Acta Technica 62 (2017), No. 6B, 837–848

Process reasoning method based on HDP-HMM and its computer simulation analysis

Hui Wang[1], Wuqiang Zhang[2]

Abstract. In order to improve the accuracy of analysis for the impact of FDI on green innovation capability, this paper proposes a method of analyzing the impact of FDI on green innovation capability based on HDP-HMM process reasoning. First, the model description of impact analysis of FDI on the green innovation capacity is given, and green innovation resources of China's manufacturing green innovation system are regarded as intermediary variable, to construct the conceptual model of impact of FDI inflow on green innovation capability of manufacturing green innovation system; secondly, HDP-HMM algorithm based on hierarchical Dirichlet process is proposed, to automatically infer the impact analysis state of FDI on the green innovation capacity; finally, through simulation experiments, the effectiveness of the proposed method is verified numerically.

Key words. HMM, Process reasoning, FDI factor, Green innovation.

1. Introduction

China has become the world's largest developing country to attract foreign investment in the context of global increasing emphasis on sustainable development and environmental regulation. FDI has not only brought a strong financial support for China's manufacturing industry to carry out green innovation activities, but also become one of the main sources for China's manufacturing industry to obtain advanced green innovation resources. With the expansion of the total amount and scale of FDI inflow, FDI has become the main driving force to promote the development of China's manufacturing green innovation system. However, because of excessive reliance of China's manufacturing industry on foreign investment and green innovation resources of foreign enterprises in the green innovation, some negative effects

[1]College of Economics & Management, Anhui Agricultural University, Hefei city of Anhui Province , China

[2]Department of Money and Credit Management, Hefei Central Sub-branch of the People's Bank of China, Hefei city of Anhui Province, China

of FDI inflow appear gradually, such as the "Pollution Haven" hypothesis and inhibition of green innovation in China's manufacturing industry by FDI inflow, etc. Therefore, the researches on whether the inflow of FDI promotes the development of green innovation system of China's manufacturing industry and the improvement of green innovation capability, and what impact of FDI inflow on China ecological environment, etc., have become the focus of scholars increasingly.

China's manufacturing industry is a typical technology and innovation-driven industry, and in order to achieve efficient, energy-saving, environmental and recycling new manufacturing industry, we must increase input in green innovation resources of China's manufacturing industry, and actively carry out green innovation activities, so as to promote the green innovation capability of green innovation system. However, green innovation is developed on the basis of technological innovation, which is a complex innovation process, and it is difficult to achieve results relying solely on independent green innovation of China's manufacturing industry in the short term. Therefore, one of the more reasonable methods is to make full use of the role of FDI in the green innovation of China's manufacturing industry. Obviously, the research on how China's manufacturing industry take effective and reasonable advantage of innovation resources brought by FDI inflow to make it give full play to its role in the green innovation capability of green innovation system according to practical conditions of China's manufacturing industry; and how to avoid the negative impact of FDI on China's manufacturing economy, technology and ecological environment, has important theoretical and practical significance.

In order to achieve the purpose of improving the accuracy of the impact analysis algorithm of FDI on green innovation capability, the paper puts forward a kind of analysis method of impact of FDI on green innovation capability based on HDP-HMM process reasoning, and takes green innovation resources of China's manufacturing green innovation system as intermediary variable, to construct the conceptual model of impact of FDI inflow on green innovation capability of green innovation system, and achieve the automata inference of impact analysis state of FDI on green innovation capability.

2. Conceptual model

The green innovation capability of manufacturing green innovation system belongs to the category of strategic management of manufacturing enterprises, which is the key to realize the manufacturing green innovation and enhance the market competitiveness of manufacturing enterprises. As one of the most direct and effective ways to improve the green innovation capability of China's manufacturing green innovation system, the impact of FDI on the green innovation capability is not only the result of direct action. Based on the theoretical analysis of open innovation, the improvement of green innovation capability of manufacturing green innovation system must be realized by coordinating the internal and external green innovation resources in the green innovation system of manufacturing industry. FDI is one of the main external sources of green innovation resources of China's manufacturing green innovation system, which has an important impact on the input of green innovation resources of China's manufacturing green innovation system. This research, based on previous theoretical study of relevant literatures and the relevant concepts of FDI and innovation resource, takes green innovation resources of China's manufacturing green innovation system as intermediary variable, to construct the conceptual model of impact of FDI inflow on green innovation capability of China's manufacturing green innovation system, as shown in Figure 1.

