Journal of Financial Regulation and Compliance

A review of credit scoring research in the age of big data

Ceylan Onay, Elif Ozturk,

A Review of Credit Scoring Research in the Age of Big Data

Abstract

Purpose – This paper surveys the credit scoring literature in the past 41 years (1976-2017) and presents a research agenda that addresses the challenges and opportunities big data bring to credit scoring.

Design/methodology/approach – Content Analysis methodology is used to analyze 258 peer-reviewed academic papers from 147 journals from two comprehensive academic research databases to identify their research themes and detect trends and changes in the credit scoring literature according to content characteristics.

Findings –We find that credit scoring is going through a quantitative transformation, where data-centric underwriting approaches, usage of non-traditional data sources in credit scoring and their regulatory aspects are the up and coming avenues for further research.

Practical implications – The paper's findings highlight the perils and benefits of using big data in credit scoring algorithms for corporates, governments and non-profit actors who develop and use new technologies in credit scoring.

Originality/value – This paper presents greater insight on how big data challenges traditional credit scoring models and addresses the need to develop new credit models that identify new and secure data sources, and convert them to useful insights that are in compliance with regulations.

Paper type: Research Paper

Keywords: Credit scoring, Big data, Data Privacy, Discriminatory scoring, Financial inclusion, Access to credit

1. Introduction

Credit scoring helps lenders evaluate the potential risk of new customers and also assess future behavior of existing customers by using statistical models to transform relevant data into numerical measures that guide credit decisions (Abdou and Pointon, 2011). Credit scoring traditionally relies on consumers' financial history to generate a credit score, which indicates the borrowers' credit risk.[1] However, big data is bringing disruptive change to credit scoring. Campbell-Verduyn, Goguen & Porter (2017) discuss that big data is penetrating to financial services industry via credit bureaus and fintechs, who are using big data in their algorithms.[2]

Big data is initially defined as information assets with "3Vs": "high-volume, high velocity and high-variety", which stress the importance of (i) the volume of the data, (ii) the speed it is collected, stored, analyzed and (iii) the diversity of data sources, in generating insights for better decision making. This definition is then extended to include two additional V's (i) veracity, referring to the quality of the data and (ii) value referring to the usefulness of the data (Laney, 2001; Frizzo-Barker et al. 2016). At the age of big data "relevant data", once defined mainly as the payment history of borrowers, is now extended to include data from social networks (Wei et al., 2016; Ge et al., 2017) as well as data from mobile phones and digital footprints of users from smart apps (Jenkins, 2014; Dwoskin, 2015; Lohr, 2015).[3] In fact, Wei, Yildirim, Van den Bulte, & Dellarocas (2016) and Kshetri (2016) show that big data enables creditworthiness assessment of potential borrowers with limited financial history

1 For example FICO scores of Fair Isaac Corporation uses consumers' debt level, length of credit history, and regular and on-time payments in assessing creditworthiness of borrowers (Wei et al., 2016)

2 For a discussion of Fintech industry, please refer to Lee and Shin (2017).

3 "Digital footprint refers to one's unique set of traceable digital activities, actions, contributions and communications that are manifested on the Internet or on digital devices." Source: Wikipedia

and thereby increases access to financial services particularly for low-income borrowers and micro-enterprises.

Yet, usage of big data and associated algorithms raise concerns on the enforcement and adequacy of regulations that aim to prevent discriminatory scoring, to protect consumers' rights to question their scores and consumers' privacy via regulations such as US Fair Credit Reporting Act, Equal Credit Opportunity Act, Fair and Accurate Credit Transactions Act of 2003 and Privacy Guidelines of OECD (Campbell-Verduyn et al., 2017). These algorithms are criticized for being "black boxes" due to their opacity, for producing arbitrary results and for furthering discrimination (Citron and Pasquale, 2014). Big data also poses challenges to privacy and security of personal information as revealed by the recent Equifax data breach, where approximately 143 million Americans personal data was stolen by hackers. In a recent statement Senator Mark Warner, the Senate Cybersecurity Caucus co-founder, called the breach "a real threat to the economic security of Americans." and mentioned the need to "rethink data protection policies." (Mathews, 2017).

The interplay between finance, technology and regulation is not new. The development of Information and Communication Technologies (ICT) has contributed to financial innovation and globalization of financial services, accompanied with deregulations and re-regulations over time (Cerny, 1994). In fact, Perez (2009, 2013) discuss that technology revolutions create major technology bubbles during the transition to the new paradigm. However, once the bubble collapses a golden age could be unleashed if the financial system is restructured accordingly and institutional governance and regulations are adequately developed. Big data is now revolutionizing how financial services, particularly credit scoring, are created and delivered. The very actors harnessing these new credit scoring technologies that use big data are banks, credit bureaus, fintech companies, and other non-bank financial service providers such as telecom companies. While big data may enable these actors to develop more accurate algorithms to assess creditworthiness, predict failure and develop tailored pricing and products/services, it at the same time brings challenges regarding data privacy and security as in the example of Equifax. However, there is a research-practice gap as the academic research in this field is scarce. Accordingly, our study is motivated by the on-going developments regarding new data sources, technologies and regulations in the credit scoring field.

Our objective is to gain a better understanding of the main themes of credit scoring as they relate to technological change and associated regulations over time. Accordingly, we conducted a content analysis of "credit scoring" across Proquest and Emerald research databases over the past 41 years (1976-2017). Content analysis is a systematic review of literature to make valid inferences about texts for knowledge building (Weber, 1990; Finfgeld-Connett, 2014). Accordingly, we reviewed 258 articles that appeared in peer-reviewed 147 different academic journals from 1976 to 2017, ranging from law journals to financial services, computer engineering and operations research journals. Our main research questions are: What are the main research themes in credit scoring literature? What is the direction and progression of credit scoring themes over time? What is the relative proportion of application, behavior and other scoring types? What types of statistical techniques and models are used in credit scoring? Which countries are represented by the research? In which journals these researches appear most frequently?

We find that credit scoring literature has evolved around 6 main themes, which are (i) statistical techniques and classification accuracy, (ii) new determinants in credit scoring, (iii) credit scoring technology adoption, (iv) review of credit scoring, (v) regulation in credit scoring and (vi) credit scoring in a different domain. The most studied themes are "statistical techniques and classification accuracy" and "new determinants in credit scoring" triggered by the development of algorithms over years, advancements in computational capabilities and new data sources, particularly recently big data. While research on these themes continues to increase at a significant pace, we also observe an increase in regulation studies that elaborate on the protection of consumer rights, data privacy and security given the underlying changes big data invoke in how data is collected, stored and used in scoring. The rule of thumb model is still the logistic regression in single model studies, while more advanced machine learning models such as SVM, neural networks, genetic algorithms and heterogeneous assemblers are studied as either hybrid models or in benchmark models. While majority of the studies develop application scoring mainly using samples from developed economies, a new stream of research focuses on usage of big data for increasing access to credit in developing and emerging economies. Finally, credit scoring research appears majorly in operations research journals followed by finance journals.

Leyshon and Thrift (1999) discuss that development of information technologies has transformed lending industry via the advent of credit scoring. Our research shows that lending industry is going through another transformation. This time credit scoring methodologies are evolving and data sources are changing led with the advent of big data. However, these non-traditional data sources, data-centric underwriting approaches and their regulatory aspects are addressed by only a handful of studies suggesting that they are the up and coming avenues for further research. This paper is organized as follows: section 2 presents the research methodology, section 3 gives the main findings and section 4 discusses and concludes.

2. Research methodology

2.1 Content Analysis

Academic journals are mediums of information dissemination and sharing between academia and business world. In this context investigating the published research articles on a given topic, one can analyze the trends and changes in that area and detect further research gaps.

Accordingly, this paper uses content analysis methodology to study the evolution of credit scoring by examining the articles published in academic peer-reviewed journals over a 41year period.

"Content analysis is a research method that uses a set of procedures to make valid inferences from the text" in an objective and systematic way (Weber, 1990). The content analysis starts with the identification of the texts and selection of the sample. In the second stage the unit of analysis is specified, after which theme categories and category schemes are determined by coders in the third stage. In the fourth stage, final categories are selected by judges, who code theme of each article according to category schemes. Finally, reliability analysis is conducted to assess the agreement level of the judges

Selecting Sample

To review the direction and progression of the credit scoring over time, this study searched for peer-reviewed articles in academic journals on ProQuest and Emerald Research Databases. Both databases provide an extensive coverage of academic journals from different disciplines including business, computer engineering and law journals (Table 1). The databases are searched for published, peer-reviewed, English-language research articles from 1976 to 2017, which included "credit scoring", "financial scoring", "consumer scoring" or "digital scoring" in its title or its abstract.

