Chin. Geogra. Sci. 2018 Vol. 28 No. 4 pp. 571–583
https://doi.org/10.1007/s11769-018-0975-1

Springer Science Press
www.springerlink.com/content/1002-0063

Regional Clustering and Synchronization of Provincial Business Fluctuations in China

SONG Tao[1], ZHENG Tingguo[2, 3], XIA Kai[3]

(1. *Institute of Economics*, *Xiamen University*, *Xiamen* 361005, *China*; 2. *Department of Statistics*, *School of Economics*, *Xiamen University*, *Xiamen* 361005, *China*; 3. *Wang Yanan Institute for Studies in Economics*, *Xiamen University*, *Xiamen* 361005, *China*)

Abstract: In this article, we propose a novel, multilevel, dynamic factor model, to determine endogenously clustered regions for the investigation of regional clustering and synchronization of provincial business fluctuations in China. The parameter identification and model estimation was conducted using the Markov Chain Monte Carlo method. We then conducted an empirical study of the provincial business fluctuations in China (31 Chinese provinces are considered except Hong Kong, Macau, and Taiwan due to the data unavailability), which were sampled from January 2000 to December 2015. Our results indicated that these provinces could be clustered into four regions: leading, coincident, lagging, and overshooting. In comparison with traditional geographical divisions, this novel clustering into four regions enabled the regional business cycle synchronization to be more accurately captured. Within the four regional clusters it was possible to identify substantial heterogeneities among regional business cycle fluctuations, especially during the periods of the 2008 financial crisis and the 'four-trillion economic stimulus plan'.

Keywords: regional division; business cycle synchronization; multilevel dynamic factor model; variance decomposition

Citation: SONG Tao, ZHENG Tingguo, XIA Kai, 2018. Regional Clustering and Synchronization of Provincial Business Fluctuations in China. *Chinese Geographical Scien*ce, 28(4): 571–583. https: //doi.org/10.1007/s11769-018-0975-1

1 Introduction

In recent decades, researchers have shown an increasing interest in the regional synchronization of provincial business fluctuations in China. It has been shown that with the deepening of economic system reform and the elimination of local trade barriers, the economies of Chinese provinces have exhibited a high degree of business cycle synchronization (Gerlach-Kristen, 2009; Huang et al., 2011). The synchronization observed in the eastern China was comparable to that observed in the United States at the end of the 1990s, while in the northwestern provinces the synchronization appears to differ from the process throughout the rest of China (Poncet, 2004). Gatfaoui and Girardin (2015) reported that that all of China's coastal provinces, except Hainan, were synchronized with the national cycle. Empirically, the multilevel dynamic factor model of Kose et al. (2003) is often adopted to extract region-specific factors from provincial economic indicators. Several studies, such as Guo and Jia (2005), Zhang and Tong (2011), and Guo and Zhao (2012), found that the factors estimated with this model could explain many of the variations in provincial data.

Although business cycle synchronization has been the focus of regional economic studies, there are at least two limitations in the method currently used to describe the characteristics of regional business cycle fluctuations. First, the traditional geographical division of Chinese provinces into eastern, intermediate, and western regions based on geographical location (Fig. 1) can not

Received date: 2017-11-10; accepted date: 2018-03-09
Foundation item: Under the auspices of the National Natural Science Foundation of China (No. 71371160), the Program for Changjiang Youth Scholars (No. Q2016131), the Program for New Century Excellent Talents in University (No. NCET-13-0509)
Corresponding author: ZHENG Tingguo. E-mail address: zhengtg@gmail.com

reflect the regional clustering characteristics well. Huang et al. (2015) showed that this division can only roughly reflect the differences in economic development between regions in China. However, in terms of business cycle fluctuations, Poncet (2004) and Gerlach-Kristen (2009) found that the remote northwestern provinces appear to differ from each other. Herrerias and Ordóñez (2012) further found that even the eastern coastal provinces were not necessarily in the same regional cluster. Therefore, it would be more appropriate to cluster the provinces endogenously according to their similarities in terms of business fluctuations. Similar problems have been found in some studies of Asia (He and Liao, 2012), Europe (Lee, 2012) and the United States (Crone and Clayton-Matthews, 2005; Stock and Watson, 2010), and in an international business cycle analysis of multiple countries (Francis et al., 2012).

Second, in the existing method it is difficult to adequately model the interactions among regions. To capture regionally common properties, a common practice is to use the orthogonal multilevel dynamic factor model proposed by Kose et al. (2003) to extract the national and region-specific factors from provincial business fluctuations in China, see for example Guo and Jia (2005), Zhang and Tong (2011), and Guo and Zhao (2012). However, this approach only estimates the region-specific factors that are uncorrelated with national factors, and 'not factors for those regions per se' (Moench et al., 2013). In addition, these factors are orthogonal among regions, which contradicts the evidence of the local links among regions, such as foreign direct investment (FDI) spillover effects (see Ouyang and Fu, 2012; Huang and Chand, 2015). The empirical results obtained from this model (Guo and Jia, 2005; Zhang and Tong, 2011) also suggest that estimated regional factors are strongly correlated with each other.

To address the problems raised above, we propose a novel multilevel dynamic factor model with an endogenous regional grouping strategy, which could be regarded as an extension of the multilevel dynamic factor model in Bai and Wang (2015), by introducing the 'data-driven' regional clustering approach in Francis et al. (2012). There are several advantages of employing such a model. First, it enables us to simultaneously estimate the common regional factors and regional clusters, avoiding any possible incorrect specifications that could result from using existing geographical regions. Second, regional factors are allowed to be correlated with national factors, which gives a more comprehensive measurement of regional business fluctuations than

Fig. 1 Traditional geographical division of regions in China

the orthogonal factors in the model proposed by Kose et al. (2003). Third, variance decomposition based on the vector autoregressive (VAR) process allows us to analyze the integration of provinces within regions.

After putting forward the multilevel factor model together with its Bayesian Markov Chain Monte Carlo (MCMC) estimation procedure, we carry out an empirical investigation of the estimated clusters (new divisions based on unknown regions) and the corresponding regional characteristics.

2 Methodologies

2.1 Multilevel dynamic factor model

Our model was based on the multilevel dynamic factor model in Bai and Wang (2015). We further incorporated a 'data-driven' approach (Francis et al., 2012) to decide the regional grouping of provinces endogenously, to avoid the misspecifications of using given regions.