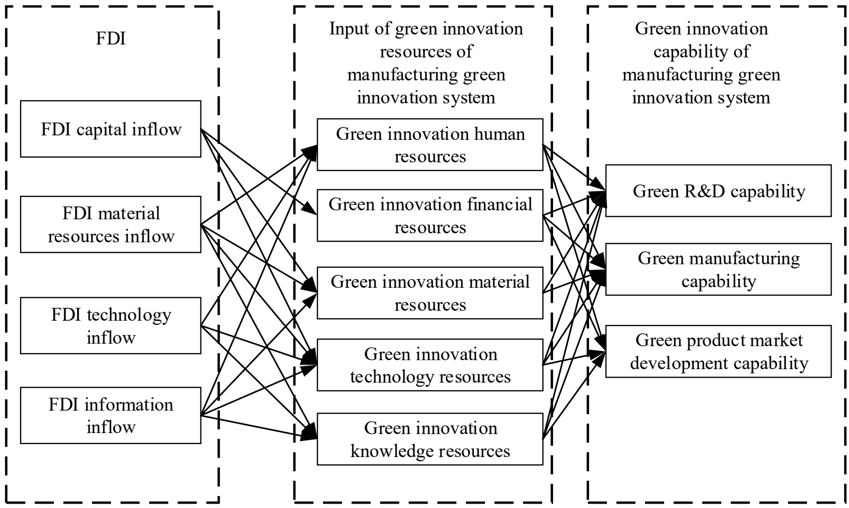


Fig. 1. Conceptual model of impact of FDI inflow on green innovation capability of China's manufacturing green innovation system

As the basis of green innovation activities in China's manufacturing industry, green innovation resources are the key elements to enhance the green innovation capability of China's manufacturing green innovation system. However, China's manufacturing industry is generally faced with the shortage of innovation resources, especially the green innovation resources, FDI inflow is the key factor to make up for the shortage of green innovation as external innovation resources of China's manufacturing industry. According to Slaughter (2002), FDI inflows provide the necessary innovation resources for the technological innovation of the host country, which has a positive impact on the innovation system resources input of the host country.

3. Algorithm theory description

3.1. Nonparametric Bayesian algorithm

HMM data segmentation needs to preassign the number of states, and uses EM algorithm to estimate parameters. For this reason, HDP-HMM algorithm based on hierarchical Dirichlet process is proposed to realize the automatic inference of state number. The model can be interpreted as a dynamic update of group numbers by using the hierarchical Dirichlet process based on state assignment. During processing, FDI impact data is divided into fixed 20 s fragment sequence, for each fragment, the statistics of total number is first made at all data positions, and then the space is divided into 10 * 10 uniform grid. After vectorization, 100 dimensional feature vectors can be obtained for each FDI data segment. In the model selection phase, the hourly feature vector in the window is calculated by the total active level within – hours, and then added to obtain the total length of vectors, so as to get total active scalar value. Infinite HMM model can be segmented by mixed active level model in Gaussian state. Once the active level of the segmented activity time is obtained, a separate anomaly detector can be run for each model.

Dirichlet process $DP\,(\gamma, H)$ is the distribution of discrete random capability measure G in $(\mathcal{Q}, \mathcal{B})$. An alternative definition [12~13] is given here:

$$G = \sum_{k=1}^{\infty} \beta_k \delta_{\varphi_k} \,. \tag{1}$$

$$\varphi_k \overset{iid}{\sim} H, k = 1, \cdots, \infty \,. \tag{2}$$

$$\beta = (\beta_k)_{k=1}^{\infty} \,. \tag{3}$$

Where, β is the weight constructed through the "sticky breaking" structure, denoted as $\beta \sim G\mathrm{EM}\,(\gamma)$. The Dirichlet process has been widely used for the prior distribution mixture measure of Bayesian mixture model, and the Dirichlet process mixture model [14] is relatively classical.

