Does FinTech Innovation Improve Traditional Banks' Efficiency and Risk Measures? A New Methodology and New Machine-Learning-Based Evidence from Patent Filings*

Libing Fang,† Xindan Li,‡ Avanidhar Subrahmanyam,§ and Ke Zhang¶

March 6, 2023

Abstract

We develop a new methodology that goes all-in machine learning to identify FinTech innovation, and, in turn, to construct a bank-specific proxy of such innovation in China. Since China stringently separates commercial (traditional) and investment banking services, this allows us to study the effects of FinTech innovation on the efficiency and risks of traditional banking. After mitigating endogeneity via propensity score matching and difference-in-differences, we show that FinTech innovation significantly improves banks' efficiency in terms of profit, cost, interest income, and noninterest income. FinTech innovation also improves banks' risk measures—including the overall risk (Z score), capital asset ratio, liquidity ratio, and the nonperforming loan ratio. Heterogeneity analysis further shows that FinTech has a greater positive impact on efficiency and risk measures for banks with greater labor intensity and higher managerial ability.

Keywords: FinTech; Bank Efficiency; Operation Risk; Textual Analysis
JEL Classification: G21, G22, G23, M15, O32, O33

*This is a working paper circulated for comment. Xindan Li acknowledges support by the National Natural Science Foundation of China (No. 71720107001, U1811462). Libing Fang acknowledges support by the National Natural Science Foundation of China (No. 72071103).

†Corresponding author, School of Management and Engineering, Nanjing University, 22 Hankou Road, Nanjing, Jiangsu 210093, China. Email: lbfang@nju.edu.cn.

‡School of Management and Engineering, Nanjing University, 22 Hankou Road, Nanjing, Jiangsu 210093, China.

§Anderson Graduate School of Management, University of California at Los Angeles, Los Angeles, CA 90095, USA. Email: asubrahm@anderson.ucla.edu.

¶School of Management and Engineering, Nanjing University, 22 Hankou Road, Nanjing, Jiangsu 210093, China. Email: kezhang@nju.edu.cn.

Does FinTech Innovation Improve Traditional Banks' Efficiency and Risk Measures? A New Methodology and New Machine-Learning-Based Evidence from Patent Filings

March 6, 2023

Abstract

We develop a new methodology that goes all-in machine learning to identify FinTech innovation, and, in turn, to construct a bank-specific proxy of such innovation in China. Since China stringently separates commercial (traditional) and investment banking services, this allows us to study the effects of FinTech innovation on the efficiency and risks of traditional banking. After mitigating endogeneity via propensity score matching and difference-in-differences, we show that FinTech innovation significantly improves banks' efficiency in terms of profit, cost, interest income, and noninterest income. FinTech innovation also improves banks' risk measures—including the overall risk (Z score), capital asset ratio, liquidity ratio, and the nonperforming loan ratio. Heterogeneity analysis further shows that FinTech has a greater positive impact on efficiency and risk measures for banks with greater labor intensity and higher managerial ability.

Keywords: FinTech; Bank Efficiency; Operation Risk; Textual Analysis
JEL Classification: G21, G22, G23, M15, O32, O33

1 Introduction

Over the last decade, there has been a huge expansion in banks' financial technology (FinTech), particularly in areas such as artificial intelligence, blockchain, cloud computing, big data, and the internet.[1] The Financial Stability Board defines FinTech as "technologically enabled financial innovation that could result in new business models, applications, processes, or products with an associated material effect on financial markets and institutions, and the provision of financial services."[2] Indeed, FinTech is blurring industry boundaries, facilitating strategic disintermediation, and democratizing access to financial services. These changes are emerging in multiple business contexts, including credit, deposits, raising capital, payments, settlements, investment management, and insurance (Philippon 2016, Thakor 2020).

What is the impact of FinTech innovation on banking? A worldwide survey of banking customers (CapGemini's 2017 FinTech Report) finds that more than half of them have interacted with with at least one FinTech provider, indicating the wide reach of FinTech.[3] In an important paper, Chen, Wu, and Yang (2019) show that FinTech innovations have yielded substantial market value to U.S. innovators. However, the U.S. is a universal banking country. So, FinTech innovation in the U.S. is related to both traditional and investment banking.

We consider the impact of FinTech innovation on traditional banking (such as loans, deposits, transactions, and settlements). For our analysis, we use data from China, which is a separated banking country.[4] China's banks are ranked at the top in worldwide rankings, with 144 Chinese banks included in the world's top 1000.[5] In addition, China leads in the worldwide adoption of FinTech lending for 89% of small- and medium-sized enterprises (SMEs), as reported in EY's "Global FinTech Adoption Index 2019."[6] For these reasons, China is uniquely suited to our goal of linking the efficiency of traditional banking to FinTech innovation. The analysis is important because in most developing countries, traditional banking is the way in which the bulk of the

[1] See the Global FinTech Report 2017 (https://tinyurl.com/57xxfvzz) and 2019 (https://tinyurl.com/3je4r5nk).

[2] https://tinyurl.com/3p4ytfha.

[3] See Goldstein, Jiang, and Karolyi (2019), and https://tinyurl.com/4cr9er35.

[4] Allen and Rai (1996), Thakor and Boot (2008), and Chen, Yang, and Yeh (2017) define universal banking countries as having banks that provide a gamut of financial services, including investment, brokerage, real estate, and insurance activities; and define separated banking countries as ones that prohibit the functional integration of commercial and investment banking. Since 1999, the US has been a universal banking country—after the Gramm-Leach-Bliley Act (GLBA) repealed restrictions preventing commercial banks from offering investment banking services.

[5] See https://www.thebanker.com/Top-1000

[6] For details, see https://tinyurl.com/2swnh2u8.

population accesses financial services.[7] Our investigation first develops a new methodology for measuring FinTech innovation, and then conducts a series of tests to reliably identify the causal impact of such innovation on bank efficiency and risk.

Many important papers have focused on measuring FinTech innovation. This literature has measured FinTech in several different ways: by matching FinTech glossaries with online searches (e.g., Cheng and Qu 2020, Dong et al. 2020, Wang, Liu, and Luo 2020), by focusing on a particular aspect of FinTech adoption (e.g, the online channel in He, Ho, and Xu 2020), and via industry-level measures of FinTech development (Lee et al. 2021, Zhao et al. 2022).[8] We instead assess bank-specific FinTech innovation by measuring the cosine similarity of patent filing texts to standard FinTech policy documents.[9] Our FinTech innovation measure presents four novel contributions. First, measures based on FinTech-related glossaries could be influenced by search engine prioritization, whereas the FinTech patents we use more directly reflect the bank's innovation capacity. Second, we construct our proxy using the patents specifically identified as FinTech ones, which allows us to precisely evaluate FinTech's impact. Third, our data from the Chinese Patent Retrieval System have information on both granted patents and patent applications, so that our work does not suffer from the truncation issue raised by Lerner and Seru (2022).[10] Finally, our approach differs from text analysis procedures that partly rely on manual coding (e.g., Chen, Wu, and Yang 2019), in that we develop a new procedure that almost exclusively uses machine learning.[11]

Equipped with our proxy, we define traditional banks in the high FinTech group—that is, the banks with higher numbers of FinTech patents than the 75th percentile of all banks in each year—as the treatment group, and define other traditional banks as the control group. We then estimate bank efficiency measures based on stochastic frontier analysis that is derived from production theory. Because the method corresponds parsimoniously with cost and profit efficiency concepts (Kumbhakar, Wang, and Horncastle 2015, Berger and Mester 1997), it is well-suited to evaluate

[7] See, for example, https://tinyurl.com/dte7avf.

[8] We provide an extensive analysis of the relation between our work and the extant literature in Appendix A.

[9] In China, 2014 is generally considered the first year of FinTech. However, some banks had FinTech patents prior to this year. Therefore, our sample covers 2010—2019, which enables us to reliably perform difference-in-differences (DID) analysis to mitigate endogeneity.

[10] Appendix B expands on this issue.

[11] Chen, Wu, and Yang (2019) use manually separate patents corresponding to financial services, and then manually classify a random sample of such patents into FinTech categories. They then train their machine-learning procedure on this classification. We directly match patents to a standard FinTech document using an automated procedure to measure similarity. Thus we are able to significantly mitigate human subjectivity in our FinTech innovation proxy.

bank efficiency.[12] The primary advantage of this method is that it separates the impact of FinTech on banks' efficiency from secular technological change that also affects efficiency.

After propensity score matching (PSM) to reduce endogeneity, we show that FinTech innovation significantly improves banks' profit, cost, interest income, and noninterest income efficiencies. Our results are remarkably consistent, and economically as well as statistically significant. They accord with arguments which posit that FinTech innovation positively impacts bank efficiency and reduces risk by facilitating customer access to financial services, broadening business coverage, increasing customer numbers, and expanding qualified loans. We also consider FinTech's impact on the non-performing loan ratio and operating risk measures, such as the capital asset ratio and the liquidity ratio. We find that FinTech innovation generally mitigates these measures of operating risk.

To further address identification concerns, we next perform a difference-in-differences (DID) analysis. In this analysis, we use a bank's first-time move into the treatment group as the identifying event. The findings demonstrate significant improvement in banks' efficiency and risk during the years after first moving into the high FinTech group, confirming the regression results. We conduct other robustness checks such as using the number of grant applications instead of patents, using principal component analysis to address multicollinearity in the control variables, and including the Digital Financial Inclusion Index as an additional control variable.[13] Finally, our heterogeneity analysis shows that FinTech innovations may offer more advantages to banks that have better managerial ability (to adopt new technologies) and greater labor intensity (to gather pertinent information and perform due diligence).

A paper closely related to ours is the important U.S.-patent-based work by Chen, Wu, and Yang (2019). Our work departs from theirs in the following ways: (i) by using Chinese patent data on both patent applications and granted patents, we sidestep the truncation issue arising from omitting non-granted applications; (ii) the focus on China allows us to isolate the impact of FinTech on traditional banking (a mainstay of financial services in developing countries); (iii) rather than focus on market valuation alone, as they do, we disaggregate the impact of FinTech

[12]For applications, see Berger, Hasan, and Zhou (2009), Jiang, Yao, and Feng (2013) for Chinese banks, Huber (2021) for U.S. banks, and Park, Han, and Lee (2022) for non-bank firms.

[13]Banks in China are greatly challenged by emerging digital financial services, such as those provided by Ant Financial Services Group, a spin-off from the Alibaba Group; WeBank, initiated by Tencent Company; and Du Xiaoman Financial, which is Baidu's FinTech arm. Even though none of these firms have deposit services, we include the shock that these challenges bring as an additional control variable in a robustness check.

innovation on specific measures of bank efficiency and risk; (iv) we propose a new methodology to isolate FinTech innovation almost exclusively using machine learning and thus reducing human intervention;and (v) we use DID/PSM to reliably identify tests for the issue of how FinTech affects traditional banks' operating efficiency.

Our work also relates to another significant paper by Zhao et al. (2022), which considers the impact of FinTech innovation in China. However, their paper relates banks' efficiency to an aggregate level of FinTech innovation, where aggregate FinTech activity is measured by classifying companies into areas such as online banking, online brokerages, online fund sales, and online asset management, and then aggregating financing activity for these companies.[14] As such, like Chen, Wu, and Yang (2019), the paper's method conflates investment banking with commercial banking activities in the measurement of FinTech activity. In contrast, our measure of FinTech innovation is bank-specific, focuses only on traditional banks, and uses machine learning. Further, unlike Zhao et al. (2022), who exclusively use regression analysis, we use DID/PSM to identify our economic pathway.

In sum, to our knowledge, our work is the first to consider the impact of FinTech innovation on *traditional* banks using a machine-learning-based *bank-specific* innovation measure, and an identified DID/PSM approach. Our results provide important support for the notion that FinTech innovation has a positive operational impact on traditional banks. These banks do not focus on issues such as individuals' asset allocation across risky assets, or automated stock-picking, but on traditional issues such loans, deposits, and clearing and settlement activities. The latter types of activities are the ways in which most customers in the developing world interact with the financial system (Beltratti and Stulz 2012), and our paper indicates that these activities are facilitated via FinTech. Thus, our findings suggest that efforts to regulate the democratization of commercial banking via FinTech should be approached with great circumspection.

We organize the paper as follows: Section 2 reviews related literature and develops our hypotheses. Section 3 describes our data sources. Section 4 describes the construction of key variables. Section 5 documents our research design. Section 6 presents our baseline results and robustness checks. Section 7 analyzes cross-sectional heterogeneities in our main results. Section 8 concludes the paper.

[14] See Wang, Sui, and Zhang (2021) for a similar application that focuses on traditional banks, but nonetheless considers an aggregate FinTech development index. Further, the details of how they construct their index are hard to discern.

2 Literature Review and Hypothesis Development

We first review the history of how the banking industry has been at the forefront of incorporating technology, and then develop arguments on how the advent of FinTech might be expected to impact banks' operating efficiency and risk.

2.1 Technological Innovation and the Banking Industry

The banking industry has a history of responding to technological innovations. For example, the telegraph's first commercial use was in 1838 (Barbiroli 1997), and the first successful transatlantic cable was laid in 1866 (Hills 2002). These innovations improved cross-border financial links and created efficient communication of financial information, transactions, and payments (Arner, Barberis, and Buckley 2015). In the 1950s, the credit card eliminated geographical restrictions on lending, and allowed people to borrow money more conveniently. Since 1967, credit cards, handheld calculators, fax machines, ATMs, and other technologies have improved banks' operating methods and efficiency (Markham 2002). More specifically, the rise of ATMs in 1967 enhanced banks' cost efficiency by saving on counter staff expenses (Shu and Strassmann 2005, Arner, Barberis, and Buckley 2015).

After the 1970s, technological applications in the banking industry accelerated. For example, the Inter-Computer Bureau's establishment created a basis for the Bank Automatic Clearing Service. Instituting the Society of Worldwide Interbank Financial Telecommunications (SWIFT) met the need for cross-border interconnection for bank payments. In particular, Internet banking opened up a channel to improve bank operating efficiency in the 1990s. However, technological use came with risks. For example, the collapse of Herstatt Bank in 1974 following SWIFT's establishment highlighted the risks of increasing international financial interconnections, particularly through new payment system technology (Yadav 2020).

Overall, technology has typically been presumed to improve bank efficiency, although some new risks have emerged. For example, the pervasive use of credit cards has increased interest and noninterest income efficiencies at the cost of increased default risk. The case of Herstatt Bank's collapse (Mourlon-Druol 2015) implies that emerging technology's application could increase competition and thus lead to a higher risk of bankruptcy.

The latest addition to the technology arsenal has been FinTech, broadly described as the application of innovations such as blockchain technology, artificial intelligence, and advanced computing power (e.g., via cloud computing) to the provision of banking services. The worldwide financial industry is flourishing with FinTech innovation—so much so that its impact has inspired a new branch of literature (e.g., Philippon 2016, Chen, Wu, and Yang 2019, Goldstein, Jiang, and Karolyi 2019).

2.2 Bank Efficiency with FinTech

The literature has provided a multitude of arguments on how FinTech affects cost, interest income, and noninterest income efficiencies. We describe the proposed mechanisms below.

Jakšič and Marinč (2019) suggest that FinTech enhances cost efficiency. They argue that FinTech broadens bank operations' business coverage and increases customer numbers. Empirical evidence supporting this argument is that an online banking channel reduces transaction costs and increases cost efficiency via scale effects (Hernando and Nieto 2007, He, Ho, and Xu 2020). The results of Chen, Wu, and Yang (2019) accord with the notion that blockchain technology may offer substantial cost savings in financial services.

Regarding interest income efficiency, FinTech gives banks a significant advantage in lending by improving loan approval procedures. Specifically, Fuster et al. (2019) finds that FinTech enables banks to process mortgage applications 20% faster. And DeYoung, Lang, and Nolle (2007) argue that banks could acquire more hard information for credit evaluation by adopting FinTech, which could expand the scale of credit loans.

One might also expect FinTech to improve noninterest income efficiency by strengthening communication with customers and promoting sales (Subramani and Walden 2001, Lee and Grewal 2004). On the other hand, Jakšič and Marinč (2019) uncover evidence that FinTech breaks through bank operations' geographical limitations, which leads to intense competition between banks for customer resources. So, the FinTech "arms race" may offset the positive impact of FinTech's profitability. Hence the overall effect of FinTech on noninterest income remains an open question.

With regard to profit efficiency, although China has taken steps to liberalize the interest rate—for example, the People's Bank of China (PBoC) removed a loan rate floor in 2013, and a deposit rate ceiling in 2015—in effect, bank deposit and lending rates still remain under scrutiny by the PBoC through instructions and window guidance. In this architecture, most banks compete on relatively similar room for revenue growth, and thus resort to improving profit efficiency via technological progress, such as FinTech innovation.

Overall, the arguments indicate that FinTech innovation should improve banks' cost, profit, and interest income efficiencies, but there are opposing arguments for the effect of FinTech on noninterest income efficiencies. In our data analysis, we are able to identify the unambiguous empirical direction of the relevant effects.

2.3 Bank Risk with FinTech

Previous studies imply that FinTech innovation can reduce banks' overall risk by promoting internal governance and control (Philippon 2016, Navaretti, Calzolari, and Pozzolo 2017). For example, FinTech can help streamline banks' business procedures, and thus provide customers with easier access to funds and more convenient loan rollovers, which may lead to lower default rates (Lin, Prabhala, and Viswanathan 2013). Moreover, using FinTech is an efficient way to improve credit products' and customers' diversification, as it facilitates access to services and broadens business coverage. Such diversification could reduce banks' overall credit risk (Demirgüç-Kunt and Huizinga 2010). More importantly, FinTech can help banks evaluate enterprises' quality more accurately, especially SMEs (Huang et al. 2020). So, FinTech should improve loans' overall quality, thereby decreasing banks' risk.

However, we also note that FinTech innovation is changing the competitive landscape and redrawing the financial services industry's lines. FinTech use might also increase banks' risk, because it increases the chances of customer short-termism (Jakšič and Marinč 2019). Specifically, FinTech innovation provides more accessible, convenient, and inexpensive financial services. This, coupled with advertising and peer pressure, might encourage more customers to apply for such services. This volume of customer acquisition and rapid subsequent attrition due to short-termism can lead to herding effects, bank runs, and decreased operation stability. DeYoung, Lang, and Nolle (2007) raise another concern: a crowding-out effect wherein an online channel popularization will shift core deposits to the money market, which makes it difficult for banks to obtain loyal customers. Thus, FinTech might aggravate overall risk in the banking industry.

For liquidity risk metrics, such as the capital asset ratio (CAR) and the current ratio (or liquidity ratio), a possible mechanism is as follows. In transforming economies, SMEs depend on bank lending for financing. Banks prefer short-term loans when dealing with SMEs because of their high risk. Using FinTech, banks can evaluate SMEs' quality more accurately and, therefore, increase the amount and duration of SME loans. However, the People's Bank of China (PBoC) recommends that banks follow the requirements set forth in the Bank for International Settlements framework: classifying loans into risk-adjusted categories to meet the comprehensive banking supervision requirement outlined in the "Basel Core Principles for Effective Supervision" (Berger, Hasan, and Zhou 2009). Bank managers are obligated to meet regulators' requests when seeking to maximize revenue. In this scenario, bank managers might prefer to meet the supervisory requirements as thresholds that they apply to the CAR and liquidity ratio, regardless of FinTech. Such a threshold approach implies that these ratios may or may not increase with FinTech innovation.

For credit risk measures, such as the NPL ratio, FinTech provides a well-known advantage, as it allows banks to acquire more quantifiable hard information (Huang et al. 2020, DeYoung, Lang, and Nolle 2007). This channel has two effects. On the one hand, more quantifiable information helps evaluate credit and thus make financial services more inclusive. As Huang et al. (2020) shows, for borrowers in China's Tier-4 cities, 62% have lower predicted default probabilities using the FinTech model, compared with 52% for borrowers in Tier-1 and Tier-2 cities. But on the other hand, FinTech businesses might suffer from regulatory arbitrage and moral hazards (Philippon 2016, Navaretti, Calzolari, and Pozzolo 2017, Buchak et al. 2018). An excessively light approach to FinTech's regulation could lead to a gradual deregulation pattern aimed at enhancing competition and expanding off-balance sheet activities. This might increase risk-taking and result in enhanced credit risk.

Overall, the above reasoning indicates that the effect of FinTech innovation on bank risk measures is an open question, and data analysis is necessary for a final answer.

3 Data Description

To investigate the link between FinTech innovation and banks' efficiency as well as risk measures, we construct an unbalanced panel dataset containing 222 listed and unlisted Chinese commerical banks from 2010–2019, which includes 113 city, 12 joint-stock, 36 foreign, 56 rural, and 5 state banks. The banks' fiscal reports come from the CSMAR financial database. We obtain some related control variables—such as the total number of branches, employees, and cities in which a bank operates—from financial statements and from a widely used retrieval system for due diligence in Chinese financial institutions (see https://qcc.com [in Chinese]). Our dataset includes 1,592 bank-year samples. The dataset is representative of the Chinese banking industry because it covers all primary banks, including state commercial and joint-stock banks, along with city and rural commercial banks. The cumulative total assets of banks in our dataset account for 75% of the Chinese banking industry's total assets.

To measure banks' FinTech innovation, we obtain the original patent filings from the China Patent Retrieval System (http://cprs.patentstar.com.cn). We collect all the patent filings of applicants that are banks or IT firms with bank blockholders.[15] The fields of the patent filing include the title, International Patent Classification (IPC) code, application date, grant date, patent abstract, applicant name, and applicant institution (identical to an applicant's name if the applicant is an institution, such as a bank). We only retain the patents belonging to IPC Classes G and H because these relate to digital computing, which is the basis for potential FinTech innovation. We identify the initial applications for patents (in process, which hope to be granted) as well as the patents that made it through the application process and were granted (this can be determined by whether the grant date field is filled). The total number of patent records is 26,168, with 9,048 coming from banks and 17,120 from IT firms.

4 Measuring FinTech Innovation, Bank Efficiency, and Risk

In this section, we present the methodology for computing our measure of FinTech innovation at the bank level, as well as the measures of bank efficiency and risk that we use in our analysis.

4.1 FinTech Innovation

Despite the increasing interest in banks' FinTech in academia and practice, a lack of consensus remains regarding its measurement. The literature considers a variety of proxies for FinTech—for example, the frequency with which the bank name matches FinTech-related glossaries in online searches (e.g., Cheng and Qu 2020, Dong et al. 2020, Wang, Liu, and Luo 2020), one aspect of

[15] These IT firms do not have licenses to do banking activities but generally provide software development that integrates financial services with emerging technologies, such as cloud-based solutions, artificial intelligence, machine-learning, and business analytics. We include IT firms whose block ownership by a sample bank exceeds 5%. This definition may raise the concern that multiple banks could be blockholders of the same IT firm. However, in our sample, while there are 16 IT firms with multiple bank blockholders, only one files digital computing patents. Therefore, this issue has negligible impact on our work.

FinTech adoption (such as the online channel in He, Ho, and Xu 2020), and the development of industry-level FinTech innovation (Lee et al. 2021, Zhao et al. 2022). To the best of our knowledge, few proxies in the literature comprehensively measure bank-specific FinTech innovation activities.

We adopt a new machine learning procedure to identify FinTech patents and, in turn, to construct a proxy for each bank's Fintech innovation. Our procedure has two stages. First, we identify FinTech patents via natural language processing (NLP) models with the help of three reference documents. Second, we count the number of FinTech patents as a measurement of FinTech innovation for each bank in each year. Our work differs from Chen, Wu, and Yang (2019), as that paper relies on manual coding to prepare batches of patent data that either have or have not been classified as FinTech. In contrast, as we see below, our procedure almost exclusively relies on machine learning to reduce human subjectivity.

The first stage of our approach has three steps. In step one, we collect three auxiliary documents on FinTech, the commercial banking business, and computer science. These documents are used as a reference for identifying whether a patent is more likely to be FinTech than one pertinent to the banking business or computer science. In step two, we use *doc2vec*—an NLP algorithm to vectorize each reference document and each patent filing text. The *doc2vec* algorithm outputs a vector to represent the text's meaning after applying a given text. This output consists of a numerical vectors representing the meanings of each patent filing text and three numerical vectors for the meaning of the reference documents FinTech (FT), banking business (BB), and computer science (CS), respectively.

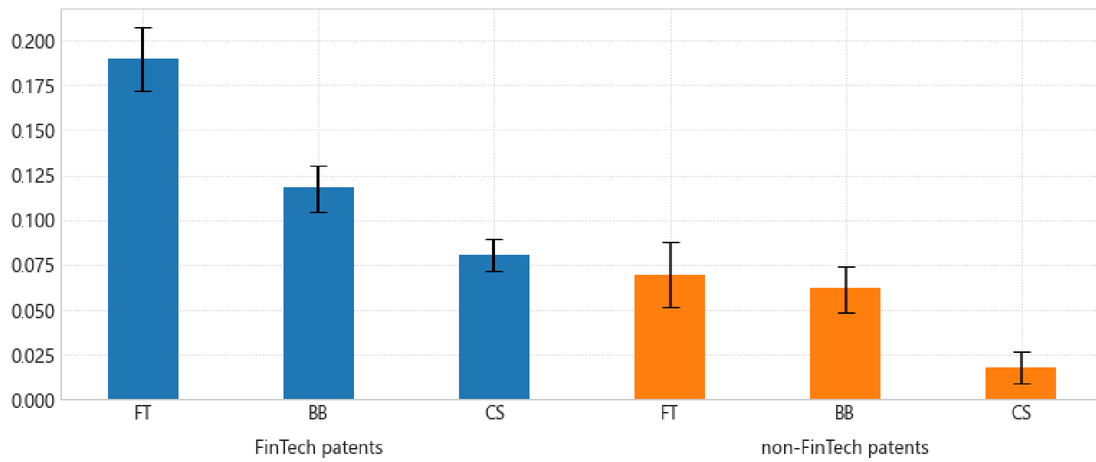
In step three, we label a patent as FinTech if its vector representation (PF) has a higher similarity to FT than to the other two vectors (BB and CS). Recall that we only use the patents that are pertinent to digital computing (classes G and H), so that this process merely determines whether these digital computing patents are more strongly related to FinTech than to the banking business or computer science in general. For a patent filing's vector, PF, its similarity to a reference document vector—say FT (or BB and CS)—is estimated by the two vectors' cosine, the so-called *cosine similarity*. It is a scalar measuring the distance between the meaning of two texts. It takes the value of 1 for two texts with the same meaning, and -1 for antonymous texts.[16] We

[16]Guzman and Li (2022) use this distance to measure the semantic similarity between online founding strategies for startup firms and incumbent firms around the time of a startup's founding. Our approach also closely resembles the test for measuring corporate innovation in Bellstam, Bhagat, and Cookson (2021).

refer readers to Appendix C for more details of the three steps. The procedure yields 7,181 initial FinTech patent applications (that are still in process) and 1,396 granted ones.

Figure 1 presents a summary of the cosine similarities for the FinTech and non-FinTech patents to FT, BB and CS. From the figure, we can see that for FinTech patents, the lower bound

Figure 1: **Summary of cosine similarities.**
The figure has two groups of bars with bootstrapped 95% confidence bands for FinTech and non-FinTech patents. The bars indicate the average cosine similarities of patent filings' text to three reference documents, respectively, on FinTech (FT), banking business (BB), and computer science (CS). The bands are 95% confidence intervals based on the 2.5/97.5 percentiles of a bootstrap involving 500 replications, as in Figure 2 of Bellstam, Bhagat, and Cookson (2021).



of 95% confidence bands for $\langle PF, FT\rangle$ is higher than that of $\langle PF, BB\rangle$ and $\langle PT, CS\rangle$, where $\langle x, y\rangle$ indicates the cosine of vectors x and y. This is consistent with the method for identifying FinTech patents as the ones whose filing text vectors have the highest cosine similarity to the vector of the FT reference compared to BB and CS. For the non-FinTech patents, we can see the average $\langle PF, BB\rangle$ and $\langle PF, FT\rangle$ are similar but both are higher than $\langle PF, CS\rangle$. This implies that the similarity of non-FinTech patents to FinTech (FT) may stem from their similarity to the banking business (BB).

The banks' non-FinTech patents have a meaning that is more similar to the banking business. This finding is consistent with the notion that the patents' applicants are traditional banks in our sample. That $\langle PF, CS\rangle$ for non-FinTech patents is the lowest among all three pairwise distances implies that our non-FinTech patents from the IPC G& H classes have little relation to CS (computer science), and this further shows that our non-FinTech patents relate to traditional banking.[17]

4.2 Bank Efficiency

Drawing on production theory from economics, the literature often has used the stochastic frontier approach (SFA) to evaluate banks' (e.g., Berger and Mester 1997, Huber 2021) and other firms' efficiency (e.g. Park, Han, and Lee 2022). This method estimates an aggregate efficiency frontier (that is, the best practice) using input and output data for all banks, and defines each bank's efficiency as an inverse of the distance to the frontier (the closer, the better). SFA's primary advantage is isolating bank-specific efficiency measures from secular technological change.

We follow Chapter 10 of Kumbhakar, Wang, and Horncastle (2015) in our SFA estimation. More specifically, let us denote bank i's profit (or a variable that contributes positively to profit) for a given year t as π^a_{it}, and the time-invariant and time-variant inefficiency as η_i and u_{it}, respectively, where $\eta_i, u_{it} > 0$. We define the overall technology efficiency (OTE) index as

$$OTE \equiv \exp\left(-\eta_i - u_{it}\right) = \frac{\pi^a_{it}}{\tilde{\pi}_{it}}, \tag{1}$$

where $\tilde{\pi}_{it}$ is the overall best-practices profit to be estimated with the same inputs used for π^a_{it}. Given this definition, OTE lies in the interval $[0, 1]$. The closer the value is to 1, the higher the overall technology efficiency. Empirically, however, for specific observations of some bank in some year, the estimated profit in best practice $\tilde{\pi}_{it}$ may be less than the actual profit π^a_{it} via noise induced by measurement error or chance. To accommodate such statistical concerns, we include two additional components: bank-specific fixed effects μ_i and random noise v_{it}. Thus, we write Equation (1) in transcendental logarithmic (translog) formation as

$$\ln \pi^a_{it} = \ln \tilde{\pi}_{it} - \eta_i - u_{it} + \mu_i + v_{it}, \tag{2}$$

where there are four stochastic components: η_i and u_{it} are positive and indicate the ith bank's time-invariant and time-variant inefficiency, respectively; and μ_i and v_{it} are bank-specific fixed effects and random shocks (noise), respectively, that are unrelated to efficiency.

The overall objective is to estimate η_i and u_{it} and thus calculate OTE via Equation (1). For

[17] To lend more intuition, Appendix D provides detailed examples of a FinTech and a non-FinTech patent identified by our procedure.

this purpose, we proceed as follows. First, following Berger and Mester (1997) and He, Ho, and Xu (2020), to potentially control for a bank's economic environment and changes, we add two terms to Equation (2):

$$\ln \pi_{it}^{a} = \ln \tilde{\pi}_{it} + \rho_1 PNPL_{it} + \rho_2 PNPL_{it}^2 + dYear_t - \eta_i - u_{it} + \mu_i + v_{it} \tag{3}$$

where $PNPL$ is the weighted average ratio of nonperforming loans (NPL) to gross loans in the province where bank i operates, and $Year_t$ denotes year dummies. The weights in $PNPL$ are the proportions of loans issued by banks in the same province. Such a weighted NPL in a province appears to be completely exogenous and is a useful control variable for a bank's economic environment. We include year dummies to avoid estimation biases that may arise with potential changes in bank performance stemming from technological progress or changes in the economic and regulatory environments over years.

