MACHINE LEARNING METHODS FOR SYSTEMIC RISK ANALYSIS IN FINANCIAL SECTORS

Gang KOU², Xiangrui CHAO¹, Yi PENG¹ *, Fawaz E. ALSAADI³, Enrique HERRERA-VIEDMA⁴

¹School of Management and Economics, University of Electronic Science and Technology of China, No. 2006, Xiyuan Ave, West Hi-Tech Zone, Chengdu, 611731, China
²School of Business Administration, Southwestern University of Finance and Economics, No. 555, Liutai Ave, Wenjiang Zone, Chengdu, 611130, China
³Department of Information Technology, Faculty of Computing and IT, King Abdulaziz University, Jeddah, Saudi Arabia
⁴Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Calle Periodista Daniel Saucedo Aranda, s/n, Granada, 18014, Spain

Received 01 August 2018; accepted 16 December 2018

Abstract. Financial systemic risk is an important issue in economics and financial systems. Trying to detect and respond to systemic risk with growing amounts of data produced in financial markets and systems, a lot of researchers have increasingly employed machine learning methods. Machine learning methods study the mechanisms of outbreak and contagion of systemic risk in the financial network and improve the current regulation of the financial market and industry. In this paper, we survey existing researches and methodologies on assessment and measurement of financial systemic risk combined with machine learning technologies, including big data analysis, network analysis and sentiment analysis, etc. In addition, we identify future challenges, and suggest further research topics. The main purpose of this paper is to introduce current researches on financial systemic risk with machine learning methods and to propose directions for future work.

Keywords: financial systemic risk, machine learning, big data, network analysis.

JEL Classification: G2, C8.

Introduction

In finance, systemic risk is a crisis that leads to the collapse of an entire financial system or entire market of an area or country, even global markets. The greatest impact of the global financial crisis in 2008, with strong economic destructive power and causing a huge chain reaction to destroy the financial industry, enabled systemic risk to be regarded as a crucial factor in relation to financial safety. Even today, the global economy has not fully recovered
from the aftermath of these events derived from the destructive effects of systemic financial risk. Thus, over the past decade, a large amount of ground-breaking academic research has focused on systemic financial risks, including the study of the financial ecosystem, financial supervision, monitoring cross-border capital flows, etc. However, systematic risk is always potentially hidden in modern large-scale financial systems, so that intelligent and automatic machine learning methods become a concerned tool to assess and detect the systemic risk from increasingly complex financial network, big data of financial transactions, and market sentiments together with risk proclivity, etc.

Systemic financial risk has become a hot concept over recent years as it has been used by researchers and government officials aiming to judge the possibility of damage to customers and financial markets, even economics. The pioneer of the financial risk research is Keynes, and he illustrated the systemic financial risk and highlighted regulation in his famous work “The General Theory of Employment, Interest, and Money” (Keynes, 1936). Generally, systemic financial risk should be a shock for stability of financial systems or public confidence (Billio, Getmansky, Lo, & Pelizzon, 2012), negative effects for economy (Bernanke, 2009; ECB, 2010), information failure of the financial market (Mishkin, 2007), and adverse impacts to financial institutions by spreading risk (Milne, 2014; Blancher et al., 2013), and so on. Despite all of this, the most important feature which differs from traditional financial risk is the issue of risk transmission and system failure across the entire financial system, caused by the continuous reaction of internal and external factors. Therefore, the results following systemic risk lead to “cascading” failures and affect the individuals in the financial system so that banking system cannot respond to liquidity and payment. Simplified, systemic financial risk should be a chain crisis triggering liquidity risk of a global or regional financial system (Silva, Kimura, & Sobreiro, 2017).

Generally, risk and its related research is to perform a generic process of theoretical and applied developments including concepts, frameworks, approaches and models to understand, assess, characterise and manage/govern risk (Aven, 2016). Therefore, risk assessment and management are the main tasks found in risk research. The topics of systemic financial risk are concerned with how we can understand, assess and manage the systemic issues focusing on risk identification and regulation. Therefore, to a large extent, assessment for systemic financial risk naturally forms a basis of responding financial crisis, and becomes the building blocks for financial institution, central bank and finance stability committee or other government organization practices.

Modern finance, similar to interdependent economies, is already composed of an interdependent system with a broad and interconnected network, which follows the increasingly global nature of human society (Leitner, 2005; Helbing, 2013; May, Levin, & Sugihara, 2008). Thus, responding to systemic risk in current finance new tools should be developed that can quickly identify and discover risk clues and objectives using information technology methods by means of automatically detecting abnormal risk behaviour and promptly handling large-scale financial data from financial markets. For example, the interconnected network is broadly distributed in bank systems, and risk can be spread in financial networks since risk exposure to one of them is induced (Giudici, Sarlin, & Spelta, 2017). Therefore, lots of machine learning methods have been implemented to try to analyse financial systemic risk and many improvements for real-world financial regulation have been acquired. In this regard,
surveying these methods combined with financial fields and proposing directions for further research is urgently required.

The aim of this paper is to introduce recent advances made in systemic financial risk using machine learning methods, focusing specially on assessment methods based on data mining and the traditional statistical and econometrics model, as well as the primary results in financial regulation for systemic risk. Therefore, this survey reviews the literature and methodologies used in the study of measurement and regulation for systemic risk in banking systems. We index related articles included in Web of Science dated from 2010 to 2017. The keywords were “financial”, “banking”, “systemic risk”, “financial stability”, “financial regulation”, “financial supervision”, “measure”, “indicator”, and “index”. Many well-known results pertaining to systemic risk are also included in order to extend a broader perspective of research, such as several important passages published in Nature, Science, American Economic Review, MIS Quarterly, and Management Science. For comprehensive results for technology, especially big data, EiCompendexWeb also index the words “big data” “financial stability,” “financial regulation,” and “financial supervision”. As shown in the references, it is interesting to note that a lot of journals from different areas have published related papers concerning financial systemic risk, such as Information Science, Operational Research and Social Science, as well as the Financial Journal. It can be judged as a classic interdisciplinary direction which includes not only interdependence within financial institutions but also has a wide range of external connections with the environment, geography, ecology, social science, technology and other factors.