3.2. HDP-HMM data stream segmentation

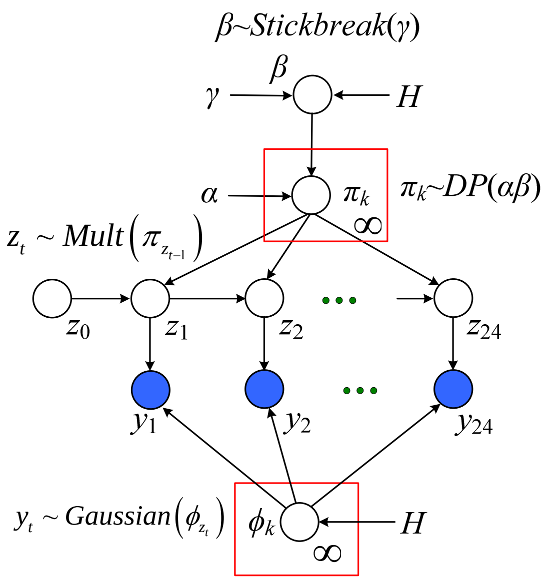
In the process of HDP, the method of replacing the number of states with Dirichlet prior number is provided. Therefore, the HDP method can be used to define the number of unknown states in HMM. Using HDP as - kinds of nonparametric building block, the HDP-HMM stochastic process can be described as:

$$\begin{cases} G_0 \sim DP(\gamma, H \times S), \theta_t \overset{iid}{\sim} G_k \,, \\ G_k \overset{iid}{\sim} DP(\alpha, G_0), y_t \sim F(\theta_{t-1}) \,, \end{cases} \tag{4}$$

Where, $k = 1, 2, \cdots \infty$, $t = 1, 2, \cdots, T$. There is T time steps in total, and T can be defined as the number of hours of data collected within a day. The HDP-HMM process is shown in Figure 2.

The relevant parameters in Figure 2 meet the following distribution:

$$\begin{cases} \beta \sim G\mathrm{EM}\,(\gamma)\,, \pi_k \sim DP(\alpha, \beta)\,, \\ \varphi_k \sim H, Z_t \sim \pi_{Z_{t-1}}, y_t \sim F(\varphi_{Z_t})\,. \end{cases} \tag{5}$$



(b) Data presentation

Fig. 2. HDP-HMM model representation

3.3. HDP-HMM process reasoning

Firstly, the HMM is used to divide the data into coherent slices in the - stage, which can fully consider the temporal dynamics of the data. The number of coherent slices of FDI impact data is unknown and needs to be estimated. The primary goal is to perform rough data segmentation at hourly intervals, with 24 input data points obtained, and corresponding to the observed variables $\{y_t\}$ and $\{y_t\}$ observed correspondingly play a role of potential state variable in the standard HMM. H is basic measure, which can sample for the parameter $\{\varphi_k\}$. In the proposed model, y_t is modeled as the Gaussian process of single variable, and φ_k is a tuple $\{\mu_k, \sigma_k^2\}$, of which μ_k and σ_k^2 are unknown and regarded as random variables. H is used as a conjugate prior, thus satisfying the Gaussian invGamma distribution in the case.

In HMM, the Gibbs reasoning is used to sample the potential state z_t and the weight β_k in order, and obtain the emission probability and transition probability through explicit integration of the parameters $\{\varphi_k\}$ and $\{\pi_k\}$. For example, given $z_{t-1} = i$ and $z_{t+1} = j$ according to the previous iteration, the conditional Gibbs distribution form of z_t sampling is:

(1) z_t sampling. Consider the conditional probability of z_t:

$$\begin{aligned} &p(z_t = k|z_{-t}, y, \beta, H) \propto, \\ &p(y_t|z_t = k, z_{-t}, y_{-t}, H) \times p(z_t = k|z_{-t}, \alpha, \beta)\,. \end{aligned} \tag{6}$$

In the formula, $p(y_t|z_t = k, z_{-t}, y_{-t}, H)$ is the likelihood value of the observed value y_t, which can be expressed as another form:

$$\int_{\varphi_k} p(y_t|z_t = k, \varphi_k)p(\varphi_k|y_{-t}, z_{-t}, H)d\varphi_k\,, \tag{7}$$