The proposed multilevel factor model for the *i*-th provincial growth y_{it} ($i = 1,\ldots, N$) is given by:

$$y_{it} = \lambda_{yf,i} f_t + \lambda_{yg,i} \sum_{b=1}^{B} \gamma_{ib} g_{bt} + e_{it}, \quad e_{it} \sim N(0, \sigma_i^2) \tag{1}$$

where f_t is a national factor, g_{bt}, $b = 1,\ldots, B$, are the region-specific factors, and e_{it} is an idiosyncratic error term. In Equation (1), γ_{ib} is a dummy variable, such that $\gamma_{ib} = 1$ if province i belongs to region b and $\gamma_{ib} = 0$ otherwise, $b = 1,\ldots, B$. In addition, the national factor f_t represents the national comovement of provincial business fluctuations, and the region-specific region g_{bt}, depending on which region it belongs to, stands for the regional comovement of provincial business fluctuations for region b, $b = 1,\ldots, B$.

To consider the dynamic interactions among the national and regional factors, we assumed the following VAR process for the national and region-specific factors:

$$\begin{pmatrix} f_t \\ g_{1t} \\ \vdots \\ g_{Bt} \end{pmatrix} = \Phi \begin{pmatrix} f_{t-1} \\ g_{1,t-1} \\ \vdots \\ g_{B,t-1} \end{pmatrix} + \begin{pmatrix} v_{f,t} \\ v_{g1,t} \\ \vdots \\ v_{gB,t} \end{pmatrix}, \quad \begin{pmatrix} v_{f,t} \\ v_{g1,t} \\ \vdots \\ v_{gB,t} \end{pmatrix} \sim N(0, \Sigma) \tag{2}$$

where Φ is a $(B+1) \times (B+1)$ matrix, and Σ is a $(B+1) \times (B+1)$ symmetric positive definite matrix. Note that the above model nests several multilevel factor models used in the literature. For example, the model introduced by Kose et al. (2003) is a special case of the model given by Equation (1) and Equation (2) when all γ_{ib}, $b = 1,\ldots, B$, are given and both Φ and Σ are diagonal. Moreover, the model becomes the multilevel factor model proposed by Bai and Wang (2015), when all γ_{ib}, $b = 1,\ldots, B$, are given.

2.2 Parameter identification

Parameter identification is an important issue that various factor models need to handle. Let $y_t = (y_{1t},\ldots, y_{Nt})'$, $F_t = (f_t, g_{1t},\ldots, g_{Bt})'$:

$$\Gamma = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1B} \\ \vdots & \ddots & \vdots \\ \gamma_{N1} & \cdots & \gamma_{NB} \end{pmatrix},$$

$$\Lambda_\Gamma = \begin{pmatrix} \lambda_{yf,1} & \lambda_{yg,1}\gamma_{11} & \cdots & \lambda_{yg,1}\gamma_{1B} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{yf,N} & \lambda_{yg,N}\gamma_{N1} & \cdots & \lambda_{yg,N}\gamma_{NB} \end{pmatrix},$$

and $$e_t = \begin{pmatrix} e_{1t} \\ \vdots \\ e_{Nt} \end{pmatrix} \tag{3}$$

Then Equation (1) can be rewritten as:

$$y_t = \Lambda_\Gamma F_t + e_t \tag{4}$$

It can be seen that the loading coefficient matrix Λ_Γ and the factor vector F_t are unidentifiable. When Γ is given, Bai and Wang (2015) show that the model given by Eqs. (4) and (2) is fully identified if Σ is an identity matrix together with $B + 1$ sign restrictions on the loading matrix Λ_Γ. Moreover, although Σ is diagonal, the dynamic interactions among factors can be captured by the off-diagonal elements of Φ.

When Γ is unknown, some extra identification conditions are required as follows:

$$0 < \sum_{i=1}^{N} \gamma_{ib} < N, \quad \sum_{b=1}^{B} \gamma_{ib} = 1 \tag{5}$$

The above conditions exclude two extreme cases in which $\sum_{i=1}^{N} \gamma_{ib} = 0$ and $\sum_{i=1}^{N} \gamma_{ib} = N$, which imply the presence of invalid factors. For example, if $\gamma_{bi} = 0$ for all i, the factor g_{bt} can not be estimated, and thus it is invalid. If $\gamma_{ib}=1$ for all *i*, the factors f_t and g_{bt} are the same and the factor g_{bt} is invalid.

2.3 Bayesian computation

Notice that the complicated multi-dimension matrix in the likelihood function usually results in many local maxima and very flat surfaces near the optimal solution, which would cause problems in the maximum likelihood estimation procedure. Therefore, we implemented our multilevel factor model given by Eqs. (1) and (2), using the MCMC procedure. The convergence and limiting properties of the MCMC method are referred to in Geweke and Keane (2001).

Denote $u_t=(v_{f,t},v_{g1,t},\cdots,v_{gB,t})'$, $Y=(y_1',\cdots,y_T')'$, $F=(F_1',\cdots,F_T')'$, and $\Sigma_e=\mathrm{diag}\{\sigma_1^2,\cdots,\sigma_N^2\}$. Furthermore, suppose that $\Sigma=I_{B+1}$ for parameter identification, where I_{B+1} is a (B +1) × (B +1) identity matrix. Based on observational data $Y=(y_1',\cdots,y_T')'$, we used the MCMC for simulation of the full joint posterior $p(F,\Phi,\Lambda,\Sigma_e\mid Y)$. The MCMC computation consists of the following three steps:

Step 1. Sampling $F^{(k+1)}$ given $\Lambda_\Gamma^{(k)}$, $\Sigma_e^{(k)}$, and $\Phi^{(k)}$. Because $e_t\sim N(0,\Sigma_e)$ and $u_t\sim N(0,I_{B+1})$, $F^{(k+1)}$ can be directly sampled under the linear and Gaussian state space model given by Eqs. (1) and (2) using the forward looking backward sampling approach based on the Kalman filter (Carter and Kohn, 1994; Kim and Nelson, 1999).