Before describing further steps in the OTE estimation, we note here that we can also estimate an analogous equation to Equation (3) for cost instead of profit. When estimating the OTE of cost given the restriction $\eta_i > 0$, and $u_{it} > 0$, we adjust Equation (1) and calculate OTE as:

$$OTE \equiv \exp\left(-\eta_i - u_{it}\right) = \frac{\tilde{C}_{it}}{C_{it}^{a}}, \tag{1$'$}$$

where

$$\ln C_{it}^{a} = \ln \tilde{C}_{it} + \rho_1 PNPL_{it} + \rho_2 PNPL_{it}^2 + dYear_t + \eta_i + u_{it} + \mu_i + v_{it}. \tag{3$'$}$$

Here, C_{it}^{a} denotes the actual total cost of bank i observed in year t; $\tilde{C}_{it}$ is the total cost in the best practice to be estimated. Comparing Equation (1$'$) for cost to Equation (1) for profit, the difference is that the OTE in this case implies that the actual cost statistically larger than that in best practice, while in the profit case, the actual profit is less than that of the best practice. We replace the corresponding terms in Equation (1) by Equation (1$'$) when estimating cost OTE.

The next step in estimating OTE involves specifying the profit (cost) in a best-practice that substitutes for $\tilde{\pi}_{it}$ in Equation (3) and $\tilde{C}_{it}$ in Equation (3$'$). Let ν_{it} denote, in turn, profit, interest income, noninterest income, or cost of a bank in the best-practice—that is, on the frontier. Following the literature (Kumbhakar, Wang, and Horncastle 2015, Park, Han, and Lee 2022, He, Ho, and Xu 2020), after omitting the subscript (i, t), we write the normalized frontier $\ln \tilde{\nu}$ in translog form as

$$
\begin{aligned}
\ln \tilde{v} = \delta_0 &+ \sum_{j=1}^{3} \delta_j \ln\left(\frac{x_j}{z}\right) + \frac{1}{2}\sum_{j=1}^{3}\sum_{k=1}^{3} \delta_{jk} \ln\left(\frac{x_j}{z}\right) \ln\left(\frac{x_k}{z}\right) \\
&+ \beta_1 \ln\left(\frac{w_2}{w_1}\right) + \frac{1}{2}\beta_2 \left[\ln\left(\frac{w_2}{w_1}\right)\right]^2 \\
&+ \sum_{j=1}^{3} \gamma_j \ln\left(\frac{x_j}{z}\right) \ln\left(\frac{w_2}{w_1}\right)
\end{aligned} \tag{4}
$$

where w_1 and w_2 are input factors, x_1 to x_3 are outputs, z denotes total assets, and $\tilde{v} \equiv \dfrac{v}{w_1 z}$.[18] The normalization by w_1 enforces cost homogeneity; that is a doubling of input costs doubles total costs; for consistency, the same normalization is applied to profits as well. The normalization by z accounts for bank scale.

As in He, Ho, and Xu (2020), the inputs w_j, $j = \{1, 2\}$, are defined as follows. First, w_1 is the average funding price, which is the ratio of interest expenses to the sum of deposits and other interest-bearing liabilities. w_2 denotes the average nonfunding input price, which is the ratio of noninterest expenses (e.g., for labor and other costs) to total fixed assets. Typically, the bulk of the noninterest expense consists of employee and administrative expenses, where the latter mostly involves marketing costs and client maintenance expenses. Both are closely related to tangible assets such as office buildings and the geographical allocation of employees and branches across regions. For output variables x_j, $j = \{1, 2, 3\}$, we consider total deposits x_1, total loans x_2, and other income-earning assets x_3. We normalize the three outputs (as well as profits) by total assets (z), again to control for bank scale.[19] Annual reports directly provide data on all of the input and output variables.

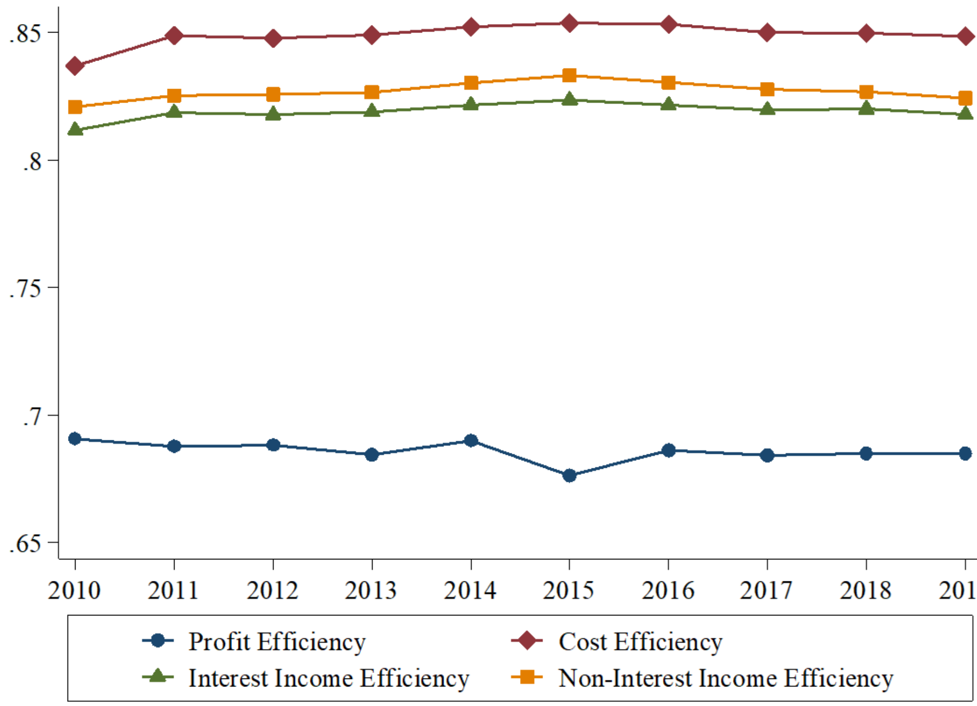
Combining Equations (3) and (4), we can estimate η_i and u_{it}, and, in turn, the profit/cost efficiency from (1), following the steps in Chapter 10 of Kumbhakar, Wang, and Horncastle (2015). Estimation details appear in Appendix E. We also estimate two alternative efficiency measures—interest income and noninterest income by replacing profit (π_{it}^a) in Equation (3), in turn, with these quantities. We estimate the average profit, cost, interest income, and noninterest income efficiencies to be 69.8%, 85.0%, 82.0%, and 82.8%, respectively. Figure 2 plots the yearly mean

[18] See Equation (A1) in Appendix 3 of He, Ho, and Xu (2020). This equation emanates from the optimization problem solved in Chapter 4 of Kumbhakar, Wang, and Horncastle (2015).

[19] Note from Equation (3) that the inefficiency terms are derived from residuals. Since the largest banks' costs and profits are many times larger than those of the smallest banks, normalization is necessary to make residual variance scale-invariant.

efficiencies.

Figure 2: **Yearly mean of four bank efficiency measures.**
This figure shows the average trend in banks' four efficiency measures, namely, profit ($Prof\_eff$), cost ($Cost\_eff$), interest income (Int_eff), and noninterest income ($Nint\_eff$). The four time series are the yearly cross-sectional averages of the four efficiency measures over the sample period from 2010–2019. The average number of banks in each year is 73.2.



The results in Figure 2 are consistent with earlier work in Berger, Hasan, and Zhou (2009) as well as Jiang, Yao, and Feng (2013); for example, the relative magnitudes are similar. We find that the average efficiency of noninterest income is slightly higher than that for interest income. This suggests that the noninterest profitability of banks is higher during our sample period when both traditional financial service providers and Internet-based firms provided financing, payment, investment, and intermediary information services by leveraging information and communication technologies.

4.3 Bank Risk

To measure bank risk, we borrow four indicators from the CAMELS rating system (Berger, Hasan, and Zhou 2009, Demirgüç-Kunt and Huizinga 2010, Jiang, Yao, and Feng 2013). First, we adopt a bank's capital asset ratio, $CAR = capital/risk\ weighted\ assets$. It measures how much capital a bank holds for covering its solvency risk. A bank's *risk weighted assets* are total assets weighted by risk assessment for each type of asset. A document titled "International Convergence of Capital Measurement and Capital Standards" (Basel II) provides suggested methodologies for calculating risk-weighted assets.[20] Chinese banks use these methods to calculate CAR and document it in their annual reports. We directly obtain CAR from CSMAR.

The Z_Score is an indicator of overall bank risk, defined as $Z\_Score = (ROA + CAR)/sd(ROA)$, where ROA is the return on assets and $sd(ROA)$ is the full sample standard deviation of ROA for a specific bank. A higher Z_Score indicates that a bank has lower bankruptcy risk. Next, the NPL ratio, defined as non-performing loans divided by total loans, is a common measure of bank credit risk. The higher this ratio, the greater the credit risk borne by a bank. Last, we use the liquidity (or current) ratio—the ratio of liquid assets to the sum of deposits and short-term debts—to measure how sufficient a bank's high-quality liquid assets are for meeting short-term obligations. A higher liquidity ratio indicates less liquidity risk.

5 Research Design

We use propensity score matching (PSM) to reduce the endogeneity concern arising from banks' jointly selecting FinTech innovation and efficiency/risk levels. To do this, we define banks in the high FinTech group (those having FinTech indicators higher than the 75% quantile of all banks in each year) as the treatment group, and the remainder as the control group. Then, we perform 1:1 nearest-neighbor (propensity score) matching between the two groups and build our final sample. Subsequently we perform a regression examining the impact of FinTech innovation on banks' efficiency and risk measures. Since the control variables in the regression are also used to construct the PSM sample, we describe the regression framework first, and provide details on the PSM in Section 5.2 to follow.

[20] For details, please see *Part 2: The First Pillar–Minimum Capital Requirements* in the framework of Basel II, https://www.bis.org/publ/bcbs107.htm.

5.1 Regression Framework

We run the following regression to estimate FinTech's impact on banks' efficiency and risk:

$$y_{it} = \beta_0 + \beta_1 FinTech_{it} + \sum_j \delta_j Control_{ijt} + \varepsilon_{it}\,, \tag{5}$$

where y_{it} represents the efficiency or risk of bank i in year t. For banks' efficiency, y_{it} alternatively denotes profit, cost, interest income, and noninterest income efficiency (as estimated in Section 4.2). For banks' risk, y_{it} alternatively denotes the Z_score, CAR, NPL, and the liquidity ratio, Liq (as defined in Section 4.3).

The independent variable $FinTech_{it}$ is a dummy indicating whether bank i in year t is in the treatment group. For each bank, we use the number of FinTech patent applications and grants as a measure of FinTech innovation. We base our main results on the number of FinTech patent applications. We use the granted number to check robustness in Section 6.3. Also, to confirm the role of FinTech patents, in another robustness check, we replace $FinTech_{it}$ in Equation (5) with ALL_PAT_{it}, the number of all patents for bank i in year t.

$Control_{ijt}$ is a set of control variables that might influence bank efficiency or risk. We use three types of control variables. We include controls for bank scale, which involves total assets ($SIZE$) as well as the total number of branches scaled by assets ($BRANCH$). Second, to control for geographical diversification, we add the number of unique cities in which the bank has at least one branch ($NCITY$), because prior literature indicates that geographical diversification improves banks' market share and noninterest income while also increasing operating costs (Cai, Xu, and Zeng 2016).

Next, we include a series of macro variables. We compute the Hirschman-Herfindahl index, HHI, for a particular bank using the number of bank branches in the principal city of operation. As a common measurement of banking market concentration, HHI has a relationship with banks' cost efficiency. We also include the logarithms of per capita GDP ($PGDP$) and population (POP), and as well as the number of firms ($NFIRM$) in banks' main operating provinces to control for the implications of economic development on bank efficiency and risk. We also include dummies for time (years) and bank types (as described at the beginning of Section 3) to capture fixed effects across banks and years. Appendix F provides detailed definitions of the controls.

5.2 Propensity Score Matching

Prior to estimating the regression, we perform PSM with 1:1 matching for the treatment banks (in the group with high FinTech innovation). We start with the set of all control variables in Equation (5) along with some additional variables, namely, the logs of total deposits, net income, and total loans. However, we select the final variables for PSM by following Imbens and Rubin (2015).[21] After PSM, the final data contains 704 bank-year observations (352 each for the treatment and control groups), with 7,231 FinTech patent applications and 2,486 granted ones, with each granted and applied-for patent matched with a non-FinTech patent. Table 1 shows

Table 1: **Diagnostic test for the final propensity score matching (PSM).**
This table reports the results of diagnostic testing of 1:1 PSM between the treated and control banks in each year from 2010–2019. The treated banks are the ones with a number of FinTech patent applications above the 75th percentile of that of all banks in each year; otherwise, the banks belong to the control sample. We initially use all the control variables in Equation (5) along with some additional ones, including the log of total deposits, net income, and total loans. We select the final necessary variables by following Imbens and Rubin (2015). $SIZE$ is total assets, $PGDP$ is the logarithm of per capita GDP of each bank's main operating province, and $NCITY$ is the number of cities in which the bank has at least one branch. The rows in the table report the mean and t-stats for all selected control variables before and after PSM. The "Unmatched" and "Matched" columns denote the diagnostic testing results for the variable before and after matching, respectively. The t-test results indicate whether a significant mean difference exists between pre- and post-match. Here ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	Unmatched Mean			Matched Mean		
Variable	Treated	Control	t-stat. test	Treated	Control	t-stat. test
log (net_income)	15.039	13.622	12.98***	15.039	14.847	1.29
log($total\_loans$)	19.045	17.708	13.23***	19.045	18.821	1.55
$SIZE$	19.905	18.572	14.23***	19.905	19.712	1.45
$PGDP$	10.339	10.274	2.57***	10.339	10.322	0.54
$NCITY$	2.8834	2.0451	11.96***	2.8834	2.7864	0.93

our diagnostic test results for PSM using the final set of matching variables, which are the logs of net income and total loans, as well as $SIZE$, $PGDP$, and $NCITY$. The table confirms that the procedure works well in terms of building the control group for our treatment banks. The t-test results indicate that there are no significant differences for the included variables across treated and control samples.

[21] The procedure involves adding covariates iteratively to the regression, and then performing a likelihood ratio test for the null that an added covariate is redundant. At each stage the variable with the maximum likelihood ratio is selected, and the process repeats till the null cannot be rejected for the covariates being added. See Imbens and Rubin (2015) for full details.

Table 2: **Descriptive statistics.**
This table reports descriptive statistics for the variables used in Equation (5). The sample period covers from 2010–2019 with 704 observations after propensity score matching, with 352 each in the FinTech treatment and control groups. The treatment group consists of banks whose number of FinTech patent applications is higher than the 75th percentile of that of all banks in a given year. The control group is the propensity-score-matched sample for the treatment group. The dependent variables in the regression are, alternatively, four efficiency measures and four risk measures. The efficiency measures include profit ($Prof\_eff$), cost ($Cost\_eff$), interest income (Int_eff), and noninterest income ($Nint\_eff$). The risk measures include overall risk (Z_score), capital asset ratio (CAR), nonperforming loan ratio (NPL), and liquidity ratio (Liq). The definitions are as follows: $Z\_score = (ROA + CAR)/sd(ROA)$, where ROA is the return on assets and $sd(ROA)$ is the standard deviation of ROA during the sample period; $CAR = capital/risk\ weighted\ assets$ indicates the extent of bank shareholders' capital covering the risky assets; NPL is the ratio of nonperforming loans to total loans; and Liq is the ratio of liquid assets to the sum of deposits and short-term debts. $FinTech$ is a dummy variable which equals unity for the treatment group and zero for the control group. $FinTech2$ is the number of granted FinTech patents in an year. ALL_PAT is the log of the number of a bank's total patent applications in each year. We include control variables as follows: $SIZE$ is the log of a bank's total assets (in million RMB); $BRANCH$ is the ratio of the total number of branches divided by total assets (in million RMB); $PGDP$, POP, and $NFIRM$ are respectively the log of per capita income, the log of the population, and the number of firms in the provinces where banks principally operate; $NCITY$ is the log of the number of cities in which the bank has branches; and HHI is the Herfindahl-Hirschman Index of bank branches in each province.

Variables	Mean	SD	Min	Median	Max
$Prof\_eff$	0.728	0.103	0.026	0.735	0.958
$Cost\_eff$	0.861	0.046	0.479	0.868	0.979
Int_eff	0.838	0.046	0.546	0.837	0.942
$Nint\_eff$	0.843	0.056	0.489	0.845	0.966
Z_Score	3.840	2.982	0.000	3.084	12.895
CAR	24.558	14.151	4.696	21.024	86.497
NPL	1.532	0.629	0.056	1.624	4.990
Liq	14.368	9.470	2.192	12.795	73.756
$FinTech$	0.500	0.500	0.000	0.500	1.000
$FinTech2$	1.901	1.245	0.000	2.079	5.004
ALL_PAT	3.106	2.430	0.000	4.852	7.075
$SIZE$	19.809	1.760	14.131	19.539	24.128
$BRANCH$	0.502	0.594	0.000	0.359	6.133
$PGDP$	10.330	0.414	9.302	10.301	11.148
POP	17.546	0.672	15.656	17.594	18.463
$NFIRM$	13.508	0.722	10.799	13.452	14.955
$NCITY$	2.835	1.377	0.693	2.565	5.908
HHI	0.261	0.210	0.063	0.176	1.000

5.3 Summary Statistics

Table 2 presents descriptive statistics for dependent and independent variables, while Table 3 presents the Pearson (lower triangle) and Spearman (upper triangle) correlations between all variables listed in Table 2. These tables are constructed for the 1:1 PSM-matched sample, so that exactly 50% each are treated and control firms.

From Table 2, we find that the average profit and cost efficiencies in the sample are 72.8% and 86.1%, respectively. The interest income and non-interest income efficiencies are similar to each other, at about 84%. The mean capital asset ratio is about 25%, while the mean liquidity ratio is about 14%. The average and median non-performing loan ratios are 1.51% and 1.62%, respectively. The mean of HHI is 0.261, suggesting moderate concentration in the industry. The medians and means are, in general, quite closely matched.

Turning now to the patent variables, the average number of granted FinTech patents ($FinTech\_2$) is 1.9, and the average number of patent applications per bank is 3.1. The FinTech dummy has a mean of 0.5, which is mechanical due to the 1:1 PSM matching. Table 3 indicates that the number of overall patents (ALL_PAT) correlates positively but not perfectly with the successful FinTech patent-related variable (the magnitude is 0.59). This implies that only some of the banks' patents can be categorized as FinTech. The correlation between the FinTech dummy and ALL_PAT is 0.42. The dummy $FinTech$ is positively correlated with all measures of efficiency. All measures of efficiency correlate positively with each other.

We also find that the non-performing loan ratio correlates negatively with profit efficiency, which is consistent with intuition. Bank size correlates weakly with profit efficiency (the magnitude is only 0.04), indicating that efficiency is not strongly related to bank scale. The $BRANCH$ variable is negatively related to profit efficiency, and only weakly correlated with other efficiency measures. This indicates that banks with a larger branch presence are not necessarily more operationally efficient. A similar argument holds for the variable $NCITY$.

In addition, we find a highly positive correlation between $NCITY$ and $SIZE$. This result is consistent with the idea that large banks have branches in more cities for geographical diversification purposes (Berger, Hasan, and Zhou 2009, Cai, Xu, and Zeng 2016). A moderately high correlation also exists between $SIZE$ and $PGDP$ indicating that larger banks are located close to prosperous population centers. The $BRANCH$ variable is positively related to the market concentration index HHI. The $NFIRM$ variable, which captures the number of firms, is strongly and inversely related to HHI. Both of these are intuitive results. We address the issue of potential multicollinearity between the control variables in Section 6.3.

Table 3: **Correlations between variables.**

This table presents the Pearson (lower triangle) and Spearman (upper triangle) correlations between variables after propensity score matching. $Prof\_eff$, $Cost\_eff$, Int_eff and $Nint\_eff$ are the four efficiency measures—profit, interest income, noninterest income, and cost efficiency, estimated by stochastic frontier analysis (SFA) with four-error components following Kumbhakar, Wang, and Horncastle (2015). Z_score, CAR, NPL, and Liq are the four risk measures—overall risk, capital asset ratio, nonperforming loan ratio and liquidity ratio. The definitions are as follows: $Z\_score = (ROA + CAR)/sd(ROA)$, where ROA is the return on assets and $sd(ROA)$ is the standard deviation of ROA during the sample period; $CAR = capital/risk\ weighted\ assets$ indicates the extent of bank shareholders' capital covering the risky assets that is reported by each bank according to regulation rules on the weight of risk assessment; NPL is the ratio of nonperforming loans to total loans; and Liq is the ratio of liquid assets to the sum of deposits and short-term debts. $FinTech$ is a dummy variable which equals unity for the treatment group and zero for the control group. The treatment group consists of banks whose number of FinTech patent applications is higher than the 75th percentile of that of all banks in a given year. The control group is the propensity-score-matched sample for the treatment group. $FinTech2$ is the number of granted FinTech patents in an year. ALL_PAT is log of the number of a bank's total patent applications in each year. The other variables are the main controls in Equation (5) defined as follows: $SIZE$ is the log of a bank's total assets (in million RMB); $BRANCH$ is the ratio of the total number of branches divided by the total assets (in million RMB); $PGDP$, POP, and $NFIRM$ are respectively the log of per capita income, the log of the population, and the number of firms in the provinces where banks principally operate; $NCITY$ is the log of the number of cities in which the bank has branches; and HHI is the Herfindahl-Hirschman Index of bank branches in each province.

	$Prof\_eff$	$Cost\_eff$	Int_eff	$Nint\_eff$	Z_Score	CAR	NPL	Liq	$FinTech$	$FinTech2$	ALL_PAT	$SIZE$	$BRANCH$	$PGDP$	POP	$NFIRM$	$NCITY$	HHI
$Prof\_eff$	–	0.360	0.361	0.601	0.214	0.080	-0.165	0.208	0.212	0.100	0.090	0.031	-0.096	0.121	-0.025	0.048	-0.011	-0.079
$Cost\_eff$	0.326	–	0.661	0.558	0.110	0.000	-0.141	-0.026	0.186	0.061	-0.004	0.035	0.025	-0.088	0.047	0.004	-0.001	0.021
Int_eff	0.305	0.657	–	0.573	0.088	0.009	-0.132	0.004	0.183	0.088	0.060	0.041	-0.002	-0.052	0.042	0.036	0.005	-0.018
$Nint\_eff$	0.562	0.598	0.595	–	0.152	0.021	-0.172	0.100	0.251	0.067	0.028	0.020	0.019	-0.031	0.048	0.021	-0.033	-0.025
Z_Score	0.160	0.102	0.098	0.148	–	0.165	-0.239	0.182	0.207	0.320	0.330	0.365	-0.267	0.265	-0.258	-0.062	0.229	-0.021
CAR	0.079	0.007	-0.003	0.002	0.119	–	-0.033	0.109	-0.003	0.040	0.051	0.158	-0.254	0.190	-0.167	-0.066	0.193	0.023
NPL	-0.253	-0.126	-0.139	-0.217	-0.166	-0.060	–	-0.219	-0.153	-0.125	-0.108	-0.069	0.121	-0.115	-0.023	-0.050	0.003	0.029
Liq	0.129	-0.010	0.039	0.081	0.156	0.136	-0.120	–	0.150	0.038	0.026	-0.039	-0.002	0.228	-0.039	0.064	-0.101	-0.066
$FinTech$	0.195	0.176	0.168	0.231	0.166	-0.005	-0.109	0.149	–	0.760	0.419	0.041	0.020	0.020	-0.020	0.037	0.016	-0.070
$FinTech2$	0.138	0.108	0.138	0.115	0.282	0.061	-0.067	0.136	0.758	–	0.558	0.392	-0.185	0.053	-0.075	0.006	0.233	-0.086
ALL_PAT	0.086	0.061	0.101	0.051	0.264	0.099	-0.032	0.121	0.421	0.589	–	0.555	-0.295	0.181	-0.142	0.046	0.451	-0.139
$SIZE$	0.037	0.045	0.061	0.042	0.375	0.183	-0.033	0.017	0.055	0.413	0.544	–	-0.555	0.362	-0.238	0.030	0.810	-0.084
$BRANCH$	-0.084	0.029	-0.009	-0.003	-0.245	-0.239	0.121	-0.099	-0.016	-0.175	-0.220	-0.394	–	-0.556	0.361	-0.022	-0.618	0.157
$PGDP$	0.123	-0.071	-0.044	-0.003	0.278	0.239	-0.125	0.164	0.020	0.079	0.168	0.362	-0.419	–	-0.414	0.227	0.284	-0.488
POP	-0.030	0.026	0.040	0.028	-0.308	-0.161	0.000	-0.056	-0.027	-0.082	-0.124	-0.233	0.227	-0.388	–	0.661	-0.311	-0.342
$NFIRM$	0.068	0.026	0.056	0.049	-0.119	0.003	-0.065	0.034	0.011	-0.014	0.022	0.037	-0.053	0.266	0.716	–	-0.106	-0.769
$NCITY$	0.021	0.001	0.021	-0.010	0.280	0.222	0.001	0.012	0.035	0.262	0.466	0.858	-0.490	0.353	-0.333	-0.082	–	0.057
HHI	-0.077	-0.003	-0.038	-0.045	-0.058	-0.049	-0.020	-0.070	-0.029	-0.064	-0.079	-0.144	0.250	-0.552	-0.285	-0.685	-0.093	–

6 Empirical Results

In this section, we first present our baseline results using the PSM-matched sample. Subsequently, to further reduce endogeneity, we perform a DID analysis. Finally, we document additional results from checking robustness with alternative measures of FinTech, bank efficiency, and risk, as well as other control variables based on principal components analysis and financial inclusion.

6.1 Baseline Regressions

We estimate Equation (5) via regression. The four efficiency measures are used alternatively as the dependent variable; Table 4 presents the results. As Section 5 shows, we use $FinTech$ and ALL_PAT as the key independent variables to compare the differences across effects from FinTech and those from overall innovation. The first four columns in Table 4 report $FinTech$'s effects, while the last four columns report the results using ALL_PAT.

According to Table 4's first four columns, $FinTech$ has a significant positive association with all four efficiency measures, including profit ($Prof\_eff$), cost ($Cost\_eff$), interest income (Int_eff), and noninterest income ($Nint\_eff$) efficiencies. The magnitudes of these coefficients indicate an economically significant influence. Specifically, compared with the banks in the control group, the banks' four efficiency indicators in the high FinTech group are larger by 0.039, 0.017, 0.021, and 0.030 on average, respectively, relative to the corresponding numbers for the low FinTech group. The magnitudes of such improvement are more than 35% of $Profit\_eff$ and $Cost\_eff$'s standard deviations, 45% of Int_eff's standard deviation, and 65% of $Nint\_eff$'s standard deviation (see Table 2). In addition, Columns (5) to (8) show that the overall innovation (ALL_PAT) measured by the number of applications for all patents (not just FinTech ones) is insignificantly associated with the efficiency indicators.

Furthermore, the effects of control variables on banks' efficiency are consistent with the existing literature. Specifically, in Columns (1) and (2) of Table 4, the $SIZE$ coefficients are significantly positive, which indicates that large banks are more profitable than small banks (Huang et al. 2020, He, Ho, and Xu 2020). The $BRANCH$ coefficients in Columns (2) and (3) are negative, which indicates that banks with more branches have higher costs and lower interest income. The $NCITY$ coefficients are significantly negative in Columns (1) to (4), which implies that geograph-

Table 4: **FinTech's impact on banks' efficiency.**
This table reports regression results from Equation (5) when the dependent variable is, alternatively, four efficiency measures, namely, profit ($Prof\_eff$), cost ($Cost\_eff$), interest income (Int_eff), and noninterest income ($Nint\_eff$). We estimate these efficiency measures by stoachastic frontier analysis (SFA) with four error components following Kumbhakar, Wang, and Horncastle (2015). $FinTech$ is a dummy indicating whether a bank's number of FinTech patent applications is higher than the 75th percentile of that of all banks in each year. ALL_PAT is the log of the number of a bank's total patent applications in each year. The main control variables are defined as follows: $SIZE$ is the log of a bank's total assets (in million RMB); $BRANCH$ is the ratio of the total number of branches divided by the total assets (in million RMB); $PGDP$, POP, and $NFIRM$ are respectively the log of per capita income, the log of the population, and the number of firms in the provinces where banks principally operate; $NCITY$ is the log of the number of cities in which the bank has branches; and HHI is the Herfindahl-Hirschman Index of bank branches in each province. We include dummy variables representing bank type and year. The estimated standard errors are in parentheses. Here ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Prof\_eff$	$Cost\_eff$	Int_eff	$Nint\_eff$	$Prof\_eff$	$Cost\_eff$	Int_eff	$Nint\_eff$
$FinTech$	0.039***	0.017***	0.021**	0.030***				
	(0.008)	(0.004)	(0.009)	(0.014)				
ALL_PAT					-0.003	-0.001	0.001	-0.002
					(0.003)	(0.005)	(0.002)	(0.003)
$SIZE$	0.019**	0.016***	0.014***	0.013***	0.024***	0.020***	0.016***	0.018***
	(0.008)	(0.003)	(0.003)	(0.004)	(0.009)	(0.004)	(0.003)	(0.004)
$BRANCH$	-0.009	-0.003**	-0.002***	-0.003	-0.008	-0.004***	-0.003**	-0.003
	(0.011)	(0.001)	(0.000)	(0.005)	(0.012)	(0.001)	(0.001)	(0.006)
$PGDP$	-0.057	-0.077***	-0.069***	-0.068***	-0.067*	-0.084***	-0.075***	-0.079***
	(0.037)	(0.016)	(0.015)	(0.018)	(0.040)	(0.017)	(0.016)	(0.020)
POP	-0.044*	-0.044***	-0.038***	-0.032***	-0.053**	-0.050***	-0.042***	-0.041***
	(0.023)	(0.010)	(0.010)	(0.011)	(0.025)	(0.011)	(0.010)	(0.012)
$NFIRM$	0.041**	0.031***	0.031***	0.023**	0.048**	0.036***	0.035***	0.030***
	(0.019)	(0.008)	(0.008)	(0.009)	(0.020)	(0.009)	(0.008)	(0.010)
$NCITY$	-0.018**	-0.003***	-0.005***	-0.005*	-0.023*	-0.005	-0.006**	-0.007
	(0.007)	(0.000)	(0.001)	(0.003)	(0.014)	(0.006)	(0.003)	(0.007)
HHI	-0.018	-0.064***	-0.049***	-0.055***	-0.015	-0.067***	-0.050***	-0.058**
	(0.044)	(0.019)	(0.018)	(0.021)	(0.046)	(0.020)	(0.019)	(0.023)
Constant	1.265**	1.753***	1.544***	1.591***	1.339**	1.795***	1.597***	1.685***
	(0.568)	(0.244)	(0.229)	(0.271)	(0.608)	(0.259)	(0.241)	(0.300)
Bank type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	704	704	704	704	704	704	704	704
$Adj.R^2$	0.124	0.194	0.145	0.181	0.138	0.163	0.178	0.111

ical diversification reduces banks' efficiencies (Berger, Hasan, and Zhou 2010).

Table 5 presents the results of Equation (5) when the dependent variables are the risk measures discussed in Section 4.3. Similar to Table 4, the key independent variable is *FinTech* in Columns (1) to (4) and *ALL_PAT* in Table 5's last four columns. We find that *FinTech*'s effects on *Z_score*, *NPL*, and *Liq* are all significant at the 1% level. In terms of economic significance, *Z_score* and *Liq* in the treatment group are larger by 0.684 and 0.129 on average, respectively, compared to the control group. The coefficient magnitudes are approximately 36% and 39% of

Table 5: **FinTech's impact on banks' risk.**
This table reports regression results from Equation (5) when the dependent variable is, alternatively, four risk measures, including overall risk (*Z_score*), capital asset ratio (*CAR*), non-performing loan ratio (*NPL*), and liquidity ratio (*Liq*). The definitions are as follows: $CAR = capital/risk\ weighted\ assets$ indicates the extent of bank shareholders' capital covering the risky assets that is reported by each bank according to regulation rules on the weight of risk assessment; $Z\_score = (ROA + CAR)/sd(ROA)$, where ROA is the return on assets and $sd(ROA)$ is the standard deviation of ROA during the sample period; NPL is the ratio of nonperforming loans to total loans; and Liq is the ratio of liquid assets to the sum of deposits and short-term debts. *FinTech* is a dummy indicating whether a bank's number of FinTech patent applications is higher than the 75th percentile of all banks in each year. *ALL_PAT* is the log of a bank's total patent applications in each year. The main control variables are defined as follows: *SIZE* is the log of a bank's total assets (in million RMB); *BRANCH* is the ratio of the total number of branches divided by the total assets (in million RMB); *PGDP*, *POP*, and *NFIRM* are respectively the log of per capita income, the log of the population, and the number of firms in the provinces where banks principally operate; and *HHI* is the Herfindahl-Hirschman Index of bank branches in each province. We include dummy variables representing bank type and year. The estimated standard errors are in parentheses. Here ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Z_score	*CAR*	*NPL*	*Liq*	*Z_score*	*CAR*	*NPL*	*Liq*
FinTech	1.069***	0.015	-0.173***	3.729***				
	(0.371)	(0.037)	(0.042)	(1.134)				
ALL_PAT					-0.059	0.012	-0.004	0.029*
					(0.062)	(0.013)	(0.017)	(0.017)
SIZE	0.583***	0.095	-0.039	0.129**	0.559***	0.034	-0.044	0.113**
	(0.149)	(0.077)	(0.041)	(0.053)	(0.158)	(0.028)	(0.037)	(0.046)
BRANCH	-0.623***	-0.061	0.190***	-0.052	-0.622***	-0.063	0.188***	-0.053
	(0.226)	(0.047)	(0.060)	(0.052)	(0.229)	(0.048)	(0.061)	(0.052)
PGDP	1.763**	0.259*	-0.636***	-0.026	1.659**	0.252	-0.609***	-0.050
	(0.741)	(0.156)	(0.202)	(0.184)	(0.751)	(0.158)	(0.206)	(0.186)
POP	-0.506	0.134	-0.236*	-0.009	-0.595	0.130	-0.208	-0.023
	(0.468)	(0.098)	(0.125)	(0.110)	(0.473)	(0.099)	(0.128)	(0.111)
NFIRM	-0.183**	-0.117***	-0.001	0.027	-0.104*	-0.114	0.025***	0.035
	(0.084)	(0.043)	(0.102)	(0.086)	(0.057)	(0.083)	(0.008)	(0.086)
NCITY	-0.220	-0.120**	0.022	-0.058	-0.245	-0.124**	0.025	-0.063
	(0.256)	(0.054)	(0.070)	(0.065)	(0.260)	(0.055)	(0.072)	(0.065)
HHI	0.199	0.317*	-0.744***	-0.082	0.269	0.333*	-0.754***	-0.069
	(0.879)	(0.184)	(0.232)	(0.197)	(0.890)	(0.186)	(0.236)	(0.199)
Constant	-12.813	-0.854	12.901***	2.085	-10.624	-0.614	12.479***	2.775
	(11.318)	(2.385)	(3.064)	(2.772)	(11.487)	(2.417)	(3.143)	(2.813)
Bank type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	704	704	704	704	704	704	704	704
$Adj.R^2$	0.298	0.174	0.192	0.188	0.282	0.172	0.111	0.183

the corresponding variables' standard deviations (from Table 2). Also, banks with high FinTech innovation have lower NPL by -0.173, which is approximately 26% of NPL's standard deviation in Table 2. These results imply that FinTech innovation leads to economically and statistically significant improvements in the Z_score, NPL, and Liq metrics.

There is no significant effect of FinTech on CAR. This makes sense based on the discussion in Section 2.3. Bank managers prefer to meet the supervisory requirements as a threshold rather than hold a larger CAR, because it closely relates to capital cost. Having to hold too much capital to cover risk assets leads to a capital cost that is too high; for this reason, bank managers prefer to maximize revenue or engage in innovation, given regulators' lower bound requirement. Thus, this result does not change our primary insight on how FinTech affects risk management. Further, as opposed to Table 4, where the overall innovation (ALL_PAT) shows no significant improvement, Table 5 shows significant improvement on Liq—that is, increasing high-quality liquid assets to meet short-term obligations.

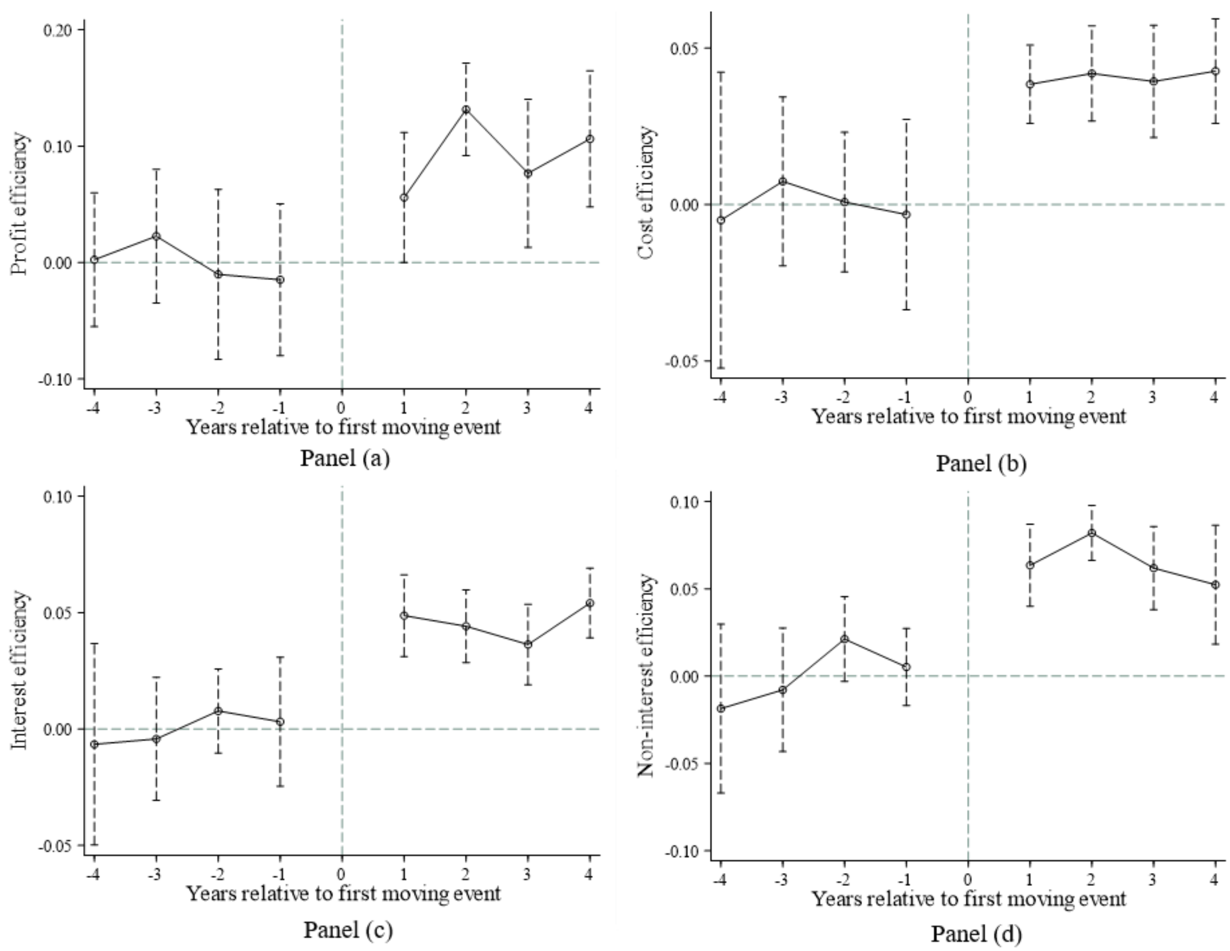
FinTech's significant positive effect on the Z-score implies that FinTech innovations decrease banks' overall risk. This is consistent with the argument that banks with higher FinTech innovation let banks collect more types of information and acquire more customers by facilitating access to funds (Buchak et al. 2018, Huang et al. 2020). The negative association between $FinTech$ and NPL is consistent with the conventional logic that FinTech innovation reduces credit risk. For example, more quantifiable hard data can aid credit evaluation, reduce information asymmetry, and optimize management processes. Also, our findings are consistent with the notion that FinTech innovation facilitates customers' access to financial services, which in turn fosters customers' preference for short-term loans (Jakšič and Marinč 2019), and thus increases banks' liquid assets. As Columns (1) and (4) of Table 5 show, the $SIZE$ coefficients indicate that large banks possess low levels of overall and liquidity risk, which is consistent with Cheng, Geng, and Zhang (2016).

6.2 Robustness Check Using DID Analysis

To reduce endogeneity concerns in Equation (5), we next perform a difference-in-differences (DID) analysis. First, we identify banks that initially move into the high FinTech group as defined in Section 5. Specifically, a bank year is defined as a *first move event* if a bank is in the high FinTech group in year t (but not in the prior years). We include these bank years in the treatment group of our DID analysis. The years t to $t+3$ are the "post-period" for each bank-year event. Second, for each bank in a first move event, we use PSM on the other banks to construct a control bank year, as Section 5 discusses. We use two event years (2014 and 2015) for our DID to allow for a sufficiently long post-period within our sample that ends in 2019. For these two years, we identify 25 and 12 banks in the treatment group, respectively.

We first conduct a parallel trend analysis for treatment and control group banks. Following Beck, Levine, and Levkov (2010), we plot the average treatment effect before and during the treatment period. The banks entering the treatment group act as the baseline. Figure 3 presents the analysis for banks' efficiency. The figure shows that there is no significant difference in effi-

Figure 3: **Parallel trend of banks' efficiency measures.**
We plot the yearly average trend of the four efficiency measures before and after the *first move event*, which is defined as a bank year such that a bank is in the high FinTech group in year t (but not prior). The high FinTech group includes the banks with FinTech patent applications above the 75th percentile of that of all banks in each year. Panels (a), (b), (c), and (d) show yearly cross-sectional averages of profit ($Prof\_eff$), cost ($Cost\_eff$), interest income (Int_eff), and noninterest income ($Nint\_eff$) efficiencies, respectively. The dashed lines passing through the circles represent 95% confidence intervals, adjusted for bank-level clustering. The vertical, dashed line extending from 0 on the horizontal axis indicates the *first moving event*.



ciency between the treatment and control group banks before the treatment years (from $t-4$ to $t-1$). However, treatment group banks (that first move into the high FinTech innovation group) reap greater profit, cost, interest income, and noninterest income efficiency than control banks in the following treatment years (from t to $t+3$).

Figure 4 presents the results of a parallel trend analysis for banks' risk. As the figure shows, before the first move event, none of the four risk measures differ significantly between treatment and control banks. However, after the event, banks in the treatment group have a significantly higher Z-score and liquidity—and lower NPL ratio (except for one observation in year $t = 3$)—than other banks. Consistent with Table 5, the results are insignificant for CAR.

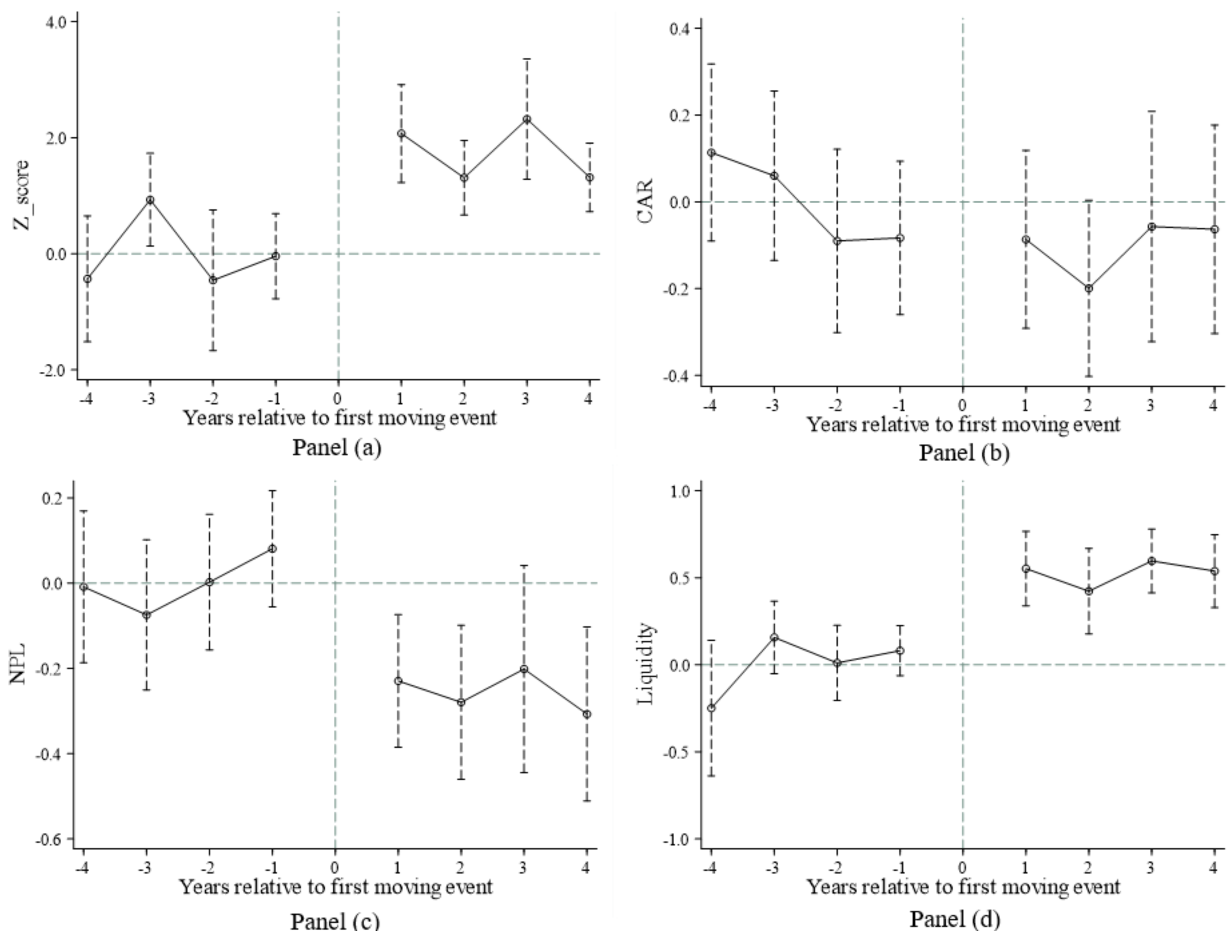
Lastly, we run a DID regression after adding three independent variables to Equation (5)—namely $Treat$, $Post$, and $Treat \times Post$. $Treat$ is a dummy variable indicating whether a bank is in the treatment group. $Post$ is the dummy variable indicating whether the year is after the bank-year event. We are interested in the DID term $Treat \times Post$. Table 6 presents the results. Consistent with Figures 3 and 4, the DID term is significant, which implies that banks' efficiency and risk (except for CAR) improve significantly after FinTech innovation, compared to their controls.

6.3 Additional Robustness Checks

Next, we discuss additional robustness checks. First, we replace the number of FinTech patent applications with the number of granted ones. The literature (see p. 2065 and the references in Lerner and Seru 2022) indicates that observed patent grants in the U.S. often undergo a review of more than two years—thus, considering only the number of granted patents could truncate the data. On the other hand, the number of patent applications could inflate measurements for innovation activities, such as in Lerner and Seru (2022). To address these concerns, we use the number of granted FinTech patents ($FinTech2$) in place of total patents. Panel A of Table 7 reports the results, which are consistent with those in Tables 4 and 5. The findings demonstrate that our results for the FinTech innovation's effect on bank efficiency and risk are robust to concerns about whether the proxy is truncated or inflated.

The variables $SIZE$, $BRANCH$, $PGDP$, POP, $NFIRM$, $NCITY$ and HHI are potentially related to each other (see Table 3). To address any multicollinearity issues, we employ PCA to extract these variables' principal components at the bank level. In untabulated results we find

Figure 4: **Parallel trend in banks' risk.**
We plot the yearly average trend of the four efficiency measures before and after the *first move event*, which is defined as a bank year such that a bank is in the high FinTech group in year t (but not prior). The high FinTech group includes the banks with FinTech patent applications above the 75th percentile of that of all banks in each year. Panels (a), (b), (c), and (d) show yearly cross-sectional averages of overall risk (Z_score), the capital asset ratio (CAR), the nonperforming loans ratio (NPL), and the liquidity ratio (Liq), respectively. The dashed lines passing through the circles represent 95% confidence intervals, adjusted for bank-level clustering. The vertical, dashed line extending from 0 on the horizontal axis indicates the *first moving event*.



that the first two components (denoted by $Comp1$ and $Comp2$) explain 94% of the total variance among the seven variables. We replace the aforementioned seven variables with $Comp1$ and $Comp2$ in Equation (5). Panel B of Table 7 reports the findings, which are consistent with those in Tables 4 and 5.

Berger, Hasan, and Zhou (2009) indicate that the efficiency level is determined by comparing actual costs to best-practice minimum costs to produce the same output under the same condi-

Table 6: **DID analysis of FinTech innovation's impact.**
This table reports the DID results for FinTech innovation's impact on banks' efficiency and risk over 2010–2019. *Treat* is a dummy variable indicating whether a bank is in the treatment group, which includes banks with high FinTech innovation. *Post* is a dummy variable indicating whether the year of the observation is after the *first move event*, defined as a bank year such that a bank is in the high FinTech group in year t, but not in this group prior. The high FinTech group includes banks with a higher number of FinTech patent applications above the 75th percentile of that of all banks in each year. *Treat* × *Post* is the DID term of our interest. The main control variables are as follows: $SIZE$ is the log of a bank's total assets (in million RMB); $BRANCH$ is the ratio of the total number of branches divided by the total assets (in million RMB); $PGDP$, POP, and $NFIRM$ are respectively the log of per capita income, the log of the population, and the number of firms in the provinces where banks principally operate; and HHI is the Herfindahl-Hirschman Index of bank branches in each province. We include dummy variables to control fixed effects of bank type heterogeneities and potentially linear trends of bank efficiency due to technological progress over the sample period 2010–2019. The estimated standard errors are in parentheses. Here ***, **, and * denote the significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Prof\_eff$	$Cost\_eff$	Int_eff	$Nint\_eff$	Z_score	CAR	NPL	Liq
Treat	-0.048**	-0.002	0.003	-0.007	0.507	0.141	-0.080	0.010
	(0.022)	(0.012)	(0.011)	(0.011)	(0.385)	(0.094)	(0.063)	(0.084)
Post	0.021	0.027	0.015	0.013	0.218	0.192	0.121*	0.256*
	(0.034)	(0.019)	(0.018)	(0.014)	(0.527)	(0.117)	(0.069)	(0.145)
Treat × *Post*	0.102***	0.025**	0.026**	0.045***	1.240***	-0.083	-0.426***	0.330***
	(0.027)	(0.012)	(0.012)	(0.012)	(0.439)	(0.107)	(0.080)	(0.098)
$SIZE$	0.020**	0.013***	0.012***	0.010**	0.365***	0.091***	-0.036	0.009
	(0.009)	(0.004)	(0.004)	(0.004)	(0.116)	(0.028)	(0.024)	(0.029)
$BRANCH$	-0.004	-0.003	-0.001	-0.003	-0.674***	0.034	-0.119***	-0.124**
	(0.012)	(0.006)	(0.006)	(0.006)	(0.210)	(0.043)	(0.040)	(0.062)
$PGDP$	-0.041	-0.047***	-0.053***	-0.042*	3.507***	0.074	-0.290	-0.133
	(0.054)	(0.016)	(0.015)	(0.024)	(1.056)	(0.204)	(0.190)	(0.159)
POP	-0.006	-0.018**	-0.020**	0.000	1.158**	-0.000	-0.071	-0.074
	(0.039)	(0.009)	(0.009)	(0.014)	(0.559)	(0.126)	(0.111)	(0.105)
$NFIRM$	0.009	0.014	0.017*	-0.005	-1.165**	-0.091	-0.149*	0.191**
	(0.033)	(0.009)	(0.009)	(0.012)	(0.481)	(0.108)	(0.089)	(0.091)
$NCITY$	0.000	-0.000	-0.000	0.000	0.002	-0.003**	-0.002*	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.007)	(0.002)	(0.001)	(0.001)
HHI	-0.047	-0.036*	-0.040**	-0.041	1.210	0.202	-0.490***	0.096
	(0.040)	(0.021)	(0.020)	(0.025)	(0.799)	(0.171)	(0.166)	(0.177)
Constant	0.803	1.212***	1.253***	1.132***	-42.838***	1.878	8.644***	2.246
	(0.830)	(0.229)	(0.227)	(0.341)	(14.780)	(2.951)	(2.766)	(2.338)
Bank type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	510	510	510	510	510	510	510	510
$Adj.R^2$	0.196	0.154	0.192	0.178	0.276	0.185	0.195	0.232

tions. However, changing economic conditions are less comparable over time and imperfectly controlled via the year dummy variables we use. We therefore use scaled bank rankings in place of the actual efficiency and risk measures. Specifically, in each year, we assign the bank with the minimum efficiency or risk level the lowest scaled rank, and vice versa. We then assign ranks in increasing order, and scale each ranking by the number of banks in each year. This measure has the benefit of being more comparable across time. We report the results in Panel C of Table 7. The results are similar to those in Tables 4 and 5.

Table 7: **Additional robustness tests.**
This table reports the results of additional robustness tests. In Panel A, we change the number of FinTech patent applications ($FinTech$) to the number of grants ($FinTech2$) to address the concern of whether the FinTech proxy is truncated or inflated. In Panel B, we address the possible issue of multicollinearity among $SIZE$, $BRANCH$, $PGDP$, POP, $NFIRM$, $NCITY$ and HHI. We employ principal component analysis (PCA) and extract the first two components ($Comp1$ and $Comp2$) across these variables; these components explain 94% of the total variance. We check the robustness by replacing $(SIZE, BRANCH, PGDP, POP, NFIRM, NCITY, HHI)$ with $(Comp1, Comp2)$ in Equation (5). In Panel C, we use the scaled rankings of a bank's efficiency and risk measures as the dependent variables. We rank bank efficiency and risk within each year, and scale the ranking by the number of banks in that year. The scaled bank rankings are distributed between 0 and 1. In Panel D, we address the concern of omitted variable bias and add the Digital Financial Inclusion Index to the controls in Equation (5). This index measures financial inclusion ($Inclusion$) in other FinTech institutions and covers 2011-–2018. The estimated coefficients of other controls are similar to those in Tables 4 and 5, so we omit them here. Standard errors are in parentheses. Here ***, **, and * denote the significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Prof\_eff$	$Cost\_eff$	Int_eff	$Nint\_eff$	Z_score	CAR	NPL	Liq
	Panel A: Patent granted as the measurement of FinTech (#Obs. = 704)							
$FinTech2$	0.021***	0.005**	0.006***	0.010***	0.399***	0.017	-0.066***	1.170***
	(0.004)	(0.002)	(0.002)	(0.002)	(0.089)	(0.019)	(0.024)	(0.321)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$Adj.R^2$	0.163	0.172	0.188	0.077	0.294	0.173	0.113	0.090
	Panel B: PCA (#Obs. = 704)							
$FinTech$	0.040***	0.019***	0.018***	0.031***	0.681***	0.017	-0.168***	3.129***
	(0.008)	(0.006)	(0.003)	(0.004)	(0.171)	(0.037)	(0.042)	(1.034)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$Adj.R^2$	0.172	0.173	0.183	0.111	0.297	0.184	0.124	0.088
	Panel C: Alternative measures of bank efficiency and risk (#Obs. = 704)							
$FinTech$	0.105***	0.110***	0.111***	0.164***	0.103***	0.015	-0.090***	0.070***
	(0.021)	(0.022)	(0.022)	(0.022)	(0.020)	(0.021)	(0.019)	(0.018)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$Adj.R^2$	0.110	0.087	0.106	0.129	0.227	0.207	0.147	0.265
	Panel D: Additional control variable of financial inclusion (#Obs. = 704)							
$FinTech$	0.039***	0.023***	0.017***	0.043***	0.686***	0.019	-0.128***	2.128***
	(0.008)	(0.008)	(0.004)	(0.013)	(0.172)	(0.035)	(0.041)	(0.717)
$Inclusion$	0.034	-0.088***	0.062***	-0.054**	0.254	0.080	0.510**	-0.146*
	(0.048)	(0.021)	(0.021)	(0.025)	(1.013)	(0.208)	(0.246)	(0.083)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$Adj.R^2$	0.175	0.151	0.115	0.125	0.298	0.174	0.128	0.188

The development of nonbank financial institutions in China, including Ant Financial Services Group, WeBank of Tencent Company, and Du Xiaoman Financial (or Baidu Financial Services Group), is an important phenomenon in recent years. These institutions provide affordable financial services such as loans and savings products, and compete aggressively with traditional banks. In response to these developments, traditional banks further their financial inclusion efforts to compete (e.g., Chauvet and Jacolin 2017, Banna et al. 2020), and the impact of FinTech on banks' efficiency might be capturing this competition. To address this concern, we use the Digital Financial Inclusion Index (published by Peking University)[22] as a proxy for the development of financial inclusion in China. Panel D of Table 7 considers this index as a control variable, and shows that our main findings are unaltered.

7 Cross-Sectional Heterogeneities

Next, we consider managerial skills and labor intensity as cross-sectional hetereogeneities that could differentially impact the effect of FinTech on banks. We first present the relevant economic mechanisms and then discuss the empirical results.

First, managerial expertise helps enhance the efficiency of newly adopted technologies and the reaping of potential benefits (Shu and Strassmann 2005). More specifically, He, Ho, and Xu (2020) argue that banks with better managerial skills could better utilize online banking to save costs and reduce risk. Second, FinTech innovations can significantly streamline work procedures and thus save labor and office space costs. For example, technologically savvy employees could identify valuable depositors by acquiring more quantifiable hard information using FinTech. Indeed, Shu and Strassmann (2005) find that new technology adoption leads to significant positive marginal gains to labor. In addition, FinTech can influence bank risk via synergies between hard and soft information. Specifically, loan officers can combine hard information generated by FinTech with soft information gathered from in-person meetings with borrowers (DeYoung, Lang, and Nolle 2007). Based on these arguments, we expect that banks that make better use of their labor force will benefit more from FinTech innovations in terms of efficiency and risk management.

[22] The index is available at https://tinyurl.com/yvcekbya.

We analyze Fintech's heterogeneous effect on bank efficiency and risk as follows:

$$y_{it} = \beta_0 + \beta_1 FinTech_{it} + \beta_2 FinTech_{it} \times HeteroFactor_{it} + \sum_j \delta_j Control_{ijt} + \varepsilon_{it}\,, \tag{6}$$

where the control variables ($Control$) are the same as in Equation (5); and $HeteroFactor$ is alternatively managerial ability (MA) and labor intensity (LI), the heterogeneous factors of interest. We measure MA as a part of the banks' efficiency that cannot be explained by corporate governance and accounting characteristics (Demerjian, Lev, and McVay 2012). In particular, we perform a Tobit regression of profit efficiency on the characteristics in Equation (5), and the residuals of this regression measure MA. We measure LI as the ratio of the number of employees to total assets.

Panel A of Table 8 shows the moderating effect of managerial ability (MA) on the association of FinTech innovation with bank efficiency and risk. The regression does not include the moderating effect of MA for profit efficiency because we use it to estimate MA. In columns (2)-(4), the coefficients indicating the interaction of FinTech and managerial ability ($FinTech \times MA$) are significantly positive. Thus, the efficiency of banks with high managerial ability benefits more from FinTech innovation. For the risk measures, MA has a significant and negative moderating effect on CAR, which seems to contradict arguments surrounding FinTech's insignificant effect on CAR in Table 5 and 6. However, this makes sense because synergies between FinTech and managerial expertise can increase available capital for effective operation; in turn, this means less capital is tied up in covering risk assets. We also find a positive effect of FinTech on the Z_score and negative effect on NPL via the MA channel, and both are significant and consistent with our main idea that the higher the managerial expertise and FinTech innovation, the better the risk measures.

Panel B of Table 8 shows that with high FinTech innovation, all efficiency measures significantly increase in labor intensity, LI. This implies that banks with higher labor intensity benefit more from FinTech innovation. For the risk measures, LI increases the positive effect of FinTech innovation on Z_score and decrease its effect on NPL. Overall, the results imply that FinTech tends to reduce risk and improve operating efficiency for banks with higher labor intensity.

Table 8: **Heterogeneity analysis.**
This table reports the results of the moderating effect of managerial ability (MA) and labor intensity (LI) on the association of FinTech innovation with four efficiency measures and four risk measures. We estimate MA by performing a Tobit regression of profit efficiency on a set of bank-level characteristics as used in Equation (5). The residuals of such Tobit regression indicate MA as done in Demerjian et al. (2013). We measure LI as the ratio of the employee number to total assets. Panels A and B report the moderating effect of MA and LI, respectively, where $FinTech \times MA$ and $FinTech \times LI$ are the terms of our interest. Column (1) in Panel A is blank because we use profit efficiency to estimate MA. Similar to Tables 4 and 5, $FinTech$ is a dummy variable indicating whether a bank's number of FinTech patent applications is higher than the 75th percentile of that of all banks in each year. Also, Panels A and B both include four efficiency measures and four risk measures. The efficiency measures include profit ($Prof\_eff$), cost ($Cost\_eff$), interest income (Int_eff), and noninterest income ($Nint\_eff$). We estimate these efficiency measures by stochastic frontier analysis (SFA) with four-error components following Kumbhakar, Wang, and Horncastle (2015). The risk measures include overall risk (Z_score), capital asset ratio (CAR), nonperforming loan ratio (NPL), and liquidity ratio (Liq). The definitions are as follows: $Z\_score = (ROA + CAR)/sd(ROA)$, where ROA is the return on assets and $sd(ROA)$ is the standard deviation of ROA during the sample period; $CAR = equity/risk\ assets$, which indicates the extent of bank shareholders' capital covering its risky assets; NPL is the ratio of nonperforming loans to total loans; and Liq is the ratio of liquid assets to current total liabilities, such as the sum of deposits and short-term debts. The estimated results of the other controls are mostly similar to those in Tables 4 and 5, so we omit them here. The sample period covers 2010–2019. The estimated standard errors are in parentheses. Here ***, **, and * denote the significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Prof\_eff$	$Cost\_eff$	Int_eff	$Nint\_eff$	Z_score	CAR	NPL	Liq
	Panel A: Moderate effect of managerial ability (#Obs. = 704)							
$FinTech$	–	0.017***	0.016***	0.029***	0.673***	0.018	-0.170***	0.129***
	–	(0.003)	(0.004)	(0.004)	(0.169)	(0.021)	(0.041)	(0.034)
$FinTech \times MA$	–	0.152***	0.128***	0.285***	4.213***	-0.167*	-1.602***	0.338
	–	(0.023)	(0.024)	(0.027)	(1.151)	(0.092)	(0.278)	(0.222)
Controls	–	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank type	–	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	–	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$Adj.R^2$	–	0.153	0.153	0.143	0.252	0.308	0.184	0.185
	Panel B: Moderate effect of labor intensity (#Obs. = 704)							
$FinTech$	0.246***	0.054**	0.046**	0.079***	0.687*	0.026	-0.517*	0.616***
	(0.054)	(0.024)	(0.021)	(0.028)	(0.379)	(0.031)	(0.289)	(0.233)
$FinTech \times LI$	0.601***	0.149***	0.150*	0.229**	4.192***	0.617	-1.451**	1.169
	(0.113)	(0.050)	(0.089)	(0.102)	(1.398)	(0.499)	(0.603)	(1.486)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank type	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$Adj.R^2$	0.118	0.199	0.125	0.140	0.322	0.181	0.139	0.117

8 Conclusion

The impact of FinTech on conventional banks remains an issue of importance, particularly because most people in developing countries rely on such banks for common transactions such as deposits, loans, and settlements. Because of its separated banking policies, China provides a desirable setting to test FinTech innovation's impact on traditional banking. Additionally, measuring an individual bank's FinTech innovation is facilitated by China's institutional environment. This is because China's patent data include both patent applications as well as granted patents, and are therefore not subject to the truncation issues surrounding the latter (Lerner and Seru 2022). Keeping these observations in mind, we collect Chinese banks' patent filings and use a novel method that goes "all-in" machine learning to develop a novel proxy for FinTech innovation. Specifically, we compute our measure via the degree of cosine similarity between patent filing texts and well-established FinTech policy documents. This measure is directly related to a bank's innovation capacity and outcomes, and thus it is not prone to advertisers artificially inflating it, unlike a measure derived from online search volume.