<Table 1 >

According to these aforementioned criteria, 299 articles without duplicates are obtained from ProQuest Research Database and 36 articles without duplicates are obtained from Emerald Research Database. After dropping 24 of the articles that are in both databases and 311 articles are reviewed for content analysis. 35 articles, which are editorial comments, book reviews, discussions, critique articles and articles that are not related with credit scoring, and 18 articles, whose full-text couldn't be reached via library facilities are excluded from the sample. Accordingly, our final dataset included 258 articles collected from 147 different academic journals published over the past 41 years.

2.2 Specifying the Unit of Analysis

One of the most fundamental and important decisions in content analysis is the choice of the basic unit of analysis for the classification to have semantic validity (Weber, 1990). According to Weber (1990) some commonly used units of analysis for content analysis are "word", "word sense", "sentence", "theme", "paragraph" and "whole text". In this study, the "theme" of articles is chosen as the main unit of analysis. Accordingly, titles and abstracts of all articles were reviewed to determine what was studied in them; the "theme" of the articles. They were categorized in order to identify main research themes and analyze their development over years, across journals/disciplines and with respect to scorecard type, statistical techniques used, and data sources.

2.3 Determining the Category Scheme

Categorization of specific unit of analyses like research methodology, data source is often relatively straightforward, on the contrary; a more interpretive approach needs to be taken during the process of categorizing the purpose of the study. The researchers reviewed the articles and identified 15 research themes from 258 articles. Then, two coders, both PhD students at the Department of Management Information Systems, were trained to analyze these themes according to their scope. Both coders independently classified these themes into broader categories and came up with respectively three and five categories. As a result, eight categories were collected from the coders.

2.4 Selection of Final Categories

In this stage, researchers evaluated all the categories in the previous stage and developed mutually exclusive and exhaustive final list of six main themes which are listed in Table 2. The description of the themes is as follows:

1) "*Statistical techniques and classification accuracy*" theme contains articles which evaluate different statistical techniques to increase the accuracy of classification.
2) "*New determinants in credit scoring*" theme contains articles about assessment of creditworthiness with additional targets, and new variables.
3) "*Credit scoring technology adoption*" theme includes articles about antecedents and consequences of credit scoring technology adoption.
4) "*Review of credit scoring*" theme includes articles about credit scoring literature reviews.
5) "*Regulation in credit scoring*" theme includes articles about impact of regulation or policy changes on credit scoring.
6) "*Credit scoring in a different domain*" theme includes articles about implementing credit scoring approach into a different area.

< Table 2 >

2.5 Reliability Analysis

In this section, the aforementioned 15 themes are placed under six categories. Three judges with PhD degrees respectively from Marketing and Finance assigned each of 15 themes to the category that fits best to its content. Each article received one theme code, which represented the primary research theme of the article. Pairwise agreement between the judges is shown in Table 3.

< Table 3 >

According to Zimmer and Golden (1988), the probability by chance alone of two judges assigning themes to same category is calculated by the following formula:

$$\mathrm{P(k\ successes)} = \left\{\frac{N!}{k! \times (N-k)!}\right\} \times \{p^k \times (1-p)^{N-k}\}$$

Where N is number of themes, k is number of matches between the judges and p is the probability that two judges will assign a theme to the same category by chance.

This formula is applied for each pairwise agreement and the probabilities of 13 and 12 agreements due to chance are represented as follows:

Judges A and B

$$\mathrm{P}(12) = \left\{\frac{15!}{12! \times (15-12)!}\right\} \times \left\{\left(\frac{1}{6}\right)^{12} \times \left(1-\frac{1}{6}\right)^{15-12}\right\}$$

$$= 1.210 \times 10^{-7}$$

Judges A and C

$$\mathrm{P}(12) = \left\{\frac{15!}{12! \times (15-12)!}\right\} \times \left\{\left(\frac{1}{6}\right)^{12} \times \left(1-\frac{1}{6}\right)^{15-12}\right\}$$

$$= 1.210 \times 10^{-7}$$

Judges B and C

$$\mathrm{P}(13) = \left\{\frac{15!}{13! \times (15-13)!}\right\} \times \left\{\left(\frac{1}{6}\right)^{13} \times \left(1-\frac{1}{6}\right)^{15-13}\right\}$$

$$= 5.583 \times 10^{-9}$$

A z-score is calculated for the probability of obtaining 12 matches, the lowest match, using formula mentioned below.

$$z = \frac{k - E_k}{\sqrt{np(1-p)}} \text{ where } E_k \text{ is expected number of matches.}$$

Substituting values for judges B and C:

$$z = \frac{12 - (15 \times (1/6))}{\sqrt{15 \times (1/6) \times (5/6)}} = 6.5818$$

Since a z-score of 2.33 corresponds to an alpha of 0.01, the probability that 12 themes or more would be assigned to the same categories by chance is very low. Accordingly, all pairwise matches are significant at the 1% level.

However, as there are a few number of categories, the possibility for random agreement increases. To increase the robustness of this study, Cohen's κ is calculated following the approach of Grayson and Rust (2001).

$$\kappa = \frac{p_a - p_c}{1 - p_c}$$

where $p_a$is the proportion of agreed on judgments ($p_a = (n_{11} + n_{22} + \cdots + n_{77})/n_{++}$); p_c is the proportion of agreements one would expect by chance ($p_c = (e_{11} + e_{22} + \cdots + e_{77})/n_{++}$); and $e_{ii} = (n_{i+}/n_{++}) \times (n_{+i}/n_{++}) \times n_{++}$.

Given the 6 categories, 15 themes and 3 independent judges of this study, Cohen's κ is calculated as follows:

Judges A and B $$\text{Cohen's } \kappa = \frac{0.80 - 0.2622}{1 - 0.2622} = 0.7289$$

Judges A and C $$\text{Cohen's } \kappa = \frac{0.80 - 0.28}{1 - 0.28} = 0.7222$$

Judges B and C $$\text{Cohen's } \kappa = \frac{0.8667 - 0.28}{1 - 0.28} = 0.8148$$

While a Cohen's κ score of 1 indicates perfect agreement, we find substantial agreement between our pair of judges. Table 4 below shows how to interpret Kappa results.

< Table 4 >

3. Research findings

3.1 Publication Year

Credit scoring literature dates back to 1976. Due to 41 years of credit scoring research, we present our analysis in mainly four publication eras; pre 2000s, early 2000s from 2000 to 2004, late 2000s from 2005 to 2009 and post 2010. The distribution of 258 articles according to publication eras is presented in Table 4. Of the 258 articles, 23 were published before 2000 and 34 were published in early 2000s. There has been a significant increase in late 2000s, which includes the 2008-2009 crises. In this period number of credit scoring publications has increased almost twofold to 60 papers. Credit scoring research has continued its accelerated growth in the aftermath of crisis and reached to 141 articles in the post 2010 era, which also signifies the start of big data research in business studies (Frizzo-Barker et al., 2016).

< Table 5 >

3.2 Theme Analysis

The results of our content analysis reveal 6 main research themes for credit scoring literature. Figure 1 and Table 6 respectively show percentage distribution of these themes and their frequency. The two leading themes are "*Statistical techniques and classification accuracy*" (41%) and "*New determinants in credit scoring*" (29%) followed by "*Credit scoring technology adoption*" theme representing 14% of credit scoring literature. In the meantime, "*Review of credit scoring*" (7%), "*Regulation in credit scoring*" (5%) and "*Credit scoring in a different domain*" (4%) themes have attracted the least attention from academics over these years.

< Figure 1 >

< Table 6 >

3.2.1 *Statistical techniques and classification accuracy*

Statistical techniques and classification accuracy theme, which contains articles that evaluate different statistical techniques to increase the accuracy of classification, is the dominant theme in credit scoring research. While an accurate classification algorithm leads to higher profitability, bad credit can impact a lender in terms of loss in capital, lower revenues and increased losses leading to bankruptcy (Abdou and Pointon, 2011; Lessman et al., 2015). Management of credit risk has become more important especially in the aftermath of 2008 crisis, where banks use their own credit scoring models under Internal Ratings Based Approach (IRB) and face stricter capital requirements (Basel, 2009; Basel 2013a).

Credit scoring is essentially a type of a classification problem to determine whether or not the borrower will default on a loan (Mavri *et al.*, 2008). Hence, it is an assessment of the risk associated with lending to an organization or individual (Paleologo *et al.*, 2010). This assessment is made with a predictive model that is generated with repayment behaviors of previous borrowers on a loan, whose performances have been observed over a period of time (Thomas *et al.*, 2002). In simple terms, credit scoring assumes that past data is a good indicator of future performance of borrowers and the quality of the scorecards, depends on the accurately classification of a case as “good” credit if repayment on time is expected and as “bad” credit if repayment is expected to fail (Siddiqi, 2012).