Step 2. Sampling $\Phi^{(k+1)}$ given $F^{(k+1)}$. Because $u_t\sim N(0,I_{B+1})$, each equation in the VAR model can be treated as a separate Bayesian regression with independent Gaussian errors. For example, denote $\bar{Y}_b=(g_{b2},\cdots,g_{bT})'$, $\bar{X}_b=(F_1,\cdots,F_{T-1})'$, and $\beta_b=(\Phi^{(b\cdot)})'$, where $\Phi^{(b\cdot)}$ is the b-th row of Φ. The conjugate prior for β_b was set as $N(\psi_0,\Omega_0)$, then β_b was sampled from the posterior distribution $N(\psi^*,\Omega^*)$ with $\Omega^*=(\Omega_0+1/(\sigma^2)\bar{X}_{b'}\bar{X}_b)^{-1}$ and $\psi^*=\Omega^*(\Omega_0^{-1}\psi_0+1/(\sigma^2)\bar{X}_{b'}\bar{Y}_b)$.

Step 3. Sampling $\Lambda_\Gamma^{(k+1)}$ and $\Sigma_e^{(k+1)}$ given Y and $F^{(k+1)}$. For $i=1,\cdots,N$, denote $\lambda_i=(\lambda_{yf,i},\lambda_{yg,i})'$, $\gamma_i=(\gamma_{i1},\dots,\gamma_{iB})'$, $\tilde{Y}_i=(y_{i1},\dots,y_{iT})'$, and $\tilde{X}_i=((f_1,\cdots,f_T)',(G_1,\cdots,G_T)'\gamma_i)'$. The conjugate prior for λ_i is assumed to be $N(\lambda_{i,0},\Omega_{i,0})$. First, generate γ_i^* from a proposed distribution of γ_i. Then given γ_i^*, the posterior distribution of λ_i is $N(\lambda_{i,1}^*,\Omega_{i,1}^*)$, where $\Omega_{i,1}^*=(\Omega_{i,0}+1/(\sigma_i^2)^{(k)}\tilde{X}_{i'}\tilde{X}_i)^{-1}$ and $\lambda_{i,1}^*=\Omega_{i,1}^*(\Omega_{i,0}^{-1}\lambda_{i,0}+1/(\sigma_i^2)^{(k)}\tilde{X}_{i'}Y_i)$. Denote $\lambda_{i,1}$ and $\Omega_{i,1}$ as the posterior parameters of γ_i in the last iteration, define the following acceptance rate:

$$\alpha_{i,\gamma}=\min\left\{1,\frac{|\Omega_{i,1}^*|^{1/2}\exp\left(\frac{1}{2}(\lambda_{i,1}^*)'\left(\Omega_{i,1}^*\right)^{-1}\lambda_{i,1}^*\right)}{|\Omega_{i,1}|^{1/2}\exp\left(\frac{1}{2}(\lambda_{i,1})'\left(\Omega_{i,1}\right)^{-1}\lambda_{i,1}\right)}\frac{p(\gamma_i^{(k)}\mid\gamma_i^*)}{p(\gamma_i^*\mid\gamma_i^{(k)})}\right\} \tag{6}$$

Similar to Francis et al. (2012), we assumed that γ_i follows a discrete uniform distribution. $\alpha_{i,\gamma}$ is determined by the first term in Equation (6). Then we sampled λ_i^* from $N(\lambda_{i,1}^*,\Omega_{i,1}^*)$, and accepted λ_i^* and γ_i^* with a probability of $\alpha_{i,\gamma}$, otherwise $\lambda_i^{(k+1)}=\lambda_i^{(k)}$, $\gamma_i^{(k+1)}=\gamma_i^{(k)}$.

In addition, we set the conjugate prior for σ_i^2 as the inverse Gamma distribution $IG(\frac{1}{2}v_0,\frac{1}{2}\delta_0)$, and the posterior distribution will then be $IG(\frac{1}{2}v_i^*,\frac{1}{2}\delta_i^*)$ with $v_i^*=v_0+\frac{1}{2}T$, $\delta_i^*=\delta_0+(Y_i-\tilde{X}_i\lambda_i^{(k+1)})'(Y_i-\tilde{X}_i\lambda_i^{(k+1)})$. Then we can sample $(\sigma_i^2)^{(k+1)}$ from $IG(\frac{1}{2}v_i^*,\frac{1}{2}\delta_i^*)$.

In the above MCMC computation, the priors of all factor loadings were assumed to be N (0, 1), while all priors for the variance parameters were assumed to be the inverted Gamma distribution $IG(6,0.1)$.

2.4 Determination of regional clusters

The number of regional clusters can be determined by a Bayesian model comparison using the deviance information criteria (DIC) suggested in Celeux et al. (2006). We followed Bai and Wang (2015) to apply the computationally simple approximation of Li et al. (2013) as follows:

$$DIC(B)\approx-\frac{4}{M}\sum_{m=1}^{M}\log p(Y\mid\Theta^{(m)}(B))+2\log p(Y\mid\hat{\Theta}(B)) \tag{7}$$

where Y denotes the observed data, $\Theta^{(m)}(B)$ denotes the parameter sampled in the m-th iteration given B, and $\hat{\Theta}(B)$ denotes the Bayesian posterior parameter estimates given B.

2.5 Variance decomposition

Based on Diebold et al. (2006), the variance decomposition can be conducted in the following way. To begin with, we represent Equation (2) by:

$$F_t = (I - \Phi L)^{-1} v_t = \sum_{k=0}^{\infty} \Phi^k v_{t-k} \tag{8}$$

where $F_t = (f_t, g_{1t}, \ldots, g_{Bt})'$ and $v_t = (v_{f,t}, v_{g1,t}, \cdots, v_{gB,t})'$. The mean-square-error (MSE) for the s-step ahead prediction of F_t is given by:

$$\begin{aligned} MSE(F_{t+s|t}) &= \mathrm{Var}\left(\sum_{k=0}^{s-1} \Phi^k v_{t+s-k}\right) = \sum_{k=0}^{s-1} \Phi^k \Sigma (\Phi^k)' \\ &= \sum_{j=1}^{B+1}\sum_{k=0}^{s-1} \Phi^k p_j p_j' (\Phi^k)' = \sum_{j=1}^{B+1} MSE_j(F_{t+s|t}) \end{aligned} \tag{9}$$

where $MSE_j(F_{t+s|t}) = \sum_{k=0}^{s-1} \Phi^k p_j p_j' (\Phi^k)'$ denotes the effect of the j-th factor on the MSE, and p_j represents the j-th column of identity matrix Σ. Let $\Lambda_i = (\lambda_{yf,i}, \lambda_{yg,i}\gamma_{i1}, \ldots, \lambda_{yg,i}\gamma_{iB})'$. Then the MSE for the s-step ahead prediction of y_{it} is given by:

$$\begin{aligned} MSE(y_{i,t+s|t}) &= \Lambda_i' MSE(F_{t+s|t}) \Lambda_i + \mathrm{Var}(e_{it}) \\ &= \Lambda_i' \sum_{j=1}^{B+1} MSE_j(F_{t+s|t}) \Lambda_i + \mathrm{Var}(e_{it}) \end{aligned} \tag{10}$$

