



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

Ruey S. Tsay<sup>a</sup>, Tomohiro Ando<sup>b,\*</sup>

<sup>a</sup> Booth School of Business, University of Chicago, United States

<sup>b</sup> 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.

© 2010 Elsevier B.V. All rights reserved.

### 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](mailto:ruey.tsay@chicagogsb.edu) (R.S. Tsay), [andoh@hc.cc.keio.ac.jp](mailto:andoh@hc.cc.keio.ac.jp), [andoh@kbs.keio.ac.jp](mailto:andoh@kbs.keio.ac.jp) (T. Ando).

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

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات