Aug 10th, 12:00 AM

The Differential Role of Alternative Data in SME-Focused Fintech Lending

Weifei Zou
Temple University, weifei.zou@temple.edu

Anthony Vance
Temple University, anthony@vance.name

Jie (Kevin) Yan
Dalton State College, jyan@daltonstate.edu

The Differential Role of Alternative Data in SME-Focused Fintech Lending

Emergent Research Forum (ERF) Papers

Weifei Zou
Temple University
weifei.zou@temple.edu

Anthony Vance
Temple University
anthony@vance.name

Jie (Kevin) Yan
Dalton State College
jyan@daltonstate.edu

Abstract

The rapid emergence of risk management Fintech has led to increasing use of alternative data in personal and business financing. Yet there are significant risks and concerns resulting from using alternative data. We therefore seek to examine the differential role of alternative data in SME-focused Fintech lending. We compare the credit evaluation and fraud detection contexts and examine the circumstances under which alternative data are useful for both contexts. Our goal is to find a parsimonious set of traditional and alternative data types that can help facilitate risk management in SME-focused Fintech lending. In this short paper, we report some preliminary results and findings from the first phase of our data collection and analysis. We then discuss the potential contributions and future plans of the study.

Keywords

Alternative data, Fintech lending, risk management, fraud, ambidexterity, interdisciplinary research.

Introduction and Motivation

Alternative data in small and medium-sized enterprise (SME)-focused Fintech lending has been defined as data that are gathered from non-traditional data sources and not typically included in the traditional credit process (Liberti and Petersen, 2019). Alternative data may include any unstructured and structured data. For example, an SME's shipping, turnover and other transaction data have been used as structured alternative data, and public records in texts, social media and owner's personal data have been used as unstructured alternative data (Owens and Wilhelm, 2017). By contrast, traditional data in SME lending largely relate to firm and relationship characteristics (Berger and Udell, 1995). For instance, data on an SME's financial performance such as revenue and profits are often used to represent firm characteristics. And relationship characteristics often reflect the length and strength of a current SME-bank relationship. Alternative data are used when there is a lack and unavailability of traditional data. And the use of a broad variety and vast amount of structured and unstructured alternative data to mitigate information friction and augment risk management is at the heart of Fintech lending (Parrish and Fishman, 2018).

Nevertheless, there are significant risks and concerns resulting from using alternative data. Specifically, some types of alternative data (e.g., bank accounts) are sensitive and borrowers may not know the data were collected for credit decisions, raising privacy concerns. And accuracy concerns would occur when the alternative data types are incomplete and inconsistent. Also, using some types of alternative data may result in discrimination issues because the data involves categories (e.g., gender and race) protected under fair lending laws (Miller et al., 2018). Indeed, a recent survey shows that financial institutions seek to use fewer types of alternative data not only because of the cost but also because of the compliance risk concerns (Parrish and Fishman, 2018). And the World Bank has called for the minimization of alternative data collection (Miller et al., 2018).

Therefore, with the abundance of alternative data, it is imperative to ask what types of alternative data to use and how to use them (Owens and Wilhelm, 2017). This motivates us to investigate the differential role of alternative data in the risk management of SME-focused Fintech lending. We suggest that not all alternative data types are created equal in assessing credit risk while detecting fraud. Some types of alternative data, for example, may be useful in underwriting but rarely valuable in fraud prevention. And some alternative data may be used as substitutes rather than complements for traditional data. In other words, what is still unclear is how alternative data types are different in credit evaluation and fraud detection, and how traditional and alternative data should be combined for better managing the *convenience-fraud risk conflict*. We refer to the convenience-fraud risk conflict as the conflicting needs of evaluating credit risk and thereby providing convenient online services while also addressing fraud risks. Our research questions therefore are: (1) *Under what circumstances are alternative data useful, specifically comparing the credit evaluation and fraud detection contexts?* (2) *How should traditional and alternative data be combined to better manage the convenience-fraud risk conflict?*

To this end, we take two steps to examine the differential role of alternative data types in Fintech lending. Specifically, in the first phase, we focus on traditional data of firm and relationship characteristics and corresponding alternative data of transaction and social media data (discussed in detail below). We examine the performance differences of these traditional and alternative data in the risk management of SME-focused Fintech lending. We then in the second phase, focus on multiple types of alternative data including mobile App analysis and locational data, individual (owners/founders) data, online reviews data and industry data, and examine the performance differences of these alternative data. In this ERF paper and below, we report some preliminary results and findings from the first phase of data collection and analysis.