The following sections will cover more aspects of financial systemic risk: main machine learning methods and core thoughts are introduced in Section 1. Assessment methods and the main results are respectively illustrated including network analysis, big data and sentiment analysis (Section 2). Furthermore, we propose future research topics based on the current literature in order to uncover more extensive applications of data mining (Section 3). Finally, this paper is concluded in the last section.

1. Machine learning methods combined with financial risk

Machine learning has been a hot topic in information and computer science fields that aims to establish an automatic algorithm to improve management tasks, making them more effective and accurate. Thus, machine learning for systemic financial risk is used to improve the assessment capacity and early warning for risk exposure using computer science approaches. The main research objects and representative literatures are listed at Table 1.

1.1. The classification of research objects

**Financial network.** The interconnected nature of financial systems is an essential feature of modern financial institutions. Thus, the financial network is the most studied topic in these fields in which the edges (connection relation) in the network among banks and countries with sovereign funds are loan related. Before the global crisis in 2008, the market-oriented financial point of view did not fully expose the destructive potential of systemic financial risk. For example, Eisenberg and Noe (2001) proposed a clearing mechanism based on default
by firms, and they also considered how systemic risk leads to the bankruptcy of individual firms. Freixas, Parigi, and Rochet (2000) studied the chain reaction caused by failures of coordination within the interbank market and examined the “too big to fail” policy. With the development of modern finance, the inter-bank market has formed a complex system which is a broad, interconnected and complex network (Helbing, 2013; May et al., 2008). Thus, the systemic risk in financial network becomes one of important objectives (Hu, Schwabe, & Li, 2015; Elliott, Golub, & Jackson, 2014; Ellis, Haldane, & Moshirian, 2014).

**Market sentiment.** As we well know, the contagion of risk is always accompanied by negative market sentiments and the results lead to customer panic within the financial system. The results will form a vicious circle of risk and emotion. Thus, the mining of market sentiment is always used to forecast a fluctuating state of the financial market and can be used as a “wind vane” for financial risk.

Generally, social networks such as Twitter and Facebook, Financial market news and laws are collected as text data set to analysis market expectation and risk clues are evaluated using correlation relationships between systemic risk and financial events (O’Halloran, Maskey, Mcallister, Park, & Chen, 2015; Cerchiello & Giudici, 2016; Giudici, Cerchiello, & Nicola, 2016). Moreover, futures and stock market asset prices are another hot topic that uses text mining by means of financial reports and quarterly earnings conferences, etc. In short, machine learning methods for sentiment analysis of the financial market is a novel insight to judging the state of financial risk.

**Stability of financial industry.** This area aims to analyse factors and incentives for financial stability that are triggered by systemic risk. The accurate causes of different risk classes can improve targeted policies to respond to different levels of risk in different financial industries, such as securities, insurance, derivatives, and futures markets.

In general, the studies can be divided into two kinds of objectives, including financial innovation and financial sector. On one hand, many financial tools and their combination are frequently used in modern financial markets, which increase the uncertainty of financial systems due to mismatch and unreasonable regulation, such as credit derivative markets, hedge funds, equity price bubbles, etc. (Calistru, 2012; King & Maier, 2009; Li, Liu, Siganos, & Zhou, 2016, etc.). On the other hand, researches are concerned with the sources of risk in different organizations of the financial system. For example, shadow banking in financial networks, resilient sectors in individual banks, levels of bank capitalization and so on (Liang, 2016; Gaffeo & Molinari, 2016; Gauthier, Lehar, & Souissi, 2012; Tsenova, 2014; Vallascas & Keasey, 2012, etc).

**Quantitative financial regulation.** As a result of the global financial crisis, financial regulation has become the consensus opinion of both banking managers and researchers when responding to financial crisis. However, the challenge of moderate policies is implemented with aims of effective financial risk control and improvement of financial development. Thus, the quantitative policy analyses are used to test and verify regulation strategies including benefit-cost insights and effectiveness of policies for financial safety. For instance, the researchers aim to establish a financial stability measurement system and comparisons of different regulation mechanisms, and so on (Cao & Illing, 2010; Ashraf, Rizwan, & L’Huillier, 2016, etc).
<table>
<thead>
<tr>
<th>Objectives</th>
<th>Branch</th>
<th>Literatures</th>
<th>Methods</th>
<th>Data and areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk exposure and transmission</td>
<td></td>
<td>Stein, 2011; Hutchison, 2002; Shen, 2017; Giudici, Sarlin, and Spelta, 2017; Giudici and Spelta, 2016; Amini, Cont, and Minca, 2013; Bluhm and Krahnen, 2014; Choi, 2014; Billio et al., 2012; Cruz and Lind, 2012; Ladley, 2013; Betz, Hautsch, Peltonen, and Schienle, 2016; Souza, 2016; Cerchielo and Giudici, 2015; Carmassi and Herring, 2016; J.P. Li, Feng, Sun, and M. L. Li, 2012; Bernardi and Romagnoli, 2016; Chinazzi and Fagiolo, 2015.</td>
<td>Multivariate network; multivariate graphical models and Bayesian graphical models; inhomogeneous directed graphs; particle movement model; conditional graphical models, distorted copula-based probability, etc.</td>
<td></td>
</tr>
<tr>
<td>Market sentiment</td>
<td>Correlation analysis of systemic risk and financial market</td>
<td>Choi, Chan, and Yue, 2017; Campbellverduyn, Goguen, and Porter, 2017; Flood, Jagadish, Kyle, Olken, and Raschid, 2011; Cui, 2015; Dong, Yang, and Tian, 2015; Smailović, Grčar, Lavrač, and Žnidarič, 2014; Cerchielo and Giudici, 2016; Sarlin, 2016a; Sarlin, 2016b; Cerchielo, Giudici, and Nicola, 2016; Brunnermeier and Pedersen, 2009; Agliardi, 2018.</td>
<td>Big data analysis; web-based Business intelligence; Bayesian method; simulation and fuzzy systems, Value-at-risk, etc.</td>
<td>Financial words base; financial market, etc.</td>
</tr>
<tr>
<td>Objectives</td>
<td>Branch</td>
<td>Literatures</td>
<td>Methods</td>
<td>Data and areas</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Market sentiment</td>
<td>Sentiment analysis in financial market</td>
<td>O’Halloran et al., 2015; Cerchiello and Giudici, 2016; Giudici et al., 2016; García, 2013; Price, Doran, Peterson, and Bliss, 2012; Tsai and Wang, 2013; Nyman, Gregory, Kapadia, Smith, and Tuckett, 2014; Chiang and Chen, 2015; Tsai and Wang, 2017; Meyer, Birkdahl, and Dai, 2017; Tsai and Wang, 2017; Tromp, Pechenizkiy, and Gaber, 2017; García, 2013.</td>
<td>Big data analysis; text mining; regression and ranking methods. lexicon and sentiment analysis; Rule-Based Emission Model algorithm, etc.</td>
<td>Financial news; reports; twitter; laws, etc.</td>
</tr>
<tr>
<td>Stability of financial industry</td>
<td>Microstructures of financial market;</td>
<td>Arora and Rathinam, 2011; Bengtsson, 2014; Jin and Nadal De Simone, 2014; Wymeersch, 2010; Bluhm and Krahnen, 2014; Calmès and Théoret, 2013; Xiong et al., 2011; Walter, 2012; King and Maier, 2009; Liang, 2016; Bongini, Nieri, Pelagatti, and Piccini, 2017; Li, Liu, Siganos, and Zhou, 2016; Acharya, 2009; Acharya, Engle, and Richardson, 2012; Brownlees and Engle, 2011; Brownlees and Engle, 2017.</td>
<td>Statistical method; Conditional-Risk and Systemic Expected Loss (SES), GARCH and SRISK methods, agent-based computational finance, etc.</td>
<td>Chinese stock market; ver-the-counter (OTC) derivatives market; European securities market; endogenous asset markets and Market-oriented banking, etc.</td>
</tr>
<tr>
<td>Financial tools, mechanisms and systemic risk,</td>
<td></td>
<td>Abedifar, Giudici, and Hashem, 2017; Calistru, 2012; Gaffeo and Molinari, 2016; Cox and Wang 2014; Acharya et al., 2012; Laeven, Ratnovski, and Tong, 2016; Calmès et al., 2013; Balogh, 2012; Calabrese and Giudici, 2015; Chen, Wang, and Yu, 2017; Hamdi, Hakimi, and Zaghoudi, 2017; Benţa et al., 2017; Crăciun, Bucerzan, Raţiu, and Manolescu, 2013; Liang, 2013.</td>
<td>Three-step importance sampling; regression analysis; logistic regression; stress testing, simulation analysis, etc.</td>
<td>Credit derivative markets, hedge fund, capital network and Shadow Banking, Tunisian banks, Romanian companies, etc.</td>
</tr>
</tbody>
</table>
1.2. The classification of methods

The most popular explanation of a machine learning task is described as: A computer program that is said to learn from experience \( E \) with respect to some class of tasks \( T \) and performance measure \( P \), if its performance at tasks in \( T \), as measured by \( P \), improves with experience \( E \). (Mitchell, 1997). The general process is comprised of understanding a task, data processing, algorithm selection and model generalization, etc. Based on the above research objectives, many methods have been employed in financial systemic risk, and the main types can be classified into three main categories including structure analysis, knowledge mining and interconnection relations corresponding to different issues.

**Network model.** One of the important tasks of macro prudential financial regulation aims to detect and regulate systemically important financial institutions. However, it is worth noting that the modern banking system and global financial market have formed a complex network with securitization and debt relation, and so on. The main methods have a centrality degree analysis, dual network, and network structure, Bayesian graphical models, etc. These methods are always used to identify important nodes in complex networks or to explore risk exposure in financial networks (Giudici et al., 2017; Battiston et al., 2016; May et al., 2008; Huang, Zhuang, Yao, & Uryasev, 2016; Dong et al., 2018a; Dong, Zhan, Kou, Ding, & Liang, 2018b, etc).

**Big data analysis.** Big data has become a hot topic in business and research. The related methods are applied to mining the correlation of different variables among massive data, such as liquidity of interbank and global capital flow. For example, the relationship that exists between web contracts and financial markets (Flood et al., 2011) and the fact that stock price is impacted along with public emotion (Smailović et al., 2014) are two topics of great interest. The main task of big data analysis aims to fuse heterogeneous financial information and acquire more knowledge in order to perceive and discern systemic risk. Thus, the big data analysis isn't different to other application areas which are composed of the information fusion of heterogeneous data and classic data mining algorithms including classification methods (Chao & Peng, 2017; Huang, Kou, & Peng, 2017; Wu & Kou, 2016; Kou, Peng, & Wang, 2014; Kou, Lu, Peng, & Shi, 2012; Chao, Kou, Li, & Peng, 2018) and correlation analysis, etc.

