not depend on the normality assumption. Indeed, as shown in Bai and Ng (2002), the estimates can still be used to estimate consistently the true number of factors without the Gaussian assumption.

In practice, the true number of factors r is unavailable. But the aforementioned estimation method can be carried out for a hypothesized number of factors k. In other words, assuming that there are k factors, one can apply the asymptotic principal component analysis to obtain the estimates $\mathbf{F}_0$ and $\boldsymbol{\Lambda}_0$ as

$$\widehat{\boldsymbol{\Lambda}}_k = \sqrt{N}\widehat{\mathbf{A}}_k, \qquad \widehat{\mathbf{F}}_k = \widehat{\mathbf{X}}_k\widehat{\mathbf{L}}_k/\sqrt{N}, \tag{5}$$

where $\widehat{\mathbf{X}}_k$, $\widehat{\mathbf{A}}_k$ and $\widehat{\mathbf{L}}_k$ are defined in the same way as their counterparts in (4) with r replaced by k.

The asymptotic principal component approach to estimating factor models is widely used, but it has some weaknesses. In particular, the approach does not address the uncertainty in parameter estimation. Parameter uncertainty has long been recognized as an important issue in statistical inference and may have some adverse effects if overlooked. See, for example, the portfolio selection problem in Bawa et al. (1979). Bayesian approach, which explicitly includes parameter uncertainty in the analysis, provides an alternative to statistical inference. Factor analysis is no exception. The goal of this paper is, therefore, to explore the advantages of Bayesian approach to factor analysis. We estimate the unknown parameters under a Bayesian framework and consider the parameter uncertainty in selecting the number of factors. Our Bayesian estimation procedure is given in the next section.

3. Bayesian estimation

Estimation of factor models has been investigated by many authors in the literature when the sample size T is much larger than the cross-section dimension N; see, for example, Aguilar and West (2000), Lopes and West (2004), Press and Shigemasu (1989), Bishop (1999), and Smidl and Quinn (2007). In this paper, we focus on the situation when N is close to, or even larger than, the sample size T. This is a challenging estimation problem. We employ a Bayesian approach with Markov chain Monte Carlo (MCMC) methods in this paper to overcome the difficulty in estimation and to consider parameter uncertainty in factor selection.

Under the Bayesian framework, we need to specify the likelihood function of the data. For the factor model in (1), this means specifying of the distribution function of $\boldsymbol{\varepsilon}_t$. Based on the objective function in (3), a natural choice is to entertain a class of elliptical distributions for $\boldsymbol{\varepsilon}_t$ that includes the Normal, Student-t, Laplace, Bessel, Kotz, and Pearson-type distributions; see Gupta and Varga (1993). Thus, to carry out the proposed Bayesian inference, we also employ the following assumption.

For simplicity in explanation, we shall use Gaussian likelihood to obtain the associated posterior distributions for estimation. However, other elliptical distributions can also be used, and our proof of consistency of the proposed factor selection criteria continues to apply for a class of elliptical distributions. Furthermore, simulation results of Section 5 show that the proposed criteria evaluated under the Gaussian likelihood continue to work well when the true distribution has fat-tails. In what follows, we introduce our Bayesian estimation method.

For the factor model considered, the estimate of the covariance matrix,

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{T}(\mathbf{Y} - \widehat{\mathbf{F}}_k\widehat{\boldsymbol{\Lambda}}_k')'(\mathbf{Y} - \widehat{\mathbf{F}}_k\widehat{\boldsymbol{\Lambda}}_k'),$$

is singular when $T \leq N$. To handle this singularity problem, we entertain a singular multivariate normal distribution:

$$\pi(\boldsymbol{\varepsilon}_t|\boldsymbol{\Sigma}^-) = \frac{|\boldsymbol{\Sigma}^-|_+^{1/2}}{(2\pi)^{-(N-d)/2}} \exp\left\{-\frac{1}{2}\boldsymbol{\varepsilon}_t'\boldsymbol{\Sigma}^-\boldsymbol{\varepsilon}_t\right\}, \tag{6}$$

where d is a non-negative integer denoting the number of zero eigenvalues of $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}^-$ is a generalized inverse matrix and $|\boldsymbol{\Sigma}^-|_+$ is the product of the $(N-d)$ non-zero eigenvalues of $\boldsymbol{\Sigma}^-$. The natural estimate of $\boldsymbol{\Sigma}^-$ is then

$$\widehat{\boldsymbol{\Sigma}}^- = \widehat{\mathbf{X}}_{\widehat{\Sigma}}\widehat{\mathbf{L}}_{\widehat{\Sigma}}^{-1}\widehat{\mathbf{A}}_{\widehat{\Sigma}}',$$

where the $N \times (N-d)$ matrices $\widehat{\mathbf{X}}_{\widehat{\Sigma}}$ and $\widehat{\mathbf{A}}_{\widehat{\Sigma}}$ are given by the singular value decomposition of $\widehat{\boldsymbol{\Sigma}}$, namely the first $N-d$ columns of the matrices $\mathbf{X}$ and $\mathbf{A}$ of the singular value decomposition $\widehat{\boldsymbol{\Sigma}} = \mathbf{XLA}'$, and the $(N-d) \times (N-d)$ diagonal matrix $\mathbf{L}_{\widehat{\Sigma}}$ consists of the first $(N-d)$ non-zero singular values of $\widehat{\boldsymbol{\Sigma}}$. Note that other singular elliptical distributions can also be used.

Suppose that the number of factors is k. When the noises $\boldsymbol{\varepsilon}_t$ are independent, we use the following likelihood to facilitate the posterior sampling:

$$L(\mathbf{Y}|\mathbf{A}_k, \mathbf{L}_k, \mathbf{X}_k, \boldsymbol{\Sigma}^-) = \frac{|\boldsymbol{\Sigma}^-|_+^{T/2}}{(2\pi)^{-kT/2}} \exp\left[-\frac{1}{2}\mathrm{tr}\{K(\mathbf{Y}, \mathbf{X}_k, \mathbf{L}_k, \mathbf{A}_k)\boldsymbol{\Sigma}^-\}\right] \tag{7}$$

with $K(\mathbf{Y}, \mathbf{X}_k, \mathbf{L}_k, \mathbf{A}_k) = (\mathbf{Y} - \mathbf{X}_k\mathbf{L}_k\mathbf{A}_k')'(\mathbf{Y} - \mathbf{X}_k\mathbf{L}_k\mathbf{A}_k')$.

3.1. Prior specification

Using the likelihood, the posterior distribution of $\mathbf{A}_k$, $\mathbf{L}_k$, and $\mathbf{X}_k$ can be obtained via the Bayes rule as

$$\pi(\mathbf{A}_k, \mathbf{L}_k, \mathbf{X}_k, \boldsymbol{\Sigma}^-|\mathbf{Y}) \propto L(\mathbf{A}_k, \mathbf{L}_k, \mathbf{X}_k, \boldsymbol{\Sigma}^-|\mathbf{Y})\pi(\mathbf{A}_k, \mathbf{L}_k, \mathbf{X}_k, \boldsymbol{\Sigma}^-).$$

For ease in computation, we assume that the priors of the parameters are mutually independent, i.e.,

$$\pi(\mathbf{A}_k, \mathbf{L}_k, \mathbf{X}_k, \boldsymbol{\Sigma}^-) = \pi(\mathbf{A}_k)\pi(\mathbf{L}_k)\pi(\mathbf{X}_k)\pi(\boldsymbol{\Sigma}^-).$$

From $\mathbf{A}_k'\mathbf{A}_k = \mathbf{I}_k$, $\mathbf{A}_k$ belongs to the unit hyperball in N dimensions and its support is restricted to be the Cartesian product of the N-dimensional unit hyperballs. Furthermore, because of the orthogonality requirement, its support is then reduced to the Stiefel manifold $S_{N,k}$ (Khatri and Mardia, 1977; Smidl and Quinn, 2007). Simply speaking, the Stiefel manifold $S_{N,k}$ is the set of ordered k-tuples of orthonormal vectors in R^N. Therefore, each of the column vectors of $\mathbf{A}_k$ spans a N-dimensional plane in the k-dimensional coordinate system. Similar argument applies to $\mathbf{X}_k$, which belongs to the Stiefel manifold $S_{N,k}$ and thus each column vector of $\mathbf{X}_k$ spans a N-dimensional plane in the k-dimensional coordinate system. See also Strachan (2004) that gives a clear treatment of this issue with respect to cointegration models.

In our implementation, the priors of $\mathbf{X}_k$ and $\mathbf{A}_k$ are flat priors over the Stiefel manifold corresponding to orthogonal transformations and, hence, are invariant with respect to the orthogonal group. The prior for the diagonal matrix $\mathbf{L}_k$ is also flat over certain spaces. Specifically, the priors of $\mathbf{A}_k$ and $\mathbf{X}_k$ are

$$\pi(\mathbf{A}_k) = \frac{1}{C(N,k)} \cdot 1(\mathbf{A}_k \in S_{N,k}), \qquad \pi(\mathbf{X}_k) = \frac{1}{C(T,k)} \cdot 1(\mathbf{X}_k \in S_{T,k}),$$

where $1(\cdot)$ is the indicator function and

$$C(T,k) = \frac{2^k \pi^{kT/2}}{\pi^{k(k-1)/4} \prod_{j=1}^{k} \Gamma\{(T-j+1)/2\}}$$

is the normalizing constant with $\Gamma(.)$ being the Gamma function.

Because of the constraint on its elements, the support of $\mathbf{L}_k$ is $T_k = \{\mathbf{L}_k | l_1 > l_2 > \cdots > l_k > 0\}$. Therefore, the prior of $\mathbf{L}_k$ is

$$\pi(\mathbf{L}_k) = \frac{1}{C(k)} \cdot 1(\mathbf{L}_k \in T_k),$$

where $C(k)$ is the normalizing constant. Note that these prior settings are consistent with the model assumptions of Section 2.

It is noticed that the model in Eq. (1) is subject to the normalization condition either $\mathbf{F}_r'\mathbf{F}_r/T = \mathbf{I}_r$ or $\boldsymbol{\Lambda}_r'\boldsymbol{\Lambda}_r/N = \mathbf{I}_r$. Also, from the results of the asymptotic principal component analysis in (5), there exist the following relationships $\boldsymbol{\Lambda}_k = \sqrt{N}\mathbf{A}_k$ and $\mathbf{F}_k = \mathbf{X}_k\mathbf{L}_k/\sqrt{N}$. Therefore, the above prior corresponds to the restriction $\boldsymbol{\Lambda}_k'\boldsymbol{\Lambda}_k/N = \mathbf{A}_k'\mathbf{A}_k = \mathbf{I}_k$ on $\boldsymbol{\Lambda}_k$. Thus, $\pi(\boldsymbol{\Lambda}_k) = C(T,k)^{-1} \cdot 1(\boldsymbol{\Lambda}_k \in S_{T,k})$ in the original space. Noting that $\mathbf{F}_k'\mathbf{F}_k = \mathbf{L}_k\mathbf{X}_k\mathbf{X}_k\mathbf{L}_k/N = \mathbf{L}_k^2/N$, the prior density of $\mathbf{F}_k$, $\pi(\mathbf{F}_k)$ in the original space assumes that a set of k orthogonal column vectors of $\mathbf{F}_k$ spans a T-dimensional plane.