After controlling for other variables and using propensity-score matching, we show that FinTech innovation significantly improves banks' profit, cost, interest income, and noninterest income efficiencies. We also show that such innovation is associated with a decreased non-performing loan ratio and and an increased liquid asset ratio. The regression results consistently point to the beneficial effects of FinTech, and are economically and statistically significant. For stronger identification, we perform a DID analysis that confirms our main findings. Thus, overall, our results provide strong and reliable support to the notion that FinTech innovation improves traditional banks' operational efficiency and reduces the riskiness of their assets. Our heterogeneity analysis shows that the beneficial effects of FinTech on banks' efficiency and risk are more pronounced for banks with higher levels of managerial skill and labor intensity.

There have been a number of papers discussing the pros and cons of FinTech and its regulation (e.g., Magnuson 2018, Omarova 2020, and Huang 2021). We find compelling evidence that such innovation increases operational efficiency and reduces operating risks of traditional banking, which is the mainstay of the worldwide financial sector (Dietrich and Wanzenried 2014). Our results thus support the conclusion of Vives (2017), that efforts to rein in FinTech should be approached cautiously and deliberately.

References

Allen, L. and A. Rai (1996). Operational efficiency in banking: An international comparison, *Journal of Banking & Finance* **20**(4): 655–672.

Arner, D. W., J. Barberis, and R. P. Buckley (2015). The evolution of FinTech: A new post-crisis paradigm?, *Georgetown Journal of International Law* **47**: 1271.

Banna, H., M. R. Alam, R. Ahmad, and N. M. Sari (2020). Does financial inclusion drive the Islamic banking efficiency? A post-financial crisis analysis, *Singapore Economic Review* pp. 1–26.

Barbiroli, G. (1997). *The dynamics of technology: A methodological framework for techno-economic analyses*, Vol. 25, Springer Science & Business Media.

Beck, T., R. Levine, and A. Levkov (2010). Big Bad Banks? The winners and losers from bank deregulation in the United States, *Journal of Finance* **65**(5): 1637–1667.

Bellstam, G., S. Bhagat, and J. A. Cookson (2021). A text-based analysis of corporate innovation, *Management Science* **67**(7): 4004–4031.

Beltratti, A. and R. Stulz (2012). The credit crisis around the globe: Why did some banks perform better?, *Journal of Financial Economics* **105**(1): 1–17.

Berger, A. N. and L. J. Mester (1997). Inside the black box: What explains differences in the efficiencies of financial institutions?, *Journal of Banking & Finance* **21**(7): 895–947.

Berger, A. N., I. Hasan, and M. Zhou (2009). Bank ownership and efficiency in China: What will happen in the world's largest nation?, *Journal of Banking & Finance* **33**(1): 113–130.

Berger, A. N., I. Hasan, and M. Zhou (2010). The effects of focus versus diversification on bank performance: Evidence from Chinese banks, *Journal of Banking & Finance* **34**(7): 1417–1435.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru (2018). Fintech, regulatory arbitrage, and the rise of shadow banks, *Journal of Financial Economics* **130**(3): 453–483.

Cai, W., F. Xu, and C. Zeng (2016). Geographical diversification and bank performance: Evidence from China, *Economics Letters* **147**: 96–98.

Chauvet, L. and L. Jacolin (2017). Financial inclusion, bank concentration, and firm performance, *World Development* **97**: 1–13.

Chen, M. A., Q. Wu, and B. Yang (2019). How valuable is FinTech innovation?, *Review of Financial Studies* **32**(5): 2062–2106.

Chen, R.-R., T.-H. Yang, and S.-K. Yeh (2017). The liquidity impact on firm values: The evidence of Taiwan's banking industry, *Journal of Banking & Finance* **82**: 191–202.

Cheng, M. and Y. Qu (2020). Does bank FinTech reduce credit risk? Evidence from China, *Pacific-Basin Finance Journal* **63**: 101398.

Cheng, M., H. Geng, and J. Zhang (2016). Chinese commercial banks: Benefits from foreign strategic investors?, *Pacific-Basin Finance Journal* **40**: 147–172.

Demerjian, P., B. Lev, and S. McVay (2012). Quantifying managerial ability: A new measure and validity tests, *Management Science* **58**(7): 1229–1248.

Demerjian, P. R., B. Lev, M. F. Lewis, and S. E. McVay (2013). Managerial ability and earnings quality, *Accounting Review* **88**(2): 463–498.

Demirgüç-Kunt, A. and H. Huizinga (2010). Bank activity and funding strategies: The impact on risk and returns, *Journal of Financial Economics* **98**(3): 626–650.

DeYoung, R., W. W. Lang, and D. L. Nolle (2007). How the internet affects output and performance at community banks, *Journal of Banking & Finance* **31**(4): 1033–1060.

Dietrich, A. and G. Wanzenried (2014). The determinants of commercial banking profitability in low-, middle-, and high-income countries, *The Quarterly Review of Economics and Finance* **54**(3): 337–354.

Dong, J., L. Yin, X. Liu, M. Hu, X. Li, and L. Liu (2020). Impact of internet finance on the performance of commercial banks in China, *International Review of Financial Analysis* **72**: 101579.

Fuster, A., M. Plosser, P. Schnabl, and J. Vickery (2019). The role of technology in mortgage lending, *Review of Financial Studies* **32**(5): 1854–1899.

Geng, Z., H. Yan, X. Qiu, and X. Huang (2021). fasthan: A BERT-based multi-task toolkit for Chinese NLP, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, pp. 99–106.

Goldstein, I., W. Jiang, and G. A. Karolyi (2019). To FinTech and beyond, *Review of Financial Studies* **32**(5): 1647–1661.

Guzman, J. and A. Li (2022). Measuring founding strategy, *Management Science,* **forthcoming**.

He, D., C.-Y. Ho, and L. Xu (2020). Risk and return of online channel adoption in the banking industry, *Pacific-Basin Finance Journal* **60**: 101268.

Hernando, I. and M. J. Nieto (2007). Is the Internet delivery channel changing banks' performance? The case of Spanish banks, *Journal of Banking & Finance* **31**(4): 1083–1099.

Hills, J. (2002). *The struggle for control of global communication: The formative century*, University of Illinois Press.

Hoberg, G. and G. Phillips (2016). Text-based network industries and endogenous product differentiation, *Journal of Political Economy* **124**(5): 1423–1465.

Huang, R. H. (2021). *Fintech Regulation in China: Principles, Policies and Practices*, Cambridge University Press.

Huang, Y., L. Zhang, Z. Li, H. Qiu, T. Sun, and X. Wang (2020). Fintech credit risk assessment for SMEs:

Evidence from China, *Technical report*, IMF.

URL: *https://www.elibrary.imf.org/view/journals/001/2020/193/article-A001-en.xml*

Huber, K. (2021). Are bigger banks better? Firm-level evidence from Germany, *Journal of Political Economy* **129**(7): 2023–2066.

Imbens, G. W. and D. B. Rubin (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge University Press, New York.

Jakšič, M. and M. Marinč (2019). Relationship banking and information technology: The role of artificial intelligence and FinTech, *Risk Management* **21**(1): 1–18.

Jiang, C., S. Yao, and G. Feng (2013). Bank ownership, privatization, and performance: Evidence from a transition country, *Journal of Banking & Finance* **37**(9): 3364–3372.

Kumbhakar, S. C., H.-J. Wang, and A. P. Horncastle (2015). *A Practitioner's Guide to Stochastic Frontier Analysis Using Stata*, Cambridge University Press, New York.

Le, Q. V. and T. Mikolov (2014). Distributed representations of sentences and documents, *Proceedings of the 31st International Conference on Machine Learning,* Beijing.

Lee, C.-C., X. Li, C.-H. Yu, and J. Zhao (2021). Does fintech innovation improve bank efficiency? Evidence from China's banking industry, *International Review of Economics & Finance* **74**: 468–483.

Lee, R. P. and R. Grewal (2004). Strategic responses to new technologies and their impact on firm performance, *Journal of Marketing* **68**(4): 157–171.

Lerner, J. and A. Seru (2022). The use and misuse of patent data: Issues for finance and beyond, *Review of Financial Studies* **35**(6): 2667–2704.

Li, K., F. Mai, R. Shen, and X. Yan (2021). Measuring corporate culture using machine learning, *Review of Financial Studies* **34**(7): 3265–3315.

Lin, M., N. R. Prabhala, and S. Viswanathan (2013). Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending, *Management Science* **59**(1): 17–35.

Magnuson, W. (2018). Regulating fintech, *Vanderbilt Law Review* **71**: 1167.

Markham, J. W. (2002). *A Financial History of the United States: From Christopher Columbus to the Robber Barons (Volume 1)*, M. E. Sharpe.

Mikolov, T., K. Chen, G. S. Corrado, and J. Dean (2013). Efficient estimation of word representations in vector space, *ICLR*.

Mourlon-Druol, E. (2015). "Trust is good, control is better": The 1974 Herstatt Bank crisis and its implications for international regulatory reform, *Business History* **57**(2): 311–334.

Navaretti, G. B., G. Calzolari, and A. F. Pozzolo (2017). FinTech and banks: Friends or foes?, *European Economy Banks, Regulation, and the Real Sector* (2): 9–30.

Omarova, S. T. (2020). Technology v technocracy: Fintech as a regulatory challenge, *Journal of Financial Regulation* **6**(1): 75–124.

Park, J., K. Han, and B. Lee (2022). Green cloud? An empirical analysis of cloud computing and energy efficiency, *Management Science,* **forthcoming**.

Philippon, T. (2016). The FinTech opportunity, *Technical report*, Bank for International Settlements.
URL: *https://www.bis.org/publ/work655.pdf*

Qiu, X., H. Pei, and H. Yan (2020). Multi-criteria Chinese word segmentation with transformer.
URL: *https://arxiv.org/abs/1906.12035v1*

Shu, W. and P. A. Strassmann (2005). Does information technology provide banks with profit?, *Information & Management* **42**(5): 781–787.

Subramani, M. and E. Walden (2001). The impact of e-commerce announcements on the market value of firms, *Information Systems Research* **12**(2): 135–154.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance* **62**(3): 1139–1168.

Thakor, A. V. (2020). Fintech and banking: What do we know?, *Journal of Financial Intermediation* **41**: 100833.

Thakor, A. V. and A. W. A. Boot (2008). *Handbook of Financial Intermediation and Banking*, Elsevier B.V., San Diego, chapter 15, pp. 483–554.

Vives, X. (2017). The impact of fintech on banking, *European Economy* (2): 97–105.

Wang, R., J. Liu, and H. R. Luo (2020). Fintech development and bank risk taking in China, *European Journal of Finance* **27**(4-5): 397–418.

Wang, Y., X. Sui, and Q. Zhang (2021). Can FinTech improve the efficiency of commercial banks? —an analysis based on big data, *Research in International Business and Finance* **55**: 101338.

Yadav, Y. (2020). Fintech and international financial regulation, *Vanderbilt Journal of Transnational Law* **53**: 1109.

Zhao, J., X. Li, C.-H. Yu, S. Chen, and C.-C. Lee (2022). Riding the FinTech innovation wave: FinTech, patents and bank performance, *Journal of International Money and Finance* **122**: 102552.

Appendix

A How our Work Fits into Related Literature

In this Appendix, we present a detailed set of relations between our work and previous literature along various dimensions. These encompass ways to measure FinTech proxies as well as other aspects of research methodology. For each dimension, we list how our work differs from other papers (our work appears last on each of these lists). We omit the details of Dong et al. (2020) and Wang, Liu, and Luo (2020), because their FinTech proxies resemble Cheng and Qu (2020).

- **FinTech Proxy**—Level

  [1] Zhao et al. (2022): An industry-level index achieved by aggregating 14 categories, such as digital bank, online brokerages, online insurance, online fund sales, person-to-person (P2P), financial information services, and crowdfunding.

  [2] Lee et al. (2021): Similar to Zhao et al. (2022).

  [3] Cheng and Qu (2020): Each bank's individual level.

  [4] He, Ho, and Xu (2020): Each bank's individual level.

  [5] Chen, Wu, and Yang (2019): Each financial institution's individual level.

  [6] Our work: Each bank's individual level.

- **FinTech Proxy**—Methodology

  [1] Zhao et al. (2022): There are two proxies at the aggregate level. One is the first principal component across the number of total established companies, registered capital, number of financing events, and amount of financing for FinTech. The other is the cumulative number of patent applications (not distinguishing FinTech or non-FinTech patents as Chen, Wu, and Yang (2019) define) by all FinTech companies and banks in each year.

  [2] Lee et al. (2021): This is the same as the first one in Zhao et al. (2022)

  [3] Cheng and Qu (2020): They calculate FinTech-related glossaries' frequency at the bank-year level in retrieved articles from an online search engine, and then extract the principal components that have eigenvalues greater than unity.

[4] He, Ho, and Xu (2020): They use a dummy variable for whether an entity adopts an online channel.

[5] Chen, Wu, and Yang (2019): These authors identify U.S. FinTech patents by combining manual and machine learning methods. Specifically, they manually identify financial services patents, and then manually classify a random sample of such patents into FinTech categories. They then train their full-sample machine-learning procedure on this classification.

[6] Our work: We identify FinTech patents via almost exclusive use of machine learning. We directly match each patent to a standardized FinTech document using an automated similarity measure.

- **FinTech Proxy**—Data Source

[1] Zhao et al. (2022): There are two proxies. First, they measure FinTech development using the total number of established FinTech companies, registered capital, number of financing events, and amount of financing; the data comes from the China FinTech Enterprise Database (FinTech Beta). Second, they collect patent data from the Chinese State Intellectual Property Office, which is similar to our work.

[2] Lee et al. (2021): This is the same as the first data source in Zhao et al. (2022).

[3] Cheng and Qu (2020): The data includes retrieved articles with three parts of keywords from the Baidu search engine. The first part, the second the bank name, and the third is overlap with FinTech-related glossaries, as in the following categories: artificial intelligence technology, blockchain technology, cloud computing technology, big data technology, and Internet technology.