**Text mining.** Text mining could enable financial managers to discover risk factors or market sentiments from large-scale financial information including news, financial reports, tweets or other social network contexts. Existing literatures include the construction of a word base, popular word extraction and sentiment index analysis, etc. Many methods have been developed for financial text mining. For example, the classification method with ranking method (Tsai & Wang, 2017, 2013); the emotional words such as analysis of the fraction of positive and negative words (García, 2013) and so on. The detailed techniques and applications for sentiment analysis can also be found in Feldman (2013) and Li, Liang, Li, Wang, and Wu (2009).

**Statistical methods and Econometrics.** Classic methods such as statistical methods and economics are most commonly used in financial evidence research. The relations among variables of financial items can be studied using causal analysis, correlation analysis and
regression analysis. Moreover, more econometric methods, including Conditional Value as Risk, Conditional-Risk and Systemic Expected Loss (SES), and so on, have always been used to evaluate systemic risk with values of risk exposure. Also, the policy evaluations can be implemented with benefit-cost analysis using econometrics.


Assessment and regulation of systemic financial risk are regarded as the two main well-established branches of fundamental risk research strategies, which recognise risk and regulate it. Naturally, there are different insights for studying systemic financial risk, including loss evolution, identification of “too big to fail” institutions and impacts on the financial system and even the economy. Although our concerns in this paper are detection methodologies for the recognition process using machine learning methods, this section will introduce related developments of systemic risk research with network-based insights, big data analysis, and sentiment mining and econometrics methods.

2.1. Network based systemic risk in financial institutions

Differing from traditional financial risk for individual institutions, network-based banking systems have been studied with a macro insight of the risk conduction of the financial network for general financial systems. It has gradually become more popular due to historical events such as the Asian financial crisis, the global financial risk in 2008 and the European debt crisis.

2.1.1. Evolution of financial network

Before the financial crisis in 2008, we found that the study of regulatory issues in the context of systemic financial risk appeared to have been neglected over many years, and that early studies had tried to respond to the risk of a single bank run. While the importance of financial regulation is widely recognized, in the face of complex networks being formed in current banking systems, dealing with systemic risk has been found to be very difficult, and is now a critical topic in current financial research (Battiston et al., 2016).

The general view is that complex networks are formed by the extensive interconnection of financial institutions so there tends to be a probability of risk transmission. On the one hand, the need for liquidity in the banking system requires direct connections between banks; on the other hand, homogeneous products and risk aversion lead to a wide range of indirect connections among banks. According to Battiston et al. (2016), economic and financial policy needs to apply network analysis, behavioural modelling and complex systems theory. The importance of monitoring and management in highly connected financial networks has attracted widespread attention. In fact, the study of financial regulatory issues is facing the challenge of dealing with complex networks. Haldane (2015) suggests that current policy research should learn from economic theory, complexity theory, the social sciences, ecology, epidemiology and finance. The tipping point and the problem of stability in a complex financial network will decide the direction of developing financial policies. A more complete
network will enhance stability under small shocks, which is analogous to the storage of water in a wetland, but risk-taking and failure will happen beyond certain points (Acemoglu et al., 2015). Haldane and May (2011) believe that, apart from the environment, conflict, and other external factors, trading in derivatives within the financial system is one of the main sources of financial network hazard. In addition, network-based financial systems and the contagion lead to amplified exposure in financial networks (Prasanna et al., 2011; Hu et al., 2012), and the liquidity of the interbank system in England exposes the instability of financial networks (Ferrara et al., 2016). All of them are hot topics for theoretical and empirical research.

2.1.2. Network-based models

On the basis of the above results, modern financial innovation has led to a complex interbank network constructed to avoid risk and assure liquidity. From this point of view, the systemic financial risk originating from a higher connected banking network is apt to contagion and to spread. In parallel with the development of this network, bank supervision and market regulation should be extended to monitor its stability. Recently, a large amount of research has studied the issue of network-based regulation of financial risk and has obtained many theoretical results.

2.1.2.1. Risk exposure and transmission in financial network

More detailed discussions are given to show how the network model is used to reveal exposure of financial risk. These financial networks are always established as a matrix of risk exposure which can simulate the process of bank default. Giudici et al. (2017) assessed multivariate network structures among different countries and risk exposure processes, as well as measures of interconnectedness based on BIS data. The proposed model can be used as an early warning and forecasting contagion. Giudici and Spelta (2016) proposed and proved that the distinct groups of countries have different actions as central nodes in the financial network by means of multivariate graphical models and Bayesian graphical models. Amini et al. (2013) studied the cascade process and asymptotic contagions in financial networks using inhomogeneous directed graphs, proposing minimal ratios of capital that should be set with respect to contagion exposure in regulation policy. Bluhm and Krahnen (2014) developed a network model to expose systemic risk on the basis of interconnected bank balance sheets and proposed systemic value at-risk macro-prudential methods. Choi (2014) recognized that a larger institution has a greater positive impact on the financial network, and therefore, bolstering the strong rather than the weak, will effectively help to maintain the stability of banking systems.

Other research discloses travel paths of risk in networks and several corresponding countermeasures can be implemented by means of counter-cyclical supervision. Billio et al. (2012) proved that inter-correlation always exists in financial networks experiencing exogenous shock. Cruz and Lind (2012) used a particle movement model to show how a chain of insolvencies is triggered by an agent in a trading network. Ladley (2013) believed that larger shocks lead to financial system risk, while the systematic risks cannot be observed from a smaller banking default. In the context where risk diversification and risk transmission are
permitted, there is no optimal market structure. In contrast, a deposit insurance scheme reduces interdependence within the interbank network so that the probability of the occurrence of systemic risk is reduced. Betz et al. (2016) provided a framework for the evaluation of systemic risk within different periods. The framework revealed a process of fragmentation of the banking sector and linkages to sovereign-banks. Souza (2016) found, in the Brazilian banking system, that a larger bank will suffer more severe losses. Cerchiello and Giudici (2015) evaluated conditional graphical models to study systemic risks among the largest European banks by identifying central institutions, based on joint market data and balance sheet data. Shen (2017) proposed a Bayesian networks approach to modeling the financial risks of e-logistic investments.