Following Box and Tiao (1972), a noninformative prior is used for $\boldsymbol{\Sigma}^-$:

$$\pi(\boldsymbol{\Sigma}^-) = |\boldsymbol{\Sigma}^-|_+^{-\frac{k+1}{2}}.$$

Consequently, the joint posterior distribution of the parameters is

$$\pi(\mathbf{A}_k, \mathbf{L}_k, \mathbf{X}_k, \boldsymbol{\Sigma}^-|\mathbf{Y}) \propto |\boldsymbol{\Sigma}^-|_+^{\frac{T+k+1}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left\{K(\mathbf{Y}, \mathbf{X}_k, \mathbf{L}_k, \mathbf{A}_k)\boldsymbol{\Sigma}^-\right\}\right] 1(\mathbf{A}_k \in S_{N,k}) \cdot 1(\mathbf{X}_k \in S_{T,k}) \cdot 1(\mathbf{L}_k \in T_k).$$

One can consider other forms of prior specification. We employ the prior specification above in this paper, because we are considering a situation that information available is limited.

The factor model discussed belongs to the class of reduced rank models and we adopt the singular value decomposition approach in this paper. This method has been applied in the econometric literature in Bayesian estimation of cointegration and instrumental variable models. See, for example, Kleibergen and van Dijk (1994, 1998), Kleibergen and Paap (2002, 2006) and Hoogerheide et al. (2007).

Remark. Since $\mathbf{A}_k$ belongs to the Stiefel manifold, that is, $\mathbf{A}_k \in S_{N,k}$, each of its column vectors spans an N-dimensional plane in the k-dimensional coordinate system. For a non-zero $N \times k$ matrix $\mathbf{P} \notin S_{N,k}$, one can use the transformation $\widetilde{\mathbf{P}} = \mathbf{P}(\mathbf{P}'\mathbf{P})^{-1/2}$ to obtain a matrix in $S_{N,k}$. Obviously, each column vector of $\mathbf{P}$ cannot be a zero vector and should not be dependent, either. This Stiefel manifold was employed in several studies before. For example, Strachan and Inder (2004) use the method in estimating the posterior probability density of the cointegrating rank of a multivariate error correction model, Strachan (2007) employs it in Bayesian inference of cointegrated $I(2)$ systems, Koop et al. (2006) apply it in Bayesian analysis of cointegrated models with priors on the cointegration space, and Smidl and Quinn (2007) use it in Bayesian principal component analysis.

3.2. Posterior sampling

Parameter estimation using the proposed posterior distribution can be achieved by the Markov chain Monte Carlo (MCMC) algorithm, which is commonly used in the statistical literature. We summarize the MCMC algorithm used in this paper as follows.

Sampling algorithm:

Step 1. Initialize the parameters $\mathbf{A}_k$, $\mathbf{L}_k$, and $\mathbf{X}_k$.
Step 2. Sample $\mathbf{A}_k$ from $\mathbf{A}_k|\mathbf{L}_k, \mathbf{X}_k, \boldsymbol{\Sigma}^-, \mathbf{Y}$.
Step 3. Sample $\mathbf{L}_k$ from $\mathbf{L}_k|\mathbf{A}_k, \mathbf{X}_k, \boldsymbol{\Sigma}^-, \mathbf{Y}$.
Step 4. Sample $\mathbf{X}_k$ from $\mathbf{X}_k|\mathbf{A}_k, \mathbf{L}_k, \boldsymbol{\Sigma}^-, \mathbf{Y}$.
Step 5. Sample $\boldsymbol{\Sigma}^-$ from $\mathbf{X}_k|\mathbf{A}_k, \mathbf{L}_k, \mathbf{X}_k, \mathbf{Y}$.
Step 6. Repeat Step 2–Step 5 for a sufficiently large number of iterations.

In the above algorithm, $\mathbf{Z}|\mathbf{W}$ denotes the conditional posterior distribution of $\mathbf{Z}$ given $\mathbf{W}$. Since the conditional posterior distributions of $\mathbf{A}_k$, $\mathbf{L}_k$, and $\mathbf{X}_k$ do not have a closed form, we use the Metropolis–Hasting (MH) algorithm to conduct the random draws. Specifically, using random draws from proposal distributions, we apply the MH algorithm to implement Steps 2–4. For instance, in Step 2, the conditional posterior density function of $\mathbf{A}_k$ is

$$\pi(\mathbf{A}_k|\mathbf{L}_k, \mathbf{X}_k, \boldsymbol{\Sigma}^-, Y) \propto L(\mathbf{Y}|\mathbf{A}_k, \mathbf{L}_k, \mathbf{X}_k, \boldsymbol{\Sigma}^-)\pi(\mathbf{A}_k).$$

After the jth iteration, we make a candidate draw of $\mathbf{A}_{k,*}^{(j+1)}$ using the proposal density function $p(\mathbf{A}_k|\widehat{\mathbf{A}}_k)$ centered at the posterior mode $\widehat{\mathbf{A}}_k$. We then accept the candidate draw with the probability

$$\alpha = \min\left\{1, \frac{\pi(\mathbf{A}_{k,*}^{(j+1)}|\mathbf{L}_k, \mathbf{X}_k, \boldsymbol{\Sigma}^-, \mathbf{Y})/p(\mathbf{A}_{k,*}^{(j+1)}|\widehat{\mathbf{A}}_k)}{\pi(\mathbf{A}_k^{(j)}|\mathbf{L}_k, \mathbf{X}_k, \boldsymbol{\Sigma}^-, \mathbf{Y})/p(\mathbf{A}_k^{(j)}|\widehat{\mathbf{A}}_k)}\right\}.$$

That is, for the $(j+1)$th iteration, $\mathbf{A}_k^{(j+1)}$ is determined by

$$\mathbf{A}_k^{(j+1)} = \begin{cases} \mathbf{A}_{k,*}^{(j+1)} & \text{with probability } \alpha, \\ \mathbf{A}_k^{(j)} & \text{with probability } 1-\alpha. \end{cases}$$

Many proposal densities are available to draw $\mathbf{A}_k$. Here we use the matrix-variate normal distribution

$$p(\mathbf{A}_k|\widehat{\mathbf{A}}_k) \propto \exp\left\{-\frac{1}{2}\boldsymbol{\Sigma}_{A_k}^{-1}(\mathbf{A}_k - \widehat{\mathbf{A}}_k)'(\mathbf{A}_k - \widehat{\mathbf{A}}_k)\right\}$$

with constraint $\mathbf{A}_k'\mathbf{A}_k = \mathbf{I}_k$, where the covariance matrix $\boldsymbol{\Sigma}_{A_k}$ should be set by user so that MH algorithm constructs a Markov chain that has the desired distribution as its equilibrium distribution. In other words, an desirable covariance matrix $\boldsymbol{\Sigma}_{A_k}$ depend on the problem and therefore are set by user's hand. For more details on MCMC method, we refer to Carlin and Louis (1996), Gilks et al. (1996) and Tierney (1994). In our Monte Carlo study, $\boldsymbol{\Sigma}_{A_k}$ is a diagonal matrix defined as $\boldsymbol{\Sigma}_{A_k} = b\mathbf{I}$ for some positive real number b. The normalization constants of the probability densities are not needed in calculating α.

The same MH algorithm is used in Steps 3 and 4. The proposal densities used for $\mathbf{X}_k$ and $\mathbf{L}_k$ are also matrix-variate normal distributions given by

$$p(\mathbf{X}_k|\widehat{\mathbf{X}}_k) \propto \exp\left\{-\frac{1}{2}\boldsymbol{\Sigma}_{X_k}^{-1}(\mathbf{X}_k - \widehat{\mathbf{X}}_k)'(\mathbf{X}_k - \widehat{\mathbf{X}}_k)\right\},$$

$$p(\mathbf{L}_k|\widehat{\mathbf{L}}_k) \propto \exp\left\{-\frac{1}{2}\boldsymbol{\Sigma}_{L_k}^{-1}(\mathbf{L}_k - \widehat{\mathbf{L}}_k)'(\mathbf{L}_k - \widehat{\mathbf{L}}_k)\right\}$$

with constraints $\mathbf{X}_k'\mathbf{X}_k = \mathbf{I}_k$ and $\{\mathbf{L}_k|l_1 > l_2 > \cdots > l_k > 0\}$, where the covariance matrices are given by $\boldsymbol{\Sigma}_{X_k} = \boldsymbol{\Sigma}_{L_k} = b\mathbf{I}$ with b being a positive number. Once random draws of $\mathbf{A}_k$, $\mathbf{L}_k$ and $\mathbf{X}_k$ are available, the posterior samples of $\mathbf{F}_k$ and $\boldsymbol{\Lambda}_k$ can be obtained by using the expression of the asymptotic principal components estimates (4).

In Step 5, the conditional posterior density function of $\boldsymbol{\Sigma}^-$ is available analytically. From the functional form of the joint posterior distribution of the parameters $\pi(\mathbf{A}_k, \mathbf{L}_k, \mathbf{X}_k, \boldsymbol{\Sigma}^-|\mathbf{Y})$, we can easily see that the conditional posterior density function of $\boldsymbol{\Sigma}^-$ is a generalized inverted Wishart distribution with parameter $K(\mathbf{Y}, \mathbf{X}_k, \mathbf{L}_k, \mathbf{A}_k)$, $T+k+1$ and N (See, e.g. Diaz-Garcia and Gutierrez-Jaimez, 2006). Therefore, we employed the Gibbs sampling algorithm in Step 5.

The outcomes of the MCMC–MH algorithm can be regarded as a random sample from the joint posterior density function after a burn-in period. In our implementation, we set all initial parameter values to the posterior modes of the data, $\widehat{\mathbf{X}}_k$, $\widehat{\mathbf{A}}_k$ and $\widehat{\mathbf{L}}_k$, given in the asymptotic principal component estimators (5). We then run the MCMC algorithm for 2000 iterations with the first 1000 iterations as the burn-in period. In other words, our inference is based on the random draws of the last 1000 iterations of the MCMC algorithm.

Note that, in principal, we can also treat k as a random variable in Bayesian analysis and obtain its posterior draws via MCMC methods. However, for computational simplicity, we take an empirical Bayes approach in this paper and determine the number of factors k using model selection criteria.

The remaining problem is to evaluate the goodness-of-fit of an estimated model with k common factors. To this end, we assess the closeness of the estimated model to the true data generating model from a predictive point of view in the next section.

Remark. To obtain a draw on the Stiefel manifold we can make use of the property discussed in the prior remark. Specifically, to make a draw $\mathbf{A}_k$ from the proposal density $p(\mathbf{A}_k|\widehat{\mathbf{A}}_k)$ with constraint $\mathbf{A}_k'\mathbf{A}_k = \mathbf{I}_k$, we perform the following: First, generate a draw $\widetilde{\mathbf{A}}_k = \widehat{\mathbf{A}}_k + \mathbf{Z}$ with $\mathbf{Z}$ following the matrix-variate normal density with mean O matrix and variance matrix $\boldsymbol{\Sigma}_{A_k}$. Second, perform a simple transformation $\mathbf{A}_k = \widetilde{\mathbf{A}}_k(\widetilde{\mathbf{A}}_k'\widetilde{\mathbf{A}}_k)^{-1/2}$. Consider, for example, $T = 2$ and $k = 1$. The matrix $\mathbf{A}_k$ reduces to the 2×1 vector $\boldsymbol{a} = (a_1, a_2)'$. We make a draw $\tilde{\boldsymbol{a}} = \hat{\boldsymbol{a}} + \boldsymbol{z}$ with $\boldsymbol{z} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{a_k})$. Then we obtain a draw $\boldsymbol{a}$ by using a simple transformation $\boldsymbol{a} = (\tilde{a}_1/\sqrt{\tilde{a}_1^2 + \tilde{a}_2^2}, \tilde{a}_2/\sqrt{\tilde{a}_1^2 + \tilde{a}_2^2})'$. In a similar manner, we can make a draw $\mathbf{X}_k$ from the proposal density $p(\mathbf{X}_k|\widehat{\mathbf{X}}_k)$ with constraint $\mathbf{X}_k'\mathbf{X}_k = \mathbf{I}_k$.