Therefore, the variance portion of $y_{i,t+s}$ due to the national and regional factors can be computed by a ratio of $\Lambda_i' MSE_j(F_{t+s|t})\Lambda_i$ over $MSE(y_{i,t+s|t})$, while the variance portion of $y_{i,t+s}$ due to its own variation can be computed by a ratio of $\mathrm{Var}(e_{it})$ over $MSE(y_{i,t+s|t})$.

3 Data

We used the provincial industrial value added to briefly analyze the regional characteristics of provincial business fluctuations in China, and further addressed the insufficiency of traditional divisions using the geographical regions. For regional economic data, the industrial value added is a commonly used indicator reflecting economic activity, and the data are officially available at a monthly frequency. Although the GDP growth rate would be a good proxy, provincial GDP data are quarterly or annual, with fewer observations than the industrial value added. Moreover, there is also no official agency or institution that extracts provincial economic climate indices in real time.

We selected the year-on-year monthly growth rates of provincial industrial value added for 31 provinces (Hong Kong, Macau, and Taiwan were not included due to the data unavailability) in China, sampled from January 2000 to December 2015. The data were sourced from the CEIC database (available at the website: https://www.ceicdata.com). Fig. 2 is a thermodynamic diagram for the data series. All series are standardized over time for better comparison. The list of provinces was sorted geographically into eastern, intermediate and western regions, separated by horizontal blue lines. The colors in the figure indicate the standardized growth rates: for example, the dark red denotes a high growth rate, while the deep green indicates a low growth rate. In addition, the vertical dashed lines mark the dates of the turning points of national business cycles based on Zheng and Wang (2013) and Zheng and Xia (2016), which include January 2003 (from recession to expansion), August 2008 (from expansion to recession), October 2009 (from recession to expansion), and October 2011 (from expansion to recession).

We can make the following observations from Fig. 2. First, the 31 provincial business fluctuations exhibited an obvious similarity. For example, most provinces were in recession at the beginning of 2002, while they were in expansion in the first half of 2014; and in early 2009, early 2010, and 2015, there was a high degree of business cycle synchronization across the provinces. Second, the provincial business fluctuations in the intermediate region had a high level of synchronization. For example, near the turning points of 2008 and 2009, the business fluctuations in the intermediate region were almost completely synchronized, indicating the presence of regional clustering. Third, the eastern and western regions were less synchronized in some periods. For example, from 2006 to 2007 and around 2010, the

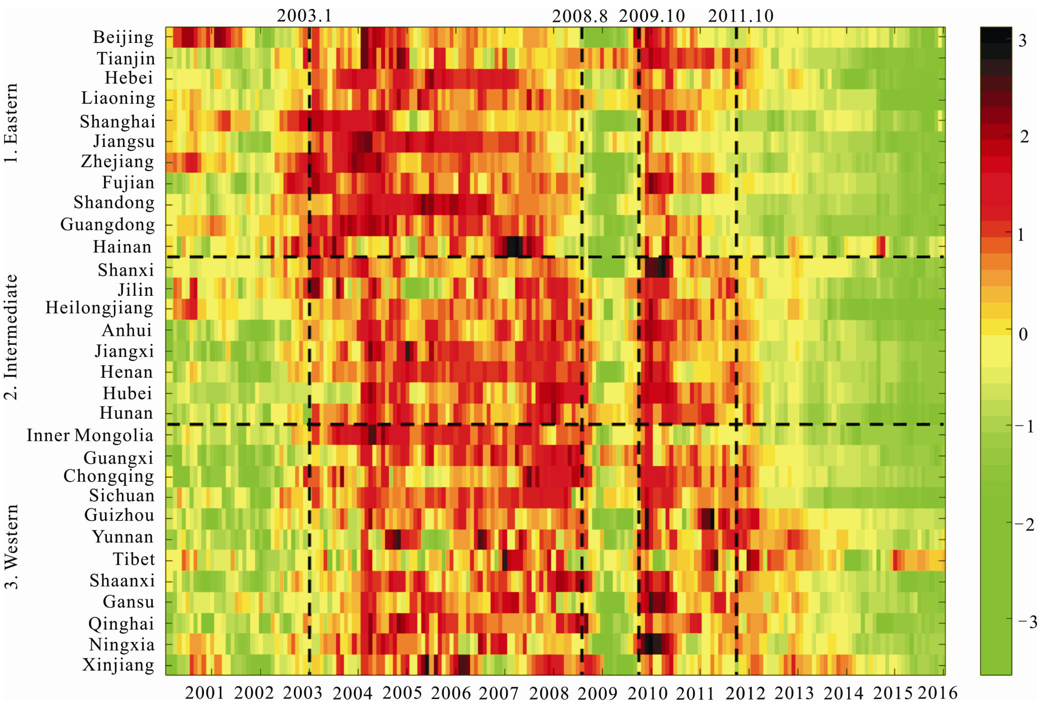

Fig. 2 Monthly growth of the industrial value added for 31 provinces

provincial business fluctuations in the eastern region had an obvious differentiation. The indication of regional comovement was even weaker in the western region. Fourth, around the turning points of the national business cycle, the eastern region generally preceded the intermediate area for most cyclical fluctuations. There was also slight evidence that economic changes in the intermediate area preceded those in the western area.