2.1.2.2. Financial network and the real economy

The impact on the real economy is another evidence-based topic in terms of special network with different nodes established in them, such as a network between the interbank market and the real economy (Gabbi et al., 2014), and financial markets (Diebold & Yilmaz, 2014). Helwege and Zhang (2016) argued that an information contagion had a greater impact on competitors from same market, and caused more profound harm in a crisis than in the bankruptcy stage. Diebold and Yilmaz (2014) used connectedness in a weighted directed graph to study daily-varying connectedness of volatilities in major US financial institutions’ stock returns in the financial crisis between 2007 and 2008. The financial market was damaged after systemic shock due to inter-conductions of assets. Giudici and Parisi (2017) used correlated stochastic processes to investigate correlation of the Debt and the GDP, and the results that provide sovereign risk are always affected by debt evolution and GDP growth.

2.1.2.3. The structure of financial network

Multilayer bank networks are one of the important features in financial systems which are built by common exposures to common assets. A multilayer network always leads to a decreasing capability to defend against financial risk and increase the cost of a financial crisis (Battiston, Farmer, & Flache, 2016). Poledna et al. (2015) recognized that a financial network should be treated as a multilayer network connected by credit, derivatives, foreign exchange, and securities, and that the expected loss will be underestimated by up to 90% of the total risk if networks are only focused on a single layer. Another result is that the financial multilayer network presents nonlinearity events where the sum of the systemic risk of all layers is lower than the total risk. Fukuyama and Weber (2015) have studied the network performance of Japanese credit cooperatives via two-stage production.

In this area, the main methodologies used to study the banking network include random graph theory, particle physics, game theory, benefit-cost analysis, i.e. a range of mathematical models and other methods to represent complex networks are brought into the study of banking systems. Also, other methods such as epidemic models and stochastic optimal models have been applied to research into financial networks. However, current studies are confined to the interbank system within a single economy such as the US banking system, the Bank of England, the European Union, Brazil, and other emerging economies. As a global research network, there is still very little research on cross-border financial institutions and
Technological and Economic Development of Economy. Article in press

the global financial system. The overall network has been not extensively researched. For example, the existing literature is largely focused on the joint financial and economic network, with very little research on issues linked to networks related to geography, humanities and society, and other external factors, which are an urgent problem to be solved in future network-based systematic financial risk research.

Admittedly, network based financial systemic risk should consider big data since the network will become larger and produce mass data. Thus, big data in financial systemic risk must be of concern to researchers and should turn into a hot topic in these areas. We will review the existing methods with big data analysis in the next section.

2.2. Big data analysis for financial risk management

In this section, we will review several recent developments for financial risk analysis using big data insight. Nowadays, big data has become a buzzword, not only in theoretical research by means of data mining technologies and business intelligence, but also in operational management extensively applied in areas which contain financial risk analysis, political economics, business data resource services, e-commerce, supplier chain risk etc. (Campbellverduyn et al., 2017; Wu & Olson, 2010; Kim, Chung, Jung, & Park, 2013; McAfee & Brynjolfsson, 2012; Koyuncugil & Ozgulbas, 2012; Bottani, Bertolini, Montanari, & Volpi, 2009; Xu, He, & Li, 2014). The related works will be surveyed following two issues corresponding to big data with algorithms and data.

2.2.1. Business intelligence and systemic financial risk

Generally, strategies and technologies of business intelligence (BI) can be divided into five parts: internet, radio frequency identification, cloud service, decision support systems and web-based BI (Choi, Chan, & Yue, 2017). The essential supports of current BI come from two key sections. Firstly, the big data resource serves as the base of decision making systems. Secondly, knowledge discovery using data mining from big data is another BI task.

The financial big data involves related financial information including interbank liquidity and global capital flow, and can be used to carry out tasks of early-warning risk and discerning risk to respond to risk by means of financial regulation. Current/recent research includes big data algorithms (Campbellverduyn et al., 2017), correlations between web contracts and financial markets (Flood et al., 2011), evaluated financial credit risk (Cui, 2015), volatility analysis of the Chinese stock market (Dong, Yang, & Tian, 2015) and relations between stock price and public emotion (Smailović et al., 2014).

For systemic financial risk, big data has been used to analyse the interrelationships between risk source and diversification of risk based on financial big data. Cerchiello and Giudici (2016) proclaimed that they built the first systemic risk model based on big data where data is selected by two heterogeneous sources. The results combined different data structures using Bayesian method and tested correlations between financial risk and public emotion. Sarlin (2016a) introduced main intelligent algorithms such as machine learning, network analytics, simulation and fuzzy systems for systemic risk. Sarlin (2016b) proposed a visual tool for financial regulation with three modules: plots, maps and networks. Cerchiello et al.
(2016) studied a risk contagion model based on financial twitter and financial markets, their correlation is predicted when shocks happen.

Currently, research using big data insights aim to fuse heterogeneous financial data, however, the amount of data is also limited to acquisition from global financial markets.

2.2.2. Data challenges in financial risk

Finance is one of the potential areas to in big data. However, the financial risk analysis depends on high quality data. Because of this, it is difficult to extend these issues due to unavailable data resources of local and global financial markets, even the data quality isn't guaranteed. Thus, research in this field is not common, although the topic has become very popular.