4. Model selection

In this section we consider the important issue of selecting the number of factors to adequately describe the information contained in a panel data set. In the literature, several methods have been proposed to select the number of factors, e.g., Bai and Ng (2002). However, these methods are basically constructed by using point estimates of the model parameters. The Bayesian estimation discussed in Section 3 enables us to consider parameter uncertainty in the model selection process. An example of Bayesian model selection criterion is Ando (2007), which considers the maximization of the posterior mean of the expected log-likelihood. In this paper, we combine the ideas of Bai and Ng (2002) and Ando (2007) to derive criteria for selecting the number of factors. Thus, our approach combines recent developments in the econometric and statistical literature.

Theorem 1. *Suppose that the model Assumptions A–D hold and that the k factor model is estimated by the Bayesian principal component procedure. Let $\hat{k}$ be the minimizer of the following C_p-type criterion for panel data*

$$PDC_p(k) = \overline{S(k, \mathbf{F}_k)} + k \times g(T, N),$$

where $\overline{S(k, \mathbf{F}_k)} = \int S(k, \mathbf{F}_k)\pi(\mathbf{F}_k, \boldsymbol{\Lambda}_k, \boldsymbol{\Sigma}^-|\mathbf{Y})d\mathbf{F}_k d\boldsymbol{\Lambda}_k d\boldsymbol{\Sigma}^-$ is the posterior mean of

$$S(k, \mathbf{F}_k) = \frac{1}{NT}\mathrm{tr}\{(\mathbf{Y} - \mathbf{F}_k\boldsymbol{\Lambda}_k')'(\mathbf{Y} - \mathbf{F}_k\boldsymbol{\Lambda}_k')\},$$

with given value of $\mathbf{Y}$. Then $\lim_{T,N\to\infty} P[\hat{k} = r] = 1$, if (a): $g(T, N) \to 0$ and (b): $D_{TN} \times g(T, N) \to \infty$ as $T, N \to \infty$, with $D_{TN} = \min\{T^{1/2}, N\}$.

The derivation of the theorem is given in Appendix provided the Supplemental file.

An example of the function $g(T, N)$ that satisfies the conditions (a) and (b) of the theorem is

$$g_1(T, N) = \left(\frac{\sqrt{T+N}}{\sqrt{TN}}\right)\log\left(\frac{\sqrt{TN}}{\sqrt{T+N}}\right).$$

Putting $g_1(T, N)$ into the information criterion function, we have the following model selection criterion:

$$\mathrm{PDC}_p(k) = \overline{S(k, \mathbf{F}_k)} + k \times \left(\frac{\sqrt{T+N}}{\sqrt{TN}}\right)\log\left(\frac{\sqrt{TN}}{\sqrt{T+N}}\right). \tag{8}$$

We can choose the number of factors by minimizing the $PDC_p(k)$ over a specified range of k. We refer to the criterion in Eq. (8) as the Panel Data C_p (PDC_p) criterion. In our implementation, the posterior mean $\overline{S(k, \mathbf{F}_k)}$ of $S(k, \mathbf{F}_k)$ is estimated by the sample mean of the realizations $S(k, \mathbf{F}_k)$ over the last 1000 MCMC iterations.

Recently, Hallin and Liska (2007) use a frequentist approach to factor selection and pointed out that a problem of information criteria such as those in our paper and Bai and Ng (2002) is that the penalty can be multiplied by an arbitrary constant without hurting the consistency. This criticism is well taken. In this paper, we also consider parameter uncertainty and the finite-sample performance of the criteria via simulation. In Theorem 1, the condition (b) differs slightly from that of Bai and Ng (2002). This is due to a technical issue in our proof.

Next, by using similar arguments of Bai and Ng (2002), we obtain the following theorem.

Theorem 2. *Under the Assumptions of Theorem 1, the panel data information criterion defined by*

$$PDIC(k) = \log\left\{\overline{S(k, \mathbf{F}_k)}\right\} + k \times g(T, N),$$

also provides a consistent estimate of the true number of factors r.

Again, the derivation of the theorem is given in the Appendix. Using Theorem 2, we can immediately obtain the following criterion

$$PDIC(k) = \log\{\overline{S(k, \mathbf{F}_k)}\} + k \times \left(\frac{T+N}{TN}\right) \log\left(\frac{TN}{T+N}\right), \tag{9}$$

and choose the number of factors by minimizing the $PDIC(k)$ score over a specified range of k. This criterion is referred to as the Panel Data Information Criterion (PDIC). Because the goodness-of-fit functions of the proposed criteria PDC_p and PDIC are different, it is sensible to use the different penalty terms. Note that the penalty term

$$g_2(T, N) = \left(\frac{T+N}{TN}\right) \log\left(\frac{TN}{T+N}\right)$$

also satisfies the conditions (a) and (b) of the theorem.

The PDIC of Eq. (9) is very close to the IC_{BN}^1 criterion of Bai and Ng (2002); see the definition in the next section. However, the most important difference between PDIC and the criteria of Bai and Ng (2002) is that when we evaluate the goodness-of-fit measure, $S(k, \mathbf{F}_k)$, PDIC uses a set of posterior samples whereas Bai and Ng (2002) just employs $S(k, \widehat{\mathbf{F}}_k)$, which is evaluated at $\widehat{\mathbf{F}}_k$. Our criteria, therefore, explicitly include the parameter uncertainty in the analysis, i.e., the uncertainty of $S(k, \mathbf{F}_k)$, whereas the criteria of Bai and Ng (2002) do not. We expect that our criteria to be more stable and have better performance. See the performance of the criterion $\widetilde{PC}_{BN}^1(k)$ of Section 5.

Remark. In the Bayesian model selection context, several other criteria are available, including the deviance information criterion (DIC) of Spiegelhalter et al. (2002), the pseudo Bayes factor with resampling approach of Gelfand (1996) and Gelfand and Dey (1994), and a minimum posterior predictive loss approach of Gelfand and Ghosh (1998). However, as shown in the derivation of DIC (Spiegelhalter et al., 2002, p. 604) a normal approximation to the likelihood is assumed. It should be noticed that the number of model parameters in the factor model of Eq. (1) might greatly exceed the sample size T. This situation is considered in the simulation study, e.g., $(T, N) = (60, 1000)$. In such a case, the above assumption of normal approximation to the likelihood cannot hold, resulting in the unavailability of DIC. This argument also applies to the BPIC of Ando (2007) and the predictive likelihood approach of Ando and Tsay (2010).

Also, one might employ the pseudo Bayes factor with the resampling approach. However, the computation becomes very intensive when the sample size T is large. Again, this situation is considered in the simulation study, e.g., $(T, N) = (1000, 60)$. Although this approach is robust to assumption violations, it might not be practically feasible since we run MCMC many times due to the use of cross-validation.

5. Numerical results

To evaluate the performance of the proposed factor selection criteria in Eqs. (8) and (9), we conduct some simulation studies. The true data generating process used in the simulation starts with a simple Gaussian factor model, but is extended to include cases with non-homoscedastic noises, serially correlated noises, dynamic factor models with and without serially correlated noises, and heavy-tailed noises.

5.1. Other selection criteria

For comparison purpose, we also employ the following ten criteria of Bai and Ng (2002) and Onatski (2005) in the study:

(1) Bai and Ng (2002)

$$PC_{BN}^1(k) = S(k, \widehat{\mathbf{F}}_k) + k \cdot S(k_{\max}, \widehat{\mathbf{F}}_{k_{\max}}) \left(\frac{T+N}{TN}\right) \log\left(\frac{TN}{T+N}\right).$$

(2) Bai and Ng (2002)

$$PC_{BN}^2(k) = S(k, \widehat{\mathbf{F}}_k) + k \cdot S(k_{\max}, \widehat{\mathbf{F}}_{k_{\max}}) \left(\frac{T+N}{TN}\right) \log C_{TN}^2.$$

(3) Bai and Ng (2002)

$$PC_{BN}^3(k) = S(k, \widehat{\mathbf{F}}_k) + k \cdot S(k_{\max}, \widehat{\mathbf{F}}_{k_{\max}}) C_{TN}^{-2} \log C_{TN}^2.$$

(4) Bai and Ng (2002)

$$IC_{BN}^1(k) = \log S(k, \widehat{\mathbf{F}}_k) + k \cdot \left(\frac{T+N}{TN}\right) \log\left(\frac{TN}{T+N}\right).$$

(5) Bai and Ng (2002)

$$IC_{BN}^2(k) = \log S(k, \widehat{\mathbf{F}}_k) + k \cdot \left(\frac{T+N}{TN}\right) \log C_{TN}^2.$$

(6) Bai and Ng (2002)

$$IC^3_{BN}(k) = \log S(k, \widehat{\mathbf{F}}_k) + k \cdot C_{TN}^{-2} \log C_{TN}^2.$$

(7) Akaike (1974)

$$AIC(k) = S(k, \widehat{\mathbf{F}}_k) + k \cdot S(k_{\max}, \widehat{\mathbf{F}}_{k_{\max}}) \left(\frac{2}{T}\right).$$

(8) Schwarz (1978)

$$BIC(k) = S(k, \widehat{\mathbf{F}}_k) + k \cdot S(k_{\max}, \widehat{\mathbf{F}}_{k_{\max}}) \left(\frac{\log(T)}{T}\right).$$

(9) Bai and Ng (2002)

$$BIC3(k) = S(k, \widehat{\mathbf{F}}_k) + k \cdot S(k_{\max}, \widehat{\mathbf{F}}_{k_{\max}}) \left(\frac{(N+T-k)\log(NT)}{NT}\right).$$

(10) Onatski (2005)

$$\tilde{r}(\delta, k_{\max}, \xi_N) = \#\{i \le N; \lambda_i > (1+\delta)\hat{u}\},$$

where $k_{\max}$ is a prespecified maximum number of factors and $C_{TN} = \min(\sqrt{T}, \sqrt{N})$. For Onatski's method, $\#\{\cdot\}$ means the number of elements that satisfy the stated condition, and λ_i is the ith largest eigenvalues of $T^{-1}\mathbf{Y}'\mathbf{Y}$. Following the suggestion of Onatski (2005), we set $k_{\max} = \xi_N = [8 \cdot (T/100)^{1/4}]$, where $[\cdot]$ is the integer part of the number. Onatski (2005) reported that these choices work well for all sample sizes. For the values of $\hat{u}$ and δ, we refer to Onatski (2005). Bai and Ng (2002) used the above AIC and BIC criteria, for their penalty terms are the standard ones in the time series analysis. Our study indicates that both $AIC(k)$ and $BIC(k)$ fare poorly compared with other criteria in selecting the number of factors.

It is clear from the definitions that the criteria of Bai and Ng (2002) do not take into account the parameter uncertainty. To demonstrate the importance of considering parameter uncertainty in factor selection, we therefore employ the following modified version of Bai and Ng (2002)'s PC criteria:

(11)

$$\widetilde{PC}^1_{BN}(k) = S(k, \widehat{\mathbf{F}}_k) + k\left(\frac{\sqrt{T+N}}{\sqrt{TN}}\right) \log\left(\frac{\sqrt{TN}}{\sqrt{T+N}}\right),$$

where the penalty term of $PC^1_{BN}(k)$ is replaced by the penalty term of the proposed PDC_p. By comparing the performance between PDC_p and $\widetilde{PC}^1_{BN}(k)$ in the simulation, we can observe the effect of parameter uncertainty.