This confirms the need to reconsider the division of regions when studying regional business cycle fluctuations. On the one hand, a division based on geographical locations can not adequately capture the regional clustering characteristics of provincial business fluctuations. It can not be ascertained if the unsynchronized points across these provinces are the result of idiosyncratic components or mistaken clusters. On the other hand, the leading-lag relationships of some provincial business fluctuations near the turning points imply possible dynamic interactions among different regions. Figure 2 shows that some eastern provinces led the cyclical movements, relative to the intermediate provinces. This addresses the importance of considering the dynamic interactions among regional factors instead of using an orthogonal factor model such as that proposed by Kose et al. (2003) or hierarchical factor models (Moench et al., 2013).

4 Results and Empirical Analysis

4.1 Results of regional clustering

Using the 31 provincial data of business flutuations, we carry out the MCMC estimation for the multilevel dynamic factor model. Under different b, $b = 1,\ldots, B$, the MCMC procedures were repeated 10 000 times. The first 2000 iterations were discarded while the remaining iterations were used for posterior estimation. The estimated regional grouping of provinces was determined by $\{\gamma_{ib}^{(k)}\}_{k=1}^{M}$, such that province i belongs to region b, $b = 1,\ldots, B$.

Based on the deviance information criteria (DIC) presented in Subsection 2.4, we firstly attempt to determine the number of regional clusters. The DIC results under different values of B are reported in Table 1. From the table, the approximated DIC has a minimum value at $B = 4$. This suggests that the number of clusters or regions should be 4.

Given the optimum number of regions, we then used the MCMC method to estimate the model under the given number $B = 4$. Due to space limitations, we do

Table 1 The DIC results under different values of B

B	2	3	4	5	6	7	8
DIC(B)	10186.81	9710.03	5508.27	8697.54	8501.01	5875.71	6971.14

Fig. 3 The division of regions based on regional clustering

not report the posterior parameter estimates in this study. Based on the MCMC estimation, we obtained the four regional clusters. More specifically, the 31 provinces in mainland China were clustered into four regions: leading, coincident, lagging, and overshooting.

Fig. 3 shows the division of regions based on our regional clustering. The leading region contained the six provinces of Beijing, Shanghai, Jiangsu, Zhejiang, Guangdong and Hainan. These provinces were mainly located on the eastern coast. The coincident region consisted of 16 provinces. This region covered a much larger territory than the intermediate region shown in Fig. 1. For example, the eastern provinces of Shandong, Hebei, Liaoning, and Tianjing, and the western provinces of Sichuang, Chongqing, Guangxi, Shaanxi, Inner Mongolia, and Ningxia, were classed as the coincident region. The lagging region includes the most remotely located western provinces of Guizhou, Yunnan, Tibet, Qinghai, and Xinjiang. These provinces used to be isolated from other areas of the country and behave differently in comparison to other regions. The overshooting region contained the four provinces of Fujian, Shanxi, Gansu, and Ningxia. Although the provinces in this region were geographically separated, our later results indicated their overshooting phenomena during the periods of the 2008 financial crisis and the 'four-trillion economic stimulus plan'.

4.2 Regional synchronization and business cycle fluctuations

To reflect the regional synchronization in Chinese provinces, a thermodynamic diagram for the growth rate series obtained by rearranging the provinces according to the estimated regional clusters is shown in Fig. 4. From this figure, the following results were obtained. First, the 31 provinces were well clustered into four regions in term of their provincial business fluctuations. In comparison to the eastern provinces shown in Fig. 2, the leading provinces were better synchronized around the turning point of January 2003 and during the periods from 2005 to 2007 and 2008 to 2011. Moreover, the provincial business fluctuations in the coincident region were quite consistent with changes in the national business cycle. Most provinces in this region switched between the red and green areas synchronously around the four turning points.

Second, there were obvious lead-lag relationships among the leading, coincident, and lagging regions. This was also the reason why we named three of the four regions as leading, coincident, and lagging. Around the turning point of January 2003, most provinces in the coincident region started to enter economic expansion,

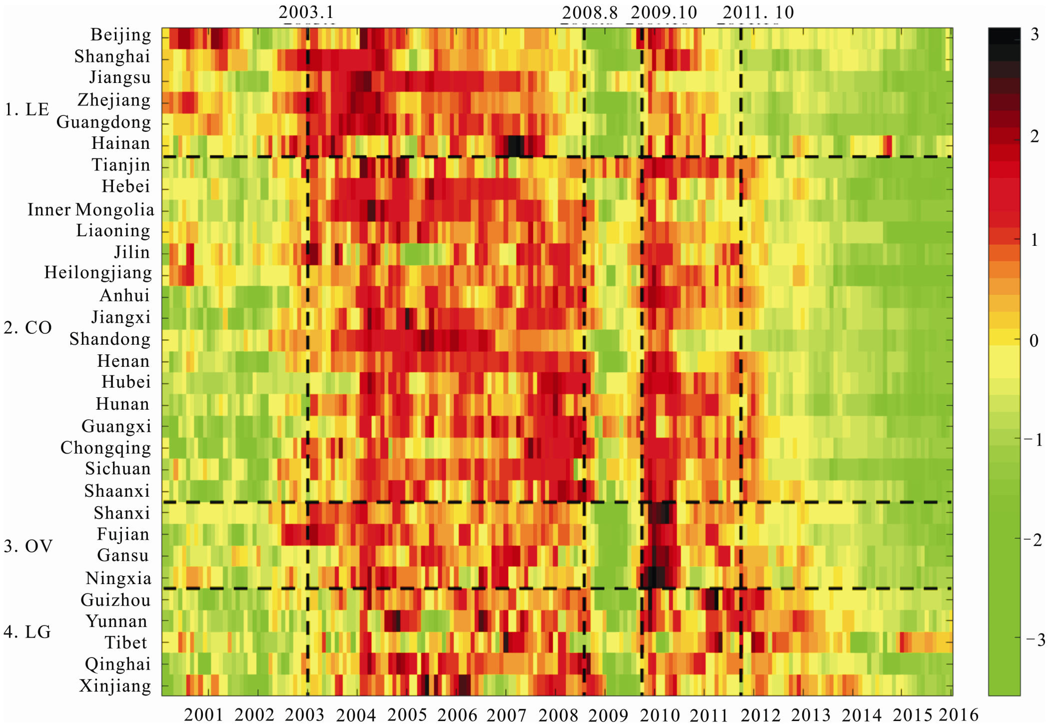

Fig. 4 Provincial growth data sorted by estimated regional clusters of provinces (LE: leading region; CO: coincident region; OV: overshooting region; LG: lagging region)

while the provinces in the leading region first entered economic expansion in mid-2002 and the provinces in the lagging region entered economic expansion in early 2004. Around the turning point of August 2008, the leading provinces started to enter economic recession first in late 2007, about one year ahead of the provinces in other regions. Around the turning point of October 2011, the lead-lag relationships were clearer, with the leading provinces entering economic recession in early 2011, while the coincident provinces entered economic recession about one year later, and the lagging provinces followed after 2012.