How to measure the effectiveness of a paper using big data is a popular topic. Some authors think that multi-source heterogeneous data integration should be served as big data (Cerchiello & Giudici, 2016). Flood et al. (2011) claimed that the risk in finance is always associated with time and data and is time-varying, thus, it is difficult to ensure effective risk analysis of the data resource. Moreover, data integration and quality are also a difficult issue to address due to the complexity of financial entities. Jagadish et al. (2014) proposed that big data analysis should include many steps, such as data collection, modeling and analysis, interpretation and deployment. However, challenges of big data including heterogeneity, timeliness, inconsistency and incompleteness are also becoming obstacles to apply in real life questions. Similarly, it is also difficult to utilize big data analysis in data resources of financial markets, especially interbank and bank networks. Yang et al. (2014) proposed a framework and classification standard in order to improve storage data. In addition, data quality also should be considered in big data analysis. Fardani Haryadi, Hulstijn, Wahyudi, van der Voort, and Janssen (2016) defined data quality of different banks and proposed detailed goals. Brammertz and Mendelowitz (2014) proposed an information extraction for financial regulation from financial big data which comes from the granularity information of financial contracts, transformed into a contingent cash flow and they also developed a mixed model for multi-level financial objectives such as institutions, systems and individuals.

It is clear that the data resource and quality are concerning topics in big data and financial data analysis. Data quality will be an important direction for research in the future (Choi et al., 2017), moreover, the evaluation index system of data quality should be further developed in order to establish a solid foundation for big data analysis. However, the volume of data hasn’t yet been discussed in related finance and management science literature, and thus it will become a new challenge in this research.

2.2.3. Sentiment analysis in financial markets

Because of the data acquisition of the interbank market, internet data such as tweets, news and laws always served as text data to study financial risk. Sentiment analysis or opinion mining aims to discover attitudes, opinions and trends of customers and markets using text mining for a certain field. First, text bases are established for a mining market sentiment. O’Halloran et al. (2015) created a news database for financial regulation extracted from U.S. federal laws in order to rank features and classify laws into various labels. The big data
analysis can help regulators select rational laws to respond to different financial markets. Also, Twitter data can be used to estimate financial systemic risk. Further, the method can be used to individuate a contagion of a financial institution (Cerchiello & Giudici, 2016; Giudici et al., 2016). García (2013) found a relation between asset price and sentiment based on the New York Times during the 20th century and showed deviation shock in recession to predict change from conditional average return. Similar results can also be seen in the quarterly earnings conference calls (Price et al., 2012).

Public emotion can be employed to determine potential trends by means of sentiment mining (Tsai & Wang, 2013), although there are many dependencies between market analysis and financial systemic risk. Nyman et al. (2014) found that financial systemic risk can be gauged by market sentiment based on text set. Chiang and Chen (2015) collected and constructed a financial text database from financial news websites and they extracted a popular word set and calculated a sentiment index. Further, they found that a sentiment index can be used to predict the stock index of the electric industry. Tsai and Wang (2017) forecasted financial risk using text information from financial banking reports, and the results show soft information that is recognized as a strong correlation to financial risk. Their discovery can prove the strong correlations between financial risk and market sentiment. Meyer et al. (2017) implemented a granularity sentiment analysis of financial news in order to reveal pricing and risk prediction models. In their model, natural language processing is used to extract syntactic sentence patterns to predict sentiment.

The sentiment analysis methods are classic machine learning algorithms, including regression and ranking methods (Tsai & Wang, 2017, 2013), lexicon and machine learning sentiment analysis (Meyer et al., 2017), Rule-Based Emission Model algorithm (Tromp et al., 2017), fractions of positive and negative words (García, 2013) and so on.

Sentiment analysis undoubtedly becomes an effective way to analysis financial risk when the data resource of a financial market hasn't currently been established with open and uniform standards. Further, more technologies and text databases should be developed for systemic risk management using text analysis.

2.3. Econometrics and statistics for systemic financial risk

Econometrics and statistics (including probability theory) have a long history for related financial research. More classic methods are used to assess systemic risk of financial institutions which aim to find the determinants and indicators of systemic risk through quantitative analysis and empirical research. Especially in finance, these methods have also occupied a large proportion and are still the basis of related studies. We will introduce mainstream models in following subsection.

Traditional models have been proposed for systemic risk, such as Conditional Value as Risk, Conditional-Risk and Systemic Expected Loss (SES), and so on. Acharya (2009) proposed that the limited liability of banks and the presence of a negative externality of one bank's failure leads to systemic risk. Acharya, Pedersen, Philippon, and Richardson (2010) proposed SES can be used to measure the capital gap of an individual bank shocked by financial crisis. Acharya et al. (2012) described a new measurement of systemic risk based on capi-
tal shortfall which is a function of leverage and expected loss. Brownlees and Engle (2011, 2017) proposed measurement of capital shortfall using GARCH and SRISK methods which can be served as warning signs of risk in financial institutions. Cross-sectional methods are used to study the interdependency of risk status among special financial institutions. Carmassi and Herring (2016) studied the relationship between the complexity of systematically important banks and systemic risks from the perspective of the structural complexity of the financial institutions. Their empirical results from panel data showed that large mergers and acquisitions are still the major source of complexity. Battaglia and Gallo (2017) believed that, in addition to the independence of the board of directors, the characteristics of the board have a significant impact on the systemic risk of banking. Capital management, the ownership structure and control of the business of banking all have some influence on this relationship.