Formula (7) can be analyzed based on conjugate properties. In the formula (7), order term $p(z_t = k|z_{-t}, \alpha, \beta)$ is a transitional process. Assumed that n_{ij} is the number of transitions from state $i$to j, and n_{*j} is the number of states from all transitions to state j. Similarly, if n_{i*} is defined as the number of all transition processes left from the state i, the Markovian properties of the transition process can be described as:

$$\begin{aligned} &p(z_t = k|z_{-t}, \alpha, \beta) \propto\,, \\ &p(z_t = k|z_{t-1}, \alpha, \beta) \times p(z_t = k|z_{t+1}, \alpha, \beta)\,. \end{aligned} \tag{8}$$

Then, the above probability can be calculated in four cases:

$$p(z_t = k|z_{-t}, \alpha, \beta) \propto \begin{cases} (n_{z_{t-1},k} + \alpha\beta_k)\dfrac{n_{k,z_{t+1}} + \alpha\beta_{z_{t+1}}}{n_{k*} + \alpha}, k \leq K, k \neq z_{t-1} \\ (n_{z_{t-1},k} + \alpha\beta_k)\dfrac{n_{k,z_{t+1}} + 1 + \alpha\beta_{z_{t+1}}}{n_{k*} + 1 + \alpha}, z_{t-1} = k = z_{t+1} \\ (n_{z_{t-1},k} + \alpha\beta_k)\dfrac{n_{k,z_{t+1}} + \alpha\beta_{z_{t+1}}}{n_{k*} + 1 + \alpha}, z_{t-1} = k \neq z_{t+1} \\ \alpha\beta_{new}\beta_{Z_t+1}, k = K + 1 \end{cases} \tag{9}$$

(2) The calculation process of weight parameter β, hyper-parameters α and γ is the same as HDP algorithm, see literature [6] for details. In order to improve the robustness of the algorithm, assume that the hyper-parameters α and γ are distributed following Gamma, and resample them at each iteration of Gibbs.

It is assumed that $X \in \mathcal{R}^{d\times n}$ has n d-dimensional central feature vectors. C is the covariance matrix, and its SVD is decomposed into:

$$C = U\sum U^T\,. \tag{10}$$

Feature vectors from $U$are divided into two groups:

$$C = [U_1, U_2]\begin{bmatrix} \sum_1 & 0 \\ 0 & \sum_2 \end{bmatrix} U^T\,. \tag{11}$$

Where, the selection of $\sum_1$ and $\sum_2$ shall meet:

$$tr\left(\sum_1\right)\Bigg/\left(tr\left(\sum_1\right) + tr\left(\sum_2\right)\right) = 0.9\,. \tag{12}$$

That is, selecting the most important feature vector, to make its total energy coverage reach 90%. U_1 is called the principal subspace and U_2 is called the residual subspace. The anomaly detection algorithm is to compare the test vector with the detection threshold by protecting it to the residual subspace U_2, and at the same time, it also is the function of non-principal eigenvalue in the residual subspace.

4. Research design

4.1. Data sources

This research, based on combination of research of domestic and foreign literatures FDI and innovation capability variables measure, is to set the measure indexes of variables in FDI inflow, green innovation resources and innovation capability of manufacturing green innovation system according to the research purpose, innovation input and output theory and limitations of index data collection.

This research mainly adopts statistical data of 2006-2010 in 28 manufacturing industries, to form - panel sample data sets for empirical research, and the data are mainly from China Statistical Yearbook, Statistical Data on Scientific and Technological Activities of Industrial Enterprises, China Statistical Yearbook of Scientific and Technological Activities, China Torch Statistical Yearbook and statistical reports released on the official website of State Intellectual Property Office and National Bureau of Statistics. Because the data needed in this research can not be found directly from the statistical yearbook, this research is based on the previous literature on the treatment of this problem by calculating the derivation. For the choice of the industry, due to the "waste resources and waste materials recycling industry, tobacco industry of the two industry in the statistical yearbook of statistics data for lack of coherence, coherence based on the consideration, in order to ensure the accuracy of data analysis excluding these two industries

4.2. Data analysis and results

Structural equation modeling (SEM) is a kind of statistical modeling technique based on the theory of cause and effect, using multiple regression analysis, confirmatory factor analysis and path analysis. PLS and LIS-REL are the two most widely used modeling techniques in SEM. Because the data collected for the structural equation model constructed in this research from FDI inflow and green innovation capability of China's manufacturing green innovation system etc. can not meet Gaussian distribution, the sample size is not up to the maximum likelihood estimation, and there are certain multiple correlations between variables, the PLS path modeling technology in structural equation is selected for data analysis. The PLS path model based on structural equation is an estimation method with combination of principal component analysis and multiple linear regression, which is composed of two parts: measure model and structural model.

