

Chapter 10

Penalized Independent Factor

Y. Chen, R.B. Chen and Q. He

Abstract We propose a penalized independent factor (PIF) method to extract independent factors via a sparse estimation. Compared to the conventional independent component analysis, each PIF only depends on a subset of the measured variables and is assumed to follow a realistic distribution. Our main theoretical result claims that the sparse loading matrix is consistent. We detail the algorithm of PIF, investigate its finite sample performance and illustrate its possible application in risk management. We implement the PIF to the daily probability of default data from 1999 to 2013. The proposed method provides good interpretation of the dynamic structure of 14 economies' global default probability from pre-Dot Com bubble to post-Sub Prime crisis.

10.1 Introduction

Sovereign default probability reflects financial vulnerability and sovereign financing or refinancing difficulties or default of advanced and emerging market economies. It is considered as a fundamental early warning indicator of financial crises and contagions of global financial markets. Thus, sovereign credit ratings and the associated sovereign default rates continue to be a major concern of international financial markets and economic policy makers. According to the current version of Basel Capital Accord 3, financial institutions will be allowed to use credit ratings and the corresponding default rates to determine the amount of regulatory capital they have to

Y. Chen (✉) · Q. He

Department of Statistics and Applied Probability, National University of Singapore,
Singapore, Singapore

e-mail: stacheny@nus.edu.sg

Q. He

e-mail: hq19861027@gmail.com

R.B. Chen

Department of Statistics, National Cheng Kung University, Tainan, Taiwan

e-mail: rbchen@mail.ncku.edu.tw

© Springer-Verlag GmbH Germany 2017

W.K. Härdle et al. (eds.), *Applied Quantitative Finance*, Statistics and Computing,

DOI 10.1007/978-3-662-54486-0_10

reserve against their credit risks. It prompts the booming research interests on the determinants and co-movements of sovereign defaults.

While the large amount of information containing in the sovereign default data makes it possible to understand the dependence among economies, the massive sample size, high dimensionality and complex dependence structure of the data create computational and statistical challenges. It turns out that data analysis in a reduced space often accompanies with improved interpretability and estimation accuracy. This possibly explains the wide adoption of factor models in literature.

Factor models try to decipher complex phenomena of large dimensional data through a small number of basic causes or factors. Though the factors are often supposed to be macroeconomic and financial determinants, our study intends to launch a new investigation into the identification of factors of sovereign default probabilities in a data-driven way. From a statistical viewpoint, understanding the dependence among these sovereign default probabilities relies on the estimation of the joint probability distribution of the multiple variables. The conventional methods such as Principal Component Analysis (PCA) and Factor Analysis (FA) extract a set of uncorrelated factors from the multivariate and dependent data within a linear framework. Under Gaussianity, non-correlation is identical to independence. With the aid of Jacobian transformation, the complex joint distribution can be obtained by using the marginal distributions of each factor in a closed form. Thus, the high dimensional statistical problem is converted to univariate cases. Independence however does not hold, if the measured variables e.g. the sovereign default probabilities are not Gaussian distributed, which is most likely in practice. In this case, the joint distribution estimation cannot be easily solved with the help of the conventional methods.

The recently developed Independent Component Analysis (ICA) method sheds lights on possible solutions. Similar to the PCA and FA methods, the ICA identifies essential factors via a linear transformation. Instead of projecting onto the eigenvectors of the covariance matrix as PCA does, the ICA directly extracts statistical independent factors from the original complex data via solving an optimization problem on statistical cross-independence. Depending on the definition of independence, various estimation methods have been proposed, including the maximization of nongaussianity (Jones and Sibson 1987; Cardoso and Souloumiac 1993; Hyvärinen and Oja 1997), the minimization of mutual information (Comon 1994; Hyvärinen 1998, 1999a), the maximum likelihood estimation (Pham and Garat 1997; Bell 1995; Hyvärinen 1999b), and the local parametric estimation with time varying loading (Chen et al. 2014).

In high dimensional space, however, ICA leads to redundant dependence by assuming each factor is associated with all the measured variables. The overparametrization is solvable by either reducing the number of factors or simplifying the structure of the loading matrix. Wu et al. (2006) proposed an ordering approach based on the mean-square-error criterion to identify the number of ICs. This dimension reduction eventually accompanies with loss of information. On the other hand, the dependence between the measured sovereign default probabilities and the factors can be sparse. A possibly more realistic situation is that each measured variable is only driven by a few factors, while others depend on a possibly different set of fac-

tors. It suggests necessity to reduce dimensionality in parameter space, with a sparse loading matrix.

Sparse estimation has been widely used especially in the regularized regression analysis. Under the sparsity assumption, unnecessary dependence is penalized and insignificant coefficients are pushed to zeros, see e.g. Lasso (Tibshirani 1996), Ridge (Frank and Friedman 1993) and the smoothly clipped absolute deviation (SCAD) penalty (Fan and Li 2001) and so on. The adoption of sparsity in independent component analysis is still new. Hyvärinen and Raju (2002) proposed sparse Bayesian ICA, where the loading matrix is assumed to be random and a conjugate sparse prior is imposed to the loading matrix. Zhang et al. (2009) incorporated adaptive Lasso in the maximum likelihood estimation method to obtain sparse loading matrix, where the statistical independent factors are assumed to follow a simple distribution family with one parameter. Theoretical properties of the estimators are unknown in the above works.

We are motivated to propose a penalized independent component analysis method, named PIF, to extract statistical independent factors via a sparse linear transformation. The sparse loading matrix is estimated under normal inverse Gaussian distributional assumption with SCAD penalty. Our main theoretical result claims that the sparse loading matrix estimator is consistent. The proposed PIF method displays appealing performance in simulation study. We implement the PIF to the daily probability of default data of Corporate Vulnerability Index from 1999 to 2013. The proposed method shows superior interpretation of the dynamic structure of 14 economies' global default probability from the pre-Dot Com bubble period to the post-Sub Prime crisis period.

The remainder of the paper is structured as follows. Section 10.2 details the sovereign default probability data. Section 10.3 presents the penalized independent factor method, the estimation procedure and statistical prosperity of the estimator. Its finite sample performance is investigated along with simulation study in Sect. 10.4. Section 10.5 implements the PIF method to the sovereign default probabilities. Section 10.6 concludes.

10.2 Data

We consider the sovereign default probabilities of 14 economies from 1st April 1999 to 31st December 2013. The data are the equally-weighted Corporate Vulnerability Index (CVI), proxies of sovereign default probability, maintained in the Credit Research Initiative, Risk Management Institute at National University of Singapore. The CVI of each economy is constructed by averaging of all the listed firms' probability of default (PD) in the corresponding exchange. It is worth mentioning that the number of firms considered over the time horizon is not fixed, given the happening of default events and IPOs. For example, on 1st Apr 1999, there were 717 firms listed in the stock exchange of China, and on 31st Dec 2013, the number of listed firms went up to 3017. The PDs were computed using the forward intensity approach in Duan

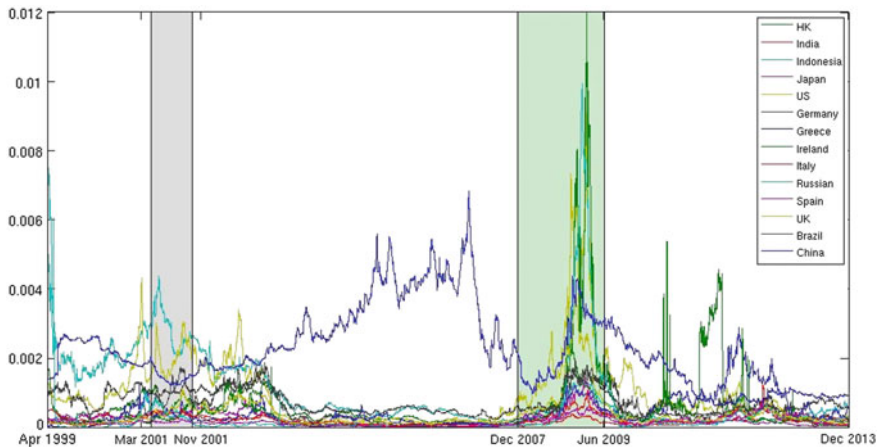


Fig. 10.1 Time series plot of the 14 economies CVI data. *Gray shadow* is the Dot Com bubble period and *light green shadow* is the Sub Prime crisis

et al. (2012) with input variables of common economic factors including e.g. stock index returns and 3-month interest rates, and firm specific factors of e.g. distance to default, ratio of cash (equivalent) to total assets, return on assets, market to book ratio and 1-year idiosyncratic volatility. The 14 economies include 9 advanced economies of Hong Kong, Japan, US, Germany, Greece, Ireland, Italy, Spain, and UK, and 5 emerging ones of China, India, Indonesia, Russian and Brazil.