5.2. Data generating processes

In our Monte Carlo study, to assess the general properties of the criteria considered, we use various configurations of T, N, and the true number of factors r. For simplicity, we shall report the results for $r = 3$ only, because other values of r have similar results. The first data generating model considered is

$$\boldsymbol{y}_t = \Lambda \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T, \tag{10}$$

where the r-dimensional factor $\boldsymbol{f}_t$ is a vector of $N(0, 1)$ variables, each element of the factor loading matrix $\boldsymbol{\Lambda}$ also follows $N(0, 1)$, and the N-dimensional noise vector $\boldsymbol{\varepsilon}_t$ is distributed as multivariate normal with mean $\mathbf{0}$ and variance $r\mathbf{I}_N$. This is the ideal case for the criteria to work well.

The second data generating model considered is

$$\boldsymbol{y}_t = \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{\varepsilon}^1_t + \delta_t \boldsymbol{\varepsilon}^2_t, \quad t = 1, \ldots, T, \tag{11}$$

where $\delta_t = 1$ if t is odd and it is zero if t is even, and the noises $\boldsymbol{\varepsilon}^1_t$ and $\boldsymbol{\varepsilon}^2_t$ follow $N(\mathbf{0}, r\mathbf{I}_N)$ and are independent. The factor vector $\mathbf{f}_t$ and the loading matrix $\boldsymbol{\Lambda}$ were generated by the same method as before. The key feature of model (11) is that the noises are not homoscedastic.

As the third example, we investigated the performance of the proposed criteria when the idiosyncratic errors have some serial correlations. The model is $\boldsymbol{y}_t = \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t + \rho \boldsymbol{\varepsilon}_{t-1}$, with $t = 1, \ldots, T$, where $\rho = 0.5$, and other variables are defined as before.

Next, consider the following data generating process:

$$\boldsymbol{y}_t = \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{f}_t = \boldsymbol{\Phi} \boldsymbol{f}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\varepsilon}_t = 0.5\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\xi}_t, \tag{12}$$

where the factors follow an AR(1) model, $\boldsymbol{\Phi}$ is a diagonal matrix with diagonal elements $\phi_{ii} = 0.5$, and $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t$ are mutually independent and serially uncorrelated noises each element of which follows a Student-t distribution with 5 degrees of freedom. The elements of the factor loading matrix $\boldsymbol{\Lambda}$ follow the standard normal distribution. This is a non-Gaussian case with heavy tails and serial correlations in the noises.

Amengual and Watson (2007) showed that the estimators of Bai and Ng (2002) can be modified to consistently estimate the number of dynamic factors in a restricted dynamic factor model. The modification is essentially to remove the serial dependence in the fitted factors. Suppose that the number of factors is k. Let $\widehat{\mathbf{F}}_k$ be the asymptotic principal component estimator of $\mathbf{F}_k$ and $\hat{\boldsymbol{f}}_t$ is the tth row of $\widehat{\mathbf{F}}_k$. Consider the vector AR(1) model $\hat{\boldsymbol{f}}_t = \boldsymbol{\phi}\hat{\boldsymbol{f}}_{t-1} + \mathbf{e}_t$. Let $\widehat{\boldsymbol{\Phi}}$ be the ordinary least squares estimator of $\boldsymbol{\phi}$. The criteria of Bai and Ng (2002) are then calculated using the residuals $\hat{\boldsymbol{y}}_t = \boldsymbol{y}_t - \widehat{\boldsymbol{\Lambda}}_k\widehat{\boldsymbol{\Phi}}\hat{\boldsymbol{f}}_{t-1}$ as the observed data. These modified criteria are used to select the number of factors. In a similar manner, the proposed criteria can also be applied to the residuals $\hat{\boldsymbol{y}}_t$ to select the number of dynamic factors in a restricted dynamic factor model.

Finally, we consider the following data generating processes:

$$\boldsymbol{y}_t = \boldsymbol{\Lambda}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{f}_t = \boldsymbol{\Phi}\boldsymbol{f}_{t-1} + \boldsymbol{\eta}_t, \tag{13}$$

where $\boldsymbol{\varepsilon}_t = 0.5\boldsymbol{\varepsilon}_{t-1} + 0.05\boldsymbol{v}_t + \sum_{j=-8, j\neq 0}^{8} 0.01\boldsymbol{v}_{t+j}$, $\boldsymbol{\Phi}$ is a diagonal matrix with diagonal elements $\phi_{ii} = 0.5$, and each element of the noise vector $\boldsymbol{\eta}_t$ and the innovation vectors $\boldsymbol{v}_t$ of the noise also follows a Student-t distribution with 5 degrees of freedom. Elements of the factor loading matrix $\boldsymbol{\Lambda}$ follow a normal distribution with mean 0 and variance $\sqrt{r}$. Thus, the true model has dynamic dependence in both the factors and the noises, and the innovations to the factors and noises all have heavy tails.

Except for Onatski's method, the minimum and maximum numbers of factors are set as $k_{\min} = 0$ and $k_{\max} = 8$, respectively, for all cases. The value of $k_{\max} = 8$ was used in Bai and Ng (2002). For Onatski's method, we follow Onatski (2005) to set the parameters. For instance, $k_{\max} = [8(T/100)^{1/4}]$.

5.3. Results

Table 1 reports the percentages of under-, correct, and over-identification for all criteria considered under the first data generating model. The percentages are based on 1000 replications for each combination of (T, N). From the table, we make the following observations. First, the proposed PDC_p criterion is capable of selecting the data generating model even when T and N are small. The method of Onatski (2005), the proposed PDIC, IC_{BN}^1 of Bai and Ng (2002), BIC3, and $\widetilde{PC}_{BN}^1$ also work well. On the other hand, all other criteria fail to select the true model in some situations. Second, as expected, the proposed PDIC criterion and the IC_{BN}^1 of Bai and Ng (2002) behave similarly, except for the case $(T, N) = (200, 25)$. These two criteria perform well in most cases. Third, both AIC and BIC fare poorly in the simulation. Fourth, the performance of PC_{BN} and IC_{BN} criteria of Bai and Ng (2002) also becomes much improved when T and N increase. Finally, for this simple model, PDC_p clearly outperforms $\widetilde{PC}_{BN}^1$, showing the importance of treating parameter uncertainty in factor selection. Overall, the table shows that the proposed criteria perform well for the data generating model of Eq. (10). They also show that the method of Onatski (2005) works well in most cases, but is not as powerful as the proposed method when T and N increase.

Table 2 summarizes the percentages of under-, correct, and over-identification of various criteria for the second data generating model over 1000 replications. From the table, the proposed PDC_p criterion, the IC_{BN}^1 of Bai and Ng (2002), and the $\tilde{r}$ of Onatski (2005) work well whereas PDIC, PC_{BN}^1 of Bai and Ng (2002), and $BIC3$ fare well when both T and N are large. On the other hand, PDIC, IC_{BN}^1 and $BIC3$ encounter some underspecification when T or N is small, and the IC_{BN}^3 and $\widetilde{PC}_{BN}^1$ have severe underspecification. Again, the inferior performance of $\widetilde{PC}_{BN}^1$ to PDC_p shows the advantages of considering parameter uncertainty in factor selection.

The results of the third data generating model are given in Table 3. Again, there are based on 1000 replications. From the table, the proposed PDIC seems to be the best with IC_{BN}^2 and $BIC3$ as close second. The proposed PDC_p works nicely when both T and N are greater than 50. The $\tilde{r}$ of Onatski (2005) continues to show selection power when both T and N are large, but it also encounters some over-specification. Finally, the PC_{BN}^1 criterion only works well when both N and T are greater than 100 in this particular case.

The results of the fourth data generating model, over 1000 replications, are summarized in Table 4. Here both the common factors and noises are serially correlated. Thus, except for the method of Onatski (2005), all criteria are applied to the transformed data $\hat{\boldsymbol{y}}_t$. From the table, the PDC_p outperforms other criteria with PDIC, $BIC3$ and the method of Onatski (2005) as second. In this particular, $\widetilde{PC}_{BN}^1$ works well when N and T are large whereas the method of Onatski (2005) suffers some power loss for all configurations of N and T. Other criteria considered encounter some severe over-specification.

Finally, Table 5 summarizes the results for the fifth data generating model over 1000 replications. In this particular case, both the common factors and noises have serial dependence and the innovations to the noises are serially correlated and have heavy tails. Again, because of the dynamic dependence, all other criteria, except the method of Onatski (2005), are applied to the transformed data $\hat{\boldsymbol{y}}_t$. From the table, only the proposed criteria are capable of identifying the true number of factors. All other criteria over-specify the number of factors in the data.

In summary, our simulation results show that the proposed criteria, especially PDC_p, outperform other criteria in selecting the number of factors in a panel data set. The method of Onatski (2005) also works for most data generating models employed but it suffers some power loss compared with PDC_p when N and T are of moderate size. Except for the most difficult model, the $BIC3$ criterion of Bai and Ng (2002) works well when N and T are large.

Table 1
Frequencies of under-, correct, and over-identification (in %), of various factor selection criteria over 1000 replications. The data are generated from the model $\boldsymbol{y}_t = \boldsymbol{\Lambda}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t$ with $r = 3$ true common factors and $s = \sqrt{3}$. The r-dimensional factor $\boldsymbol{f}_t$ is a vector of $N(0, 1)$ variables, each element of the factor loading matrix $\boldsymbol{\Lambda}$ also follows $N(0, 1)$, and the N-dimensional noise vector $\boldsymbol{\varepsilon}_t$ is distributed as multivariate normal with mean $\mathbf{0}$ and variance $s^2\mathbf{I}_N$. Because the results of PC^2_{BN} and IC^2_{BN} are similar to those of PC^3_{BN} and IC^3_{BN}, they are omitted from the table.

T/N	PDC_p			$PDIC$			PC^1_{BN}			PC^3_{BN}			IC^1_{BN}			IC^3_{BN}			AIC			BIC			$BIC3$			$\tilde{r}$			$\widetilde{PC}^1_{BN}$		
25/25	0	68	32	21	79	0	0	0	100	0	0	100	15	81	4	94	6	0	0	0	100	0	0	100	0	0	100	1	38	61	0	0	100
25/50	0	100	0	4	96	0	0	0	100	0	0	100	4	96	0	93	7	0	0	0	100	0	0	100	0	72	28	0	65	35	0	98	2
25/100	0	100	0	2	98	0	0	0	100	0	0	100	2	98	0	92	8	0	0	0	100	0	0	100	20	80	0	0	89	11	2	98	0
25/200	0	98	2	1	99	0	0	50	50	0	5	95	1	99	0	91	9	0	0	0	100	0	5	95	0	100	0	0	95	5	3	97	0
50/25	0	100	0	9	91	0	0	0	100	0	0	100	9	91	0	90	10	0	0	0	100	0	0	100	0	76	24	0	77	23	0	97	3
50/50	0	100	0	0	100	0	13	87	0	0	0	100	0	100	0	86	14	0	0	0	100	0	0	100	0	100	0	0	79	21	5	95	0
50/100	0	100	0	0	100	0	0	100	0	0	3	97	0	100	0	81	19	0	0	0	100	0	3	97	0	100	0	0	91	9	4	96	0
50/200	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	78	22	0	0	0	100	0	100	0	0	100	0	0	93	7	12	88	0
100/25	0	100	0	3	97	0	0	0	100	0	0	100	2	98	0	98	2	0	0	0	100	0	0	100	0	100	0	0	94	6	1	99	0
100/50	0	100	0	0	100	0	0	100	0	0	1	99	0	100	0	86	14	0	0	0	100	0	0	100	0	100	0	0	86	14	7	93	0
100/100	0	100	0	0	100	0	0	100	0	0	0	100	0	100	0	43	57	0	0	0	100	0	0	100	0	100	0	0	92	8	5	95	0
100/200	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	36	64	0	0	0	100	0	100	0	0	100	0	0	96	4	4	96	0
200/25	3	97	0	56	44	0	54	46	0	0	7	93	0	100	0	89	11	0	0	0	100	0	0	100	22	78	0	0	96	4	4	96	0
200/50	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	91	9	0	0	0	100	0	0	100	0	100	0	0	93	7	9	91	0
200/100	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	33	67	0	0	0	100	0	0	100	0	100	0	0	94	6	4	96	0
200/200	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	0	100	0	100	0	0	100	0	0	95	5	0	100	0
60/1000	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	0	100	0	100	0	0	100	0	0	98	2	14	86	0
1000/60	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	0	100	0	0	100	0	100	0	0	98	2	11	89	0

Table 2

Frequencies of under-, correct, and over-identification (in %) of various factor selection criteria over 1000 Replications. The data are generated from the model $\boldsymbol{y}_t = \boldsymbol{\Lambda}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t^1 + \delta_t \boldsymbol{\varepsilon}_t^2$ with $r = 3$ common factors, $s = \sqrt{3}$, $\delta_t = 1$ if t is odd and it is zero if t is even, and the noises $\boldsymbol{\varepsilon}_t^1$ and $\boldsymbol{\varepsilon}_t^2$ follow $N(\mathbf{0}, s^2\mathbf{I}_N)$ and are independent. The r-dimensional factor $\boldsymbol{f}_t$ is a vector of $N(0, 1)$ variables, and each element of the factor loading matrix $\boldsymbol{\Lambda}$ also follows $N(0, 1)$. Because the results of PC_{BN}^2 and IC_{BN}^2 are similar to those of PC_{BN}^3 and IC_{BN}^3, they are omitted from the table.