The shortage of liquidity always gives rise to the risk of bank runs, and the liquidity will decrease in financial risk so that chain reaction occurs in bank systems resulting in systemic risk. Many statistical methods have been introduced to handle systemic questions. Cox and Wang (2014) studied predictors of bank failures, including the proportion of illiquid loans and exposure to the interbank funding market, and identified the signs that can help managers forecast risk. The study of Laeven et al. (2016) found that the growth of systemic risk is proportional to the size of the bank and inversely proportional to the bank's capital. Therefore, they proposed that additional capital requirements do not contribute significantly to preventing systemic risks. Calmès and Théoret (2013) developed a new type of risk-weighted asset management index, which is more effective at capturing the level of systemic bank risk. Classic regression methods are also popular approaches to analyse systemic risk in finance. Balogh (2012) studied the factors affecting the ratio of bad assets and the capital adequacy ratio of the banking system in Europe through regression analysis, then analysed the main content of the macro-prudential regulatory framework and developed a set of monitoring and early warning indicators of systemic risks. Calabrese and Giudici (2015) applied a logistic regression method for binary data on the basis of both macroeconomic factors, which are relevant under the situation of actual defaults, and mergers and acquisitions.

Macroeconomic methods, for example, the famous Dynamic Stochastic General Equilibrium (DSGE) model, are also used to analyse the impact on the real economy with respect to financial risk. Brunnermeier and Sannikov (2014) built a DSGE to study growth rate and found it would be unstable due to the drastic fluctuation of capital price. Duca and Peltonen (2013) proposed a macro index combining M2, house prices and GDP for systemic risk. Huang, Zhou, and Zhu (2012) constructed a method of systemic risk measurement from a “historical correlation” and simulation viewpoint. Calmès and Théoret (2014) studied the effects of macroeconomic risks on systematic financial risks and identified the differences between banking risks in the downturn and up-cycles of economics.

In addition, other econometrics and statistics methods also include forward-looking (Kritzman & Li, 2010; Segoviano & Goodhart, 2009), contingent claims analysis (Merton, 1974; Jobst & Gray, 2013) and stress testing (Allen, Goldstein, Jagtiani, & Lang, 2016), credit risk distribution (Bernardi & Romagnoli, 2016), etc.

Econometric methods used in existing literature are mainly based on univariate and multivariate regressions. There are other methods, such as Li, Wang, and He (2013) that used a
Support Vector Machine (SVM) to study the forecasting of systemic financial risks. The main panel data came from the banks in Europe, the United States, Canada, China, Croatia, and the Islamic National Bank, and the variables were balance sheet ratios like debt-to-income and loan-to-value. L. Zhang, Hu, and D. Zhang (2015) studied credit risk assessment models for small and medium enterprises in supply chain finance using SVM. Calabrese, Elkink, and Giudici (2017) employed a binary spatial autoregressive model to measure contagion effects during the European sovereign debt crisis.

Although econometric methods employed to measure systemic risk have been around a long time the modern financial system has become more a complex correlation in banking. Therefore, econometric methods have become more difficult to study due to the complex network of banking.

### 2.4. Financial market risk and stability analysis

Studies in this area cover a number of different financial sectors such as securities, insurance, derivatives, and future markets. Research on microstructures is an important part of financial system risk including over-the-counter (OTC) derivatives market (Arora & Rathinam, 2011), invest fund risk (Bengtsson, 2014; Jin & Nadal De Simone, 2014), European securities market (Wymeersch, 2010), endogenous asset markets (Bluhm & Krahnen, 2014) and Market-oriented banking (Calmès & Théoret, 2013), Cross-market financial risk (Xiong et al., 2011) as well as the functional mechanism of the systemically important institutions (Walter, 2012).

Some authors attempted to explore relations between different financial tools and systemic risk, such as credit derivative markets, hedge fund, capital network, Shadow Banking, and so on. Abedifar, Giudici, and Hashem (2017) found that the conventional banks with Islamic windows have the least resilient sector under systemic risk attack in dual banking systems. Calistrulu (2012) studied risk management measures for credit derivatives and suggested that the operational efficiency of credit derivative markets may prevent the occurrence of systemic risks. King and Maier (2009) studied the effect of hedge fund linkages on systemic financial risks. Liang (2016) studied the effects of Shadow Banking on China's financial system. They believed that, on the one hand, shadow banking has promoted risk contagion to financial institutions through the financial network. On the other hand, they believed that the role of shadow banking in contributing to the real economy is weakening. The study of Gaffeo and Molinari (2016) showed that an important warning factor is the size of the inter-bank market and the level of bank capitalization. The large, tightly connected central market is beneficial for the capital network of the highly developed interbank market, but, for the primary capitalized market, it can significantly weaken the system's flexibility.

Recent representative studies also include: Bongini et al. (2017) which addressed systemic risk in the insurance industry. The study suggested that investors are sceptical about whether the new regulatory framework can reduce systemic risks in the insurance industry and whether the "too big to fail" policy can control the spread of moral hazard. Kuzubas, Saltoglu, and Sever (2016) showed how different levels of leverage can drastically change the systemic effect of defaults and the nature of interbank market contagion. The study of Li, Liu, Siganos, and Zhou (2016) showed that proper regulation reduced the likelihood of equity price bubbles caused by increased issuance of bank equities. J. P. Li, Feng, Sun, and
M. L. Li (2012) reviewed risk integration of the banking industry and proposed several topics for further research. Hamdi et al. (2017) studied bank performance and risk under a new business model in Tunisian banks via non-interest income analysis.

2.5. Quantitative analysis in financial regulation

Over the past few years, a large amount of research has made significant progress on financial regulation dealing with systemic risks (Galat & Moessner, 2013; Adrian, Covitz, & Liang, 2015; Alexander, 2011; Borio, 2003; Kara, 2016).

In response to the financial crisis, the consensus opinion of financial researchers, central banks, and international financial institutions is that the lack of financial supervision led to the spread of systemic risks. A wide range of financial regulatory research derives from the lesson of the global financial crisis. Besides proposing three main methods, which include network, big data analysis and econometrics summarized as primary technologies to measure systemic risk, the regulation design is also another important field based on the proposed analysis results in theoretical research and the practice of finance regulation. Several benefit-cost analysis pathways should be considered to evaluate the cost of financial regulation (Posner & Weyl, 2013).