Figure 10.1 displays the movements of the 14 CVIs from 1999 to 2013. To understand the dynamic structure of CVIs over time, we divide the time horizon of the 15 years into five sub-periods according to the business cycles announced by the National Bureau of Economic Research, including two recessions occurred from 1st March 2001 to 30th November 2001 (Dot Com bubble) and from 1st December 2007 to 30th June 2009 (US Sub Prime crisis). During the two recessions, the level of CVI increases on average 26 and 53% respectively. The relatively high level of the sovereign default probabilities continues after the recessions for a while and then drops to low value. China, however, behaves distinctively from the rest. The CVI of China is much larger than the others during 2002–2007, i.e. the post-Dot Com bubble period. For example, China’s CVI is 3 times of the second highest value of Indonesia. Table 10.1 reports the CVI summary statistics of each economy over the 15 years. China and US have the highest level (mean) of CVI. The level of the US’ CVI is high mainly during the two recessions, the Doc Com bubble and Sub Prime crisis. China, on the other hand, though immune to the Dot Com bubble recession, due to its constantly achieved 2-digits growth during 2003 to 2007, accompanies with high level for the “higher return higher risk” philosophy. In terms of variation, China reaches to the highest CVI variation, with a standard deviation of at least 12% larger than the rest. Moreover, all CVIs are positively skewed with extreme values and the JB statistics are all significant, indicating the deviation from Gaussianity.

Table 10.1 Summary statistics of the CVI data over the time horizon, Apr 1999–Dec 2013

	Mean(10^{-3})	SD(10^{-3})	Skewness	Kurtosis	JB-stats(10^4)
China	2.19	1.21	0.71	3.17	0.06*
Hong Kong	0.46	0.38	2.17	9.53	1.38*
India	0.21	0.11	0.70	2.70	0.05*
Indonesia	0.92	0.95	2.06	9.68	1.39*
Japan	0.26	0.20	1.98	8.54	1.04*
US	1.00	1.08	2.86	14.43	3.67*
Germany	0.53	0.44	1.13	3.23	0.12*
Greece	0.46	0.45	2.02	7.73	0.87*
Ireland	0.58	1.17	4.68	29.57	17.82*
Italy	0.22	0.16	1.83	8.00	0.86*
Russian	0.40	1.07	5.19	32.99	22.61*
Spain	0.18	0.13	1.13	3.72	0.13*
UK	0.41	0.48	3.76	19.83	7.63*
Brazil	0.72	0.33	0.79	2.46	0.06*

The conventional PCA is not able to deliver independent factors. Table 10.2 reports the correlation matrix of the CVI data during the 15 years, which are mostly positive except China. While China has either negative or weak correlations with the other economies, the US remains high positive correlations to most of the advanced economies such as Japan and UK, consistent to its influential role in the global financial markets (Tables 10.3, 10.4, 10.5 and 10.6).

More detailed summary statistics on CVIs over the 5 time periods can be found in Tables 10.7, 10.8, 10.9, 10.10, 10.11, 10.12, 10.13, 10.14, 10.15 and 10.16 in the Appendix.

10.3 Penalized Independent Factor

Consider p -dimension random vector $\mathbf{X} = (X_1, \dots, X_p) \in \mathbb{R}^p$. The penalized independent factor analysis is to factorize the variables into a linear combination of latent independent random factors $\mathbf{Z} = (Z_1, \dots, Z_p) \in \mathbb{R}^p$:

$$\mathbf{Z} = B\mathbf{X} \tag{10.1}$$

where B refers to a sparse and invertible loading matrix. Given the observed realizations $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})$ with $i = 1, \dots, n$, the task here is to estimate the sparse loading B as well as to obtain the independent factor \mathbf{Z}_i with $i = 1, \dots, n$, without any prior knowledge of the sparsity structure of B .

Table 10.2 Correlation matrix of the 14 economies from 1st April 1999 to 31st December 2013

	China	HK	India	Indo	Japan	US	DE	Greece	Ireland	Italy	Russian	Spain	UK	Brazil
China	1													
HK	-0.04	1												
India	0.60	1	1											
Indo	0.55	0.77	1	1										
Japan	-0.09	0.51	0.50	1	1									
US	-0.05	0.74	0.42	0.88	1	1								
DE	-0.19	0.69	0.40	0.74	0.72	1	1							
Greece	0.24	0.24	0.18	0.18	0.34	0.34	1	1						
Ireland	0.01	0.40	0.14	0.02	0.66	0.66	0.68	0.49	0.34	1				
Italy	-0.27	0.67	0.47	0.22	0.66	0.72	0.73	0.56	0.63	1	0.57			
Russian	0.08	0.54	0.32	0.23	0.66	0.70	0.43	0.26	0.71	0.78	1	0.47		
Spain	-0.43	0.41	0.32	0.00	0.42	0.49	0.68	0.46	0.75	0.76	1	1	0.58	
UK	0.03	0.76	0.32	0.16	0.81	0.87	0.34	0.73	0.75	0.76	1	1	1	0.61
Brazil	-0.08	0.79	0.70	0.66	0.77	0.70	0.68	0.04	0.35	0.54	0.44	0.31	0.61	1

Table 10.3 Summary statistics of simulation results. The results of the PIF is marked in bold and reported in the first row of each scenario. The results of the NIG-ICA and ICA are reported in the second and third row of each scenario. Detection of zeros is the percentage of zeros correctly estimated, and miss-detection is the percentage of non-zero entries estimated as zero. λ is the optimal penalty parameter obtained via minimizing BIC. Each measurement is given in the form of sample average (sample standard deviation)

	Method	ED	MN	RMSE	Det and Miss-det of zeros %	λ
Non-sparsity	PIF	4.80(2.65)	3.97(2.50)	0.09(0.03)	-/0	0
	NIG-ICA	4.80(2.65)	3.97(2.50)	0.09(0.03)	-/0	0
	ICA	14.51(8.25)	10.26(5.71)	0.19(0.14)	-/0	-
Sparsity	PIF	6.67(3.98)	5.54(3.65)	0.09(0.03)	100/0	0.04
	NIG-ICA	6.96(3.99)	5.75(3.65)	0.10(0.03)	0/0	0
	ICA	27.19(17.47)	20.40(13.61)	0.20(0.14)	0/0	-
Highly-sparsity	PIF	3.17(1.83)	2.69(1.62)	0.08(0.03)	99/0	0.07
	NIG-ICA	3.90(1.86)	3.20(1.64)	0.11(0.04)	0/0	0
	ICA	13.81(10.72)	10.19(7.28)	0.28(0.86)	0/0	-

Table 10.4 True Loading matrix. Zero entries are left blank

17	14	10		-13			21			-15			
	-10	15		13	19		-10	12	17			-14	
	-34	-11	-23	27			7	27	43	37	-24	27	18
		-44		-14	13	-97	50	-12	-127		172	-65	
			11										
	7			-79	-29	146	20	-40	-206		-74	223	9
	33	28	-6	-51	-19	71		182	-32		-99	-89	10
	18	-109		-51		-28		-33	54		70	-62	28
				-77	-20	88	-66	30	-95		-144	308	9
	40	43	-13	-108		-49	-14	-79	10		-51	177	-21
	-32	-92	7	57	6	-21	16	14	40	8	-114	29	-9
			64		90	6	-10	-11	-22		-153	-29	28
	-86	-25	17	-116		-24	26	-9	123		-47	108	10
		7	-73		34	-26	-47	-17	25	-32		163	372

Table 10.5 Simulation results for large dimension loading matrix. Each measurement is given in the form of mean(std). The penalty parameter is $\lambda = 0.08$ by minimizing BIC. #0s is the percentage of zero elements estimated correctly by the method. Mis-detection is the number of elements that are wrongly pushed to zero