(1) The relicapability is mainly tested through Cronbach's a coefficient and Composite Relicapability (CR). It is believed that a is greater than 0.7, indicating that the latent variables in this research have better stability; Cronbach's a coefficient

and composite relicapability of all variables in Table 1 are higher than 0.7, indicating that the measured variables used in this research have better relicapability.

Table 1. Relicapability test of measurement model

Latent variable	Observed variable	Loading coefficient of crossed factor	Cronbach's a coefficient	Composite relicapability	AVE
FZ	FZ1	0.976	0.9496	0.9753	0.9519
	FZ2	0.974			
FW	FW1	0.982	0.9636	0.9820	0.9647
	FW2	0.983			
FJ	FJ1	0.942	0.8832	0.9448	0.8953
	FJ2	0.945			
FX	FX1	0.923	0.8628	0.9348	0.8776
	FX2	0.952			
CR	CR1	0.956	0.8943	0.9498	0.9044
	CR2	0.948			
CC	CC1	0.876	0.7051	0.8715	0.7723
	CC2	0.882			

(2) Validity test

In the structural equation model, the square root of crossed factor loading coefficient and AvE value is used to test convergent validity and discriminant validity of the measurement model respectively. As shown in Table 2, the crossed factor load of the measure index is higher than 0.7 and the AVE value is higher than 0.5, which indicates that the measured variables can effectively explain the latent variables, and have good convergent validity. According to Table 2, the square root of AVE is greater than the correlation coefficient of other measure indexes, which shows that the variables in this model have good discriminant validity.

Based on the mediating role of green innovation resources in the manufacturing green innovation system, the hypothesis test of FDI inflow and green innovation capability of manufacturing green innovation system is carried out. In order to verify the mediating role of manufacturing green innovation resources in the impact of FDI inflow on green innovation capability of green innovation system, this paper compares the path coefficient including intermediary variable model with the path coefficient excluding intermediary variable model, when there is great difference, it indicates significant mediating effect. The results of the research show that, before green innovation resources for the manufacturing green innovation system are not excluded, as shown in Table 3, the sum of path coefficients of impact of FDI flow into innovation resources on green innovation capability is 0.192703, 0.054111 and -0.049331 respectively. Among them, the impact of FDI capital inflow on green innovation capability is 0.188817, 0.126589 and 0.037751 respectively, that of FDI material inflow is 0.017383, -0.213435 and 0.152835; that of FDI technology inflow is -0.109424, -0.035465 and -0.273235; and that of FDI information inflow is 0.096196, 0.113890 and 0.109608. However, after excluding, it is found that the sum of path coefficients of impact of FDI flow into innovation resources on green innovation capability is 1.042872, 1.01969 and 0.896262, which is much greater than the impact path coefficient of the both in the original model.

Table 2. Validity test of measurement model

Latent variable	Observed variable	Loading coefficient of crossed factor	Cronbach's a coefficient	Composite relicapability	AVE
CW	CW1	0.952	0.9532	0.9693	0.9612
	CW2	0.947			
CJ	CJ1	0.987	0.9473	0.9752	0.9613
	CJ2	0.983			
CZ	CZ1	0.952	0.8256	0.9529	0.8892
	CZ2	0.955			
LY	LY1	0.934	0.8759	0.9275	0.9017
	LY2	0.941			
LZ	LZ1	0.952	0.8276	0.9456	0.8866
	LZ2	0.953			
LS	LS1	0.881	0.7152	0.8647	0.7899
	LS2	0.886			

Table 3. Path coefficient and test of the impact of FDI inflow on green innovation capability of China's manufacturing green innovation system

Relationship among variables	Path coefficient	Test results
FZ→CC	0.8291	Support
FZ→CW	0.1952	Support
FZ→CJ	0.0969	Support
FW→CR	0.2328	Support
FW→CW	0.3307	Support
FW→CZ	0.0678	Support
FJ→CR	0.3715	Support
FJ→CW	0.3245	Support
FJ→CJ	0.3149	Support