However, Abdou and Pointon (2011) discuss that there is no overall best statistical technique that fits all. Yet state-of-the-art research has focused on (i) non-parametric and computational intelligence techniques like artificial neural networks, support vector machines, decision trees and genetic algorithms (Marques et al., 2013; Tsai and Hung, 2014; Lessman et al., 2015;

Ozturk et al., 2016), (ii)) dynamic programming approaches to find the optimal time to rebuild or readjust scorecards (Jung et al., 2015), (iii) dynamic incremental modeling (Sun et al., 2015) and model-dependent back testing approaches (Bravo and Maldonado, 2015) that deal with concept drift and (iv) scorecard performance measures (Rezáč, 2015; Řezáč and Řezáč, 2011;Oliver, 2013).

3.2.2. *New determinants in credit scoring*

"*New determinants in credit scoring*" theme contains articles about assessment of creditworthiness with additional targets, and new predictor variables. Most recent research under this theme has focused on (i) profit-based scoring systems that estimate profitability of loans rather than probability of default with measures like internal rate of return (Marron, 2007 (Marron, 2007; Finlay, 2008; Stewart, 2011; Verbraken *et al.*, 2014; Serrano-Cinca and Gutiérrez-Nieto, 2016), (ii) introduction of new target variables like "indeterminates" (Řezáč, 2013) or differentiating defaulters with game theory approach as "Can't Pays (borrowers who do not pay because of cash flow problems) and Won't Pays (borrowers that do not pay because of lack of willingness to pay)" (Bravo *et al.*, 2015) and (iii) new predictor variables such as inclusion of spatial risk factors as an indicator of local economy characteristics for SME lending (Fernandes and Artes, 2016) or adding borrowers' psychological traits for micro lending (Baklouti, 2014). State-of-the-art research under this theme has investigated the impact of big data in credit scoring via analysis of P2P lending platforms and credit bureaus that make use of social network data to improve their credit models (Wei et al., 2016; Ge et al., 2017).

3.2.3 Credit scoring technology adoption

"*Credit scoring technology adoption*" theme includes articles about antecedents and consequences of credit scoring technology adoption. A recent stream of research in this theme has focused on (i) the impact of credit scoring on access to finance and financial inclusion particularly for SMEs and microfinance institutions (De Young et al, 2011; Bumacov, Ashta and Singh, 2014) (ii) on adverse selection and debt monitoring (Ono et al, 2014; Hasumi and Hirata, 2014; Micucci and Rossi, 2017) and (iii) the impact of credit scoring adoption on risk management during crisis (Feess and Hege, 2012; Demma, 2017).

3.2.4. Review of credit scoring

"*Review of credit scoring*" theme includes articles about credit scoring literature reviews. Recent surveys of the literature have focused on (i) different types of data mining methods given the complexity and volume of data that needs to be analyzed (Hooman et al., 2016), (ii) review of evolutionary computing methods in credit scoring (Marques et al., 2013), (iii) use of experimental design in credit scoring model assessment and comparison (García et al., 2015), (iv) evolution of credit scoring as related to its usage in microfinance (Bumacov et al., 2017) and (v) usage of big data in credit scoring in relation to financial inclusion of unbanked segments (Aitken, 2017).

3.2.5. Regulation in credit scoring

"Regulation in credit scoring" theme includes articles about impact of regulation or policy changes on credit scoring. A main topic of interest has been on preventing discrimination in credit scoring. A recent stream of literature has studied (i) the effectiveness of anti-discrimination laws on preventing inclusion of the variables that represent protected features in credit models (Chan and Seow, 2013), (ii) developing regulatory oversight for fairness and accuracy of artificially intelligent scoring systems (Citron and Pasquale, 2014), and (iii) the impact of big data and associated algorithms on credit discrimination (Zarsky, 2014), governance (Campbell-Verduyn et al., 2017) and data privacy (Roderick, 2014).

3.2.6. Credit scoring in a different domain

"*Credit scoring in a different domain*" category includes articles about implementing credit scoring approach into a different area. The recent research on this theme has investigated (i) the relationship between cardiovascular disease and credit scores (Israel et al., 2014), (ii) the relationship between energy-efficient house adoption and credit scores (Sanderford et al., 2015), (iii) development of a compliance framework to assess money laundering and terrorism financing risk (Sathye and Islam, 2011), and (iv) modeling container stacking and reshuffling operations (Gharehgozli et al., 2017).

3.3 Direction and Progression of Themes over time

To get a better understanding of the direction and progression of the credit scoring literature over time, themes of articles over years is analyzed in detail in Table 7, which present the number of articles per theme per publication era. We discuss findings according to Figure 2, where we calculate percentage distribution of themes in terms of eras.

In pre2000s era "Statistical techniques and classification accuracy" appears as the most dominant theme representing 35% of the credit scoring research. In this period "New Determinants in credit scoring" (22%) was the second most important theme followed by "Credit scoring technology adoption" and "Regulation in credit scoring" themes each representing 13% of research articles published in this period. This period dates back to 1970s and represents the initial development stage of credit scoring methods coinciding with the introduction of FICO scores by Fair Isaac Company and adoption of these scores by

Fannie Mae and Freddie Mac.[4] The extent and nature of information collected and consumers' rights to access were regulated mainly by US Fair Credit Reporting Act of 1970 and 1980 Privacy Guidelines developed by the Organisation for Economic Co-operation and Development (Campbell-Verduyn et al., 2017). "Review of credit scoring" and "credit scoring in a different domain" themes individually represented 9 percent of the research in this era.

In the early 2000s, "*Statistical techniques and classification accuracy*" and "*New Determinants in credit scoring*" themes continued to be the two leading themes, while popularity of "*Credit scoring technology adoption*" theme increased slightly. This was a period when new variables were introduced to credit scorecards and the impact of adoption on the performance of the companies was gaining attention. In this era while research on "*Review of credit scoring*" increased, "*Regulation in credit scoring*" and "*Credit scoring in a different domain*" themes have been relatively neglected.

In the late 2000s era, research on "Statistical techniques and classification accuracy" and "Credit scoring technology adoption" themes accelerated respectively twofold and threefold to 42% and 23% of the total publications. This era covers the 2008-09 crisis period, during which Basel Committee of Banking Supervision addresses the need for better governance of credit risk of banks and releases enhanced standards of Basel II and starts working towards Basel III.[5] Accordingly increases in these themes signifies the need of companies to manage their credit risks better particularly in the credit crisis periods. "Regulation in credit scoring" theme also became popular by researchers in this period following the release of Fair and Accurate Credit Transactions Act of 2003, while research on "New Determinants in credit scoring", "Review of credit scoring", and "Credit scoring in a different domain" themes experienced a relative decline.

[4] Source: http://www.wikizero.org/index.php?q=aHR0cHM6Ly9lbi53aWtpcGVkaWEub3JnL3dpa2kvRklDTw

[5] Please refer to https://www.bis.org/bcbs/basel3.htm for details of Basel III.

As of 2010 "*New determinants in credit scoring*" theme enjoyed more than threefold increase reaching 32% of publications while "*Statistical techniques and classification accuracy*" maintained its dominancy with 42%. This was a period during which big data and non-traditional data sources are being introduced into the credit scoring algorithms, while regulations were also revised to incorporate risk management challenges big data brings as Basel addresses the IT risk and governance with big data (Mitchell, 2013; Campbell-Verduyn et al., 2017). Accordingly, "*Regulation in credit scoring*" theme increased fourfold to 6% of publications in this period. Having left behind more than 30 years of research we also see an increase in the number of articles under the "*Review of credit scoring*" theme qualitatively assessing the evolution of the field particularly on data mining and evolutionary computing, while "*Credit scoring in a different domain*" theme also increased to 5% of total publications. On the contrary, "*Credit scoring technology adoption*" theme saturated to 9% of publications, as the benefits of credit scoring adoption is understood over the past years.

< Table 7 >

< Figure 2>

3.4 Details of Scoring Type

In this section, we analyze the research on three main types of credit scoring, which are application, behavioral and collection scoring (Paleologo *et al.*, 2010). In application scoring, the subject is a new loan; in behavior scoring the subject is an existing loan and in collection scoring it is a delinquent loan. Essentially, dynamics of these three types of credit scoring differ in terms of performance and sample windows, target definitions, exclusion criteria and observed characteristics.