The lead-lag relationships among the leading, coincident, and lagging regions can be further explained from empirical facts. Due to the geographical location of the east coast of China, the higher degree of economic openness, and the greater benefits from national economic policies, the provincial economics in the leading region was always first to react. However, they were also more easily influenced by external and policy shocks, such as the 2008 financial crisis and the 'four- trillion economic stimulus plan'. In contrast, the provinces in the lagged region were mainly in the remote western area and were insensitive to external shocks due to their low degree of economic openness. They were also largely unaffected by several major regional economic policies established in the country. However, provincial business fluctuations in the lagged region were also affected by national economics and the spillover effects from the leading and coincident regions.

Third, there was an overreaction or overshooting phenomenon in the overshooting region due to the financial crisis and the four-trillion economic stimulus plan. During the period from 2008 to 2010, the overshooting provinces were recorded as a deeper green and deeper red than the other provinces, indicating a deeper trough after the subprime mortgage crisis and a higher peak after the 'four-trillion economic stimulus plan'. In addition, the overshooting provinces displayed similar trends to the leading, coincident, and lagging regions. By comparing provincial business fluctuations, Fujian matched the leading region, Shanxi matched the coincident region, and Gansu and Ningxia matched the lagging region. The reason why these provinces were originally classed as overshooting was because their economies are mostly dependent on the output of resources and the corresponding industrial structure was relatively singular. If their economic system encountered an external shock, the adjustment would be particularly intense.

4.3 Analysis of regional factors

Through the MCMC computation, we also obtained four estimated regional factors, $\{g_{1t},\cdots,g_{4t}\}$, which were the common factors of provincial business fluctuations within regions and could be used to represent the indicators of regional business cycle fluctuations.

Fig. 5 plots the time series paths of these regional factors for the four estimated regions. The leading-lag relationships were clearly visible among the leading, coincident, and lagging factors, and the different time-series dynamics of regional factors were also apparent. The leading factor followed an overall downward trajectory. Before the financial crisis of 2008, the leading region was in expansion, with values larger than zero. It then began to decline as the regional economy contracted sharply due to the financial crisis. It rapidly recovered under the 'four-trillion economic stimulus plan', but then entered recession after early 2011. The coincident factor initially rose and then decreased. Before 2008, the coincident factor continued to increase from a low negative value to a large positive value. It fell to a near zero value during the financial crisis and then recovered to a high positive value due to the 'four-trillion economic stimulus plan'. After the mid-2010s, the factor continued to decrease and became negative after late 2012. The lagging factor followed a largely upward trajectory. Since 2001, this region experienced a gradual increase and the pattern was similar to that of the coincident region during the period from 2004 to 2008. The regional economy also slowed down in 2009 but recovered quickly and continued to grow until early 2012. A significant feature of this region was that it succeeded in maintaining a high level of growth after 2012. The overshooting factor behaved more or less the same as the coincident factor. The overshooting region experienced a significantly deep trough in 2009 and a high peak in 2010, indicating that it was severely affected by the financial crisis and benefitted substantially from the 'four-trillion economic stimulus plan'. Its successful rebound was largely due to the high level of growth before the outbreak of the subprime crisis, although the overshooting effect due to the fiscal stimulus