	ED	MN	RMSE	Detection of zeros %	λ
PIF	88.60(26.11)	60.00(24.63)	0.20(0.10)	99.85	0.08
NIG-ICA	90.23(27.74)	61.50(25.68)	0.22(0.10)	0.00	0
ICA	419.24(56.11)	204.00(36.54)	1.29(0.05)	0.00	—

Table 10.6 Number of factors participated by each economy. Sparsity is reflected by of the percentage of zeros in the loading matrix

Country	China	HK	India	Indo	Japan	US	DE	Greece
1999:4–2001:2	1	6	9	4	12	6	8	6
2001:3–2001:11	3	9	9	5	10	6	9	12
2001:12–2007:11	3	10	12	8	13	10	11	9
2007:12–2009:6	7	11	12	9	12	7	9	11
2009:7–2013:12	6	11	11	10	11	9	6	8
Country	Ireland	Italy	Russian	Spain	UK	Brazil	Total	Sparsity%
1999:4–2001:2	9	10	1	11	12	4	99	49
2001:3–2001:11	9	11	9	12	11	7	122	38
2001:12–2007:11	13	13	12	12	12	11	149	24
2007:12–2009:6	7	13	6	9	11	12	136	31
2009:7–2013:12	2	12	9	11	11	9	126	36

Table 10.7 Summary statistics of the CVI data, Apr 1999–Feb 2001

	Mean(10^{-3})	SD(10^{-3})	Skewness	Kurtosis	JB-stats(10^3)
China	2.08	0.59	-1.94	7.55	1.04*
Hong Kong	0.55	0.33	1.33	4.91	0.31*
India	0.33	0.08	0.61	3.05	0.04*
Indonesia	2.27	1.04	3.18	14.04	4.73*
Japan	0.32	0.05	0.24	2.35	0.02*
US	1.13	0.63	2.62	10.94	2.64*
Germany	0.23	0.13	2.11	6.99	0.98*
Greece	0.25	0.27	1.08	3.14	0.14*
Ireland	0.20	0.06	1.57	4.71	0.37*
Italy	0.15	0.07	1.47	5.07	0.38*
Russian	0.39	0.97	3.51	14.63	5.38*
Spain	0.15	0.04	0.47	1.94	0.06*
UK	0.17	0.07	1.08	3.77	0.15*
Brazil	0.95	0.26	-0.69	2.55	0.06*

Table 10.8 Summary statistics of the CVI data during DOT COM bubble, Mar 2001–Nov 2001

	Mean(10^{-3})	SD(10^{-3})	Skewness	Kurtosis	JB-stats(10^3)
China	1.40	0.27	-2.56	16.33	2.34*
Hong Kong	0.78	0.22	-0.03	2.11	0.01*
India	0.44	0.04	-0.90	3.88	0.05*
Indonesia	2.85	0.60	0.54	2.41	0.02*
Japan	0.40	0.06	0.07	1.64	0.02*
US	2.08	0.45	0.40	2.10	0.02*
Germany	1.07	0.34	0.39	1.96	0.02*
Greece	0.42	0.11	0.62	3.45	0.02*
Ireland	0.36	0.16	1.20	3.31	0.07*
Italy	0.40	0.08	-0.53	2.09	0.02*
Russian	0.44	0.18	0.49	2.25	0.02*
Spain	0.18	0.03	1.09	3.83	0.06*
UK	0.58	0.20	0.78	2.77	0.03*
Brazil	1.05	0.09	0.98	4.00	0.06*

The loading matrix and independent factors are only identifiable up to scale. For any constant $c \neq 0$, one obtains another set of loading matrix cB and independent factors denoted cZ satisfying (10.1). To avoid the identification problem, we assume that the independent factors have unit variance. Moreover, we set the number of independent factors to p , as the primary goal of our study is to convert the multivariate problem into a number of univariate ones with sparsity such that it eases the

Table 10.9 Summary statistics of the CVI data, Dec 2001–Nov 2007

	Mean(10^{-3})	SD(10^{-3})	Skewness	Kurtosis	JB-stats(10^3)
China	2.96	1.26	0.12	2.80	0.01*
Hong Kong	0.43	0.32	1.20	3.43	0.54*
India	0.17	0.09	1.21	3.60	0.57*
Indonesia	0.80	0.62	1.24	3.33	0.57*
Japan	0.21	0.18	1.21	3.12	0.54*
US	0.62	0.69	1.59	4.44	1.11*
Germany	0.46	0.46	1.27	3.23	0.59*
Greece	0.22	0.14	1.88	8.01	3.58*
Ireland	0.23	0.28	1.98	6.71	2.70*
Italy	0.15	0.09	0.80	2.35	0.27*
Russian	0.05	0.04	8.03	126.92	1425.52*
Spain	0.08	0.05	0.93	2.40	0.35*
UK	0.30	0.24	1.38	3.45	0.71*
Brazil	0.68	0.33	0.90	2.34	0.34*

Table 10.10 Summary statistics of the CVI data during Sub Prime crisis, Dec 2007–Jun 2009

	Mean(10^{-3})	SD(10^{-3})	Skewness	Kurtosis	JB-stats(10^3)
China	2.29	1.02	0.23	1.98	0.03*
Hong Kong	0.87	0.66	1.11	3.38	0.12*
India	0.20	0.10	0.25	1.73	0.04*
Indonesia	0.74	0.39	0.33	1.55	0.06*
Japan	0.52	0.35	0.64	2.19	0.06*
US	2.64	1.97	1.15	3.25	0.13*
Germany	0.92	0.51	0.22	1.38	0.07*
Greece	0.59	0.33	0.34	1.63	0.06*
Ireland	2.12	2.61	1.60	4.71	0.32*
Italy	0.42	0.20	0.54	2.55	0.03*
Russian	1.97	2.54	1.33	3.70	0.18*
Spain	0.36	0.10	0.04	2.52	0.01*
UK	1.30	0.94	0.94	2.55	0.09*
Brazil	0.99	0.42	0.23	1.33	0.07*

understanding of the dependence with reduced parameter space and simultaneously an improved estimation accuracy.

Denote the probability density function of each independent factor to be $f_j(z)$ for $j = 1, \dots, p$. The log-likelihood is defined as:

Table 10.11 Summary statistics of the CVI data, Jul 2009–Dec 2013

	Mean(10^{-3})	SD(10^{-3})	Skewness	Kurtosis	JB-stats(10^3)
China	1.30	0.68	1.02	3.18	0.29*
Hong Kong	0.27	0.13	0.87	2.54	0.22*
India	0.18	0.08	-0.09	1.70	0.12*
Indonesia	0.25	0.10	2.96	12.52	8.62*
Japan	0.18	0.08	0.64	3.55	0.13*
US	0.70	0.43	1.38	4.93	0.78*
Germany	0.53	0.27	1.18	4.49	0.53*
Greece	0.83	0.58	1.12	3.70	0.38*
Ireland	0.70	0.96	2.43	8.19	3.46*
Italy	0.24	0.15	2.54	13.16	8.85*
Russian	0.31	0.25	1.47	6.31	1.34*
Spain	0.25	0.14	0.73	2.65	0.15*
UK	0.31	0.13	1.11	3.69	0.37*
Brazil	0.54	0.16	1.00	3.73	0.31*

$$l(B) = \sum_{i=1}^n \sum_{j=1}^p \log f_j(b_j^\top \mathbf{X}_i) + n \log |det(B)| \tag{10.2}$$

where b_j^\top denotes the j -th row of B . To achieve the sparsity of the loading matrix B , a penalty function, denoted as ρ_λ is added to the log-likelihood, where λ is a tuning parameter. The penalized log-likelihood is defined as:

$$\mathbf{P}(B) = \sum_{i=1}^n \sum_{j=1}^p \log f_j(b_j^\top \mathbf{X}_i) + n \log |det(B)| - n \sum_{j=1}^p \sum_{k=1}^p \rho_\lambda(|b_{jk}|) \tag{10.3}$$

where b_{jk} denotes the (j, k) -th element of the loading matrix B . Take the gradient of the penalized likelihood function with respect to the loading matrix, we obtain:

$$\frac{\partial \mathbf{P}}{\partial B} = \sum_{i=1}^n \begin{bmatrix} \frac{f'_1(b_1^\top \mathbf{X}_i)}{f_1(b_1^\top \mathbf{X}_i)} \\ \frac{f'_2(b_2^\top \mathbf{X}_i)}{f_2(b_2^\top \mathbf{X}_i)} \\ \dots \\ \frac{f'_p(b_p^\top \mathbf{X}_i)}{f_p(b_p^\top \mathbf{X}_i)} \end{bmatrix} \mathbf{X}_i^\top + n[B^\top]^{-1} - n\Omega$$

where $[B^\top]^{-1}$ is the inverse of transpose matrix of B , $\Omega_{jk} = sgn(b_{jk})\rho'_\lambda(|b_{jk}|)$ is the first derivative of the penalty function with respect to each element of the loading matrix, and $f'_i(s)/f_i(s)$ is the first derivative of log-density function of each