T/N	PDC_p			$PDIC$			PC_{BN}^1			PC_{BN}^3			IC_{BN}^1			IC_{BN}^3			AIC			BIC			$BIC3$			$\tilde{r}$			$\widetilde{PC}_{BN}^1$		
25/25	0	0	100	52	44	4	0	0	100	0	0	100	26	38	36	99	1	0	0	0	100	0	0	100	0	0	100	1	19	80	0	0	100
25/50	0	0	100	29	71	0	0	0	100	0	0	100	30	70	0	99	1	0	0	0	100	0	0	100	15	85	0	0	49	51	0	60	40
25/100	0	10	90	13	87	0	0	0	100	0	0	100	12	88	0	100	0	0	0	0	100	0	0	100	20	80	0	0	83	17	4	96	0
25/200	1	54	45	5	92	3	0	0	100	0	0	100	8	92	0	100	0	0	0	0	100	0	0	100	100	0	0	0	99	1	13	87	0
50/25	0	0	100	43	57	0	0	0	100	0	0	100	22	78	0	100	0	0	0	0	100	0	0	100	0	88	12	1	66	33	1	87	12
50/50	0	74	26	11	89	0	0	0	100	0	0	100	7	93	0	100	0	0	0	0	100	0	0	100	11	89	0	0	72	28	26	74	0
50/100	0	100	0	3	97	0	0	85	15	0	0	100	3	97	0	100	0	0	0	0	100	0	0	100	16	84	0	0	87	13	57	43	0
50/200	0	100	0	0	100	0	0	100	0	0	91	9	0	100	0	100	0	0	0	0	100	0	91	9	28	72	0	0	98	2	66	34	0
100/25	0	86	14	41	59	0	0	0	100	0	0	100	17	83	0	100	0	0	0	0	100	0	0	100	11	89	0	0	86	14	12	88	0
100/50	0	100	0	5	95	0	0	97	3	0	0	100	0	100	0	100	0	0	0	0	100	0	0	100	33	67	0	0	83	17	62	38	0
100/100	0	100	0	0	100	0	0	100	0	0	0	100	0	100	0	100	0	0	0	0	100	0	0	100	0	100	0	0	86	14	80	20	0
100/200	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	100	0	0	0	0	100	0	100	0	0	100	0	0	90	10	74	26	0
200/25	0	100	0	87	13	0	0	26	74	0	0	100	10	90	0	100	0	0	0	0	100	0	0	100	55	45	0	0	94	6	27	73	0
200/50	0	100	0	14	86	0	0	100	0	0	100	0	0	100	0	100	0	0	0	0	100	0	0	100	22	78	0	0	95	5	69	31	0
200/100	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	100	0	0	0	0	100	0	0	100	0	100	0	0	90	10	73	27	0
200/200	0	100	0	0	100	0	0	100	0	0	94	6	0	100	0	68	32	0	0	0	100	0	95	5	0	100	0	0	92	8	53	47	0
60/1000	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	100	0	0	0	0	100	0	100	0	30	70	0	0	99	1	69	31	0
1000/60	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	100	0	0	0	0	100	0	0	100	15	85	0	0	97	3	72	28	0

Table 3
Frequencies of under-, correct, and over-identification (in %) of various factor selection criteria over 1000 replications. The data are generated from the model $\boldsymbol{y}_t = \boldsymbol{\Lambda}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t + \rho\boldsymbol{\varepsilon}_{t-1}$ with $r = 3$ common factors, $s = \sqrt{r}$ and $\rho = 0.5$, and the noise $\boldsymbol{\varepsilon}_t^1$ follows $N(\mathbf{0}, s^2\mathbf{I}_N)$. The r-dimensional factor $\boldsymbol{f}_t$ is a vector of $N(0, 1)$ variables, and each element of the factor loading matrix $\boldsymbol{\Lambda}$ also follows $N(0, 1)$. Because the results of PC_{BN}^2 and IC_{BN}^2 are similar to those of PC_{BN}^3 and IC_{BN}^3, they are omitted from the table.

T/N	PDC_p			$PDIC$			PC_{BN}^1			PC_{BN}^3			IC_{BN}^1			IC_{BN}^3			AIC			BIC			$BIC3$			$\tilde{r}$			$\widetilde{PC}_{BN}^1$		
25/25	0	0	100	8	62	30	0	0	100	0	0	100	1	1	98	94	6	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
25/50	0	0	100	4	63	33	0	0	100	0	0	100	0	10	90	99	1	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
25/100	0	0	100	2	64	34	0	0	100	0	0	100	0	6	94	100	0	0	0	0	100	0	0	100	0	0	100	0	1	99	0	0	100
25/200	0	0	100	1	50	49	0	0	100	0	0	100	0	0	100	100	0	0	0	0	100	0	0	100	0	39	61	0	0	100	0	1	99
50/25	0	2	98	27	72	1	0	0	100	0	0	100	9	53	38	100	0	0	0	0	100	0	0	100	0	33	67	0	9	91	0	5	95
50/50	0	8	92	5	95	0	0	0	100	0	0	100	0	84	16	100	0	0	0	0	100	0	0	100	0	100	0	0	23	77	3	97	0
50/100	0	50	50	0	100	0	0	0	100	0	0	100	0	92	8	100	0	0	0	0	100	0	0	100	0	100	0	0	45	55	7	93	0
50/200	0	91	9	0	100	0	0	0	100	0	0	100	0	97	3	100	0	0	0	0	100	0	0	100	0	100	0	0	86	14	11	89	0
100/25	0	63	37	43	57	0	0	0	100	0	0	100	5	95	0	100	0	0	0	0	100	0	0	100	0	100	0	0	42	58	0	100	0
100/50	0	99	1	3	97	0	0	0	100	0	0	100	0	100	0	100	0	0	0	0	100	0	0	100	0	100	0	0	53	47	16	84	0
100/100	0	100	0	0	100	0	13	87	0	0	0	100	0	100	0	89	11	0	0	0	100	0	0	100	0	100	0	0	71	29	23	77	0
100/200	0	100	0	0	100	0	33	69	0	0	0	100	0	100	0	99	1	0	0	0	100	0	0	100	0	0	100	0	93	7	16	84	0
200/25	0	100	0	56	44	0	0	0	100	0	0	100	3	97	0	100	0	0	0	0	100	0	0	100	0	40	60	0	81	19	13	87	0
200/50	0	100	0	2	98	0	0	99	1	0	12	88	0	100	0	99	1	0	0	0	100	0	0	100	0	100	0	0	86	14	39	61	0
200/100	0	100	0	0	100	0	0	100	0	0	0	100	0	100	0	95	5	0	0	0	100	0	0	100	0	100	0	0	83	17	19	81	0
200/200	0	100	0	0	100	0	0	100	0	0	0	100	0	100	0	31	69	0	0	0	100	0	0	100	0	100	0	0	91	9	6	94	0
60/1000	0	100	0	0	100	0	0	100	0	0	0	100	0	100	0	100	0	0	0	0	100	0	0	100	0	100	0	0	95	5	9	91	0
1000/60	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	100	0	0	0	0	100	0	0	100	0	100	0	0	94	6	42	58	0

Table 4
Frequencies of under-, correct, and over-identification (in %) of various factor selection criteria over 1000 replications. The data are generated from the model $\boldsymbol{y}_t = \boldsymbol{\Lambda}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t$ with $\boldsymbol{f}_t = \boldsymbol{\Phi}\boldsymbol{f}_{t-1} + \boldsymbol{\eta}_t$, and $\boldsymbol{\varepsilon}_t = 0.5\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\xi}_t$, so that both the factors and noises have dynamic dependence. Each element of $\boldsymbol{\xi}_t$ and $\boldsymbol{\eta}_t$ follows a Student-t distribution with 5 degrees of freedom, and each element of the factor loading matrix $\boldsymbol{\Lambda}$ follows the standard normal distribution. Because the results of PC_{BN}^2 and IC_{BN}^2 are similar to those of PC_{BN}^3 and IC_{BN}^3, they are omitted from the table.

T/N	PDC_p			$PDIC$			PC_{BN}^1			PC_{BN}^3			IC_{BN}^1			IC_{BN}^3			AIC			BIC			$BIC3$			$\tilde{r}$			$\widetilde{PC}_{BN}^1$		
25/25	2	32	66	0	10	90	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	3	78	19	0	0	100
25/50	0	21	79	0	1	99	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	87	13	0	0	100
25/100	0	17	83	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	1	87	12	0	0	100
25/200	0	7	93	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	99	1	0	0	100
50/25	0	60	40	0	56	44	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	55	45	0	72	28	0	0	100
50/50	0	74	26	0	23	77	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	67	33	0	81	19	0	0	100
50/100	0	88	12	0	6	94	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	73	27	0	93	7	0	8	92
50/200	0	95	5	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	80	20	0	98	2	0	16	84
100/25	0	99	1	0	97	3	0	0	100	0	0	100	0	10	90	0	0	100	0	0	100	0	0	100	0	100	0	0	61	39	0	7	93
100/50	0	100	0	0	97	3	0	0	100	0	0	100	0	3	97	0	0	100	0	0	100	0	0	100	0	100	0	0	70	30	0	100	0
100/100	0	100	0	0	98	2	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	100	0	0	81	19	0	100	0
100/200	0	100	0	0	99	1	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	100	0	0	90	10	0	100	0
200/25	0	100	0	0	100	0	0	0	100	0	0	100	0	82	18	0	0	100	0	0	100	0	0	100	0	100	0	0	76	24	0	98	2
200/50	0	100	0	0	100	0	0	0	100	0	0	100	0	90	10	0	0	100	0	0	100	0	0	100	0	100	0	0	78	22	0	100	0
200/100	0	100	0	0	100	0	0	0	100	0	0	100	0	55	45	0	0	100	0	0	100	0	0	100	0	100	0	0	82	18	0	100	0
200/200	0	100	0	0	100	0	0	0	100	0	0	100	0	1	99	0	0	100	0	0	100	0	0	100	0	100	0	0	86	14	0	100	0
60/1000	0	100	0	0	100	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	100	0	0	96	4	0	100	0
1000/60	0	100	0	0	100	0	0	0	100	0	0	100	0	100	0	0	0	100	0	0	100	0	0	100	0	100	0	0	95	5	0	100	0

Table 5
Frequencies of under-, correct, and over-identification (in %) of various factor selection criteria over 1000 Replications. The data are generated from the model $\boldsymbol{y}_t = \boldsymbol{\Lambda f}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{f}_t = \boldsymbol{\Phi f}_{t-1} + \boldsymbol{\eta}_t$, where $\boldsymbol{\varepsilon}_t = 0.5\boldsymbol{\varepsilon}_{t-1} + 0.05\boldsymbol{v}_t + \sum_{j=-8,j\neq 0}^{8} 0.01\boldsymbol{v}_{t+j}$. $\boldsymbol{\Phi}$ is a diagonal matrix with diagonal elements $\phi_{ii} = 0.5$. Each element of the noise vector $\boldsymbol{\eta}_t$ and the innovation vectors $\boldsymbol{v}_t$ follows a Student-t with 5 degrees of freedom. Each element of the factor loading matrix $\boldsymbol{\Lambda}$ follows a normal distribution with mean 0 and variance $\sqrt{r}$. Because the results of PC_{BN}^2 and IC_{BN}^2 are similar to those of PC_{BN}^3 and IC_{BN}^3, they are omitted from the table.

