Measuring and Predicting Systemic Risk in the Chinese Banking System

Yibing Chen[1, 2], Yong Shi[1, 2,*], Cheng-Few Lee[3], Minqiang Li[4], Yuewen Liu[5]
[1]Research Centre on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing 100190, China
[2]The Key Research Lab on Big Data Mining and Knowledge Management, Chinese Academy of Sciences
[3]Department of Finance and Economics, Rutgers University, Piscataway, NJ 08854, USA
[4]College of Management and Economics, Tianjin University, Tianjin 300072, China
[5]School of Management, Xi'an Jiaotong University, Xi'an 710049, China
chenyibing10@mails.ucas.ac.cn; yshi@ucas.ac.cn; cflee@business.rutgers.edu

***Abstract*—This paper highlights the importance of measuring systemic risk of commercial banks. Conditional Value-at-Risk (CoVaR) is used to measure the degree of "risk externalities" that a specific bank contributes to the whole banking system. Our analysis not only presents current levels of systemic risk of individual banks but also the changes with time passes. There is some evidence that larger banks contribute more to systemic risk, but size is far from being a dominant factor. We further explore to use some determinant balance-sheet factors to predict forward CoVaR for regulatory purpose. We extend modified Support Vector Regression (SVR) specifically for panel data, and apply the new model to predict systemic risk of commercial banks. The results show that the model is suitable for this problem.**

Keywords—systemic risk; conditional Value-at-Risk; Chinese banking system; modified Support Vector Regression; panel data

I. INTRODUCTION

This paper attempts to estimate the level of systemic risk of Chinese commercial banks, and explore method to link previous balance sheet data to forward systemic risk for prediction purpose.

Systemic risk has received much attention since the recent U.S. financial crisis[1]. Systemic risk is defined as the danger of one specific bank being in stress amplifying the panic in the whole banking system, leading to the failure of other banks, and consequently to the financial crisis[2]. Therefore, measuring and identifying level of this kind of risk for each bank is essential for bank supervisors and policy makers as well.

Most of the existing literatures of measuring systemic risk used the credit default swap (CDS) data[3]. This type of estimation method reflects the dependence between banks, however, it can only reveal credit risk rather than other types of risk. What's more, CDS data is difficult to find in emerging market like China. Recently, Adrian and Brunnermeier (2011) proposed a new methodology to measure and estimate systemic risk for each bank based on public data from stock market, which overcame the disadvantages of previous methods based on CDS data[4]. Using stock market data, they find that financial institutions in the U.S, those banks with large sizes, high levels of leverage, big maturity mismatch, and large market to book value contributed more to systemic risk.

Following Adrian and Brunnermeier's method, this paper attempts to quantify the level of systemic risk in the Chinese banking system using stock market data. Besides finding these important bank balance-sheet characteristics to explain the systemic risk, this paper innovatively adopts data mining approach—Support Vector Machines to help predict forward CoVaR for regulatory purpose.

The rest of this paper is organized as follows. Section 2 describes our methodology. We firstly present how to use quantile regression to estimate conditional VaR as systemic risk measurement for commercial banks. Then we extend modified Support Vector Regression (SVR) specifically for panel data. We derive the dual problem and list the algorithm procedure. In Section 3, data used for the study is described. Section 4 presents the current level as well as historical level of systemic risk for all the listed commercial banks. We then use our modified SVR model for panel data to link the previous balance-sheet data to the forward systemic risk. Section 5 concludes the paper and points out our future research directions.

II. METHODOLOGY

A. Conditional VaR Estimation

Recall that Value-at-Risk (VaR) is defined as the q quantile.

$$\Pr(X^i \leq VaR_q^i) = q \tag{1}$$

where X^i is the variable of bank i for which VaR is defined, shown as Equation (1). Please note that VaR_q^i is typically a negative number.

The key insight of Adrian and Brunnermeier's method is that one can estimate a banking system's VaR conditional on the event that a specific bank is under distress (i.e. the bank's market return reaches its VaR level), defined as Equation (2).

$$\Pr(X^{sys} \leq CoVaR_q^{sys|X^i=VaR_q^i} \mid X^i = VaR_q^i) = q \tag{2}$$

CoVaR is used to measure how much is the bank's risk spillover to the whole banking system. When the specific bank i contributes a lot to the system, its CoVaR, denoted as $CoVaR_q^{sys|i}$ would be a large negative number. This indicates a

978-1-4799-4274-9/14 $31.00 © 2014 IEEE
DOI 10.1109/ICDMW.2014.13

high potential loss of the system when this specific bank falls in distress.