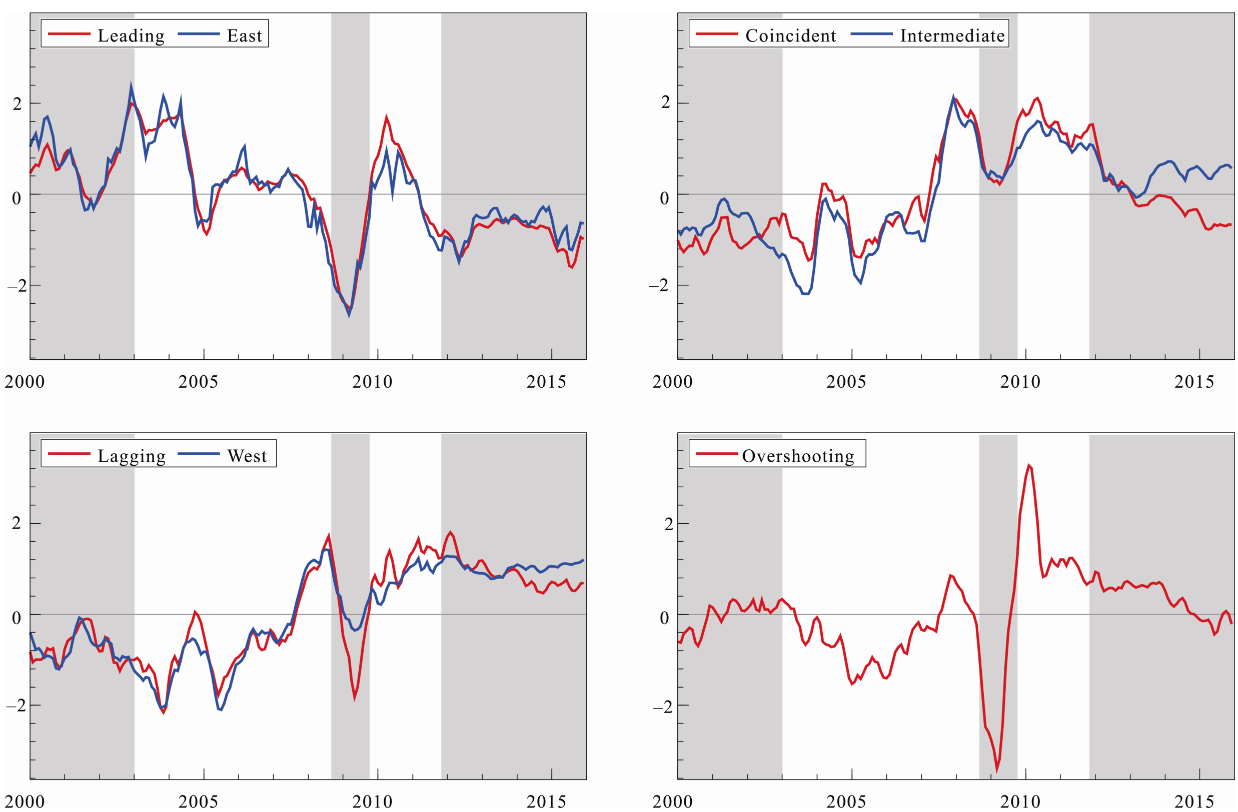

Fig. 5 Regional factors for the estimated and given geographical regions. Shaded areas represent economic recessions

lasted for a very short period. The region maintained a relatively high level of growth during the period from 2011 to 2013. After 2014, the economy of the overshooting region gradually declined.

Although the multilevel factor model could also be estimated for given geographical regions, the corresponding regional factors could be underestimated or overestimated. For further comparison, Fig. 5 also plots the regional factors estimated by treating the eastern, intermediate, and western regions as given geographical regions. We compared these three regional factors for the given geographical regions with the leading, coincident, and lagging factors, respectively. The eastern factor was similar to the leading factor, but in some periods there were certain deviations between them. For example, the eastern factor was underestimated in 2010 because some of the provinces it included, such as Liaoning, Shandong, and Tianjing, were not well synchronized with the other provinces in the eastern region. This result was consistent with the findings of Herrerias and Ordóñez (2012). Second, the downward trend of most provincial business fluctuations in the intermediate region during the period from 2013 to 2015 (see Fig. 2) was not consistent with the rising trend of the intermediate factor. This implies that the intermediate factor could not reflect the regional synchronization of provincial business fluctuations in the intermediate region. In contrast, this may be better captured by the coincident factor. Third, the regional factors of the western and lagging regions were similar, with the exception of some disparities, such as the trough of 2009 and during the period from 2013 to 2015. In conclusion, the three estimated factors for the given geographical regions performed worse than the factors for unknown regions.

4.4 Regional integrations: a variance decomposition analysis

Variance decomposition analysis based on the multilevel factor model is a useful way to determine whether the estimated factors in Fig. 5 are 'common' enough within each region (Kose et al., 2003). If a regional factor explains a substantial share of the provincial variations in its region, then the within-region integration is strong, and this regional factor reveals the common property of the region. The higher the proportion of provincial variations in the region explained by the regional factor, the higher the similarity of provincial business fluctuations within the region is, and the more representative the common factor is.

Based on the variance decomposition shown in Subsection 2.5, the results under both unknown and given regions are reported in Table 2. From the table, the following findings are apparent. First, the four estimated factors for unknown regions can explain more variations than the three estimated factors for given geographical regions. As shown in the left panel of Table 2, on average, about 19% of variations were explained by the estimated four regional factors, which was much higher than the corresponding 8% under the geographical regions in the right panel. According to Kose et al. (2003), this result implies that the integration within regions was considerably higher than the integration across regions, which demonstrates the importance of characterizing the regional business cycle fluctuations endogeneously.

Second, the leading, coincident, and overshooting factors explained more than 15% of the variations. In terms of the proportion of factors at the average level within the regions, the regional factors explained 15% and 20% of the variations in 6 leading and 16 coincident provinces, respectively. This indicates that the provinces in the leading and coincident regions were strongly correlated. In the overshooting region, the regional factor explained 31% of the variations in the four overshooting provinces, which confirmed the dominant role of the overshooting regional factor. This strong within-region integration may be the consequence of a similar economic structure or local government strategies during the financial crisis and the 'four-trillion economic stimulus plan'. In the lagging region, the proportion was only 8%, which implies that the within-region integration was relatively weak and the regional factor played a minor role. This result was similar to that reported by Poncet (2004) and Gerlach-Kristen (2009), indicating that the remote provinces were still rather isolated from the rest of the country.

Third, the proportion of the variation explained by the leading, coincident, and lagging regional factors was considerably higher than the proportion explained by the eastern, intermediate, and western regions. Both the proportion of variation explained and the number of provinces in the coincident region were double that of the intermediate region, which demonstrates the strong trend toward integration in the coincident provinces of China.

Table 2 Variance decomposition for the estimated and given geographical regions

Estimated regions	
Leading	
Beijing	0.07
Shanghai	0.33
Jiangsu	0.01
Zhejiang	0.36
Guangdong	0.11
Hainan	0.03
Regional Ave.	0.15
Coincident	
Tianjin	0.22
Hebei	0.03
Inner Mongolia	0.04
Liaoning	0.09
Jilin	0.11
Heilongjiang	0.11
Anhui	0.37
Jiangxi	0.20
Shandong	0.04
Henan	0.15
Hubei	0.46
Hunan	0.27
Guangxi	0.28
Chongqing	0.42
Sichuan	0.16
Shaanxi	0.30
Regional Ave.	0.20
Lagging	
Guizhou	0.13
Yunnan	0.10
Tibet	0.01
Qinghai	0.10
Xinjiang	0.05
Regional Ave.	0.08
Overshooting	
Shanxi	0.58
Fujian	0.20
Gansu	0.25
Ningxia	0.22
Regional Ave.	0.31
Overall Ave.	0.19

Given geographical regions	
Eastern	
Beijing	0.12
Tianjin	0.03
Hebei	0.02
Liaoning	0.02
Shanghai	0.36
Jiangsu	0.03
Zhejiang	0.62
Fujian	0.03
Shandong	0.02
Guangdong	0.18
Hainan	0.05
Regional Ave.	0.13
Intermediate	
Shanxi	0.02
Jilin	0.04
Heilongjiang	0.04
Anhui	0.14
Jiangxi	0.07
Henan	0.06
Hubei	0.30
Hunan	0.09
Regional Ave.	0.10
Western	
Inner Mongolia	0.04
Guangxi	0.04
Chongqing	0.05
Sichuan	0.03
Guizhou	0.04
Yunnan	0.05
Tibet	0.01
Shanxi	0.05
Gansu	0.02
Qinghai	0.03
Ningxia	0.01
Xinjiang	0.01
Regional Ave.	0.03
Overall Ave.	0.08

Note: The values in the table represent the proportion of the variation explained by the regional factors or provinces

5 Conclusions

Because the traditional regional division in China cannot adequately describe regional economic synchronization, this study devised a novel specification using a multilevel factor model with an endogenous regional grouping strategy and estimations realized by the Bayesian MCMC method. Using this model, we conducted an empirical investigation of the provincial business fluctuations in the 31 mainland provinces in China.