T/N	PDC_p			$PDIC$			PC_{BN}^1			PC_{BN}^3			IC_{BN}^1			IC_{BN}^3			AIC			BIC			$BIC3$			$\tilde{r}$			$\widetilde{PC}_{BN}^1$		
25/25	0	96	4	0	97	3	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
25/50	0	99	1	0	98	2	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
25/100	0	100	0	0	100	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
25/200	0	100	0	0	100	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
50/25	0	99	1	0	98	2	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
50/50	0	99	1	0	100	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
50/100	0	100	0	0	100	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
50/200	0	100	0	0	100	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
100/25	0	100	0	0	100	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
100/50	0	100	0	0	100	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
100/100	0	100	0	0	100	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
100/200	0	100	0	0	100	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
200/25	0	100	0	0	100	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
200/50	0	100	0	0	100	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
200/100	0	100	0	0	100	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
200/200	0	100	0	0	100	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
60/1000	0	100	0	0	100	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100
1000/60	0	100	0	14	86	0	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	82	18	0	0	100	0	0	100

6. Application

We obtained daily returns of 49 industrial portfolios from the Fama and French database. The dataset assigns each NYSE, AMEX, and NASDAQ stock to an industrial portfolio at the end of June of year t based on its four-digit SIC code at that time. More details can be found at Fama and French database website. The dimension of our panel analysis is $N = 49$.

As stated before, we analyze the daily returns in the following three periods: (1) June 30, 2006 to June 29, 2007 denoting the period before the outbreak of the subprime crisis, (2) August 1, 2007 to August 29, 2008 denoting the uncertain period after the outbreak of the subprime crisis, but before the collapse of Lehman Brothers, (3) October 1, 2008 to September 30, 2009 denoting the period after Lehman's failure. Thus, the sample sizes T are 251, 274, and 252, respectively, for the three periods. We omit one month between the periods to reduce the effect of subperiod division.

Although the size of panel in this application might not be large, our goal is to explore the impact of recent financial crisis on the U.S stock market. It is clear that the proposed approach can be generalized to analyze the aggregated data with $(T, N) = (777, 49)$. But analyzing the aggregated data would require the number of common factors to be time-varying and more intensive computation. In a spatial dynamic factor model, Lopes et al. (2008) use a reversible jump MCMC scheme to traverse the space of the number of factors with $(T, N) = (312, 22)$.

Table 6 provides descriptive statistics for each of the 49 industry portfolio returns. We see that the magnitudes of volatilities have increased after the outbreak of the financial crisis. Fig. 1 shows the boxplot of the returns of selected industry portfolios in the three periods. From the plot, we again observe that the returns are less-volatile in Period 1. Similar plots are also obtained for other industries.

Since the common factors might have certain dynamic dependence, we modified the panel data so that the number of factors in a restricted dynamic factor model can be consistently estimated via various criteria; see the procedure used in the simulation study.

Given the number of factors k, let $\widehat{\mathbf{F}}_k$ be the asymptotic principal component estimator of $\mathbf{F}_k$ and $\hat{\boldsymbol{f}}_t$ the tth row of $\widehat{\mathbf{F}}_k$. We consider the vector AR(1) model $\hat{\boldsymbol{f}}_t = \boldsymbol{\phi}\hat{\boldsymbol{f}}_{t-1} + \mathbf{e}_t$ and let $\widehat{\boldsymbol{\Phi}}$ be the ordinary least squares estimator of $\boldsymbol{\Phi}$. The criteria are then calculated using the residuals $\hat{\boldsymbol{y}}_t = \boldsymbol{y}_t - \widehat{\boldsymbol{\Lambda}}_k\widehat{\boldsymbol{\Phi}}\hat{\boldsymbol{f}}_{t-1}$ as the observed data. Except for the criterion $\tilde{r}$, the other criteria used in the simulation study can also be applied to the residuals $\hat{\boldsymbol{y}}_t$ to select the number of dynamic factors in a restricted dynamic factor model.

Table 7 reports the estimated number of factors by various criteria. The candidate number of factors ranges from 0 to 20. Since Bai and Ng (2002)'s criteria $\widetilde{PC}^1_{BN}$, $\widetilde{PC}^2_{BN}$ and $\widetilde{PC}^3_{BN}$ tend to select the maximum number of factors entertained, we do not report their selection results. Also, based on their inferior performance in the simulation study, we delete AIC and BIC scores in the study. From the table, we see that the number of common factors reduced substantially after the outbreak of the subprime crisis. The proposed criterion, PDC_p, indicates that the number of common factors reduced to 5 from 7 after the outbreak. The number of common factors reduced to 4 if we use $PDIC$. The $\widetilde{IC}^1_{BN}$ criterion also reduces the number of factors in Periods 2 and 3.

To investigate further, we calculated the correlations between the posterior modes of the first 8 factors and some well-known factors in the literature, including Fama and French (1993)'s three factors, Momentum factor, Short-Term Reversal factor, and Long-Term Reversal factor. Fama and French (1993) suggested that an asset return model on a stock index can be constructed using three different weighted averages of the portfolio values: one based on size (SMB), another based on the book-to-market ratio (HML), and the third based on excess return (ER) on the market. The excess return on the market is computed as the value-weighted return on all NYSE, AMEX, and NASDAQ stocks minus the one-month Treasury bill rate. We downloaded these factors from French's Data Library, although the data originally came from the CRSP database.

Table 8 provides the correlations for each of the three periods. From the table, we make the following observations. First, the 6 observable factors are mainly related to the first two latent factors of our analysis. In particular, Factor 1 is highly related to the market excess returns (ER) of Fama and French. However, the correlations vary from one period to another. For example, the correlation between Factor 1 and HML changes from 0.35 to −0.21 to −0.58 from Period 1 to Periods 2 and 3. Second, Factor 2 has high correlations with momentum factor and long-term reversal (LTR) factor in Period 2 whereas its correlation with the momentum factor is small in Periods 1 and 3. This is interesting and understandable because Period 2, which is from August 1, 2007 to August 29, 2008, denotes a down market as the Dow Jones Industrial Index declined from 13,362.37 to 11,543.35. It seems reasonable that the momentum and long-term reversal factors play a more dominating role in a steady declining market. Third, the magnitude of correlation between the SMB factor of Fama and French and Factor 1 decreases from 0.68 to 0.13 from Period 1 to Period 3, indicating that the recent financial crisis has substantially altered the relationship between market cap and market return. This is also reasonable as large-cap blue-chip stocks tend to be less sensitive to financial crisis compared with small-cap stocks. Fourth, after the failure of Lehman Brothers, Factor 3 is correlated with the SMB and LTR factors. Fifth, the low correlations between Factors 3 and 8 and the 6 well-known observable factors show that there exist other factors important to the market yet to be discovered. In other words, the 6 commonly used factors cannot explain fully the behavior of the US financial market. Finally, treating them as dependent variables, one can use Table 8 to study the explanatory power of the 8 latent factors on the 6 commonly used factors. For instance, in all three periods, ER is mainly explained by Factor 1.

Next, we evaluate the forecasting performance of the models selected by various criteria and show that they improve the forecasting ability over a benchmark model. The benchmark model employs the six commonly used factors, namely the

Table 6
Descriptive statistics for each of the 49 industry portfolio. The sample periods are from (1) June 30, 2006 to June 29, 2007; (2) August 1, 2007 to August 29, 2008; (3) October 1, 2008 to September 30, 2009.

	Period 1		Period 2		Period 3	
	Mean	Sd.	Mean	Sd.	Mean	Sd.
Agric	9.93	87.86	3.10	174.02	23.03	300.91
Food	9.10	58.76	−1.08	98.27	13.42	207.21
Soda	7.64	115.24	−9.45	147.32	12.79	241.34
Beer	11.50	67.23	−4.53	134.79	13.85	223.42
Smoke	8.43	148.68	13.55	164.58	1.44	244.05
Toys	7.18	89.68	−17.56	143.56	17.98	281.59
Fun	6.72	85.23	−13.91	148.90	15.91	324.21
Books	5.12	78.32	−21.63	168.11	25.15	391.32
Hshld	10.12	86.04	−5.77	130.03	17.27	291.71
Clths	10.72	81.63	−8.06	153.47	14.52	313.81
Hlth	8.40	65.12	0.35	94.75	12.98	233.08
MedEq	6.96	65.21	−1.62	94.24	19.01	222.55
Drugs	6.79	84.22	−2.19	110.87	34.50	241.43
Chems	10.88	96.83	−1.39	157.71	16.37	331.03
Rubbr	8.82	84.82	−9.20	160.42	8.78	313.54
Txtls	9.37	106.33	−5.45	149.91	8.02	349.08
BldMt	4.38	84.00	−7.40	137.43	10.97	308.10
Cnstr	2.42	122.98	−11.07	235.78	37.17	500.37
Steel	15.66	133.50	0.93	184.76	11.62	461.36
FabPr	17.83	150.90	12.96	199.31	10.25	436.68
Mach	9.39	96.43	1.09	142.40	7.87	357.82
ElcEq	9.00	89.13	−0.51	131.65	18.57	290.36
Autos	9.50	99.22	−14.48	180.28	23.95	426.29
Aero	16.59	107.34	−3.04	156.35	9.09	293.34
Ships	0.99	112.46	−5.21	178.89	5.93	381.15
Guns	7.77	116.32	−8.94	183.89	26.35	279.69
Gold	−0.07	195.93	−6.75	246.05	32.09	508.06
Mines	11.82	146.56	0.80	215.83	26.26	379.52
Coal	−3.27	202.48	40.15	318.02	16.94	679.34
Oil	5.23	131.74	7.40	180.08	8.67	465.34
Util	8.68	63.48	2.71	101.80	3.22	237.14
Telcm	10.56	73.46	−12.17	142.68	26.70	314.96
PerSv	9.44	67.87	1.68	155.73	23.29	299.13
BusSv	10.23	72.33	−5.13	118.31	21.13	252.40
Hardw	8.84	80.75	−8.94	134.19	22.64	273.24
Softw	10.51	78.11	−3.77	117.07	23.25	246.91
Chips	6.63	93.16	−8.06	125.43	23.08	265.90
LabEq	10.13	78.64	−0.27	109.31	16.21	254.12
Paper	8.96	80.69	−7.77	173.53	18.65	386.48
Boxes	23.45	108.79	2.91	175.63	6.90	310.94
Trans	5.80	96.96	−5.56	183.25	13.74	302.57
Whlsl	9.02	74.55	−4.27	118.34	19.04	267.24
Rtail	5.75	82.14	−8.32	173.33	21.21	315.70
Meals	8.50	75.58	−10.42	149.87	18.03	292.20
Banks	0.35	50.69	−7.70	132.02	4.59	237.99
Insur	8.04	61.08	−3.01	146.54	18.37	349.59
RlEst	7.46	84.59	−14.36	163.80	16.05	439.38
Fin	11.03	89.23	−1.58	178.74	18.96	320.26
Other	7.23	70.89	−4.41	72.98	10.39	156.34

three factors of Fama and French (1993), Momentum factor, Short-Term Reversal factor and Long-Term Reversal factor. The criteria used to select the common factors are the PDC_p, $PDIC$, $BIC3$, $\tilde{r}$, and $\widetilde{IC}^1_{BN}$. Including the benchmark, we have a total of six models. Because the future value $\boldsymbol{f}_{t+1}$ of common factors are unknown at time t, we employ the vector AR(1) model $\boldsymbol{f}_t = \boldsymbol{\Phi}\boldsymbol{f}_{t-1} + \mathbf{e}_t$ and use its 1-step ahead prediction $\widehat{\boldsymbol{\Phi}}\boldsymbol{f}_t$ for $\boldsymbol{f}_{t+1}$ in the study, where $\widehat{\boldsymbol{\Phi}}$ is the ordinary least squares estimate of $\boldsymbol{\Phi}$.