Our research shows that majority of the research has paid attention to application scoring, while behavioral and collection scoring are neglected although they both play great importance in credit risk management as borrower's financial status may change after the initial credit assessment (Sohn and Kim,2013). Furthermore, offering a high credit limit to an existing borrower who has problems with his/her existing loans would also hamper the lender. Yet, behavior (11 articles) and collection (1 article) scoring have attracted very little interest and only three articles have studied both application and behavior scoring. On the other hand, five articles have studied credit bureau scoring, where lenders use credit bureau data, while 40 articles analyzed credit scoring in general without specifying scoring type. Frequency of scoring types is shown on Figure 3.

The distribution and representativeness of the data continues to be a main consideration in developing scorecards. In application and behavior scoring, the data collected is highly unbalanced or skewed and research on scoring imbalanced data is scarce (Siddiqi, 2012; Wang *et al.,* 2015). In this regard, Wang *et al.* (2015) suggest that research that focus on scoring imbalanced data can help improve representativeness of the credit scoring datasets. Application, behavior and collection scoring models also suffer from sample selection bias as they are developed from only granted loans. However, developing a scorecard that ignore "rejects" is not applicable to the total population and forecasting for all applicants will not be accurate and realistic (Siddiqi, 2012). While a stream of literature studies reject inference techniques that address this problem (Jacobson and Roszbach, 2003; Crook and Banasik, 2004; Bücker *et al.*, 2013), including certain types of loans that should not be in the development sample also deteriorates the model especially for behavior scorecards. Including loans for special customers such as staff and VIPs or lost/stolen cards, deceased customers, restructured payment plans in the observed dataset will change the true characteristics that predict the target (Siddiqi, 2012).

< Figure 3>

3.5 Details of Statistical Techniques

Another dimension we analyze in detail is the statistical techniques used in the literature. Studies that include a scorecard development consist of 173 articles and the details are given in Figure 4 and Table 8. Studies, which implemented only one statistical technique are listed as "Single Model" articles, studies that combine more than one statistical approach are grouped as "Hybrid Model" and studies which compare different statistical techniques for the same data set are categorized as "Benchmark Models". Overall majority of the studies focused on "Single Models". While benchmark models were prevalent in the pre2010 period, Single Model studies have been dominant with 48 articles out of 106 articles published in the post 2010 period. Logistic regression appears as the rule of thumb approach to credit scoring in Single model studies, followed by probit regression over years. Logistic regression also emerges as the industry standard for benchmark models to enhance the reliability of the findings. We also see a more than threefold increase in articles that use "Hybrid Models" (24 articles), which reaches almost equal footing with logistic regression (26 articles) as of 2010s.

< Table 8>

< Figure 4>

3.6 Details of Scoring Datasets

Another important aspect of credit scoring is the source and size of the datasets used in the scorecard development with regard to the generalizability of the results (Abdou and Pointon, 2009; Crone and Finlay, 2012; Marques *et al.*, 2013a; García *et al.*, 2015; Trinkle and Baldwin, 2016). The main requirement of a good model construction is to select a sample, which is random and representative of the population. Yet, most of the studies compared different classification methods and improved the accuracy with well-known German, Australian and Japanese data sets. However, they are not representative enough of real-life credit scoring applications (Marques *et al.*, 2013a) where larger sample sizes are needed (Crone and Finlay, 2012).

We analyze under two main categories; (i) researches that use real credit datasets from developed versus developing and emerging economies and (ii) research that rely on publicly available datasets on the internet. Out of the 258 articles 180 articles mentioned the source of their dataset and in Table 9 we present frequency of dataset sources over publication eras. The majority of the credit-scoring research has been on developed economies, which is the dominant source over all publication eras. However, in the post-2010 era we observe a significant increase on research that has paid attention to developing and particularly emerging economies. Majority of the research that uses emerging economies data have focused on "New Determinants in credit scoring" theme, particularly on credit scoring for micro-finance institutions and SMEs. Similarly, usage of Internet available datasets has also increased in the post-2010 period. Out of 101 articles published in post-2010 era 45 articles used developed country datasets, 25 articles used developing and emerging economy datasets, 27 articles used internet available datasets and 4 articles used cross-country datasets.

<Table 9>

3.7 Journal Analysis

The results of the journal analysis are presented in Figure 5, which shows the journal names that published more than two articles about credit scoring over the years. Journals that published less than two articles is grouped as "Other". Operations research journals lead the credit scoring research followed by finance journals. "Journal of the Operational Research Society" and "European Journal of Operational Research" have published respectively 27 and 21 articles over years and hence have the highest share among 258 articles. Journal of

Banking and Finance take the third place with 9 publications and Intelligent Systems in Accounting and Finance take the fourth place with 6 publications.

<Figure 5>

4. Discussion and conclusion

In this research, we perform a content analysis of credit scoring literature in the past 41 years and identify the six main research themes in academic research. We show that "*Statistical techniques and classification accuracy*" theme and "*New determinants in credit scoring*" theme are the two most prevalent research avenues. We find that the trending research topics on these themes have focused on the impact of big data via (i) development of classification algorithms that rely on machine learning and artificially intelligent scoring systems and (ii) introduction of non-traditional data sources into algorithms. "*Regulation in credit scoring*" theme has regained popularity with the need to redefine data privacy guidelines and to ensure enforcement of anti-discrimination laws in the age of big data. The research on "*Credit scoring technology adoption*" theme has relatively matured, while we find an increase in the research on "*Credit scoring in a different domain*" theme and "*Review of credit scoring*" theme that survey data mining and evolutionary computing methods and usage of big data for increasing financial inclusion.

Credit scoring game is changing for the traditional lender. Credit scoring is moving from traditional data sources to non-traditional data sources particularly via big data. Social media activities, telecom and utility bills, and psychometrics are becoming the new sources to identify the behavioral patterns and creditworthiness of the borrowers. Large datasets are not new for the financial services companies. What is new is the digitized sources of data that were once incompatible and inefficient to analyze with limited computing power, and the governance of these artificially intelligent algorithms. Big data in this sense refers to new data sources from which insights can be derived for creating innovative products and solutions as well as the technology that makes it possible. Non-traditional data such as digital footprints and activities of users on online shopping, gaming or social networks such as Twitter, Linkedin, Google, Facebook are used by credit bureaus and fintech companies to predict creditworthiness of borrowers' (Roderick, 2014; CFSI, 2015; Baer *et al.*, 2013; PWC, 2015; Wei et al., 2016). Fintech start-ups are relying on machine learning to leverage big data capabilities to predict creditworthiness and market customized products by using data from non-traditional credit information.[6] With these new data sources borrowers, who either do not have a sufficient financial history to get a credit score or deemed too risky to lend, may become creditworthy as their behavior becomes more predictable by constant monitoring. Gabor and Brooks (2017) discuss that this digital revolution, based on feeding digital footprints to algorithms, may accelerate access to finance particularly for the unbanked, who are excluded from the financial system in developing countries. Kshetri (2016) show that in fact big data enables access to finance for low-income families and micro-finance enterprises in China by (i) reducing transaction costs via "digitizing the activities and/or minimizing the physical intervention between the borrower and the lender", and (ii) reducing the information opacity regarding the identity, ability to pay and willingness to pay of the borrower by incorporating non-financial information such as data from government agencies, hobbies, time of day the person is shopping online and type of items purchased into their credit scoring algorithms. Furthermore, Chinese banks are launching their own e-commerce sites to access retail transaction data to incorporate into their own scoring models in response to the competition fintechs bring. More recently, Lenddo, a P2P lending platform that uses social network data, and Entrepreneurial Finance Lab, which was a research project in Harvard

[6] https://www.ft.com/content/7933792e-a2e6-11e4-9c06-00144feab7de

University to increase financial inclusion using psychometric data, have merged into "LenddoEFL" with an objective of increasing financial inclusion by reaching 1 billion people all over the world.[7] Considering that there are approximately 2.5 billion unbanked people in the world (Demirguc-Kunt and Klapper, 2012), financial inclusion is one of the most promising benefits of big data in credit scoring as it may make "opaque" borrowers more transparent. Nevertheless, Gabor and Brooks (2017) stress that this information technology based financial inclusion projects commodify the personal data of "newly included" yet "risky" populations. This "digital legibility" to access finance depends on digital footprints used in proprietary "black-box" algorithms, whose workings is unknown (Pasquale, 2015) and borrowers' privacy rights are on hold.