The difference of CoVaR between when the underlying bank is on the distress and when it is on the median state can capture the risk externality that the underlying bank imposes on the system, which can be defined as Equation (3).

$$\Delta CoVaR_q^{sys|i} = CoVaR_q^{sys|X^i=VaR_q^i} - CoVaR_q^{sys|X^i=median^i} \quad (3)$$

The CoVaR measure can be computed from models with time-varying second moments from measures of extreme events, or by bootstrapping past returns. This kind of models can be estimated via maximum likelihood using a stochastic volatility or GARCH model if assumption about error term is made[4]. In this paper, we use quantile regression[5] to estimate CoVaR, shown in Equation (4) and (5).

$$\begin{aligned} q(\tau) &= \mathrm{argmin}_\xi \left\{ \tau \int_{y>\xi} | y-\xi | \, \mathrm{dF}(y) + (1-\tau) \int_{y<\xi} | y-\xi | \, \mathrm{dF}(y) \right\} \\ &= \mathrm{argmin}_\xi \left\{ \int \rho_\tau (y-\xi) \mathrm{dF}(y) \right\} \end{aligned} \quad (4)$$

$$X^{sys} = \alpha^i + \beta^i X^i + \varepsilon^i \quad (5)$$

Where Equation (4) shows the estimation principle, and Equation (5) is the equation for estimation.

In principle, quantile regression is extended to allow for nonlinearity by introducing higher order dependence of the system return as a function of a specific bank's return. Under this framework, the specific CoVaR measure can be obtained conditional on X^i is at its VaR, given in Equation (6).

$$CoVaR_q^{sys|X^i=VaR_q^i} = \widehat{\alpha}_q^i + \widehat{\beta}_q^i VaR_q^i \quad (6)$$

The $\Delta CoVaR_q^{sys|i}$ is then given by

$$\Delta CoVaR_q^{sys|X^i=VaR_q^i} = \widetilde{\beta}_q^i (VaR_q^i - VaR_{50\%}^i) \quad (7)$$

Equation (6) and (7) are the most simple version for estimating CoVaR when it is constant over time. To allow for time variation in the joint distribution of X^i and X^{sys}, the conditional distribution would be a function of lagged macroeconomic state variables. We run the following quantile regressions to estimate time varying $CoVaR_t$ and VaR_t, shown in Equation (8) and (9).

$$X_t^i = \alpha^i + \gamma^i M_{t-1} + \varepsilon_t^i \quad (8)$$

$$X_t^{sys} = \alpha^{sys|i} + \beta^{sys|i} X_t^i + \gamma^{sys|i} M_{t-1} + \varepsilon_t^{sys|i} \quad (9)$$

Note that the systematic state variables M_{t-1} are lagged. They should not be interpreted as systematic risk sources but rather as conditioning variables that are influencing conditional mean and volatility of the risk measures.

Then we obtain predicted value from Equation (8) and (9) as shown in Equation (10) and (11).

$$VaR_t^i(q) = \widehat{\alpha}_q^i + \widehat{\gamma}_q^i M_{t-1} \quad (10)$$

$$CoVaR_t^i(q) = \widehat{\alpha}^{sys|i} + \widehat{\beta}^{sys|i} VaR_t^i(q) + \widehat{\gamma}^{sys|i} M_{t-1} \quad (11)$$

Finally, under this kind of model specification, we can compute $\Delta CoVaR_t^i$ for each bank, as in Equation (12).

$$\Delta CoVaR_t^i(q) = \widehat{\beta}^{sys|i}(VaR_t^i(q) - VaR_t^i(50\%)) \quad (12)$$

This method has the well-known virtue that it does not require pre-assumption about the distribution. Specifically, normality is not required[5].

B. *Modified Support Vector Regression Model for Panel Data*

Support vector machines (SVMs), with their roots in Statistical Learning Theory and optimization methods, are proved to be effective and promising techniques for data mining[6, 7]. Large number of researches have shown that SVMs have demonstrated promising empirical performance, and have been successfully applied in many fields[8-10].