For each of the three periods, we conduct an out-of-sample forecasting exercise. Specifically, we start the estimation with $T_o = 75$ and $N = 49$ to select the number of common factors via various criteria. For a selected common factor series $\boldsymbol{f}_t$, we fit a multiple linear regression to each of the 49 industrial portfolios and apply the VAR(1) model to obtain factor prediction. From the predicted factor and the fitted regression models, we obtain 1-step ahead forecast for each index of the 49 industrial portfolios. Factor regression model is often used in the papers (Aguilera et al., 2006; Ando and Tsay, 2009, forthcoming; Heij et al., 2007; Koch and Naito, 2010; Serneels and Verdonck, 2009; Stock and Watson, 2002a).

The observed returns at $T_o + 1$ are then used to calculate forecasting errors. The estimation window in then advanced by 1 and the estimation-forecasting procedure iterated. The forecasting exercise is repeated until the end of each period. For

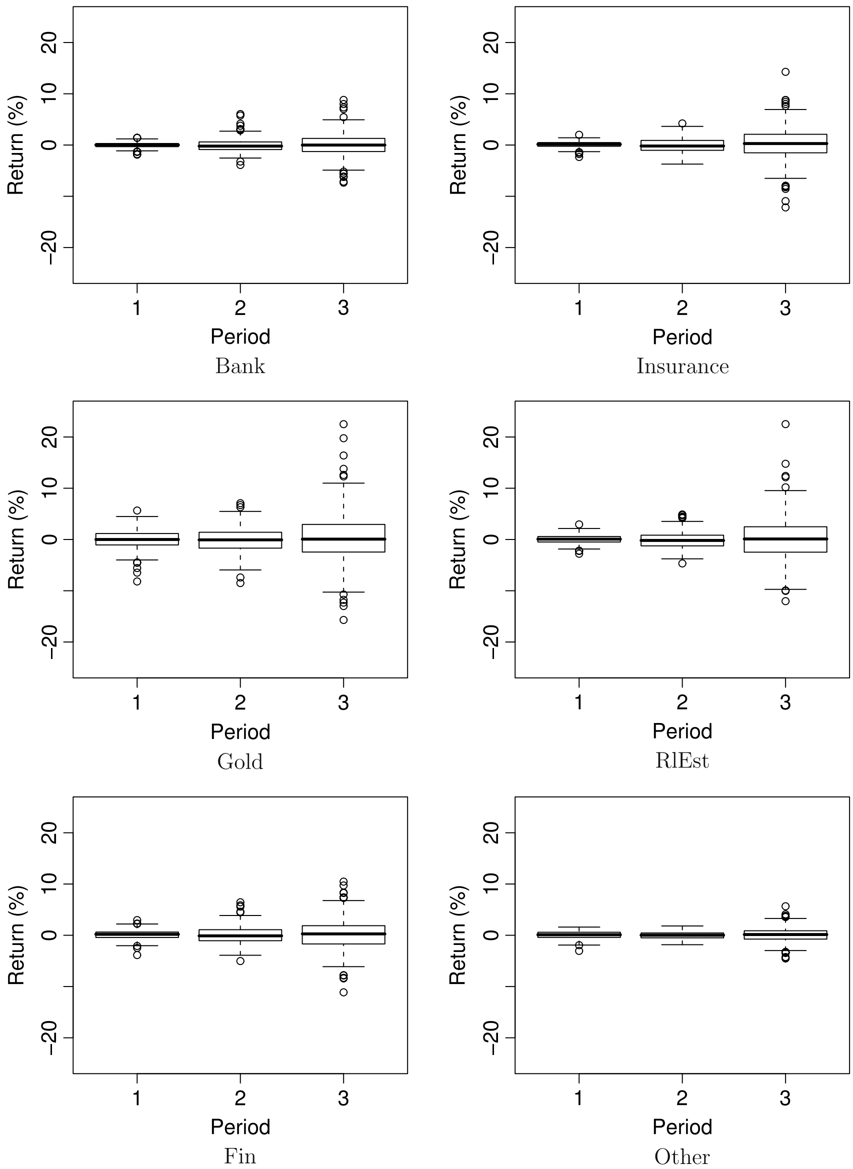

Fig. 1. Boxplot of the returns of industry portfolios for the three periods.

Table 7
Selected number of factors via various criteria. The sample periods are from (1) June 30, 2006 to June 29, 2007; (2) August 1, 2007 to August 29, 2008; (3) October 1, 2008 to September 30, 2009.

	Period 1	Period 2	Period 3
PDC_p	7	5	5
$PDIC$	8	4	4
$BIC3$	8	7	8
$\tilde{r}$	8	6	7
$\widetilde{IC}_{BN}^{1}$	8	4	5

each of the 49 portfolios, we compute the out-of-sample mean squares of forecast errors (MSFE)

$$\text{MSFE} = \frac{1}{T-75}\sum_{t=76}^{T}(y_t - \hat{y}_t)^2,$$

where T denotes the sample size of each period, to assess the forecasting performance. Since the proposed methods of this paper are Bayesian, we use the posterior mean as the forecast $\hat{y}_t$.

Table 8
The correlations between the posterior modes of the first 8 factors and the Fama and French (1993)'s Factors (SMB, HML, ER), Momentum Factor (MM), Short-Term Reversal Factor (STR), Long-Term Reversal Factor (LTR).

	ER	SMB	HML	MM	STR	LTR
Period 1: June 30, 2006 to June 29, 2007						
Factor 1	−0.93	−0.68	0.35	−0.39	−0.12	0.45
Factor 2	−0.14	−0.20	0.24	0.18	0.14	−0.60
Factor 3	0.01	−0.04	0.01	0.04	0.04	0.10
Factor 4	−0.03	−0.03	0.01	0.04	−0.02	−0.20
Factor 5	0.03	0.09	−0.04	−0.13	−0.02	−0.02
Factor 6	0.00	0.07	−0.03	0.01	0.05	−0.12
Factor 7	0.05	0.03	0.05	−0.02	−0.04	−0.08
Factor 8	−0.00	−0.10	0.04	−0.01	0.07	−0.02
Period 2: August 1, 2007 to August 29, 2008						
Factor 1	−0.95	−0.30	−0.21	0.43	−0.34	0.15
Factor 2	−0.01	0.22	0.43	−0.68	0.07	0.84
Factor 3	0.00	0.03	0.03	−0.02	−0.05	−0.11
Factor 4	−0.02	0.12	0.12	−0.24	−0.01	0.08
Factor 5	−0.01	0.01	−0.08	0.16	−0.12	−0.05
Factor 6	−0.01	−0.02	0.33	−0.23	0.16	0.13
Factor 7	−0.01	−0.04	−0.04	0.12	−0.04	−0.04
Factor 8	0.06	−0.07	−0.03	0.05	−0.04	−0.01
Period 3: October 1, 2008 to September 20, 2009						
Factor 1	−0.95	−0.13	−0.58	0.81	−0.36	0.27
Factor 2	−0.00	0.26	0.24	−0.06	−0.13	0.47
Factor 3	0.05	−0.22	−0.16	−0.06	0.04	−0.44
Factor 4	0.08	−0.14	−0.10	0.10	0.06	−0.15
Factor 5	−0.09	0.20	0.17	−0.21	−0.28	0.18
Factor 6	0.10	−0.32	−0.01	0.10	0.09	−0.10
Factor 7	−0.04	0.05	−0.03	−0.00	0.12	0.00
Factor 8	0.00	−0.10	−0.03	−0.06	−0.06	−0.01

Table 9
Rankings of forecasting performance of various criteria based on the 49 industrial portfolios of Fama and French database with Rank 1 being the best, where Bench denotes the benchmark that uses 6 observable common factors.

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6
Period 1: June 30, 2006 to June 29, 2007						
PDC_p	37	2	0	0	7	3
$PDIC$	8	37	4	0	0	0
$BIC3$	8	37	4	0	0	0
$\tilde{r}$	8	37	4	0	0	0
$\widetilde{IC}^1_{BN}$	8	37	4	0	0	0
Bench	4	2	0	0	1	42
Period 2: August 1, 2007 to August 29, 2008						
PDC_p	5	0	35	7	2	0
$PDIC$	41	5	3	0	0	0
$BIC3$	2	1	1	1	42	2
$\tilde{r}$	0	2	4	38	5	0
$\widetilde{IC}^1_{BN}$	41	5	3	0	0	0
Bench	3	1	0	1	0	47
Period 3: October 1, 2008 to September 30, 2009						
PDC_p	42	7	0	0	0	0
$PDIC$	7	0	41	0	0	0
$BIC3$	0	0	0	4	45	0
$\tilde{r}$	0	0	1	44	4	0
$\widetilde{IC}^1_{BN}$	42	7	0	1	0	0
Bench	0	0	0	0	0	49

For each period, each portfolio has 6 MSFEs. To summarize the performance of the 6 methods, we rank them based on their MSFE with Rank 1 being the best and Rank 6 the worst. We then tabulate the rankings over the 49 industrial portfolios for the 6 methods. Table 9 gives the measure of forecasting performance with Rank 1 denoting the smallest MSFE. The numbers in the table denote the counts of each rank over the 49 portfolios. For example, the proposed criterion PDC_p was ranked first 40 times out of the 49 portfolios in Period 1 and ranked twice as the worst in the same period. From the table, no single method dominates the others, but the proposed criteria PDC_p and $PDIC$ work well. The method $\widetilde{IC}^1_{BN}$ also works reasonably well whereas the benchmark appears to perform poorly.

Table 10
Sharpe ratio (SR) and terminal wealth (TW) value for models selected via various criteria. Sharpe ratio of the market index (MI) based on the excess returns is also reported. The sample periods are from (1) June 30, 2006 to June 29, 2007; (2) August 1, 2007 to August 29, 2008; (3) October 1, 2008 to September 30, 2009. Bench denotes the benchmark that uses 6 observable common factors.

	Period 1		Period 2		Period 3	
	SR	TW	SR	TW	SR	TW
PDC_p	0.158	1.167	−0.038	0.934	0.163	1.514
$PDIC$	0.156	1.165	−0.030	0.945	0.159	1.494
$BIC3$	0.156	1.165	−0.058	0.927	0.126	1.422
$\tilde{r}$	0.156	1.165	−0.075	0.907	0.114	1.364
$\widetilde{IC}_{BN}^{1}$	0.156	1.165	−0.030	0.945	0.163	1.514
Bench	0.145	1.113	−0.025	0.950	0.160	1.511
MI	0.065	1.063	−0.043	0.879	0.141	1.454

Based on the results of Table 9, it is clear that dynamic factor models outperform the benchmark model in forecasting, where the latter uses observable factors. The table also shows that the two proposed criteria work well with $\widetilde{IC}_{BN}^{1}$ being a close second. Among the dynamic factor models, the proposed criteria work well in selecting the number of common factors for forecasting. The criterion $\widetilde{IC}_{BN}^{1}$ also works reasonably well in our empirical study.