While the utopian view of big data discusses that it may enhance the well-being of both consumers and companies via increased efficiency and tailored solutions, there remains the negative effects of big data as the opacity of algorithms that use it, their power to increase inequalities and discrimination and to hamper data privacy via constant data "surveillance" by industry players and governments (Kshetri, 2014; Wang and Yu, 2015; Campbell-Verduyn et al., 2017). How big data is collected and used creates data privacy and ethical infringement risks (Frizzo-Barker et al., 2016). As algorithms that rely on big data analytics may reveal sensitive and personally identifiable information (PII) (Kshetri, 2014, 2016; King and Forder, 2016), data privacy and regulations to govern it has become one of the most important upcoming issues in credit scoring. Hurley and Adebayo (2016) highlight the risks of these alternative credit-scoring approaches as data accuracy, transparency, unfairness and discrimination.[8] However, Citron and Pasquale (2014) discuss that laws do not sufficiently protect how scores are mined from big data and suggest that artificially intelligent scoring

[7] https://include1billion.com

[8] According to Equal Credit Opportunity Act (1975), information including race, ethnicity, national origin, gender, marital status and receipt of public assistance cannot be used as a variable in credit granting decisions.

systems need to be closely regulated to prevent stigmatization of borrowers so that they provide fair and unbiased assessments of creditworthiness. Accordingly, more academic research is needed to identify which pieces of information is in fact useful and are in compliance with regulations by analyzing the components of credit scores big data analytics produce. In other words, make algorithms more transparent.

Big data may also cause significant losses due to possible security breaches as high data volume may attract cybercriminals, as high variety of data may reveal more PII and make security breach detection harder, as organizations may lack adequate data storage and management skills to respond to high variability of data flows, and as complexity of data may make re-identification possible (Kshetri, 2014). These privacy and security concerns regarding how big data is collected, stored and used in credit scoring by industry players needs to be addressed by academic research with respect to their compliance with existing international regulations as well as in design of new regulations to address new privacy concerns that come with big data. In this regard, the General Data Protection Regulation (GDPR) of EU, which will come to effect as of May 2018 addresses several key issues: (i) "right to be forgotten", which requires companies to justify storing personal data, (ii) "right to access", which enables customers to learn how their personal data is used, (iii) "breach notification", which requires a notification within 72 hours of the breach, (iv) "data portability", where customers can share their data with other service providers and (v) "privacy by design", where compliance to data protection directive becomes a legal requirement in the design of the systems.[9] Furthermore, Revised Payment Service Directive (PSD2) of EU enables customers to share their account and financial payment data with other third-party service providers, through open Application Programming Interface (API) of

[9] https://www.eugdpr.org/key-changes.html

banks.[10] These regulatory developments bring challenges to banks and credit bureaus, who used to be main owners of customers' accounts and financial payment data, as well as bringing new opportunities to collaborate with fintechs, particularly alternative lending platforms.[11] Accordingly, the means companies develop credit scoring systems using big data has to be investigated thoroughly given this new paradigm shift and more research is needed to develop more effective scorecards and more transparent algorithms that are in compliance with anti-discrimination laws and privacy rights on use of personal data (Ferretti, 2006; Chan and Seow, 2013; CFSI, 2015).

Amoore and Piotukh (2015) discuss that big data analytics and algorithms are changing how we ingest data, how we partition it and finally how we act upon it through real-time analytics. They discuss that while machine learning algorithms using unstructured data are "instruments of perception" and may reveal hidden insight, at the same time they may be indifferent to the heterogeneity of the underlying data in pursuit of extracting "what is of interest"; people, objects or patterns. Accordingly, governance systems regarding social and economic life must fully comprehend the workings of advanced analytics and algorithms behind big data. Announcement of the Social Credit System of China's State Council[12], which would come into effect as of 2020 and where every citizen would have a credit score that relies on financial information, criminal records, social media behavior as well as political opinions, shows how big data may be used to design and govern a social system. Furthermore, Chinese online financial services regulator has recently announced establishment of "Internet Finance Industry Credit Information Sharing Platform" as a joint platform with industry players in an

[10] https://ec.europa.eu/info/law/payment-services-psd-2-directive-eu-2015-2366/law-details_en

[11] The Future of Financial Services, World Economic Forum Report, 2015

[12] https://www.wsj.com/articles/chinas-new-tool-for-social-control-a-credit-rating-for-everything-1480351590
http://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion

effort to regulate the internet finance industry in China.[13] These developments extend the definition of credit scoring from a mere credit rating concept to a citizen rating concept, where new metrics include social, political, and environmental factors. The repercussions of these new technologies and scorecards on consumer privacy, whether or not these are enabling or further constraining access to finance or what should be the involvement and role of governments needs to be addressed by academic research.

Going back to Leyshon and Thrift (1999) paper another "quantitative revolution" is taking place in credit scoring via big data, (i) where we still question new boundaries of inclusion and exclusion with non-traditional data sources given the opacity of algorithms, (ii) where organizational architectures are changing as we observe fintech startups forcing banks to change their business models[14], and (iii) where competition is still dependent on data and software but this time data is big and algorithms are self-learning. The most important role academic studies play is bridging the gap between corporate, governmental, and non-profit actors, who are developing credit scoring systems, by supporting them assess the perils and benefits of new technologies that paves way for innovation.

[13] https://www.reuters.com/article/us-china-regulations-loans/china-online-finance-regulator-launches-credit-rating-platform-china-daily-idUSKBN1DS02L
http://www.nifa.org.cn/nifaen/2955875/2955895/2964303/index.html

[14] Please refer to Kshetri (2016) for a discussion of how banks are shifting their business to online and offering new e-commerce platforms to access retail transaction data to incorporate into their credit scoring.

References

Abdou, H. A., & Pointon, J. (2009). Credit scoring and decision making in Egyptian public sector banks. *International journal of managerial finance, 5*(4), 391-406.

Abdou, H. A., & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: A review of the literature. *Intelligent Systems in Accounting, Finance and Management, 18*(2-3), 59-88.

Aitken, R. (2017). 'All data is credit data': Constituting the unbanked. *Competition & Change, 21*(4), 274-300.

Amoore, L., & Piotukh, V. (2015). Life beyond big data: Governing with little analytics. *Economy and Society, 44*(3), 341-366.

Bücker, M., Van Kampen, M., & Krämer, W. (2013). Reject inference in consumer credit scoring with nonignorable missing data. *Journal of Banking & Finance, 37*(3), 1040-1045.

Baer, T., Goland, T., & Schiff, R. (2013, April). *http://www.mckinsey.com.* Retrieved April 10, 2017, from New credit-risk models for the unbanked: http://www.mckinsey.com/business-functions/risk/our-insights/new-credit-risk-models-for-the-unbanked

Baklouti, I. (2014). A psychological approach to microfinance credit scoring via a classification and regression tree. *Intelligent Systems in Accounting, Finance and Management, 21*(4), 193-208.

Basel Committee on Banking Supervision. (2009, July). *Enhancements to the Basel II framework.* Retrieved October 19, 2017, from https://www.bis.org/publ/bcbs157.pdf

Basel Committee on Banking Supervision. (2013a). *Principles for Effective Risk Data Aggregation and Risk Reporting, January.* Retrieved November 19, 2017, from https://www.bis.org/publ/bcbs239.pdf

Basel Committee on Banking Supervision. (2013b). *Progress in Adopting the Principles for Effective Risk Data Aggregation and Risk Reporting, December.* Retrieved November 19, 2017, from https://www.bis.org/publ/bcbs268.pdf

Bravo, C., & Maldonado, S. (2015). Fieller Stability Measure: a novel model-dependent backtesting approach. *Journal of the Operational Research Society, 66*(11), 1895-1905.

Bravo, C., Thomas, L. C., & Weber, R. (2015). Improving credit scoring by differentiating defaulter behaviour. *Journal of the Operational Research Society, 66*(5), 771-781.

Brill, J. (2012, March 2). *Big Data, Big Issues.* Retrieved 10 01, 2017, from www.ftc.gov: https://www.ftc.gov/public-statements/2012/03/big-data-big-issues

Bukovina, J. (2016). Social media big data and capital markets—An overview. *Journal of Behavioral and Experimental Finance, 11*, 18-26.

Bumacov, V., Ashta, A., & Singh, P. (2014). The use of credit scoring in microfinance institutions and their outreach. *Strategic Change, 23*(7-8), 401-413.

Bumacov, V., Ashta, A., & Singh, P. (2017). Credit scoring: A historic recurrence in microfinance. *Strategic Change, 26*(6), 543-554.

Campbell-Verduyn, M., Goguen, M., & Porter, T. (2017). Big Data and algorithmic governance: the case of financial practices. *New Political Economy, 22*(2), 219-236.

Cerny, P. G. (1994). The dynamics of financial globalization: Technology, market structure, and policy response. *Policy Sciences, 27*(4), 319-342.

CFSI. (2015). *Big Data, Big Potential: Harnessing Data Technology for the Underserved Market.* Center for Financial Services Innovation Report.