Table 10.12 Correlation matrix of the 14 economies from 1st April 1999 to 28th Feb 2001

	China	HK	India	Indo	Japan	US	DE	Greece	Ireland	Italy	Russian	Spain	UK	Brazil
China	1													
HK	-0.49	1												
India	0.49	1	1											
Indo	-0.40	0.81	0.67	1										
Japan	0.06	0.11	-0.03	0.07	1									
US	-0.28	0.61	0.24	0.13	0.81	1								
DE	-0.22	0.56	0.02	0.19	0.36	0.81	1							
Greece	-0.42	0.59	0.19	0.14	0.38	0.68	0.69	1						
Ireland	-0.20	0.28	0.08	0.04	0.01	0.61	0.32	0.48	1					
Italy	-0.37	0.69	0.25	0.32	0.35	0.61	0.79	0.77	0.29	1				
Russian	-0.37	0.68	0.63	0.91	-0.22	0.10	0.28	0.07	0.43	0.04	1			
Spain	-0.22	0.43	0.50	0.56	-0.01	0.28	0.43	0.07	0.04	0.05	0.52	1		
UK	-0.43	0.72	0.30	0.34	0.32	0.87	0.78	0.86	0.58	0.80	0.17	0.29	1	
Brazil	-0.23	0.47	0.35	0.32	0.11	0.50	0.50	0.40	0.26	0.42	0.16	0.37	0.53	1

independent factor. The sparse loading matrix is estimated using the gradient method. Given the loading matrix estimator, the independent factors are recovered in (10.1).

10.3.1 Independent Component's Density: NIG

The density of IC is unknown. Hyvärinen (1999b) developed the maximum likelihood estimation approach of independent factor extraction under a simple but unrealistic distribution with one distributional parameter, and proved consistency of the estimator, see also Pham and Garat (1997), Bell (1995). The log-likelihood function is defined under a simple but unrealistic distribution with one distributional parameter. Financial risk factors are however neither Gaussian distributed nor the special cases of the exponential power family. Instead, the factors are often asymmetric and with extreme values. This motivates the adoption of the normal inverse Gaussian (NIG) distribution for its desirable probabilistic features. With 4 distributional parameters, the NIG distribution is able to mark data characteristics from the central locations to the tails behaviours.

In our study, each factor is assumed to be normal inverse Gaussian (NIG) distributed with individual distributional parameters. The density is of the form:

$$f_{\text{NIG}}(z_j) = \frac{\phi_j \delta_j}{\pi} \frac{K_1 \left\{ \phi_j \sqrt{\delta_j^2 + (z_j - \mu_j)^2} \right\}}{\sqrt{\delta_j^2 + (z_j - \mu_j)^2}} \exp\{\delta_j \sqrt{\phi_j^2 - \beta_j^2} + \beta_j (z_j - \mu_j)\},$$

where μ_j , δ_j , β_j and ϕ_j are NIG parameters for $j = 1, \dots, p$. $K_1(\cdot)$ is the modified Bessel function of the third type. The distributional parameters fulfill the conditions $\mu_j \in \mathbb{R}$, $\delta_j > 0$, and $|\beta_j| \leq \phi_j$. The limiting distributions of NIG have been well developed in bn (1997); Blæsild (1999) including the Normal distribution, the Cauchy distribution and the Student-t distribution.

- For $\beta = 0$, $\phi \rightarrow \infty$ and $\delta/\phi = \sigma^2$, $\text{NIG}(\phi, \beta, \mu, \delta) \rightarrow N(\mu, \sigma^2)$
- For $\phi, \beta \rightarrow \infty$, $\mu = 0$ and $\delta = 1$, $\text{NIG}(\phi, \beta, \mu, \delta) \rightarrow \text{Cauchy}$
- For $\phi, \beta \rightarrow 0$, $\mu = 0$ and $\delta = 1$, $\text{NIG}(\phi, \beta, \mu, \delta) \rightarrow \text{Student} - t_1$

See bn (1997) for more details. Moreover, all independent factors are assumed to have unit variance to avoid identification ambiguity.

10.3.2 Penalty Function: SCAD

Question remains on the selection of penalty function in the estimation. Various penalty function has been proposed in literature, including the first order norm penalty of Lasso (Tibshirani 1996), the second order norm penalty of Ridge

(Frank and Friedman 1993) and the smoothly clipped absolute deviation (SCAD) penalty (Fan and Li 2001) and so on. Among them, the SCAD penalty is theoretically desirable with oracle property and has been widely used in quantile regression, logistic regression, high dimensional data analysis, large scale genomic data analysis and many others, see Gou et al. (2014), Xie and Huang (2009). In our study, we use the SCAD penalty, which is defined in the form of its first derivative:

$$\rho'_\lambda(\theta) = \lambda\{I(\theta \leq \lambda) + \frac{(a\lambda - \theta)_+}{(a - 1)\lambda}I(\theta > \lambda)\} \tag{10.4}$$

where $\theta > 0$ and $a = 3.7$ suggested in Fan and Li (2001).

10.3.3 Estimation

Substitute the NIG density and the SCAD penalty function into (10.3):

$$\mathbf{P}(B) = \sum_{i=1}^n \sum_{j=1}^p \log f_j(b_j^\top \mathbf{X}_i) + n \log |\det(B)| - n \sum_{j=1}^p \sum_{k=1}^p \rho_\lambda(|b_{jk}|) \tag{10.5}$$

$$= \sum_{i=1}^n \sum_{j=1}^p \left\{ \log \frac{\phi_j \delta_j}{\pi} \frac{K_1 \left(\phi_j \sqrt{\delta_j^2 + (b_j^\top \mathbf{X}_i - \mu_j)^2} \right)}{\sqrt{\delta_j^2 + (b_j^\top \mathbf{X}_i - \mu_j)^2}} + \delta_j \sqrt{\delta_j^2 - \beta_j^2} + \beta_j (b_j^\top \mathbf{X}_i - \mu_j) \right\} \\ + n \log |\det(B)| - n \sum_{j=1}^p \sum_{k=1}^p \rho_\lambda(|b_{jk}|) \tag{10.6}$$

and the gradient of the log-likelihood function is:

$$\frac{\partial l}{\partial B} = \sum_{i=1}^n \begin{bmatrix} \frac{f'_1(b_1^\top \mathbf{X}_i)}{f_1(b_1^\top \mathbf{X}_i)} \\ \frac{f'_2(b_2^\top \mathbf{X}_i)}{f_2(b_2^\top \mathbf{X}_i)} \\ \dots \\ \frac{f'_p(b_p^\top \mathbf{X}_i)}{f_p(b_p^\top \mathbf{X}_i)} \end{bmatrix} \mathbf{X}_i^\top + n[B^\top]^{-1} - \Omega$$

where $\Omega_{jk} = \text{sgn}(b_{jk})\rho'_\lambda(|b_{jk}|)$ and $\frac{f'_j(s)}{f_j(s)} = \beta_j + \phi_j \frac{K'_1(\phi_j \sqrt{\delta_j^2 + (s - \mu_j)^2})}{K_1(\phi_j \sqrt{\delta_j^2 + (s - \mu_j)^2})} \frac{s - \mu_j}{\sqrt{\delta_j^2 + (s - \mu_j)^2}} - \frac{s - \mu_j}{\delta_j^2 + (s - \mu_j)^2}$.