In the field of economic and finance, several successful applications of SVMs have been reported, including financial time series forecasting[11-13] and bankruptcy prediction[14-16].

Nevertheless, to the best of our knowledge, the SVM models used in finance field are limited to cross-sectional data and time series data, leaving panel data undiscussed. However, in real life, panel data are also common data form, which incorporates both cross-section and time series dimensions. For example, when we observe financial performance of 16 Chinese listed commercial banks for 5 years. This data set is the typical example of panel data. From time-series dimension, it is better to give more weights on recent data than distant data; while from cross-section dimension, all the banks should be treated equal.

Moreover, from the application perspective, this paper would be the first to apply data mining approach to banking systemic risk prediction problem.

Recall the formulation of standard SVM regression model, shown in Equation (13). This model specification equal weight all the data points when introduce the constant penalty parameter C.

$$\begin{aligned} &\min_{w,b,\xi^{(*)}} \quad \frac{1}{2} \| w \|^2 + C\sum_{i=1}^{l} (\xi_i + \xi_i^*) \\ &s.t. ((w \cdot x_i) + b) - y_i \le \varepsilon + \xi_i, \; i = 1,2\cdots,l \\ &y_i - ((w \cdot x_i) + b) \le \varepsilon + \xi_i^*, \; i = 1,2\cdots,l \\ &\xi_i \ge 0, \; i = 1,2\cdots,l \\ &\xi_i^* \ge 0, \; i = 1,2\cdots,l \end{aligned} \quad (13)$$

Tay and Cao (2002) proposed modified support vector regression model for financial time series forecasting[17]. They argued that more weights should be given to recent data than distant data. Therefore, they increased the value of penalty parameter C as time passes, shown in Equation (14). In Equation (14), C_i is ascending for more recent data points.

$$\min_{w,b,\xi^{(*)}} \quad \frac{1}{2}\|w\|^2 + \sum_{i=1}^{l} C_i(\xi_i + \xi_i^*)$$
$$\begin{aligned} s.t.&((w \cdot x_i) + b) - y_i \le \varepsilon + \xi_i,\ i = 1,2\cdots,l \\ &y_i - ((w \cdot x_i) + b) \le \varepsilon + \xi_i^*,\ i = 1,2\cdots,l \\ &\xi_i \ge 0,\ i = 1,2\cdots,l \\ &\xi_i^* \ge 0,\ i = 1,2\cdots,l \end{aligned} \tag{14}$$

Following Tay and Cao's arguments, we insist that for panel data, more weights should be given to recent data in time series dimension, while in cross-section dimension, all the data points are treated equal. Therefore, we propose our modified support vector regression model for panel data, shown as Equation (15).

$$\min_{w,b,\xi^{(*)}} \quad \frac{1}{2}\|w\|^2 + \sum_{i=1}^{T}\sum_{j=1}^{N} C_i(\xi_{ij} + \xi_{ij}^*)$$
$$\begin{aligned} s.t.&((w \cdot x_{ij}) + b) - y_{ij} \le \varepsilon + \xi_{ij}, i = 1,2\cdots T; j = 1,2\cdots N \\ &y_{ij} - ((w \cdot x_{ij}) + b) \le \varepsilon + \xi_{ij}^*, i = 1,2\cdots T; j = 1,2\cdots N \\ &\xi_{ij} \ge 0, \qquad i = 1,2\cdots T; j = 1,2\cdots N \\ &\xi_{ij}^* \ge 0, \qquad i = 1,2\cdots T; j = 1,2\cdots N \end{aligned} \tag{15}$$

In model shown in Equation (15), it can be easily observe that, we only control for the penalty parameter C_i in time series dimension.

Our model is a typical convex optimization problem, if Lagrange function is introduced, we can easily get the dual problem of Model (15), as presented in Equation (16)[1].

$$\begin{aligned} \min_{\alpha^{(*)} \in R^{2l}} \quad & \frac{1}{2}\sum_{i,k=1}^{T}\sum_{j,m=1}^{N}(\alpha_{ij}^* - \alpha_{ij})(\alpha_{km}^* - \alpha_{km})K(x_{ij}, x_{km}) \\ &+ \varepsilon\sum_{i=1}^{T}\sum_{j=1}^{N}(\alpha_{ij}^* + \alpha_{ij}) - \sum_{i=1}^{T}\sum_{j=1}^{N} y_{ij}(\alpha_{ij}^* - \alpha_{ij}) \\ s.t. \quad & \sum_{i=1}^{T}\sum_{j=1}^{N}(\alpha_{ij}^* - \alpha_{ij}) = 0 \\ & 0 \le \alpha_{ij} \le C_i, \qquad i = 1,2\cdots T; j = 1,2\cdots N \\ & 0 \le \alpha_{ij}^* \le C_i, \qquad i = 1,2\cdots T; j = 1,2\cdots N \end{aligned} \tag{16}$$

For the function form of C_i, there are many choices as presented in [17]. Here, we choose the linear form for simplicity, defined as in Equation (17).