As a prominent topic in financial regulatory research, systemic risk regulation also concerns network analysis. Discovering knowledge and making management decisions from the results of quantitative analysis will be one of the mainstream directions for future financial regulation, for example, interdependence across policies (Bosma, 2016) and the relation of optimal intervention policy rule (Clark & Jokung, 2015). Cao and Illing (2010) compared different regulation mechanisms and derived a method for coping with systemic liquidity risk. Nucera et al. (2016) studied the classification of systemic risk, finding that the results of price-based and capital-based ratings significantly deviated from each other.

The econometric and statistical perspective is the most traditional research direction in this area. Ashraf et al. (2016) established a standard financial stability measurement system for the banking system of Islamic countries, using regression analysis, which verified the effectiveness of the Net Stable Funding Ratio (NSFR) for the financial stability of banks there.

Other insights were the relation between financial risk and economic development (Calmès & Theoret, 2014; Lupu, 2015; Fazio et al., 2015; Fernández et al., 2016).

3. Future research

It can be seen from the above literature that systemic risk has become a very active area of financial research. According to the previous analysis, we present some current trends in the financial systemic risk, along with some open questions and prospects. We identify three current trends:

1) using big data analysis to respond to systemic risk in finance;
2) data-driven systemic risk analysis;
3) Reg-tech for systemic risk in the global financial market.
3.1. What is the impact of big data?

Undoubtedly, the development of big data has opened new avenues for theoretical research as well as the application of industrial data. The use of big data in the study of systemic financial risk is bound to be one of the main directions for future research. At present, the financial regulation of systemic financial risk is mainly focused on the macro perspective and the identification of pro-cyclical risk and response. The management of systemically important institutions, data mining and the application of knowledge discovery requires further investigation in the future. Therefore, the use of big data technology must break through current research barriers to explore micro-management areas, including the detection of abnormal behaviours, classified management of micro subjects, typical case modeling, and cross-country corporate regulation modeling.

For example, the identification of “too big to fail” institutions is one of the main tasks of macro-prudential financial regulation. However, with the highly networked financial system and research on the risk contagion uncertainty, as well as financial system complexity, the importance of “too big to fail” has been gradually turning into a “too complex to fail” issue. Identifying the key nodes in the complex financial network is an effective way of promoting financial regulatory measures and policies. Therefore, mainstream research on complex financial networks should include big data analysis technologies, further, the stability of ecology of finance based on partial differential equations, etc.

3.2. Should data-driven in the financial systemic risk be stressed?

Real data in financial systemic risk using data mining is one of the main challenges since the financial data is held under strict resource and budget constraints. Big data isn’t yet thoroughly applied in financial systemic risk due to unavailable data, such as connection of interbank data. Thus, empirically grounded research for global financial markets with cross-country financial insight is further topic. It is an extremely hard task to gather data and/or to sanitize whatever data may be available. The systemic financial risk, as one of application-oriented topics, needs more communication in collecting and sharing data from different sources, such as governments, international organizations and financial systems, etc.

For example, in traditional research, the sources of risk contagion and systemic risk in the financial network were thought to arise from the systemically important financial institutions. Therefore, most countries relied on these institutions for financial regulation. However, who can guarantee that the stability of the network connecting non-financial institutions and financial institutions will not affect the stability of the system as a whole? The multi-dimensional system structure of the financial network means that the financial institutions in the network are affected by the non-financial institutions, leading to increased instability. The network relationship in current research concentrates on debt, capital lending, and cash flow between financial institutions. Therefore, the data sources of non-financial and financial institutions, especially cross-border institutions, will be one of the main directions of future research.
3.3. How to optimally transform into policies from quantitative study with data science

Responding to systemic financial risk just like to financial regulations is more concerned with building an organic connection for theoretical research and management practices which can be devoted to dealing with financial regulation questions. It is the nature of management science and an essential demand of information technology. Reg-tech, which responds to financial risk using information technology, has become a popular topic in recent years. The regulation of systemic financial risk is a financial system engineering problem requiring the use of a comprehensive set of policy tools in terms of quantitative studies in data science.

For example, a sound financial regulation study should combine macro and micro-prudential elements, as well as behavioural supervision. However, the best policy choice requires a comprehensive assessment of costs and benefits. Therefore, the evaluation of the effectiveness of the implementation of policies is one of the important areas of systemic financial risk regulation. The main directions for future research include multi-objective methods as well as the application and promotion of cost-effective analysis methods.

Conclusions

Responding financial systemic risk has become critical to modern financial market and financial safety. However, modern financial systems construct a complex ecosystem rather than independent individuals. Thus, many machine learning methods have been developed to detect and identify the systemic risk in financial markets and sectors. Additionally, lots of traditional econometrics and statistical models are also used to systemic risk analysis. Thus, there need a bibliometric research for this issue and summarize the current studies.

For introducing current researches on assessment and measurement of financial systemic risk combined with machine learning technologies, we survey existing literatures and methodologies including big data analysis, complex network analysis, sentiment analysis and classic econometric models. Moreover, stability analysis of financial markets and quantitative analysis for financial policy are also introduced as a branch of financial regulation. Finally, we proposed many further research directions, such as big data analysis, data-driven research, and policy analysis associated to data science. In summary, responding financial systemic risk should develop more accurate regulation policies using intelligent methods based on financial big data in the future.

Additionally, we hope that our insights on the many opportunities available in this research domain will help guide the potential research directions for the future development of this field.

Acknowledgements

This research has been partially supported by grants from the National Natural Science Foundation of China (#U1811462, #71874023, #71771037, #71725001, and #71433001).
References