The optimization problem is solved in two steps, where maximum is achieved by changing the loading matrix B and the NIG parameters iteratively until the algorithm converges. The algorithm starts with an initial estimator of B_0 , e.g. the estimation obtained by the conventional ICA:

1. Given the previous estimator of B , optimize the penalized log-likelihood function to obtain the NIG distributional parameters estimator. The EM algorithm is adopted for the estimation of NIG parameters, see Karlis (2002).

2. Based on the estimated NIG estimator, update the estimator of B by maximizing the penalized log-likelihood function.
3. Scale the estimator of B and the NIG parameters to have unit variance of each independent factor.
4. Repeat, until converge.

The penalized maximum likelihood estimation involves the choice of the tuning parameter λ . While too large tuning parameter leads to over sparse loading matrix, too small tuning parameter has over fitting effect to identify the true model. Cross validation (Kohavi 1995) and generalized cross validation (Li 1987) can be used. However the approaches are computational intensive. Even worse, there is a positive probability of model over-fitting by generalized cross validation (Wang et al. 2007). Alternatively, several information criteria have been proposed and widely used in time series analysis. In our study, we consider using the Schwarz–Bayesian information criterion (BIC) (Schwarz 1978) for its computation tractability and its consistency in model selection. The BIC is defined as:

$$BIC = -l(\hat{B}) + \log n \times \#\{\hat{B}_{ij} \neq 0\}$$

where \hat{B} is the estimator of B . The penalty parameter with the lowest BIC is chosen to be optimal.

10.3.4 Property of Estimator

We prove the consistency of the PIF estimator under two conditions:

- C1. The observations (X_{i1}, \dots, X_{ip}) are IID with density $(g_1(X, B), \dots, g_p(X, B))$ with respect to some measure μ . The density has a common support and is identifiable. Furthermore, the first logarithmic derivatives of g_i satisfying the equation

$$E \frac{\partial \log g_a(X, B)}{\partial B_{jk}} = 0 \tag{10.7}$$

for all a, j and k .

- C2. $E[-\Omega_a]$ is positive definite at point B with Ω_a defined as:

$$\Omega_a = \begin{pmatrix} \frac{\partial^2 g_a(B)}{\partial b_{11} \partial b_{11}} & \frac{\partial^2 g_a(B)}{\partial b_{11} \partial b_{12}} & \cdots & \frac{\partial^2 g_a(B)}{\partial b_{11} \partial b_{1p}} & \frac{\partial^2 g_a(B)}{\partial b_{11} \partial b_{21}} & \cdots & \frac{\partial^2 g_a(B)}{\partial b_{11} \partial b_{pp}} \\ \frac{\partial^2 g_a(B)}{\partial b_{12} \partial b_{11}} & \frac{\partial^2 g_a(B)}{\partial b_{12} \partial b_{12}} & \cdots & \frac{\partial^2 g_a(B)}{\partial b_{12} \partial b_{1p}} & \frac{\partial^2 g_a(B)}{\partial b_{12} \partial b_{21}} & \cdots & \frac{\partial^2 g_a(B)}{\partial b_{12} \partial b_{pp}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 g_a(B)}{\partial b_{pp} \partial b_{11}} & \frac{\partial^2 g_a(B)}{\partial b_{pp} \partial b_{12}} & \cdots & \frac{\partial^2 g_a(B)}{\partial b_{pp} \partial b_{1p}} & \frac{\partial^2 g_a(B)}{\partial b_{pp} \partial b_{21}} & \cdots & \frac{\partial^2 g_a(B)}{\partial b_{pp} \partial b_{pp}} \end{pmatrix}$$

Theorem 10.1 *Let $(X_{11}, X_{12}, \dots, X_{1p}), \dots, (X_{n1}, X_{n2}, \dots, X_{np})$ be IID measured vector, each with a density (g_1, g_2, \dots, g_p) that satisfies conditions (C1) and (C2).*

If $\max\{p''_{\lambda_n}(|B_{jk}|) : B_{jk} \neq 0\} \rightarrow 0$, then there exists a local maximizer \hat{B} of $\mathbf{P}(B)$ such that $\|\hat{B} - B\| = \mathcal{O}_p(n^{-1/2} + a_n)$, where $a_n = \max\{p'_{\lambda_n}(|B_{jk}|) : B_{jk} \neq 0\}$

Note that, though the density of the observed variables g_a is unknown, Theorem 10.1 holds as long as the two conditions hold. Detailed proof can be found in Appendix.

10.4 Simulation

Before the implementation with real sovereign default probability data, we investigate the finite sample performance of the PIF method first by performing a number of simulation studies under the known data generating processes. Our interest is on the estimation accuracy of the proposed method and its robustness under various scenarios compared to the conventional ICA approach.

We design our simulation studies so that they properly reflect the real study at hand. All the parameters are obtained from analyzing the Corporate Vulnerability Index (CVI) data from April 1999 to February 2001, before the Dot Com bubble. In the first experiment, small dimensional data are generated based on the CVIs of India, Indonesia and Japan, 3 Asia countries of both emerging and advanced economies. We consider 3 scenarios with non-sparsity, medium sparsity and high sparsity in the loading matrix. In the second experiment, large dimensional data are produced, where the parameters are learned from the CVI data of the 14 economies from April 1999 to February 2001.

In the data generation process, we follow the model setting in (10.1) and generate dependent data with the loading matrix:

$$\mathbf{X}_i = B^{-1}\mathbf{Z}_i, \quad i = 1, \dots, n.$$

The generated data are considered as the measured variables. Each experiment is repeated 100 times with $n = 200$ observations. Both the PIF and the conventional ICA methods are implemented. In addition to the two approaches, we also implement ICA with the NIG distributed source assumption, named as NIG-ICA in the following.

We evaluate the estimation accuracy of the PIF method, with focus on the factor loadings B and the identified factors \mathbf{Z}_i . We compare the estimation accuracy of the PIF method based on 3 measurements. For the loading matrix, our interests are the overall estimation accuracy and the elementary accuracy. While the Euclidean distance (ED) is used to measure the estimation error of the loading matrix estimator, the maximum norm (MN) reports the largest elementary bias of the matrix estimator. For the identified independent factors, we compute the root mean squared error (RMSE) to show the identification accuracy. The criteria are defined as follows:

$$ED = \sum_{jk} (b_{jk} - \hat{b}_{jk})^2 \quad (10.8)$$

$$MN = \max (|b_{jk} - \hat{b}_{jk}|) \quad (10.9)$$

$$RMSE = \sqrt{\frac{1}{np} \sum_{ij} (Z_{ij} - \hat{Z}_{ij})^2} \quad (10.10)$$

where b_{jk} refers to the (j, k) -th element of the matrix B , and \hat{b}_{jk} represents the corresponding element estimators.

10.4.1 Experiment 1: 3 Dimensional Data

In the low dimensioned experiment, 3 scenarios are analyzed with 3 different loading matrices that are either non-sparse, sparse, or highly sparse:

Non-sparse loading matrix:

$$\begin{bmatrix} 52.7 & -10.7 & 14.4 \\ -32.3 & -17.3 & -5.2 \\ 18.1 & -6.3 & 12.8 \end{bmatrix};$$

Sparse loading matrix:

$$\begin{bmatrix} -3.2 & 31.2 & \mathbf{0} \\ 40.1 & -96.4 & -20.9 \\ -29.4 & 18.7 & \mathbf{0} \end{bmatrix};$$

Highly-sparse loading matrix:

$$\begin{bmatrix} -3.3 & 31.2 & \mathbf{0} \\ \mathbf{0} & 10.1 & \mathbf{0} \\ \mathbf{0} & 44.2 & -25.0 \end{bmatrix}.$$

Table 10.3 reports the simulation results based on the 100 replications. For all the 3 scenarios, the PIF is better than ICA in terms of estimation accuracy for both the loading matrix and the independent factors. In the sparsity scenario, the estimation accuracy of PIF is much better with lower ED of 6.67(SD: 3.98), MN of 5.54(SD: 3.65) and RMSE of 0.09(SD: 0.03) than that of ICA with ED of 27.19(SD: 17.47), MN of 20.40(SD: 13.61) and RMSE of 0.20(SD: 0.14). The improved accuracy is mostly contributed by the adoption of the NIG distributional assumption. In the highly-sparse scenario, the PIF is remarkably better than the conventional ICA. The improvement w.r.t to the NIG-ICA becomes larger.