[1] Due to the limited space, we omit the detail of derivation.

$$C_i = \frac{i}{T(T+1)/2} C \tag{17}$$

III. Data

At the end of year 2012, China's banking sector consisted of two policy banks and China Development Bank, 5 large commercial banks, 12 joint-stock commercial banks, 144 city commercial banks, 337 rural commercial banks, etc. Overall, the number of banking institutions in China's banking system amounted to 3747[18].

Although number of banking institutions is large, a small group of them dominate the entire market. Due to data availability but also the non-substitutable roles they play, we focus all the 16 Chinese listed commercial banks.

The publicly traded banks include all the 5 large commercial banks[2], 8 of 12 joint-stock commercial banks[3], and also 3 city commercial banks[4]. According to 2012 annual report of China Banking Regulatory Commission (CBRC) and our estimates from annual reports of all the listed banks, the 16 banks take a large asset proportion, around 62%[5] of the whole banking sector.

These listed commercial banks began to adopt new accounting standards in 2007. After the year of 2007, Chinese banking sector has been developing stably without significant reforms. All the listed commercial banks were listed after 2009. Therefore we choose the time period of 2009-2013 to gather and organize our data for analysis.

Data are gathered from Wind database[6]. Most of them are originally from annual financial statements and their notes of the 16 listed commercial banks while some items are from the regulatory agency—CBRC.

As with our previous research on domestic systemically importance [19], closing prices of every trading day for all the 16 listed commercial banks are used. Daily return rate of each bank at time t is then calculated as Equation (18).

[2] Industrial and Commercial Bank of China (ICBC), Agricultural Bank of China (ABC), Bank of China (BOC), China Construction Bank (CCB), and Bank of Communications (BOCOM).

[3] China Merchants Bank (CMB), Shanghai Pudong Development Bank (SPDB), China Minsheng Banking Corporation (CMBC), China Citic Bank (CITIC), China Everbright Bank (CEBB), Industrial Bank Corporation (IBC), Huaxia Bank (HXB), and Shenzhen Development Bank (SDB, now is merged with Ping'an Bank).

[4] Bank of Beijing (BBJ), Bank of Nanjing (BNJ), and Bank of Ningbo (BNB).

[5] 2012 annual report of CBRC stated that large commercial banks include ICBC, ABC, BOC, CCB and BOCOM own 44.93% assets of the whole banking sector. Upon all the end-2012 figures of assets of the 16 listed commercial banks, we estimated that their assets take around 61.5% of the whole industry.

[6] Wind database is a widely used data service provider in China. It provides data from annual reports and footnotes from listed companies, data in report released by governments such as National Bureau of Statistics, China Banking Regulatory Commission, and also stock market prices and trading volumes from exchanges.

$$return_t = \ln \frac{closeprice_t}{closeprice_{t-1}} \quad (18)$$

IV. QUANTITATIVE RESULTS

A. VaR and CoVaR Estimation

We run quantile regression to estimate the parameters for the 16 listed commercial banks respectively. Table 1 shows part of the parameters and t-statistics at 1% confidence level in 2013.

TABLE I. QUANTILE REGRESSION RESULTS (PART)

	Intercept	X^i
ICBC	-0.0155*** (-12.08)	0.6390*** (1.24)
ABC	-0.0248*** (-12.31)	1.0540*** (2.79)
BOC	-0.0291*** (-8.94)	1.4305* (1.67)
CCB	-0.0443*** (-6.62)	0.9133 (0.61)
BOCOM	-0.0229*** (-7.14)	1.0641 (1.54)

Figure 1 presents the 1% VaR and CoVaR of all the 16 banks at 2013.