Chan, L. W., & Seow, H. V. (2013). Legally scored. *Journal of Financial Regulation and Compliance, 21*(1), 39-50.

Citron, D. K., & Pasquale, F. A. (2014). The scored society: due process for automated predictions. *Washington Law Review, 89*(1).

Crone, S. F., & Finlay, S. (2012). Instance sampling in credit scoring: An empirical study of sample size and balancing. *International Journal of Forecasting, 28*(1), 224-238.

Crook, J., & Banasik, J. (2004). Does reject inference really improve the performance of application scoring models? *Journal of Banking & Finance, 28*(4), 857-874.

Demirguc-Kunt, A., & Klapper, L. F. (2012). Measuring financial inclusion: The global findex database.

Demma, C. (2017). Credit Scoring and the Quality of Business Credit During the Crisis. *Economic Notes, 46*(2), 269-306.

DeYoung, R., Frame, W. S., Glennon, D., & Nigro, P. (2011). The information revolution and small business lending: The missing evidence. *Journal of Financial Services Research, 39*(1-2), 19-33.

Dwoskin, E. (2015, November 30). Lending Startups Look at Borrowers' Phone Usage to Assess Creditworthiness. *The Wall Street Journal*.

Equal Credit Opportunity Act (1975), U.S.C., Title 15, Sec. 1691 et. seq.

Fair and Accurate Credit Transactions Act, Pub. L. 108-159, 117 Stat. 1952 (2003).

Feess, E., & Hege, U. (2012). The Basel Accord and the value of bank differentiation. *Review of Finance, 16*(4), 1043-1092.

Fernandes, G. B., & Artes, R. (2016). Spatial dependence in credit risk and its improvement in credit scoring. *European Journal of Operational Research, 249*(2), 517-524.

Ferretti, F. F. (2006). Re-thinking the regulatory environment of credit reporting: Could legislation stem privacy and discrimination concerns? *Journal of Financial Regulation and Compliance, 14*(3), 254-272.

Finfgeld-Connett, D. (2014). Use of content analysis to conduct knowledge-building and theory-generating qualitative systematic reviews. *Qualitative Research, 14*(3), 341-352.

Finlay, S. M. (2008). Towards profitability: A utility approach to the credit scoring problem. *Journal of the Operational Research Society, 59*(7), 921-931.

Frizzo-Barker, J., Chow-White, P. A., Mozafari, M., & Ha, D. (2016). An empirical study of the rise of big data in business scholarship. *International Journal of Information Management, 36*(3), 403-413.

Gabor, D., & Brooks, S. (2017). The digital revolution in financial inclusion: international development in the fintech era. *New Political Economy, 22*(4), 423-436.

García, V., Marqués, A. I., & Sánchez, J. S. (2015). An insight into the experimental design for credit risk and corporate bankruptcy prediction systems. *Journal of Intelligent Information Systems, 44*(1), 159-189.

Ge, R., Feng, J., Gu, B., & Zhang, P. (2017). Predicting and Deterring Default with Social Media Information in Peer-to-Peer Lending. *Journal of Management Information Systems , 34*(2), 401–424.

Gharehgozli, A., Mileski, J. P., & Duru, O. (2017). Heuristic estimation of container stacking and reshuffling operations under the containership delay factor and mega-ship challenge. *Maritime Policy & Management, 44*(3), 373-391.

Grayson, K., & Rust, R. (2001). Interrater reliability. *Journal of Consumer Psychology, 10*(1), 71-73.

Hasumi, R., & Hirata, H. (2014). Small business credit scoring and its pitfalls: Evidence from Japan. *Journal of Small Business Management, 52*(3), 555-568.

Hooman, A., Marthandan, G., Yusoff, W. F., Omid, M., & Karamizadeh, S. (2016). Statistical and data mining methods in credit scoring. *The Journal of Developing Areas, 50*(5), 371-381.

Hurley, M., & Adebayo , J. (2016). Credit Scoring in the Era of Big Data. *Yale Journal of Law and Technology, 18*(1), 148.

Israel, S., Caspi, A., Belsky, D. W., Harrington, H., Hogan, S., Houts, R., . . . Moffitt, T. E. (2014). Credit scores, cardiovascular disease risk, and human capital. *Proceedings of the National Academy of Sciences, 111*(48), 17087-17092.

Jacobson, T., & Roszbach, K. (2003). Bank lending policy, credit scoring and value-at-risk. *Journal of Banking & Finance, 27*(4), 615-633.

Jenkins, P. (2014, June 23). Big data lends new zest to banks' credit judgments. *Financial Times*.

Jung, K. M., Thomas, L. C., & So, M. C. (2015). When to rebuild or when to adjust scorecards. *Journal of the Operational Research Society, 66*(10), 1656-1668.

King, N. J., & Forder, J. (2016). Data analytics and consumer profiling: finding appropriate privacy principles for discovered data. *Computer Law & Security Review, 32*(5), 696-714.

Kshetri, N. (2014). Big data′ s impact on privacy, security and consumer welfare. *Telecommunications Policy, 38*(11), 1134-1145.

Kshetri, N. (2016). Big data's role in expanding access to financial services in China. *International Journal of Information Management, 36*(3), 297-308.

Landis, J. R., & Koch , G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.

Laney, D. (2001, February 6). *3D data management: Controlling data volume, velocity and variety.* Retrieved 10 15, 2017, from www.blogs.gartner.com: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Lee, I., & Shin, Y. J. (2017). Fintech: Ecosystem, business models, investment decisions, and challenges. *Business Horizons*.

Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research, 247*(1), 124-136.

Leyshon, A., & Thrift, N. (1999). Lists come alive: eletronic systems of knowledge and the rise of credit-scoring in retail banking. *Economy and Society, 28*(3), 434-466.

Lohr, S. (2015, January 18). *Banking Start-Ups Adopt New Tools for Lending.* Retrieved October 8, 2017, from www.nytimes.com: https://www.nytimes.com/2015/01/19/technology/banking-start-ups-adopt-new-tools-for-lending.html?_r=0

Marques, A. I., García, V., & Sánchez , J. S. (2013). A literature review on the application of evolutionary computing to credit scoring. *Journal of the Operational Research Society, 64*(9), 1384-1399.

Marqués, A. I., García, V., & Sánchez, J. S. (2013). On the suitability of resampling techniques for the class imbalance problem in credit scoring. *Journal of the Operational Research Society, 64*(7), 1060-1070.

Marron, D. (2007). Lending by numbers': credit scoring and the constitution of risk within American consumer credit. *Economy and Society, 36*(1), 103-133.

Mathews, L. (2017, September 7). Retrieved from Forbes: https://www.forbes.com/sites/leemathews/2017/09/07/equifax-data-breach-impacts-143-million-americans/#25e2471f356f

Mavri, M., Angelis, V., Ioannou, G., Gaki, E., & Koufodontis, I. (2008). A two-stage dynamic credit scoring model, based on customers' profile and time horizon. *Journal of Financial Services Marketing, 13*(1), 17-27.

Medema, L., Koning, R. H., & Lensink, R. (2009). A practical approach to validating a PD model. *Journal of Banking & Finance, 33*(4), 701-708.

Micucci, G., & Rossi, P. (2017). Debt Restructuring and the Role of Banks' Organizational Structure and Lending Technologies. *Journal of Financial Services Research, 51*(3), 339-361.

Mitchell, R. (2013, September 16). *Revised OECD Privacy Guidelines Focus On Accountability, Notification of Breaches.* Retrieved September 1, 2017, from www.bna.com: https://www.bna.com/revised-oecd-privacy-n17179877087/

Nasir, V. A. (2004). E-consumer complaints about on-line stores. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior, 17*, 68-87.

Oliver, R. M. (2013). Financial performance measures in credit scoring. *EURO Journal on Decision Processes, 1*(3-4), 169-185.

Ono, A., Hasumi, R., & Hirata, H. (2014). Differentiated use of small business credit scoring by relationship lenders and transactional lenders: Evidence from firm–bank matched data in Japan. *Journal of Banking & Finance, 42*, 371-380.

Organisation for Economic Co-operation and Development (OECD). Recommendation of the council concerning guidelines governing the protection of privacy and transborder flows of personal data (September 1980).

Ozturk, H., Namli, E., & Erdal, H. I. (2016). Modelling sovereign credit ratings: The accuracy of models in a heterogeneous sample. *Economic Modelling, 54*, 469-478.

Paleologo, G., Elisseeff, A., & Antonini, G. (2010). Subagging for credit scoring models. *201*(2), 490-499.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information.* Harvard University Press.