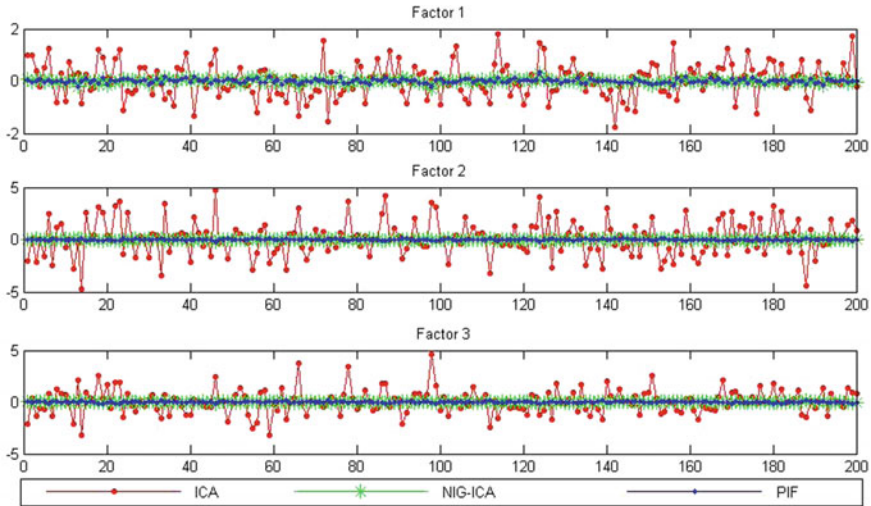


Fig. 10.2 Illustration of residuals of the factors in the sparsity setup. ICA is marked as *circle*, NIG-ICA is labeled with *star* and PIF with *dot*

Moreover, the tuning parameter λ is reasonably selected by using BIC. In the non-sparsity scenario, the optimal λ is 0, indicating non-necessity of penalty as the true loading matrix is not sparse. In the sparsity and high-sparsity scenarios, the optimal λ becomes 0.04 and 0.07 respectively, leading to a high detection rate of zero elements at 100 and 99% respectively. On the contrary, ICA and NIG-ICA are not able to detect any zero elements in the loading matrix. Furthermore, there is no mis-detection by PIF, meaning that no entries in the loading matrix are over pushed to zero.

Figure 10.2 illustrates one representation of the estimation error of the recovered independent factors by the PIF, NIG-ICA and ICA methods respectively in the high-sparsity scenario. While the ICA produces more variations with wider spread, the PIF and NIG-ICA recover the independent factor with smaller errors.

10.4.2 Experiment 2: Large Dimensional Data

In the second experiment with large dimensional data, we generate 14-dimensional dependent data with a sparse loading matrix learning from the CVI data, over a time span of April 1999 to February 2001. The loading matrix is shown in Table 10.4, where 35% of elements are zero.

The generation is repeated 100 times with $n = 200$ sample size. Table 10.5 reports the estimation result. The penalty parameter of PIF is chosen to be $\lambda = 0.08$ by minimizing BIC. The estimation accuracy of PIF is much better with ED of 88.60(SD: 26.11), MN of 60.00(SD: 24.63) and RMSE of 0.20(SD: 0.10) than that of ICA with ED of 419.24(SD: 56.11), MN of 204.00(SD: 36.54) and RMSE of 1.29(SD: 0.05) and slightly better than NIG-ICA with ED of 90.23(SD: 27.74), MN of 61.50(SD: 25.68) and RMSE of 0.22(SD: 0.10). In addition, PIF is able to detect 99.85% of zero entries in the loading matrix and without any miss-detection record of non-zeros.

The simulation study shows that the proposed PIF method has good performance compared to the alternative ICA and NIG-ICA methods with improved estimation accuracy. The good performance mostly attributes to the adoption of the NIG distribution and further by the sparsity of loading matrix. By adding the SCAD penalty function, the proposed PIF is able to identify zero entries in the sparse loading matrix and involves no miss-detection of non-zeros. Moreover, the penalty parameter can be reasonably chosen by using BIC. For example, in the non-sparse scenario, the penalty parameter is selected to be zero. The relative good performance of the PIF is stable with respect to the increase of sparsity and dimensionality.

10.5 Real Data Analysis

In this section, we analyze the sovereign default probabilities of 14 economies from April 1999 to December 2013. The sovereign default probabilities are quantified as daily equally-weighted CVI (Corporate Vulnerability Index) of each economy. The 14 economies are mixture of advanced and emerging economies including China, Hong Kong, India, Indonesia, Japan, US, Germany, Greece, Ireland, Italy, Russian, Spain, UK and Brazil. Data are obtained from the Risk Management Institute at National University of Singapore. We divide the time span into five sub-periods based on the business cycles announced by the National Bureau of Economic Research among which two recessions happened: Dot Com bubble from March 2001 to November 2001 and the US sub prime crisis from December 2007 to June 2009. Our interest is to identify the statistical independent dominant factors and investigate the cross-dependence of the sovereign defaults among the economies.

We implement the proposed PIF method. Table 10.6 summarizes the sparse structure of the loading matrices over the 5 time periods. Each economy column reports the number of non-zero elements in the column of loading matrix, representing the number of factors participated in the economies. The total number of non-zero elements in the loading matrix is summarized in the column Total. Sparsity is reflected by the percentage of zero elements in the loading matrix. It shows that there is a V-shape sparsity in terms of US default probability over time, possibly driven by the cyclical pattern of the global economy. Five advanced economies Japan, Germany, Italy, Spain and UK display relatively stable low-sparse structure across the whole time. China and Hong Kong exhibit co-movement, indicating the connection between the two economies, though Hong Kong given its higher level of globalization appears

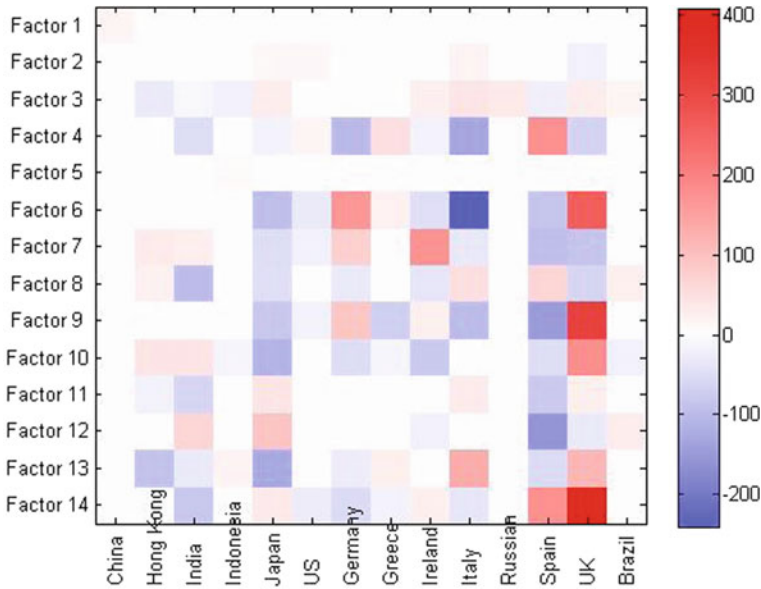


Fig. 10.3 Loading matrix: Apr 1999–Feb 2001

in more factors than China across all periods. The emerging economies of China, India and Indonesia show constant increasing in the number of participated factors along with their increased connection to the global economy especially in the fast growing export business.

Figures 10.3, 10.4, 10.5, 10.6 and 10.7 provides details of the estimated loading matrices over the five time periods. In each plot, we display the loadings of an independent factor with respect to the economies. Zero elements are colored in white. The loading matrix is interpretable. In the pre-Dot Com bubble period, the advanced economies including Japan, Germany, Ireland, Spain and UK participate the most number of factors, while the emerging economies such as China, Indonesia, Russia and Brazil are only related to a few factors. China, for example, only participates in one factor and moreover it is the only element of the factor, implying the closed market of China in the early time. During 1999 to 2001, most defaults in China happened due to the reforming of the state-owned enterprises, which were less likely affected or influenced by the global economy. On the contrary, Japan participates more than 10 factors implying its close connection to the global financial market. In the recent, the sparse inequality between the advanced and emerging economies decreases from period to period, see Fig. 10.8.