5. Conclusions

This paper proposes a kind of analysis method for the impact of FDI on green innovation capability based on HDP-HMM process reasoning and gives the model description for analysis of impact of FDI on green innovation capability, and achieves the automatic inference for analysis state of impact of FDI on green innovation capability by HDP-HMM algorithm, and the experimental results verify the effectiveness of the algorithm. Directions for future research are: (1) further expansion of impact factors, for example, select industrial agglomeration and environmental regulation, and analyze its impact on green innovation capability; (2) further improvement of the performance of HDP-HMM process reasoning algorithm.

Acknowledgement

Anhui College Humanities and social science key research project “Study on the long term mechanism of promoting green innovation of FDI in Anhui province from the perspective of low carbon economy ”(Item number: SK2017A0143);

Anhui Agricultural University Prosperity Fund key project “The measurement and analysis of the influence of foreign banks’ Greenfield Investment on the stability of China’s banking system” (Item number: 2016zs01zd);

Anhui Agriculture University introduction and stable talent project “Research on promoting green innovation mechanism of FDI from the perspective of low carbon economy” (Item number: yj2017-14).

References

[1] Hu D H, Zhang X X, Yin J, et al.: (2009) *Abnormal activity recognition based on HDP-HMM models*[C]// International Jont Conference on Artifical Intelligence. Morgan Kaufmann Publishers Inc. 2009:1715-1720.

[2] Hu D H, Zhang X X, Yin J, et al.: (2004) *Abnormal Activity Recognition Based on HDP-HMM Models.*[J]. Mathematical & Computer Modelling An International Journal, 40(1-2):11-21.

[3] Seger C J H, Bryant R E: (1995) *Formal verification by symbolic evaluation of partially-ordered trajectories*[J]. Formal Methods in System Design, 6(2):147-189.

[4] Tristan Henderson, Saleem Bhatti: (2003) *Networked games:a QoS-sensitive application for QoS-insensitive users?*[C]// ACM SIGCOMM Workshop on Revisiting Ip Qos: What Have We Learned, Why Do We Care. ACM, 2003:141-147.

[5] Berghaus A, Toplak F: (1987) *Surgical concepts for reconstruction of the auricle. History and current state of the art*[J]. Arch Otolaryngol Head Neck Surg, 112(4):388-397.

[6] Yokotani N, Sogawa K, Matsubara S, et al.: (1990) *cDNA cloning of cytochrome P-450 related to P-450p-2 from the cDNA library of human placenta. Gene structure and expression*[J]. European Journal of Biochemistry, 187(1):23-29.

[7] Plasmans J, Lukach R, Kort P M: (2005) *Innovation Strategies in a Competitive Dynamic Setting*[J]. General Information, 2005(2):1-26.

[8] Schulte B A, Müller-Schwarze D, Tang R, et al.: (2010) *Determinant Factors of Economic Growth in Malaysia: Multivariate Cointegration and Causality Analysis*[J]. European Journal of Economics Finance & Administrative Sciences, 21(24):123-137.

[9] Atkinson N S, Robertson G A, Ganetzky B: (1991) *A component of calcium-activated potassium channels encoded by the Drosophila slo locus.*[J]. Science, 253(5019):551-5.

[10] King S, Huang S, Renals S: (2009) *Analysis of Speaker Adaptation Algorihms for HMM-based Speech Synthesis and a Constrained SMAPLR Adaptation Algorithm*[J]. IEEE Audio Speech & Language Processing, 17(1):66 - 83.

[11] Yan H, Lazarian A, Draine B T: (2008) *Dust Dynamics in Compressible Magnetohydrodynamic Turbulence*[J]. 616(2):895.

[12] Montero J C, Seoane S, Ocaña A, et al.: (2011), *P-Rex1 participates in Neuregulin-ErbB signal transduction and its expression correlates with patient outcome in breast cancer*[J]. Oncogene, 30(9):1059-71.

[13] Schumann U, Hempel R, Flentje H, et al.: (2013) *Contrail study with ground-based cameras*[J]. Atmospheric Measurement Techniques, 6(2013):3597-3612.

Received May 7, 2017