Perez, C. (2009). The double bubble at the turn of the century: technological roots and structural implications. *Cambridge Journal of Economics, 33*(4), 779-805.

Perez, C. (2013). Unleashing a golden age after the financial collapse: Drawing lessons from history. *Environmental Innovation and Societal Transitions, 6*, 9-23.

Řezáč, M. (2013). Determining the Target Variable in Credit Scoring Models. *GSTF Journal of Mathematics, Statistics & Operations Research, 1*(1).

Řezáč, M. (2015). ESIS2: Information Value Estimator for Credit Scoring Models. *45*(2), 303-322.

Řezáč, M., & Řezáč, F. (2011). How to measure the quality of credit scoring models. *Finance a úvěr: Czech Journal of Economics and Finance, 61*(5), 486-507.

Roderick, L. (2014). Discipline and power in the digital age: The case of the US consumer data broker industry. *Critical Sociology, 40*(5), 729-746.

Sanderford, A. R., Overstreet, G. A., Beling, P. A., & Rajaratnam, K. (2015). Energy-efficient homes and mortgage risk: crossing the chasm at last? *Environment Systems and Decisions, 35*(1), 157-168.

Sathye, M., & Islam, J. (2011). Adopting a risk-based approach to AMLCTF compliance: the Australian case. *Journal of Financial Crime, 18*(2), 169-182.

Serrano-Cinca, C., & Gutiérrez-Nieto, B. (2016). The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending. *Decision Support Systems, 89*, 113-122.

Siddiqi, N. (2012). *Credit risk scorecards: developing and implementing intelligent credit scoring* (Vol. 3). John Wiley & Sons.

Sohn, S. Y., & Kim, Y. S. (2013). Behavioral credit scoring model for technology-based firms that considers uncertain financial ratios obtained from relationship banking. *Small Business Economics, 41*(4), 931-943.

Stewart, R. T. (2011). A profit-based scoring system in consumer credit: making acquisition decisions for credit cards. *Journal of the Operational Research Society, 62*(9), 1719-1725.

Sun, J., Li, H., Chang, P. C., & Huang, Q. H. (2015). Dynamic credit scoring using B & B with incremental-SVM-ensemble. *Kybernetes, 44*(4), 518-535.

Thomas, L. C., Edelman, D. B., & Crook, J. N. (2002). *Credit scoring and its applications.* Society for industrial and Applied Mathematics.

Trinkle, B. S., & Baldwin, A. A. (2016). Research Opportunities for Neural Networks: The Case for Credit. *Intelligent Systems in Accounting, Finance and Management, 23*(3), 240-254.

Tsai, C. F., & Hung, C. (2014). Modeling credit scoring using neural network ensembles. *Kybernetes, 43*(7), 1114-1123.

Verbraken, T., Bravo, C., Weber, R., & Baesens, B. (2014). Development and application of consumer credit scoring models using profit-based classification measures. *European Journal of Operational Research, 238*(2), 505-513.

Wang, H., Xu , Q., & Zhou, L. (2015). Large unbalanced credit scoring using Lasso-logistic regression ensemble. *PloS one, 10*(2), e0117844.

Wang, Z., & Yu, Q. (2015). Privacy trust crisis of personal data in China in the era of Big Data: The survey and countermeasures. *Computer Law & Security Review, 31*(6), 782-792.

Weber, R. P. (1990). *Basic content analysis.* UK: Sage Publications.

Wei, Y., Yildirim, P., Van den Bulte, C., & Dellarocas, C. (2016). Credit Scoring with Social Network Data. . *Marketing Science, 35*(2), 234-258.

PWC (2015, February). Retrieved April 14, 2017, from Is it time for consumer lending to go social?: http://www.pwc.com/us/en/consumer-finance/publications/social-media-in-credit-underwriting-process.html

Zarsky, T. Z. (2014). Understanding discrimination in the scored society. *Washington Law Review, 89*, 1375.

Zimmer, M. R., & Golden, L. L. (1988). Impressions of retail stores: A content analysis of consumer images. *Journal of Retailing*, 265-293.

About the authors

Ceylan Onay, Associate Professor of Finance and Deputy Director, Information Systems Research Center of Boğaziçi University, holds a PhD in finance from Department of Management, Boğaziçi University. She received her BA in management from Marmara University and her MBA from Koc University, Turkey. Her research spans a variety of areas within corporate finance, with particular emphasis on Information Technology investments, financial innovation, banking, foreign direct investments and productivity. This research has been published in top finance journals such as the *Journal of Financial Services Research, International Review of Financial Analysis, Journal of Financial Regulation and Compliance, Finance a Uver, Borsa Istanbul Review*. Dr. Onay teaches Corporate finance, IT investments and Electronic Finance Strategies in both graduate and undergraduate programs at MIS Department of Boğaziçi University. She is also a member of Association of Information Systems (AIS) and Head of Sectoral Relations of The Turkey Chapter of the Association for Information Systems (TRAIS).

Dr. Onay is the 2008 recipient of the prestigious Marie Curie Intra-European Fellowship of European Union People program and has been a Research Fellow (since 2008) at the Emerging Markets Group (EMG) Research Center of Cass Business School, London, UK.