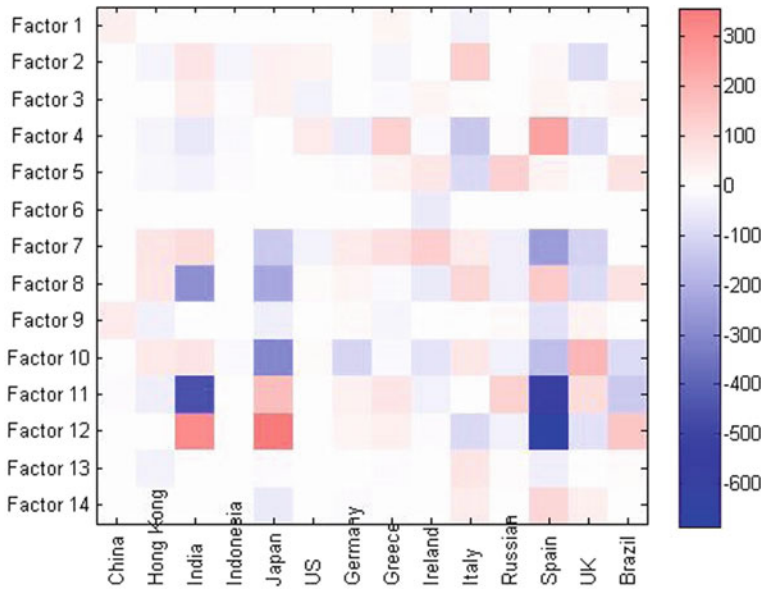


Fig. 10.4 Loading Matrix: Mar 2001–Nov 2001(Dot Com bubble)

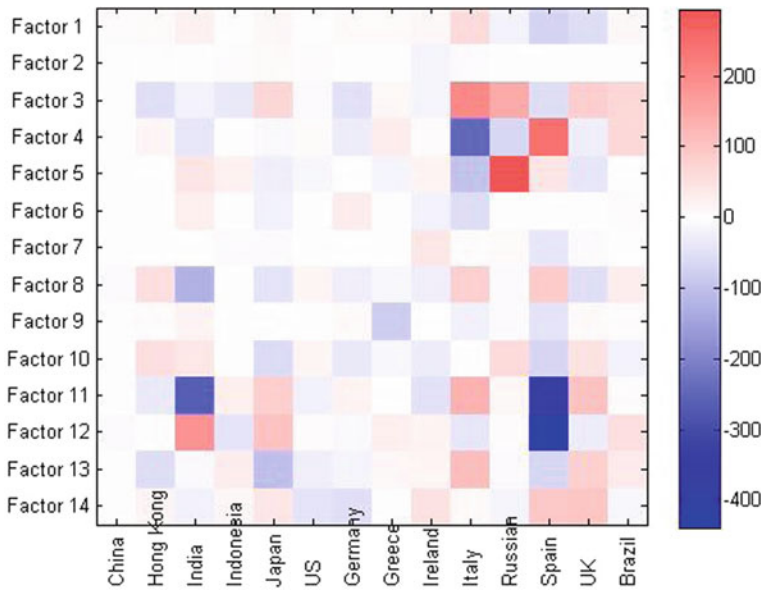


Fig. 10.5 Loading Matrix: Dec 2001–Nov 2007

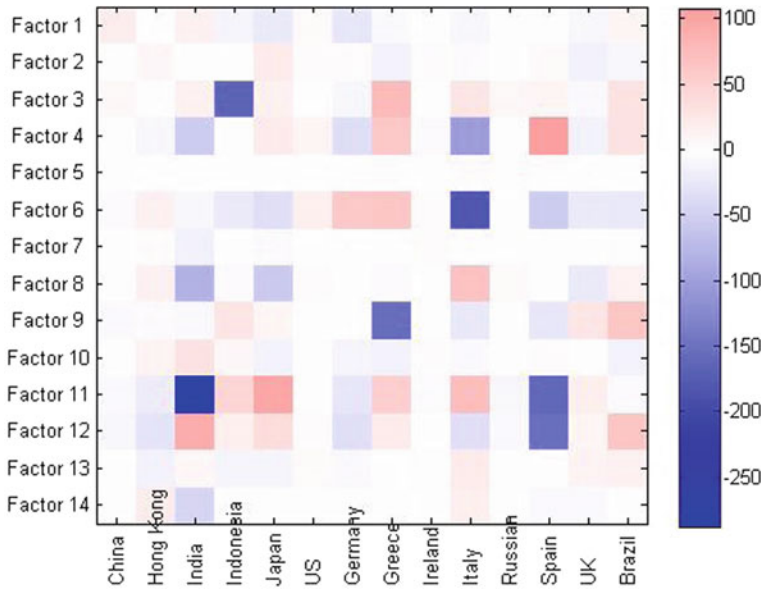


Fig. 10.6 Loading Matrix: Dec 2007–Jun 2009(Sub prime crisis)

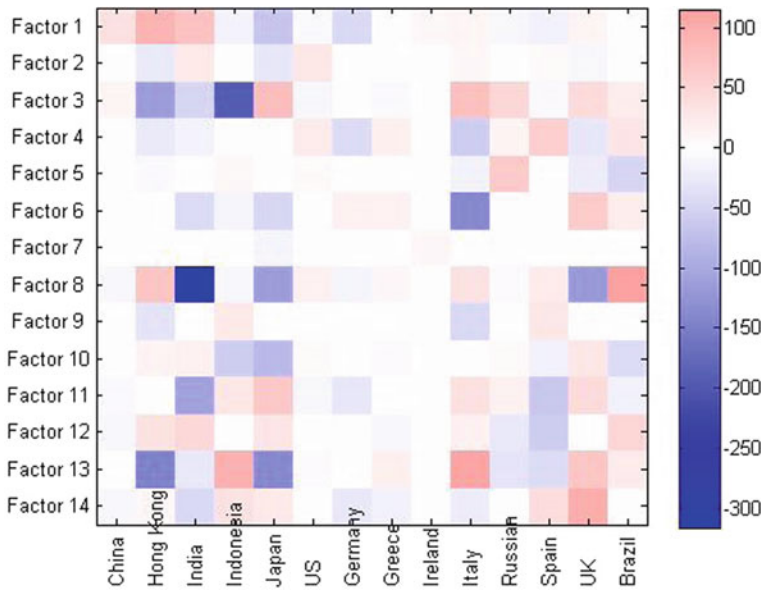


Fig. 10.7 Loading Matrix: Jul 2009–Dec 2013

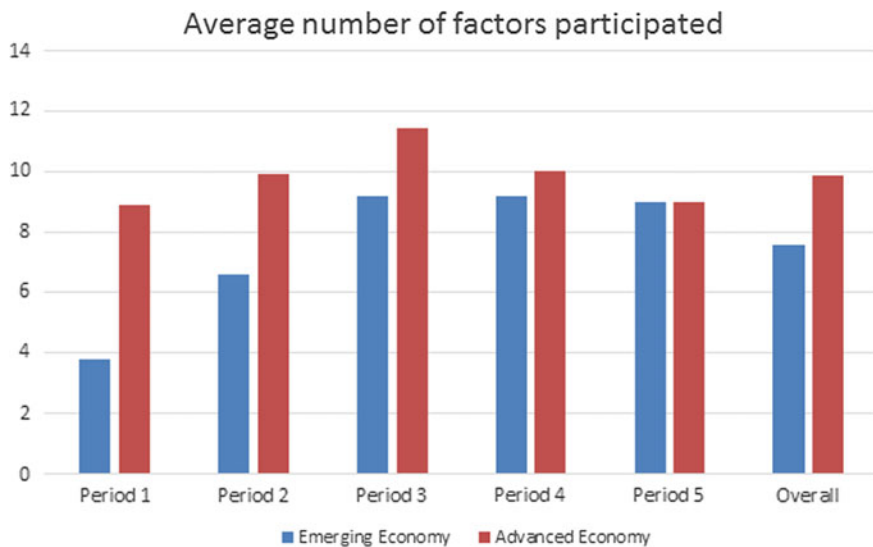


Fig. 10.8 Histogram of average number of factors participated by emerging economies and advanced economies across different period and overall