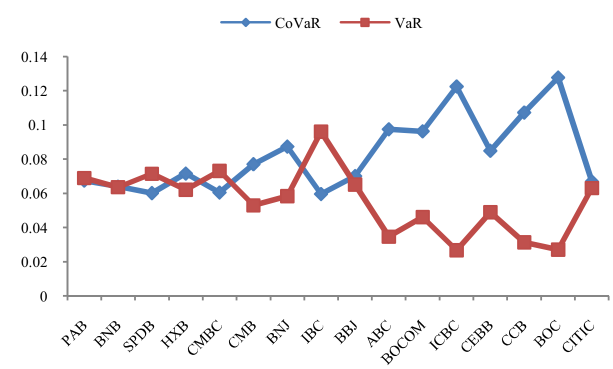

Fig. 1. 1% VaR and CoVaR of 16 listed banks at 2013

We are able to observe that, on general, large commercial banks such as ABC, ICBC, CCB, BOC exhibit lower VaR, which indicates that they tend to be less volatile. However, they also have much higher CoVaR, implying they would contribute much more to systemic risk when they are in distress.

Figure 2 is the 1% $\Delta CoVaR^{sys|i}$ of all the listed commercial banks during 2009-2013.

Figure 2 also confirm the previous conclusion that large banks tend to contribute more to systemic risk. What's more, by and large, systemic risk of each banks witness a sharp decrease during 2009-2012. However, in 2013, the average level of systemic risk increased and went back to the 2009 level. The major four banks even reached their peak in systemic risk measure, which alerts the regulatory agency.

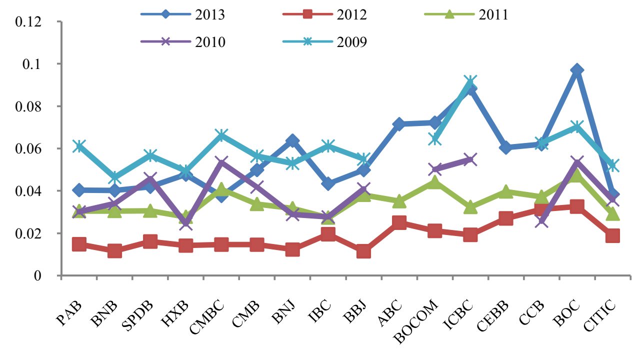

Fig. 2. 1% $\Delta CoVaR$ of 16 listed banks during 2009-2013[7]

B. Forward $\Delta CoVaR$ Prediction Using Modifed SVR

In this section, we first performed fixed-effect panel data regression using bank balance-sheet data to see whether the forward $\Delta CoVaR$ can be explained by bank balance-sheet characteristics.

The variables we collect from balance sheet are:

(1) leverage, defined as total asset divided by total equity;

(2) maturity mismatch, defined as (short term debt-cash) divided by total liabilities;

(3) mark-to-book ratio, defined as the ratio of the market to the book value of total equity;

(4) size, defined as the log of total equity.

Table 2 presents whether systemic risk contribution can be forecasted by lagged characteristics.

TABLE II. RELATIONSHIP BETWEEN $\Delta CoVaR$ AND BALANCE-SHEET CHARACTERISTICS

	1% $\Delta CoVaR$	5% $\Delta CoVaR$
VaR	-0.017*** (0.001)	-0.017*** (0.001)
Leverage	-6.762*** (2.065)	-6.324*** (3.472)
Maturity Mismatch	-43.792*** (-13.104)	-15.820* (-7.980)
Market-to-Book Ratio	-17.573*** (-3.451)	-19.480*** (-3.560)
Size	-330.056*** (-10.072)	-285.121*** (8.553)
Adjusted R^2	0.3578	0.4259

The regression results above shows that banks with higher leverage, more maturity mismatch and larger size tend to be associated with larger systemic risk contributions one year later both at 1% and 5% level.

[7] There are missing values of ABC and CEBB since they were not listed until 2010. Therefore, estimation for the two banks began in 2011.

To precisely capture the potential nonlinear relationship between these characters and forward $\Delta CoVaR$, we constructed our prediction models based on data mining approach. We used the data during 2009-2012 for training and left the data in 2013 for testing. Standard SVM regression model and modified SVM regression model for panel data are used for comparison. The results are listed in Table 3.