A Review of Credit Scoring Research in the Age of Big Data

Figures

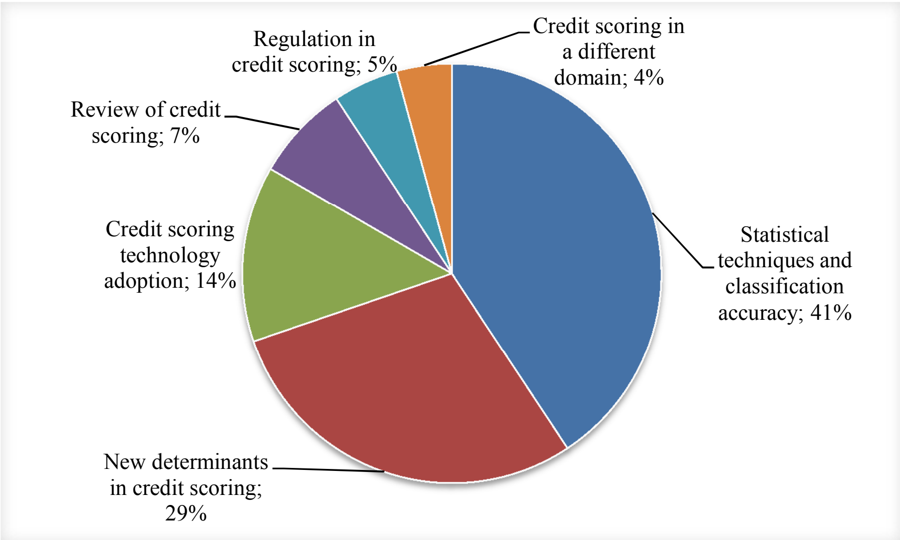

Figure II. Percentage distribution of themes

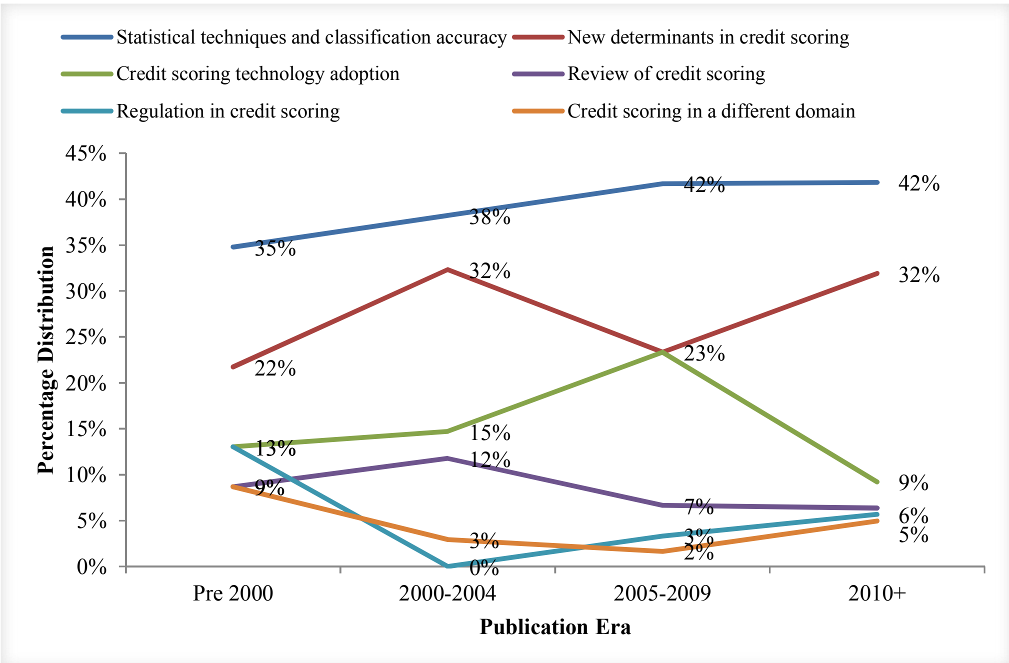

Figure III. Percentage Distribution of Themes for each publication era

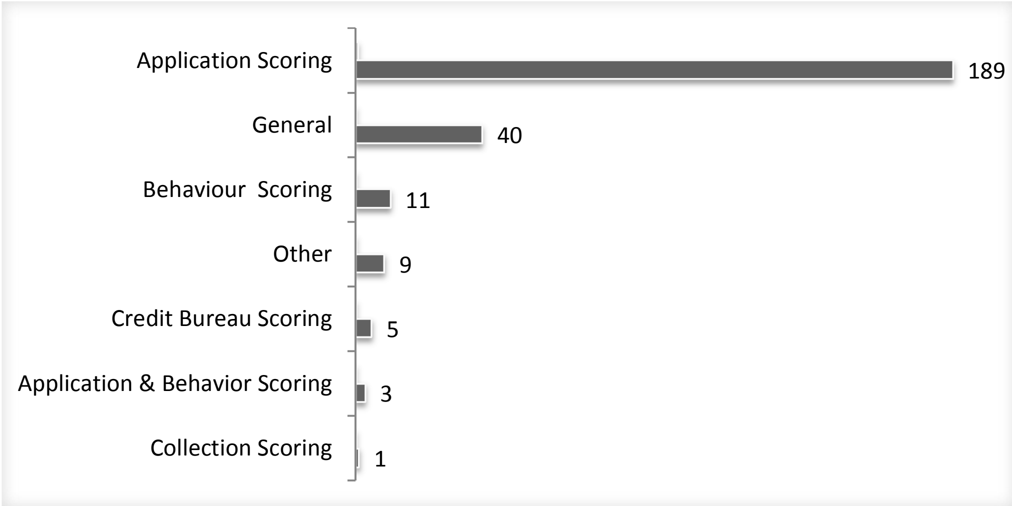

Figure IV. Frequency of scoring types

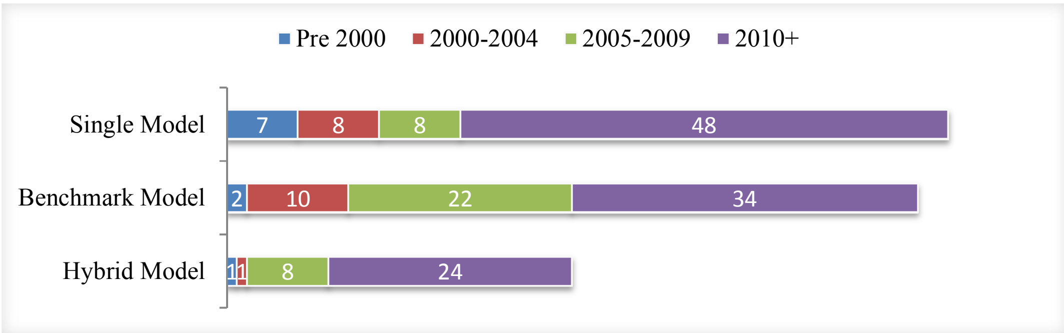

Figure V. Classification Techniques

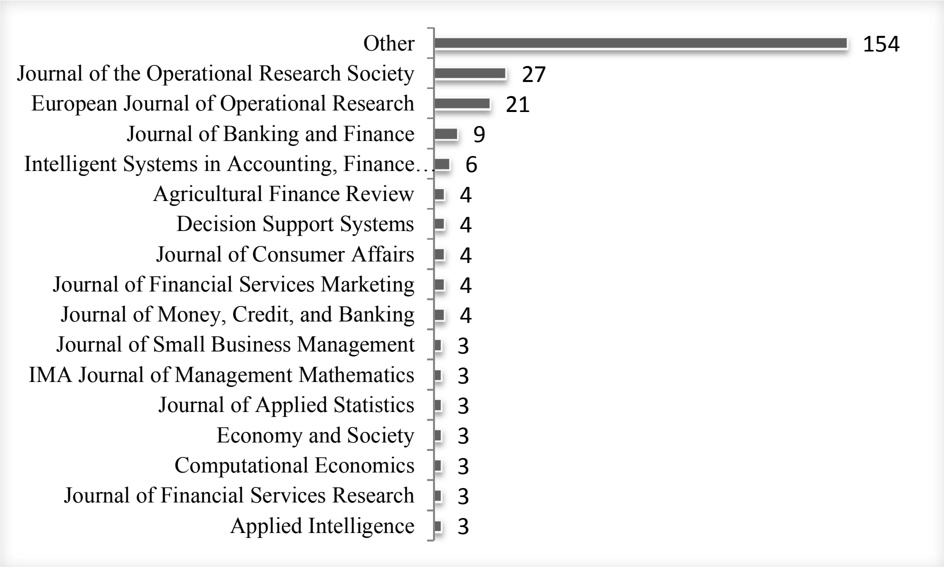

Figure VI. Number of Articles per Journal

A Review of Credit Scoring Research in the Age of Big Data

Tables

Database	Scopes
ProQuest Central	ProQuest Central is a multi-disciplinary database that provides a single source for scholarly journals, news, expert reports, working papers, thesis and datasets along with millions of pages of digitized historical primary sources. Its coverage includes but is not limited to ABI/Inform Collection and databases such as Computing, Political Science, Telecommunications, and Social Sciences databases.
Emerald	Emerald provides access to a portfolio of more than 170.000 peer-reviewed articles from more than 300 journals, more than 2,500 books and over 1,500 teaching cases from a diverse range of management disciplines such as finance, business strategy, information and knowledge management.
Source: Based on our investigation in ProQuest and Emerald Databases.	

Table I. Academic Research Databases and Scopes

1) Statistical techniques and classification accuracy
2) New determinants in credit scoring
3) Credit scoring technology adoption
4) Review of credit scoring
5) Regulation in credit scoring
6) Credit scoring in a different domain

Table II. Theme categories

Judges	Number of Matching (out of 15)	Percentage
A and B	12	80.00%
A and C	12	80.00%
B and C	13	86.67%

Table III. Percentage agreement between judges

< 0	Less than chance agreement
0.01–0.20	Slight agreement
0.21– 0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
0.81–0.99	Almost perfect agreement

Table IV. Cohen's κ Results (Landis and Koch , 1977)

Year	**Number of Articles**
Pre 2000	23
2000-2004	34
2005-2009	60
2010+	141
Grand Total	**258**

Table V. Number of articles per publication era

Main Themes	**Frequency of Articles**	**% of Articles**
Statistical techniques and classification accuracy	105	41%
New determinants in credit scoring	75	29%
Credit scoring technology adoption	35	14%
Review of credit scoring	19	7%
Regulation in credit scoring	13	5%
Credit scoring in a different domain	11	4%
Grand Total	**258**	**100%**

Table VI. Frequency of themes

Main Themes	Pre 2000	2000-2004	2005-2009	2010+	Total
Statistical techniques and classification accuracy	8	13	25	59	105
New determinants in credit scoring	5	11	14	45	75
Credit scoring technology adoption	3	5	14	13	35
Review of credit scoring	2	4	4	9	19
Regulation in credit scoring	3		2	8	13
Credit scoring in a different domain	2	1	1	7	11
Grand Total	**23**	**34**	**60**	**141**	**258**

Table VII. Theme analysis of 258 articles for each publication era

Used Technique	Pre 2000	2000-2004	2005-2009	2010+	Grand Total
Single Model	7	8	8	48	71
Logistic Regression		4	6	26	36
Probit Regression		2	1	6	9
Other	7	2	1	16	26
Benchmark	2	10	22	34	68
Hybrid Model	1	1	8	24	34
Grand Total	**10**	**19**	**38**	**106**	**173**

Table VIII. Frequency of Classification Techniques over publication eras

Data Source	Pre 2000	2000-2004	2005-2009	2010+	Grand Total
Developed/Developing/Emerging Countries	**17**	**19**	**33**	**74**	**143**
Developed	16	16	24	45	101
Developing	1	2	1	6	10
Emerging			6	19	25
Multiple countries		1	2	4	7
Internet Available Data Set		**1**	**2**	**21**	**24**
Internet Available Data Set + Real Data		**1**	**6**	**6**	**13**
Grand Total	**17**	**21**	**41**	**101**	**180**

Table IX. Data Source Classification