Based on our empirical analysis, we can draw the following conclusions. First, the 31 mainland provinces could be endogeneously clustered into four regions: leading, coincident, lagging, and overshooting, which directly reflected the dynamic interactions among them. Second, this new division of regions allowed us to more accurately capture the regional synchronization of provincial business fluctuations, compared to the traditional division into geographical regions. Empirically, the regional factors reflected the business cycle fluctuations well across the provinces within these regions, and they displayed a more intensive integration than that indicated from the traditional geographical regions. Third, the four estimated regional factors also reflected the substantial heterogeneities among regional business cycle fluctuations, especially during the periods of the 2008 subprime mortgage crisis and the "four-trillion economic stimulus plan".

References

Bai J S, Wang P, 2015. Identification and Bayesian estimation of dynamic factor models. *Journal of Business & Economic Statistics*, 33(2): 221–240. doi: 10.1080/07350015.2014.941467

Carter C K, Kohn R, 1994. On Gibbs sampling for state space models. *Biometrika*, 81(3): 541–553. doi: 10.1093/biomet/81.3.541

Celeux G, Forbes F, Robert C P, et al., 2006. Deviance information criteria for missing data models. *Bayesian Analysis*, 1(4): 651–673. doi: 10.1214/06-BA122

Crone T M, Clayton-Matthews A, 2005. Consistent economic indexes for the 50 states. *The Review of Economics and Statistics*, 87(4): 593–603. doi: 10.1162/003465305775098242

Diebold F X, Rudebusch G D, Aruoba S B, 2006. The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, 131(1–2): 309–338. doi: 10.1016/j.jeconom.2005.01.011

Francis N, Owyang M T, Savascin Ö, 2012. An endogenously clustered factor approach to international business cycles. Federal Reserve Bank of St. Louis Working Paper. St. Louis: Federal Reserve Bank.

Gatfaoui J, Girardin E, 2015. Comovement of Chinese provincial business cycles. *Economic Modelling*, 44: 294–306. doi: 10.1016/j.econmod.2014.10.015

Gerlach-Kristen P, 2009. Business cycle and inflation synchronisation in Mainland China and Hong Kong. *International Review of Economics & Finance*, 18(3): 404–418. doi: 10.1016/j.iref.2008.09.006.

Geweke J, Keane M, 2001. Computationally intensive methods for integration in econometrics. *Handbook of Econometrics*,, 5: 3463–3568. doi: 10.1016/S1573-4412(01)05009-7

Guo Q W, Jia J X, 2005. Dynamic factor analysis of provincial business cycles in China. *Management World*, (11): 50–58. (in Chinese)

Guo Q W, Zhao X J, 2012. Local governmental investment competition and business cycle fluctuation. *The Journal of World Economy*, (5): 3–21. (in Chinese)

He D, Liao W, 2012. Asian business cycle synchronization. *Pacific Economic Review*, 17(1): 106–135. doi: 10.1111/j.1468-0106.2011.00574.x

Herrerias M J, Ordóñez J, 2012. New evidence on the role of regional clusters and convergence in China (1952–2008). *China Economic Review*, 23(4): 1120–1133. doi: 10.1016/j.chieco.2012.08.001

Huang J L, Li K W, Li D F, 2011. The synchronization of regional real business cycle in China. *The Journal of World Economy*, (9): 19–41. (in Chinese)

Huang Q, Chand S, 2015. Spatial spillovers of regional wages: evidence from Chinese provinces. *China Economic Review*, 32(2): 97–109. doi: 10.1016/j.chieco.2014.12.001

Kim C J, Nelson C R, 1999. *State-Space Models With Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. Cambridge, MA: MIT Press.

Kose M A, Otrok C, Whiteman C H, 2003. International business cycles: World, region, and country-specific factors. *American Economic Review*, 93(4): 1216–1239. doi: 10.1257/000282803769206278

Lee J, 2012. Measuring business cycle comovements in Europe: evidence from a dynamic factor model with time-varying parameters. *Economics Letters*, 115(3): 438–440. doi: 10.1016/j.econlet.2011.12.125

Li Y, Zeng T, Yu J, 2013. Robust deviance information criterion for latent variable models. CAFE Research Paper No. 13.19. doi: 10.2139/ssrn.2316341

Moench E, Ng S, Potter S, 2013. Dynamic hierarchical factor models. *The Review of Economics and Statistics*, 95(5): 1811–1817. doi: 10.1162/REST_a_00359

Ouyang P M, Fu S H, 2012. Economic growth, local industrial development and inter-regional spillovers from foreign direct investment: evidence from China. *China Economic Review*, 23(2): 445–460. doi: 10.1016/j.chieco.2012.03.005

Poncet S, 2004. Are Chinese provinces forming an optimal currency area? Magnitude and Determinants of Business Cycles within China. Working Paper, 2004.

Stock J H, Watson M W, 2010. The evolution of national and regional factors in US housing construction. In: Bollerslev T, Russell J, and Watson M (eds). *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*. Oxford: Oxford University Press, 35–61. doi: 10.1093/acprof:oso/9780199549498.001.0001

Zhang W B, Tong D, 2011. Provincial industrial business cycles in China. *The Journal of Quantitative & Technical Economics*, 28(1): 104–116. (in Chinese)

Zheng T G, Wang X, 2013. Measuring China's business cycle with mixed-frequency data and its real time analysis. *Economic Research Journal*, 48(6): 58–70. (in Chinese)

Zheng T G, Xia K, 2016. Macroeconomic data releasing and methodology research on measuring China's business cycle in the real-time. *Systems Engineering-Theory & Practice*, 37(4): 817–830. (in Chinese)