[4] He, Ho, and Xu (2020): An affirmative statement on whether to adopt an online channel in the bank's fiscal report.

[5] Chen, Wu, and Yang (2019): Granted patent filings at the US Patent and Trademark Office (USPTO).

[6] Our work: Applied and granted patent filings from the Chinese Patent Retrieval System (CPRS).

- **Dependent Variable**

[1] Zhao et al. (2022): Performance measures—Capital adequacy, Asset quality, Management, Earning, Liquidity, and Sensitivity (CAMELS) indicators, including the cost-to-income ratio (CTI) and return on assets (ROA); Risk measures—CAMELS indicators, including the total capital ratio (CAR), nonperforming loans to gross loans ratio (NPL), and liquid assets to total deposits ratio (LTD).

[2] Lee et al. (2021): Performance measures—cost efficiency from stochastic frontier analysis (SFA); no risk measures.

[3] Cheng and Qu (2020): Credit risk (the nonperforming loan ratio or NPL).

[4] He, Ho, and Xu (2020): Performance measures—profit, cost, interest income, and noninterest income efficiency from SFA; Risk measures—Z-score, CAR, NPL, and liquidity ratio.

[5] Chen, Wu, and Yang (2019): Performance measures—the market value of financial institutions; no risk measures.

[6] Our work: Performance measures—profit, cost, interest income, and noninterest income efficiency; Risk measures—Z score, CAR, NPL, and liquidity ratio.

- **Testing Framework**

[1] Zhao et al. (2022): no DID, no PSM.

[2] Lee et al. (2021): no DID, no PSM.

[3] Cheng and Qu (2020): DID, no PSM.

[4] He, Ho, and Xu (2020): no DID, no PSM.

[5] Chen, Wu, and Yang (2019): no DID, no PSM.

[6] Our work: DID + PSM.

B How China's Patent Filings Address Imperfections in U.S. Data

Lerner and Seru (2022) show that patent-level data in the US can cloud inferences because of the following: (a) the truncation of patent data, especially just prior to the end date, because only successful patent applications are observed, and (b) changes in the composition of sector- or region-level innovators, and how this interacts with (a). Under the leadership of Bronwyn Hall and Jim Bessen, a number of updated versions of the NBER data have been compiled, the most recent being an extension through 2006; please refer to Lerner and Seru (2022) for more details on the imperfections in these data.

Patent filings in China suffer little from the two issues above. Specifically, data from China National Intellectual Property Administration include successful and unsuccessful applications, and have text field identifiers that uniquely flags each applicant. And more importantly, applicants' names in patent filings are consistent with their official names, typically the names on the business license. Therefore, we are able to reliably match patent applicants to the banks in our sample.

A minor issue may arise in our patent data from an employee applying for a patent in person. According to the Chinese Intellectual Property Law, applicants must submit a statement of no conflict with their employment if their application is in person. The incentives to be dishonest are low, however; specifically, the benefit that an employee receives from "stealing" such intellectual property is less applicable in our context because: (1) if a patent does have high market value, the bank managers would "keep their eyes" on the employee with the patent; (2) if it does not have high market value, the employee would not steal it compared to the benefit of retaining a steady job in the bank's technology department. Actually, to the best of our knowledge, there are no such reports of intellectual property infringement in the Chinese banking industry.

C Identifying FinTech Patents Using an NLP Algorithm

This Appendix describes details of the machine-learning procedure for identifying FinTech patents.

C.1 Text data preparation

We collect all IPC G&H classes' patents whose applicant names match banks and their subsidiary IT firms. For the natural language processing (NLP) algorithm to automatically identify FinTech patents, we need three reference documents on FinTech (FT), computer science (CS), and commercial banking business (BB).

For the FT reference document, we collect four authoritative releases:

1) *Fintech Development Plan (2019–2021)* released by People's Bank of China, that outlines guidelines, basic principles, development targets, key tasks, and supportive measures for FinTech

development.

2) *Implications of FinTech Developments for Banks and Bank Supervisors* issued by the Basel Committee on Banking Supervision on October 31, 2017.

3) *FinTech and Market Structure in Financial Services* issued by the Financial Stability Board on February 14, 2019.

4) *FinTech and the Future of Finance* issued by the World Bank Group on May 18, 2022.

We translate the English texts to Chinese and append these to the Chinese text. We then use the concatenated combination as a reference for FinTech documentation.

For the *CS* reference document, we use a document titled *Chinese Terms in Computer Science and Technology (Third Edition)*. This document was published by the China National Committee for Terminology in Science and Technology in 2018. It is an authoritative encyclopedia for computer science, and its PDF is available from our library's electronic literature database. There is no copyright issue for our purpose.

For the *BB* document, we use a book titled *Commercial Banking Operations and Management* published by Science Press in 2021. This textbook is used by the 13th Five-Year Plan for Higher Education, designed by China's Ministry of Education. We select this reference because it is relatively authoritative, and recommended at the national level to undergraduate students. This document is also available in PDF format from our library's electronic literature database, and again, there is no copyright issue for our purpose.

C.2 NLP for Identifying FinTech Patents

Following standard NLP practice, we preprocess the reference documents and banks' patent filings for vectorization. This work includes segmenting and removing stop-words. Segmenting entails partitioning words in each sentence, and is required because Chinese words are not separated by spaces, unlike in English. Stop-words are common nonsense words that are of little value for classification, and need to be deleted before applying NLP. Based on these considerations (listed in Qiu, Pei, and Yan 2020), Geng et al. (2021) share a toolkit for Chinese textual analysis at Github.[1] This tool is what we use for segmenting and deleting stop-words.

[1] See https://github.com/fastnlp/fastHan.

After preprocessing, we convert the reference documents (FT, CS, and BB) and each bank's patent filing (PF) to numerical vectors. Following the literature (Hoberg and Phillips 2016, Bellstam, Bhagat, and Cookson 2021, Guzman and Li 2022), we use *doc2vec* (Le and Mikolov 2014) to do vectorization. *doc2vec* is the state-of-the-art method for vectorization, and grew out of an older word-embedding model, *word2vec* (Mikolov et al. 2013).

As stated in Guzman and Li (2022), word-embedding models code the context of words, unlike traditional bag-of-words approaches (e.g., Tetlock 2007). In essence, *word2vec* classifications are estimated through a neural network that predicts the probability of a word's use based on other words occurring before or after. For example Kellogg, the school of management, and Kellogg, the food company would result in different vectors. However, General Mills and Kellogg, both food companies, would result in similar vectors. Therefore, *word2vec* is said to be able to learn the "meaning" of a word. We use *doc2vec*, which is the expanded version of *word2vec*, to build vectors for document-level meaning.[2]

The third step is to label which patents are FinTech by comparing how close each patent's meaning is to the FT reference document. More specifically, we calculate the cosine similarity for each bank's patent vector (PF) to each reference document vector, which is defined as $x \cdot y/(|x| \cdot |y|)$, where $x \cdot y$ is the inner product of two vectors x and y, and $|x| \cdot |y|$ is the product of the two vectors' scalar lengths. The pairwise similarity measures for PF, FT, CS, and BB are denoted as $\langle PF, FT \rangle$, $\langle PF, CS \rangle$ and $\langle PF, BB \rangle$, respectively. We label the patent as a FinTech patent if $\langle PF, FT \rangle$ is larger than $\langle PF, CS \rangle$ and $\langle PF, BB \rangle$, and non-FinTech otherwise.

D Two Detailed Examples of FinTech and Non-FinTech Patents

We present two examples of patents: The first is identified by our procedure as a FinTech patent and the second is designated as a non-FinTech patent.

For FinTech, Figure D.1 shows that the Agricultural Bank of China has applied for a patent titled "Method and Device for Assessing Credit Risk." In English, the abstract of this patent states the following: "The present application provides a method and device for assessing credit risk. The method can select features that have a significant impact on credit risk using business data via multiple rounds of screening with replacement, in order to build a model for evaluating and

[2]See Li et al. (2021) and the afore-mentioned Guzman and Li (2022) for recent examples of *doc2vec* applications in finance and management academia, respectively.

Figure D.1: **Example of a FinTech patent from the Agricultural Bank of China's application.**

Title(CN)	信用评价模型的构建方法及装置
Title(EN)	Method and device for assessing credit risk
IPC	G06Q 40/02(2012.01)
BankName(CN)	中国农业银行股份有限公司
BankName(EN)	AGRICULTURAL BANK OF CHINA LTD
BankType(CN)	国有大型商业银行
BankType(EN)	State-owned bank
ApplicationNo.	201810689255.1
ApplicationDate	2018.06.28
GrantNo.	nan
GrantDate	nan
Abstract(CN)	本申请提供了一种信用评价模型构建方法，该方法可以通过多次筛选及回选的方式，从业务数据的属性特征中选择出对评价信用评分具有影响作用的属性特征，选择出的属性特征用于构建违约概率评价模型，该模型可以计算违约概率，再获得预设的违约概率与信用评分转换模型，该模型可以将违约概率转换为信用评分，因此该两个模型可以作为信用评价模型。另外，本申请还提供了一种信用评价模型构建装置，用以保证所述方法在实际中的应用及实现。

calculating default probability. Then, it uses a predefined conversion model to convert the default probability to a credit score that can be used in business. Thus, the combination of these two models is proposed as the model for assessing credit risk in practice. In addition, this application provides a device that allows for the model's application and implementation in practice."

For an example of a non-FinTech patent, Figure D.2 shows that the Industrial & Commercial Bank of China (ICBC), the largest state-owned bank, has applied for a patent titled "Database Switching Methods, Electronic Devices, Systems, and Storage Media." We translate the patent's abstract as follows: "The present invention discloses a database switching method, electronic device, and storage medium. The invention includes: obtaining status information on execution capability of the master database and at least one standby database; judging whether to execute a switching operation between the master and standby database according to the status information of the master database; executing the switching operation between the master database and the standby database; and and performing data replenishment for the new master database. The present invention can flexibly select a suitable new master database according, and ensure the integrity of the data after the switching operation via data replenishment."

E A Four-Component Stochastic Frontier Analysis

We provide details of our efficiency estimation. We denote the frontier efficiency ($\ln \tilde{\pi}_{it}$ or $\ln \tilde{C}_{it}$) with the controls $PNPL$ and $Year$ in Equation (3) as $f(x_{it}, \beta)$, where β is the coefficient vector and

Figure D.2: **Example of a non-FinTech patent from the Industrial & Commercial Bank of China's application.**

Title(CN)	数据库切换方法、装置、系统、电子设备及存储介质
Title(EN)	Database switching methods, devices, systems, electronic devices and storage media
IPC	G06F 16/27(2019.01)
BankName(CN)	中国工商银行股份有限公司
BankName(EN)	INDUSTRIAL & COMMERCIAL BANK OF CHINA (ICBC)
BankType(CN)	国有大型商业银行
BankType(EN)	State-owned bank
ApplicationNo.	201910732853.7
ApplicationDate	2019.08.09
GrantNo.	nan
GrantDate	nan
Abstract(CN)	本发明公开了一种数据库切换方法、装置、系统、电子设备及存储介质，其中，该方法包括：分别获取主数据库和至少一个备用数据库的状态信息，状态信息包括：数据信息和执行能力信息；根据主数据库的状态信息判断是否执行主备数据库切换操作；响应于判断结果为执行主备数据库切换操作，根据预定切换策略从至少一个备用数据库中选择新的主数据库；根据获取的至少一个备用数据库的状态信息，对新的主数据库进行数据补偿操作；对主数据库和新的主数据库执行切换操作。通过本发明，可以根据实际情况灵活选择合适的新主数据库，并且通过数据补偿操作，可以保证切换操作后数据的完整性。

x_{it} is the vector of all independent variables. Then, we can rewrite Equation (3) as

$$y_{it} = f(x_{it}, \beta) - \eta_i - u_{it} + \mu_i + v_{it}, \tag{E.1}$$

where y_{it} is alternatively the i'th bank's total profit, interest income, and noninterest income in year t. As Equation (1′) shows, if y_{it} denotes the cost efficiency, we rewrite Equation (3) as

$$y_{it} = f(x_{it}, \beta) + \eta_i + u_{it} + \mu_i + v_{it}. \tag{E.1′}$$

Let us describe the estimation procedure for Equation (E.1) (a similar process applies to Equation (E.1′)). In (E.1), $\eta_i > 0$ and $u_{it} > 0$ are respectively the time-invariant and time-variant components of inefficiency, while v_{it} and μ_i respectively represent noise and a bank-level effect that are unrelated to efficiency. Following Kumbhakar, Wang, and Horncastle (2015), Equation (E.1) can be written as a combination of the following equations:

$$y_{it} = f(x_{it}, \beta) + \alpha_i + \epsilon_{it} \tag{E.2}$$

$$\epsilon_{it} = v_{it} - u_{it} \tag{E.3}$$

$$\alpha_i = \mu_i - \eta_i \tag{E.4}$$

Equation (E.3) decomposes ϵ_{it} (the residual inefficiency) into a time-variant and positive inefficiency ($u_{it} > 0$) and random noise (v_{it}). Similarly, Equation (E.4) separates α_i into a time-invariant inefficiency component under control of the bank ($\eta_i > 0$) and a bank-level random effect (μ_i).

Our estimation follows the steps suggested in Chapter 10 of Kumbhakar, Wang, and Horncastle (2015). First, we run a panel regression on Equation (E.2) to estimate bank-level heterogeneity $\hat{\alpha}_i$ and the error term $\hat{\epsilon}_{it}$.[3] Second, we isolate $\hat{u}_{it}$ from $\hat{\epsilon}_{it}$ and $\hat{\eta}_i$ from $\hat{\alpha}_i$ via maximum likelihood estimations of Equations (E.3) and (E.4) respectively. The log-likelihood functions are built on the assumptions that the component with unrestricted sign is normally distributed, while the non-negative component is half-normal. Specifically, for Equation (E.3), $v_{it} \sim$ i.i.d. $N(0, \sigma_v^2)$, and $u_{it} \sim$ i.i.d. $N^+(0, \sigma_u^2)$; while for Equation (E.4), $\mu_i \sim$ i.i.d. $N(0, \sigma_\mu^2)$, and $\eta_i \sim$ i.i.d. $N^+(0, \sigma_\eta^2)$. Finally, we calculate $OTE = \exp(-\hat{\eta}_i - \hat{u}_{it})$.

[3] In this estimation, we impose the standard homogeneity restriction $\delta_{jk} = \delta_{kj}$. See Chapter 6 of Kumbhakar, Wang, and Horncastle (2015) for estimation details.

F Control Variables

	Interpretation	Data source
Main controls		
SIZE	The logarithm of total assets on each bank's balance sheet in each year	Banks' yearly fiscal report table from CSMAR
BRANCH	The total number of each bank's branches in each year normalized by total assets	Bank branches' basic information from CSMAR
NCITY	Number of unique cities in which the bank has at least one branch	Bank branches' basic information table from CSMAR
HHI	Hirschman-Herfindahl index in each year, computed by market share using the number of branches of each bank in the city where a bank has the most branches	Bank branches' basic information table from CSMAR
PGDP	The logarithm of yearly per capita GDP of each banks' operating province in which the bank has the most branches	National Bureau of Statistics, China; Bank branches' basic information table from CSMAR
POP	The logarithm of yearly population in each banks' operating province in which the bank has the most branches	National Bureau of Statistics, China; Bank branches' basic information table from CSMAR
NFIRM	The logarithm of yearly number of firms that are registered in a province where a bank has the most branches	National Bureau of Statistics, China; Bank branches' basic information table from CSMAR
Other dummies		
Year	A set of dummy variables for each year, with the first year as the baseline	
Bank type	A set of dummy variables for bank types, encompassing city banks, joint-stock banks, foreign banks, and rural banks, with five state-owned banks used as the basline	Banks' basic information table from CSMAR