TABLE III. ACCURACY COMPARISON BETWEEN TWO SVR MODELS

	1% $\Delta CoVaR$	**5%** $\Delta CoVaR$
SVR	25.32%	22.62%
SVR-panel	25.15%	22.58%

We can observe that although no major differences between the two model, the accuracy is still improved a little when we use the modified SVM regression model for panel data. Due to lack of sufficient data under current topic research, we would include more dataset for further validation in our future work.

V. CONCLUSIONS

This paper attempts to use CoVaR to measure systemic risk of commercial banks in Chinese banking system. CoVaR measures the degree of "risk externalities" that a specific bank contributes to the whole banking system. There is some evidence that larger banks contribute more to systemic risk, but size is far from being a dominant factor.

We further explore to use some determinant balance-sheet factors to predict forward CoVaR for regulatory purpose. We extend modified Support Vector Regression (SVR) specifically for panel data. We apply the new model to predict systemic risk of commercial banks. The results show that the model is suitable for this problem.

We contribute to the existing SVM literatures by proposing the modified support vector regression model for panel data. From the application perspective, this paper would be the first to apply data mining approach to banking systemic risk prediction problem.

ACKNOWLEDGMENT

This work has been partially supported by the following Grants: Key Project (no. 71331005), Major International Joint Research Project (no. 71110107026), General Fund (no.71071151) and Youth Fund (no. 71201143) from the National Natural Science Foundation of China.

REFERENCES

[1] Acharya, V.V., et al., *Measuring systemic risk.* 2010.

[2] Roengpitya, R. , P. Rungcharoenkitkul, *Measuring systemic risk and financial linkages in the Thai banking system.* Bank of Thailand Discussion Paper, 2011. **2**: p. 2010.

[3] Chan-Lau, J., et al., *Assessing the systemic implications of financial linkages.* IMF Global Financial Stability Report, 2009. **2**.

[4] Adrian, T. , M.K. Brunnermeier, *CoVaR*, 2011, National Bureau of Economic Research.

[5] Koenker, R. , G. Bassett Jr, *Regression quantiles.* Econometrica: journal of the Econometric Society, 1978: p. 33-50.

[6] Deng, N., Y. Tian,C. Zhang, *Support vector machines: optimization based theory, algorithms, and extensions*2012: CRC Press.

[7] Qi, Z., Y. Tian,Y. Shi, *Laplacian twin support vector machine for semi-supervised classification.* Neural Networks, 2012. **35**: p. 46-53.

[8] Qi, Z., Y. Tian,Y. Shi, *Twin support vector machine with universum data.* Neural Networks, 2012. **36**: p. 112-119.

[9] Tian, Y., Y. Shi,X. Liu, *Recent advances on support vector machines research.* Technological and Economic Development of Economy, 2012. **18**(1): p. 5-33.

[10] Qi, Z., Y. Tian,Y. Shi, *Structural twin support vector machine for classification.* Knowledge-Based Systems, 2013. **43**: p. 74-81.

[11] Cao, L. , F.E. Tay, *Financial forecasting using support vector machines.* Neural Computing & Applications, 2001. **10**(2): p. 184-192.

[12] Kim, K.-j., *Financial time series forecasting using support vector machines.* Neurocomputing, 2003. **55**(1): p. 307-319.

[13] Tay, F.E. , L. Cao, *Application of support vector machines in financial time series forecasting.* Omega, 2001. **29**(4): p. 309-317.

[14] Huang, C.-L., M.-C. Chen,C.-J. Wang, *Credit scoring with a data mining approach based on support vector machines.* Expert Systems with Applications, 2007. **33**(4): p. 847-856.

[15] Min, J.H. , Y.-C. Lee, *Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters.* Expert Systems with Applications, 2005. **28**(4): p. 603-614.

[16] Fan, A. , M. Palaniswami. *Selecting bankruptcy predictors using a support vector machine approach.* in *IEEE-INNS-ENNS International Joint Conference on Neural Networks*. 2000. IEEE Computer Society.

[17] Tay, F.E. , L. Cao, *Modified support vector machines in financial time series forecasting.* Neurocomputing, 2002. **48**(1): p. 847-861.

[18] CBRC, *2012 annual report of China banking regulatory commission*, 2013, China Banking Regulatory Commission.

[19] Chen, Y., et al., *Domestic Systemically Important Banks: A Quantitative Analysis for the Chinese Banking System.* Mathematical Problems in Engineering, 2014. **2014**.