The prior comparisons focus on point forecasts. In what follows, we go 1-step further to consider portfolio construction. We adopt the standard mean–variance portfolio selection of Markowitz (1952) using the 49 industrial portfolios as available assets. Under the mean–variance selection method, an investor adopts a one-period investment horizon and allocates a wealth over m assets with allocation weight $\boldsymbol{w} = (w_1, \ldots, w_m)'$. An optimal portfolio $\boldsymbol{w}$ is determined by solving the problem

$$\text{maximize } \boldsymbol{w}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}, \quad \text{s.t. } \boldsymbol{w}'\mathbf{1} = 1,$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and the covariance matrix of asset returns, γ is the investor's risk-aversion parameter, and $\mathbf{1}$ is a column unit vector.

For a given criterion in a given period, we use the predicted return described above for $\boldsymbol{\mu}$. The covariance matrix $\boldsymbol{\Sigma}$ is estimated by the sample covariance matrix at the forecast origin. The portfolio weights are then determined and used to calculate the returns in the next trading day. The procedure is repeated until the end of the period. Table 10 reports the Sharpe ratio and terminal wealth for portfolios constructed via various criteria. The results of the benchmark and market index based on excess returns are also included. In this particular instance, the proposed criteria PDC_p and $PDIC$ and the $\widetilde{IC}_{BN}^{1}$ work well. However, they do not outperform the benchmark. It seems that the uncertainty in covariance matrix estimation affects the performance of the dynamic factor models.

7. Concluding remarks

This paper analyzed the effect of recent subprime financial crisis on the US stock market. To this end, we developed a new Bayesian panel data analysis method for identifying the common factors for stock returns when the dimension involved is high. Using Bayesian analysis, we proposed two criteria for selecting the number of factors in a panel data set. Under certain conditions, we established the consistency of the criteria when both the sample size T and the cross-section dimension N approach infinity. For finite samples, our simulation results showed that the proposed criteria outperform those proposed by Bai and Ng (2002), especially when the number of observations and the cross-section dimension are small. In addition, the PDC_p criterion also outperforms the method of Onatski (2005) in most cases considered in our simulation study, even though the latter works reasonably well in most cases. The PDC_p criterion seems to be more robust when there are heteroscedasticity, serial correlation, and fat-tailed features in the data.

Our empirical analysis indicates that the US stock market was subject to 8 common factors before the outbreak of the subprime crisis, whereas the number of common factors reduced substantially after the outbreak. After Lehman's failure, a small number of common factors has been governing the fluctuations of stock market. We found empirical evidence that the structure of US stock market has changed drastically after the subprime crisis. Finally, we also showed that the proposed method performs well in out-of-sample forecasting.

Acknowledgements

The authors would like to thank the referees for their careful suggestions in improving this work. The second author also acknowledges the support of the research fund from Inamori Foundation, Japan.

Appendix. Supplementary data

Supplementary material related to this article can be found online at doi:10.1016/j.csda.2010.11.028.

References

Aguilera, A.M., Escabias, M., Valderrama, M.J., 2006. Using principal components for estimating logistic regression with high-dimensional multicollinear data. Comput. Statist. Data Anal. 50, 1905–1924.
Aguilar, O., West, M., 2000. Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. J. Bus. Econom. Statist. 18, 338–357.
Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Control AC-19, 716–723.
Amengual, D., Watson, M.W., 2007. Consistent estimation of the number of dynamic factors in a large N and T panel. J. Bus. Econom. Statist. 25, 91–96.
Ando, T., 2007. Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. Biometrika 94, 443–458.
Ando, T., 2009. Bayesian factor analysis with fat-tailed factors and its exact marginal likelihood. J. Multivariate Anal. 100, 1717–1726.
Ando, T., Tsay, R., 2009. Model selection for generalized linear models with factor-augmented predictors (with discussion). Appl. Stoch. Models Bus. Ind. 25, 207–235.
Ando, T., Tsay, R., 2010. Predictive likelihood for the Bayesian model selection and averaging. Int. J. Forecast. 26, 744–763.
Ando, T., Tsay, R., 2011. Quantile regression models with factor-augmented predictors and information criteria. Econom. J. (forthcoming).
Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135–172.
Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.
Bawa, V., Brown, S., Klein, R., 1979. Estimation Risk and Optimal Portfolio Choice. North-Holland, New York.
Bernanke, B., Boivin, J., 2003. Monetary policy in a data rich environment. J. Monetary Econom. 50, 525–546.
Bhattacharya, A., Dunson, D., 2009. Sparse Bayesian infinite factor models. Duke Statistics Discussion Papers 2009–23, Duke University.
Bishop, C.M., 1999. Variational principal components. Artif. Neural Netw. 1, 509–514.
Box, G.E., Tiao, G.C., 1972. Bayesian Inference in Statistical Analysis. Addison-Wesley Publishing Company, Reading,.
Campbell, J.Y., Chanb, Y.L., Viceirac, L.M., 2003. A multivariate model of strategic asset allocation. J. Finan. Econom. 67, 41–80.
Carlin, B., Louis, T., 1996. Bayes and Empirical Bayes Methods for Data Analysis. Chapman and Hall, New York.
Carvalho, C.M., Chang, J., Lucas, J.E., Nevins, J.R., Wang, Q., West, M., 2008. High-dimensional sparse factor modeling: applications in gene expression genomics. J. Amer. Statist. Assoc. 103, 1438–1456.
Connor, G., Korajczyk, R., 1986. Performance measurement with the arbitrage pricing theory: a new framework for analysis. J. Finan. Econom. 15, 373–394.
Connor, G., Korajczyk, R., 1988. Risk and return in an equilibrium apt: application to a new test methodology. J. Finan. Econom. 21, 255–289.
Diaz-Garcia, J.A., Gutierrez-Jaimez, R., 2006. Distribution of the generalized inverse of a random matrix and its applications. J. Statist. Plan. Infer. 136, 183–192.
Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. J. Finan. Econom. 33, 3–56.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic factor model: identification and estimation. Rev. Econom. Statist. 82, 540–554.
Forni, M., Lippi, M., 2001. The generalized factor model: representation theory. Econom. Theory 17, 1113–1141.
Forni, M., Reichlin, L., 1998. Let's get real: a factor-analytic approach to disaggregated business cycle dynamics. Rev. Econom. Stud. 65, 453–473.
Frühwirth-Schnatter, S., Lopes, H.F., 2010. Parsimonious Bayesian factor analysis when the number of factors is unknown. Working paper, the University of Chicago Booth School of Business.
Gelfand, A., 1996. Model determination using sampling-based methods. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), Markov Chain Monte Carlo in Practice. Chapman and Hall, London, pp. 145–161.
Gelfand, A., Dey, D., 1994. Bayesian model choice: asymptotics and exact calculations. J. Roy. Statist. Soc. B 56, 501–514.
Gelfand, A., Ghosh, S.K., 1998. Model choice: a minimum posterior predictive loss approach. Biometrika 85, 1–11.
Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), 1996. Markov Chain Monte Carlo in Practice. Chapman and Hall, New York.
Gupta, A.K., Varga, T., 1993. Elliptically Contoured Models in Statistics. Kluwer Academic Publishers, Dordrecht.
Hallin, M., Liska, R., 2007. The generalized dynamic factor model: determining the number of factors. J. Amer. Statist. Assoc. 102, 603–617.
Heij, C., Groenen, P.J.F., van Dijk, D., 2007. Forecast comparison of principal component regression and principal covariate regression. Comput. Statist. Data Anal. 51, 3612–3625.
Hoogerheide, L.F., Kleibergen, F., van Dijk, H.K., 2007. Natural Conjugate priors for the instrumental variables regression model applied to the Angrist–Krueger data. J. Econometrics 138, 63–103.
Khatri, C.G., Mardia, K.V., 1977. The von mises fisher distribution in orientation statistics. J. Roy. Statist. Soc. B 39, 95–106.
Kleibergen, F., Paap, R., 2002. Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration. J. Econometrics 111, 223–249.
Kleibergen, F., Paap, R., 2006. Generalized reduced rank tests using the singular value decomposition. J. Econometrics 133, 97–126.
Kleibergen, F., van Dijk, H.K., 1994. On the shape of the likelihood/posterior in cointegration models. Econom. Theory 10, 514–551.
Kleibergen, F., van Dijk, H.K., 1998. Bayesian simultaneous equations analysis using reduced rank structures. Econom. Theory 14, 701–743.
Koch, I, Naito, K., 2010. Prediction of multivariate responses with a selected number of principal components. Comput. Statist. Data Anal. 54, 1791–1807.
Koop, G., Leon-Gonzalez, R., Strachan, R., 2006. Efficient posterior simulation for cointegrated models with priors on the cointegration space. Working Paper No. 05/13, University of Leicester, Department of Economics.
Lopes, H.F., West, M., 2004. Bayesian model assessment in factor analysis. Statist. Sinica 14, 41–67.
Lopes, H.F., Salazar, E., Gamerman, D., 2008. Spatial dynamic factor models. Bayesian Anal. 3, 759–792.
Markowitz, H., 1952. J. Finance 7, 77–91.
McLachlan, G.J., Peel, D., Bean, R.W., 2003. Modelling high-dimensional data by mixtures of factor analyzers. Comput. Statist. Data Anal. 41, 379–388.
Onatski, A., 2005. Determining the number of factors from empirical distribution of eigenvalues. Discussion Paper 0405-19, Economics Department, Columbia University.
Press, S.J., Shigemasu, K., 1989. Bayesian inference in factor analysis. In: Glesser, L.J. (Ed.), Contributions to Probability and Statistics. Springer, New York.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6, 461–464.
Serneels, S., Verdonck, V., 2009. Principal component regression for data containing outliers and missing elements. Comput. Statist. Data Anal. 53, 3855–3863.
Smidl, V., Quinn, A., 2007. On Bayesian principal component analysis. Comput. Statist. Data Anal. 51, 4101–4123.
Spiegelhalter,, Best,, Carlin,, van der Linde,, 2002. Bayesian measures of model complexity and fit (with discussion). J. Roy. Statist. Soc. B 64, 583–639.
Stock, J.H., Watson, M.W., 1998. Diffusion indexes. NBER Working Paper 6702.
Stock, J.H., Watson, M.W., 2002a. Forecasting using principal components from a large number of predictors. J. Amer. Statist. Assoc. 97, 1167–1179.
Stock, J.H., Watson, M.W., 2002b. Macroeconomic forecasting using diffusion indexes. J. Bus. Econom. Statist. 20, 147–162.
Stock, J.H., Watson, M.W., 2003. Forecasting output and inflation: the role of asset prices. J. Econom. Lit. 41, 788–829.
Stock, J.H., Watson, M.W., 2004. Forecasting output and inflation: the role of asset prices. J. Forecast. 23, 405–430.
Strachan, R.W., 2004. On priors on cointegrating spaces. Keele Economics Research Papers 2004/06, Keele University.
Strachan, R.W., 2007. Bayesian inference in cointegrated $I(2)$ systems: a generalisation of the triangular model. Econometric Rev. 26, 439–468.
Strachan, R.W., Inder, B., 2004. Bayesian analysis of the error correction model. J. Econometrics 123, 307–325.
Tierney, L., 1994. Markov chains for exploring posterior distributions (with discussion). Ann. Statist. 22, 1701–1762.
West, M., 2003. Bayesian factor regression models in the "large p, small n" paradigm. Bayesian Statist. 7, 723–732.