10.6 Conclusion

We propose the PIF method to transform the observed multivariate correlated variables into independent factors with a sparse loading matrix. We derive the consistency and convergence rate of the sparse loading matrix estimator. Based on the NIG distributional assumption, the estimation is done with a two step ML estimation algorithm by iterating NIG parameter updating and sparse loading matrix estimation. The optimal penalty parameter is chosen via minimizing BIC. We compare the performance of PIF with two alternatives, ICA and NIG-ICA in simulation. The results show the proposed PIF has good performance compared with the conventional ICA and NIG-ICA in both the loading matrix estimation and factor recovery. The estimation accuracy is much improved due to the imposing of NIG distribution. Furthermore, by adopting the SCAD penalty function in PIF, the estimation accuracy is further improved with sparse structure. Moreover, the optimal penalty parameter is reasonably selected by minimizing BIC. The performance of PIF is consistently better with respect to different level of sparse structure and dimensionality of the loading matrix. We implement the PIF to sovereign default probability using CVI data maintained at Credit Research Initiative, Risk Management Institute, National University of Singapore. The estimated loading matrix displays significant sparse structure. For example, China in the pre-Dot Com Bubble period only participates in one factor and is the only element, implying the independence of China’s closed market and the global economy. The proposed model can be easily applied to other high-dimensional data.

Appendix

Proof of Theorem 1

Proof The explicit form of the density function g_j is not required, as long as the two conditions are fulfilled. Under condition C1 and C2, Equation $\|\hat{B} - B\| = O_P(n^{-1/2} + a_n)$ is equivalent to proof that for any given $\epsilon > 0$, there exist a large C s.t.

$$P\{\sup_{\|u\|=C} Q(B + \alpha_n u) < Q(B)\} \geq 1 - \epsilon \quad (10.11)$$

where $Q(B)$ is the penalized likelihood and u is a p-by-p matrix.

Let $D_n(u) = Q(B + \alpha_n u) - Q(B)$

$I_u(B) = -E(\text{tr}(\nabla_B \text{tr}(\nabla_n^1 l(B)^\top u)^\top u)) = -E(\text{tr}(\nabla_B d_u \frac{1}{n} l(B)^\top u)) > 0$ for any $y \in R^{p \times p}$ based on condition (B)

If $D_n(u) < 0$ by choosing a sufficiently large C, then the proof is done.

$$\begin{aligned} D(u) &= l(B + \alpha_n u) - l(B) - n \sum \{\rho_{\lambda_n}(|B_{jk} + \alpha_n u_{jk}|) - \rho_{\lambda_n}(|B_{jk}|)\} \\ &\leq l(B + \alpha_n u) - l(B) - n \sum_{B_{jk} \neq 0} \{\rho_{\lambda_n}(|B_{jk} + \alpha_n u_{jk}|) - \rho_{\lambda_n}(|B_{jk}|)\} \\ &\leq \alpha_n \text{tr}(\nabla l(B)^\top u) + \frac{1}{2} \alpha_n^2 \text{tr}(\nabla_B d_u l(B)^\top u) \{1 + o_P(1)\} \\ &\quad - \sum_{B_{jk} \neq 0} [n \alpha_n \rho'_{\lambda_n}(|B_{jk}|) \text{sgn}(B_{jk}) u_{jk} + n \alpha_n^2 \rho''_{\lambda_n}(|B_{jk}|) u_{jk}^2 \{1 + o(1)\}] \\ &\leq \alpha_n \text{tr}(\nabla l(B)^\top u) - \frac{1}{2} n \alpha_n^2 I_u(B) \{1 + o_P(1)\} \\ &\quad - \sum_{B_{jk} \neq 0} [n \alpha_n \rho'_{\lambda_n}(|B_{jk}|) \text{sgn}(B_{jk}) u_{jk} + n \alpha_n^2 \rho''_{\lambda_n}(|B_{jk}|) u_{jk}^2 \{1 + o(1)\}] \quad (10.12) \end{aligned}$$

The first inequality is because $\rho_{\lambda_n}(0) = 0$ and $\rho_{\lambda_n}(\beta) \geq 0$. The next inequality is Taylor expansion. Then substitute $I_u(B)$ into the equation.

Base on condition (A), $n^{-1/2} \text{tr}(\nabla l(B)^\top u) = O_P(1)$, thus the first term of (8) is of order $O_P(n^{1/2} \alpha_n) = O_P(n \alpha_n^2)$. By choosing a sufficiently large C, the second term dominates the first term in $\|u\| = C$.

The last term in (8) is bounded by

$$\sqrt{sn} \alpha_n a_n \|u\| + n \alpha_n^2 \max\{\rho''_{\lambda_n}(|B_{jk}|) : B_{jk} \neq 0\} \|u\|^2 \quad (10.13)$$

The first part of (9) is dominated by the second term in (8) when choosing a sufficiently large C . The second term in (9) is also dominated by the second term in (8) as $\max\{\rho''_{\lambda_n}(|B_{jk}|) : B_{jk} \neq 0\} \rightarrow 0$

Proof is completed.

References

- Barndorff-Nielsen, O. (1997). Normal inverse gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, 24, 1–13.
- Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7, 1129–1159.
- Blæsild, P. (1999). Generalized hyperbolic and generalized inverse Gaussian distributions, *Working Paper, University of Århus*.
- Cardoso, J. F., & Sôuloumiac, A. (1993). Blind beamforming for non-Gaussian signals. *IEE Proceedings F Radar and Signal Processing*, 140, 362–370.
- Chen, R. B., Chen, Y., & Härdle, W. K. (2014). TVICA Time varying independent component analysis and its application to financial data. *Computational Statistics & Data Analysis*, 74, 95–109.
- Comon, P. (1994). Independent component analysis, A new concept? *Signal Processing*, 36, 287–287. Higher Order Statistics.
- Duan, J. C., Sun, J., & Wang, T. (2012). Multiperiod corporate default prediction—A forward intensity approach. *Journal of Econometrics*, 170, 191–209.
- Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96, 1348–1360.
- Frank, L. E., & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, 35, 109–135.
- Gou, J., Zhao, Y., Wei, Y., Wu, C., Zhang, R., Qiu, Y., et al. (2014). Stability SCAD: A powerful approach to detect interactions in large-scale genomic study. *BMC Bioinformatics*, 133, 140–159.
- Hyvärinen, A. (1998). Analysis and projection pursuit. *Advances in Neural Information Processing Systems*, 10, 273.
- Hyvärinen, A. (1999a). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10, 626–634.
- Hyvärinen, A. (1999b). The fixed-point algorithm and maximum likelihood estimation for independent component analysis. *Neural Processing Letters*, 10, 1–5.
- Hyvärinen, A., & Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Networks*, 9, 1483–1492.
- Hyvärinen, A., & Raju, K. (2002). Imposing sparsity on the mixing matrix in independent component analysis. *Neurocomputing*, 49, 151–162.
- Jones, M. C., & Sibson, R. (1987). What is projection pursuit? *Journal of the Royal Statistical Society. Series A (General)*, 24, 1–10.
- Karlis, D. (2002). An EM type algorithm for maximum likelihood estimation of the normal-inverse Gaussian distribution. *Statistics & Probability Letters*, 57, 43–52.
- Kohavi, R., et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai*, 14, 1137–1145.
- Li, K. C. (1987). Asymptotic optimality for C_p , CL , cross-validation and generalized cross-validation: discrete index set. *The Annals of Statistics*, 15, 958–975.
- Pham, D. T., & Garat, P. (1997). Blind separation of mixture of independent sources through a maximum likelihood approach. In *Proceedings of EUSIPCO*.
- Schwarz, G., et al. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464.

- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267–288.
- Wang, H., Li, R., & Tsai, C. L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*, 94, 553–568.
- Wu, E. H., Philip, L. H., & Li, W. K. (2006). An independent component ordering and selection procedure based on the MSE criterion. *Independent Component Analysis and Blind Signal Separation* (pp. 286–294).
- Xie, H., & Huang, J. (2009). SCAD-penalized regression in high-dimensional partially linear models. *The Annals of Statistics*, 37, 673–696.
- Zhang, K., Peng, H., Chan, L., & Hyvärinen, A. (2009). ICA with sparse connections: revisited. *Independent component analysis and signal separation* (pp. 195–202).