Methodology

Our research context relates to a collaborative partnership between a leading Fintech company, FinTell, and a joint-stock commercial bank, LoanBank (a pseudonym), in China. We chose LoanBank and FinTell as our research context because partnerships between traditional banks and external Fintech companies have become a new business model of SME-focused Fintech lending (Owens and Wilhelm, 2017). In this partnership model, FinTell does not originate and fund loans but focuses on providing LoanBank risk management Fintech with algorithm-based systems and solutions during the credit process of underwriting and fraud prevention. LoanBank, on the other hand, focuses on performing its own deliberation and then making lending decisions and rates based on the risk profiling results provided by FinTell.

Variables, Measures and Data

The dependent variables in this study include two binary variables of *default* and *fraud*. They are labeled as 1 if a default or fraud event is identified and as 0 if otherwise. The independent variables in the first phase include traditional data and corresponding alternative data. The measures of traditional data of firm and relationship characteristics include *financial ratios* as well as *the strength of an existing SME-bank relationship*. For the corresponding alternative data, we focus on *transaction data* and *social media data*. We suggest that an SME's transaction data are most suitable for replacing financial ratios. This is because transaction data such as sales and purchasing, turnover and return, and logistics and shipping can be readily processed to generate alternative financial ratios such as ratios of total sales to purchasing (weekly, monthly or yearly). Similar to traditional financial ratios in nature, these alternative ratios should give lenders granular insight into the financial performance of the SME. On the other hand, when an existing relationship is absent, we suggest that it can be substituted by an SME's social media data. Indeed, social media data have been employed in several studies as a means for lenders to know more about an SME for credit risk assessment and fraud prevention (e.g., Dong et al., 2018; Owens and Wilhelm, 2017). Below we discuss each of them and the corresponding data sources in detail.

Fraud and Default

Our data for measuring the default and fraud event is based on the business loan-level data of LoanBank. By collaborating with FinTell, we obtained access to the data of LoanBank's loan applications between June 2015 to June 2017, a total of 3,773 applications from 3,732 SMEs. The dataset covers loan applications that have been directly processed and settled by the branches/subbranches' loan officers and presidents. According to the interviews with the product managers of FinTell, loan officers of LoanBank are required to file case reports when they identify an anomaly by the SME in the loan process. A total of 63 fraud reports (of the 3,773 applications) are filed as of May 18, 2019. Among them, 45 reports are filed during loan verification, 11 during loan disbursement, and 7 during loan monitoring. And a total of 42 loans, out of 1,325 applications that were approved, are labeled as default by LoanBank. LoanBank consider a loan in default status if the scheduled payment has not been made for 180 days.

Financial Ratios and Relationship Strength

Prior IS literature has used 12 financial ratios such as asset turnover, leverage, and sales growth disclosed in the annual financial statements of public traded firms (see Abbasi et al., 2012; Dong et al., 2018). Considering the context of our study where most SMEs are private and thus do not have the stock market information to calculate all the financial ratios, we employed three ratios including *leverage*, *return on assets* (ROA), and *asset turnover* (ATO) that can represent an SME's overall firm characteristics in terms of assets, liabilities and profitability. Leverage represents the amount of debt an SME has used to finance its asset and is calculated as total debt divided by total asset. Return on assets represents the profit an SME has earned with its overall resources and is calculated as total net income divided by total asset. And asset turnover represents the revenue an SME has generated with its overall resources and is calculated as total net sales divided by total asset. We used the total asset, debt, revenue, and net income of an SME in the year prior to loan origination (as required in the loan application form by LoanBank) to calculate the three ratios. For the measure of relationship strength, in line with previous literature, we used the number of years that a current SME-bank relationship has lasted (Berger and Udell, 1995). We assume that a relationship between LoanBank and an SME has established since the year when the first loan of the SME was approved.

Transaction and Social Media Data

Previous studies have identified common transaction data employed by Fintech lenders including B2B/e-commerce sales and purchasing data, logistics and shipping data, online accounting data, online SME billing and payment data, and inventory tracking data (see Owens and Wilhelm, 2017). Our selection of transaction data was mainly based on the data mart employed on FinTell's risk management platform that was developed for credit analysis and risk assessment. FinTell relies on its external data vendors for acquiring the transaction data of SMEs that conduct business on their B2B and B2C platforms. For the partnership with LoanBank, FinTell used three types of transaction data including sales and purchasing data, shipping data and inventory data. These data were then converted into corresponding financial ratios, i.e., *sales to purchasing*, *monthly shipping volume*, and *average inventory* (year-end to year-end), for modeling and risk evaluations. We calculated these ratios for the year prior to the loan origination for each SME. SMEs with records less than 1 year or no transaction data records were excluded.

Prior research has used social media data such as Weibo.com (Ge et al., 2017) and SeekingAlpha (Dong et al., 2018) in loan default and fraud detection[1]. We used social media data of WeChat Work in this study. Like Facebook for Business, WeChat Work provides companies, especially SMEs, in China a platform to connect with their current and potential customers. WeChat allows its users to submit various types of reports on a company including false information, false or prompting activities, lewd content, violation of intellectual properties, harassing, and others (each report is limited to 200 Chinese characters). In line with previous studies (see Dong et al., 2018), we developed two variables using the social media report data. One variable of *report quantity* represents the total number of wrongdoing reports submitted on an SME prior to loan origination. And the other variable of *report quality* captures the overall negative emotion and opinion submitted on an SME prior to loan origination. We measured report quality of the SME using a ratio calculated by total number of negative emotion and opinion characters divided by total number of characters.

Preliminary Results

---

[1] As a large amount and variety of alternative data has been used by Fintech companies in China, there is increasing interest in developing a Unified Fintech Regulatory System (an initiative by National Internet Finance Association of China) for data and information privacy.

In this section we report some preliminary results, *with fraud as the only dependent variable*. Given that our dependent variables are binary, we developed logistic models to test the detective ability of each foregoing measure. We then conducted head-to-head comparisons between each traditional data measure and corresponding alternative data measure, a total of 22 comparisons for two different dependent variables. We conducted each comparison using the largest subsample of observations with no missing values in order to maximize the power and external validity of the test. We employed seemingly unrelated estimation (SUEST) to test whether the coefficients (odds ratio) across the two models are statistically different. Overall, our head-to-head comparisons show that *ATO, leverage, relationship* and *the quantity of social media report* are statistically superior in fraud detection, compared to other variables (measures). To further compare the four measures, we conducted the feature (i.e., measure) selection analysis using the logistic classifier with 10-fold cross-validation (see Dutta et al., 2017). The results in Figure 1 show that leverage, ATO and the social media report quantity were chosen for most folders in fraud detection, with 9, 10, and 9 out of 10 folds, respectively. We then focused on the three measures only and calculated the AUCs (area under the ROC curve) of different combinations – 0.68 (Leverage, ATO), 0.65 (Leverage, Social_Quan), 0.67 (ATO, Social_Quan), and 0.72 (Leverage, ATO, Social_Quan). The ROC curves of combinations of leverage and ATO as well as leverage, ATO and social media report quantity with operating points are plotted in Figure 2. Table 1 further details the operating points with the corresponding number of false alarms. It shows that all three measures should be used if users desire a high level of performance. For example, at the 90% operating point, combining the three measures would have a total of 591 false detections, compared to 680 of using leverage and ATO only. Yet the leverage and ATO combination is superior if users desire fewer false alarms and a number of missed detections is acceptable. At the 30% operating point, for example, the number of false detections for using leverage and ATO drops below 100, 70 versus 116 of combining all the three measures.

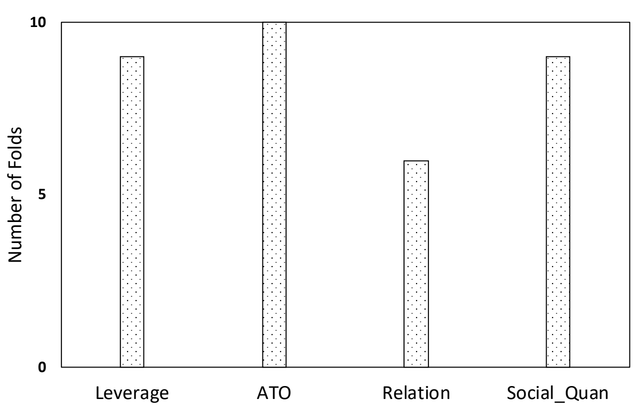

Figure 1. Fraud: Selected Measures
(# Fraud: 18; # Obs. 950)

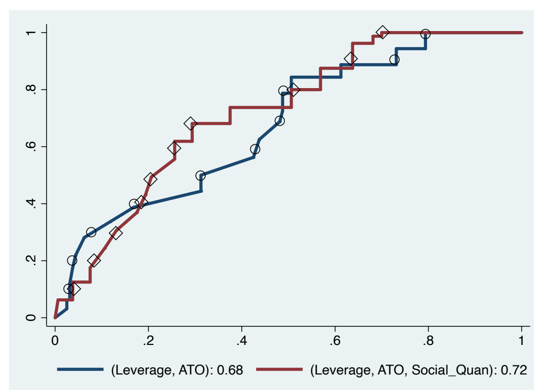

Figure 2. Fraud: ROCs and Operating Points
(# Fraud: 18; # Obs. 950)

	100%	90%	80%	70%	60%	50%	40%	30%	20%	10%
Leverage, ATO	741	680	461	447	400	288	153	70	32	23
Leverage, ATO, Social_Quan	652	591	475	274	237	186	172	116	79	32

Table 1. Operating Points and Number of False Alarms

Contributions and Future Plans

This study will make several contributions. First, our study will extend the existing Fintech research in IS by focusing on the use of risk management Fintech in the banking sector. Unlike peer-to-peer lending and crowdfunding that have been extensively studied, risk management Fintech and its applications in banking have received little attention to date. Our research is one of the first to study the risk management Fintech and develop insights on the differential role of alternative data in SME-focused Fintech lending. By categorizing alternative data into different types and comparing them with traditional data, our findings will reveal how much predictive value each type of alternative data will add to both credit evaluation and fraud detection. In addition, our findings will help answer some important questions in Fintech lending including 1) whether traditional data outperform alternative data in facilitating risk management ambidexterity, 2) whether traditional and alternative data are overall complements or substitutes in facilitating risk management ambidexterity, and 3) how and to what extent traditional and alternative data should be combined for better credit and loan decisions.

Second, our study will contribute to the interdisciplinary research in IS by creating knowledge that spans multiple disciplines. There is growing recognition that IS research should be interdisciplinary (Robey, 2003). However, a recent review of interdisciplinary research in IS shows that only 5 papers, out of the 176 papers from the AIS basket of 8 journals, are interdisciplinary (Tarafdar and Davison, 2018). By focusing on SME lending. Our study will not only enrich the finance literature but, more importantly, also address the use of alternative data and Fintech in the risk management of SME lending, a new and complex problem that is specific to the IS field.

While our paper is research-in-progress, we plan to extend the research in the following ways. First, we plan to add a literature review section to summarize related studies and thereby further identify gaps and justify our research questions. For example, prior IS research on peer-to-peer lending and crowdfunding has provided valuable insights into the use of alternative data in Fintech risk management but tends to focus on one particular type of alternative data only. Second, we plan to add a theoretical background section focusing on organizational ambidexterity and conflict management (See Raisch and Birkinshaw, 2008). We will first review existing IS studies that have examined conflicts managed by ambidexterity and thereby highlight the characteristics of organizational ambidexterity theory relevant to our study. We will then elaborate on the convenience-fraud risk conflict and how the use of alternative data helps manage it. Our goal is to extend organizational ambidexterity theory by proposing and developing the concept of *risk management ambidexterity* in the Fintech lending context. Third, we plan to complete the method section by including the second phase of data and analysis. We will add the measures of multiple alternative data types (i.e., mobile App analysis and locational data, individual data, online reviews data and industry data) and examine their performance differences in facilitating risk management ambidexterity.

REFERENCES

Abbasi, A., Albrecht, C., Vance, A., and Hansen, J. 2012. "MetaFraud: A Meta-Learning Framework for Detecting Financial Fraud," *MIS Quarterly* (36:4), pp. 1293-1327.

Berger, A. N., and Udell, G. F. 1995. "Relationship Lending and Lines of Credit in Small Firm Finance," *Journal of Business* (68:1), pp. 351–381.

Dong, W., Liao, S., and Zhang, Z. 2018. "Leveraging Financial Social Media Data for Corporate Fraud Detection," *Journal of Management Information Systems* (35:2), pp. 461-487.

Dutta, I., Dutta, S., and Raahemi, B. 2017. "Detecting financial restatements using data mining techniques," *Expert Systems with Applications* (90:4), pp. 374-393.

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

Liberti, J. M., and Petersen, M. A. 2019. "Information: Hard and Soft," *The Review of Corporate Finance Studies* (8:1), pp. 1–41.

Miller, M., Medine, D., and Montes, F. 2018. "*Data Protection and Privacy for Alternative Data*," Discussion Paper, The World Bank and CGAP.

Owens, J., and Wilhelm, L. 2017. "*Alternative Data Transforming SME Finance,*" International Finance Corporation, Washington, DC.

Parrish, L., and Fishman, J. 2018. "*Alternative Data: Across the Loan Life Cycle: How FinTech and Other Lenders Use It and Why*," Experian Information Solutions, Costa Mesa, CA.

Raisch, S., and Birkinshaw, J. 2008. "Organizational Ambidexterity: Antecedents, Outcomes, and Moderators," *Journal of Management* (34:3), pp. 375–409.

Robey, D. 2003. "Identity, Legitimacy and the Dominant Research Paradigm: An Alternative Prescription for the IS Discipline: A Response to Benbasat and Zmud's Call for Returning to the IT Artifact," *Journal of the Association for Information Systems* (4:7), pp. 352-359.

Tarafdar, M, and Davison, R. M. 2018. "Research in Information Systems: Intra-Disciplinary and Inter-Disciplinary Approaches," *Journal of the Association for Information Systems* (19:6), pp. 523-551.