An Empirical Investigation into the Influence of Big Data Technology on the Profitability of China's City Commercial Banks

Jiadong Luo[1], Jie Yang[2], and Guo Wu[1](✉)

[1] Department of Finance, Sanda University, Shanghai, China
guo.wu@sandau.edu.cn

[2] Department of Accounting, Sanda University, Shanghai, China

Abstract. Big data technology has led to significant changes in the banking industry. This study aims at evaluating the current applications of big data technology in the China's banking sector empirically using the STATA 15 software with the data from the *iWind* database between the year 2013 and the year 2019. A multi-factor statistical model is established to show the relationship between the level of bank profitability and the development of big data technology, while other bank specific variables are controlled. A static panel data model is adopted and the regression results from using different estimators and standard errors are also mathematically compared to assess the robustness of the regression results. The results suggest that big data technology currently has a negative impact on the level of bank profitability.

Keywords: big data · panel data model · STATA

1 Introduction

The incorporation of big data technology into the banking industry is an emerging trend and it is expected that big data analytics can help commercial banks grow significantly. Big data can benefit banks in many areas, such as fraud detection, customer service, regulatory compliance, products customization, and risk management [1]. Big data gives a comprehensive analysis of the business activities within a bank and it allows a bank to perform effective customer feedback analysis and then improve its service accordingly. It is likely that banks with greater applications of big data in their daily operations could become more profitable. However, China's commercial banks, especially city commercial banks (CCBs) have been reluctant to embrace big data technology in recent years [2].

CCBs are primarily operating in local cities and they seek to provide funds for local small or medium-sized companies. In the past years, CCBs have achieved a significant increase in asset size and market share [3]. Although the recent development of big data analytics has helped improve the performance of a few joint-stock commercial banks through the technology spillover effect [4], most CCBs are still left behind due to the slow progress in the integration of big data technology into their daily business [5].

G. Guan et al. (Eds.): ICBDSS 2022, AHCS 8, pp. 1007–1014, 2023.
https://doi.org/10.2991/978-94-6463-064-0_104

Big data technology is undoubtedly playing a significant role in disrupting the traditional banking model. It has been shown that city commercial banks which focus on local lending activities usually set aside more cash as loan loss provisions due to higher credit risk as loans are relatively concentrated on local small enterprises [6]. Meanwhile, the waning liquidity [7], the gradually weakening capital buffers [8], and the increasing competition between different types of lenders could also make CCBs vulnerable to future economic shocks. The use of big data technology could be a potential solution to the above issues. It is therefore important to recognize the key benefits of various big data applications in the banking sector and understand how this rapid-developing technology can be integrated into a bank's daily business activities.

2 Theoretical Analysis

2.1 Big Data Applications in Banking

It is proposed in the present work that the influence of big data applications in commercial banks can be disruptive in three different dimensions, which are customer, analysis, and process. Figure 1 shows a list of some traditional banking activities which may be affected significantly by big data technology. For example, banks can use big data for loan pricing automation and risk assessment, while predictive analytics algorithms based on big data can also be used to effectively streamline the lending decision-making process.

2.2 The Influence of Big Data on Bank Profitability

Big data technology is a fundamental component of financial technologies, which represents the use of modern technologies in the design of various financial products or services [9]. The competition for local customers among CCBs can be greatly enhanced because some giant tech-driven companies, such as Tencent and Alibaba, are also entering the local financial markets and providing products and services similar to CCBs.

Meanwhile, unlike fintech companies, CCBs are not able to expand physical branches outside their provinces due to regulatory requirements. CCBs have a relatively limited deposit base and retail customers are flexible to shift their savings to some banks due to the availability of convenient internet banking Apps. Without adequate reliable big data infrastructure, CCBs may have to increase saving rates to maintain a decent level

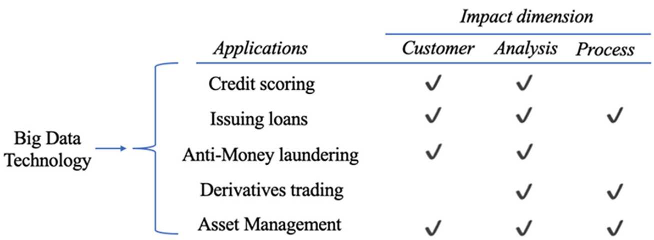

Fig. 1. The "CAP" framework of assessing big data applications in banking.

of liquidity on their balance sheets. Therefore, a hypothesis regarding the relationship between big data technology and bank profitability can be formed:

Hypothesis: *Big data technology has a negative influence on bank profitability.*

It should be noted that the bank profitability may be affected by other variables as well.

Firstly, it has been shown that the non-performing loan (NPL) ratio of CCBs is generally higher than that of state-owned commercial banks (SOBs) and joint-stock commercial banks (JSBs) over the past few years [7]. This could be due to the fact that loans issued by CCBs are more regionally concentrated and their assets are therefore not well-diversified. A variety of studies [10–12] were conducted to evaluate the effect of non-performing loans on bank profitability in various countries and most of the results show that the level of non-performing loans negatively affects the level of bank profitability. Therefore, the NPL is added as a control variable in this study.

Secondly, instead of relying on bank deposits to fund their loan business, CCBs tend to manage their liabilities mainly through interbank activities and the market borrowing rates for financial institutes like CCBs are relatively stable throughout recent years due to the regular liquidity management by China's central bank since 2013 [13]. The net interest margin of CCBs may thus primarily depend on the level of loan rate and CCBs could eventually benefit from higher loan rates in the market. Therefore, the level of LR is also treated as a control variable. Thirdly, in line with some literature studies [14, 15], the variable of capital adequacy ratio (CAR) is also included in the econometric model.

3 Empirical Analysis

3.1 Model and Data

The present study is conducted by using the bank data from nine China's listed city commercial banks during the 2013–2019 period. Return on equity (ROE) is used as a proxy for the bank profitability in this study. The index of big data [14] is calculated from a principal components analysis using the data of third-party payments, internet finance users, and the market size of big data computing. Equation 1 represents a static panel data model which indicates that the bank profitability (ROE) of CCBs depends on the big data [14], the capital adequacy ratio (CAR), the non-performing loan ratio (NPL),

Table 1. The summary of descriptive statistics.

Variable	Obs.	Mean	Std. Dev.	Min	Max
ROE	63	15.6559	3.7635	9.9300	26.2200
BD	63	9.2666	0.3173	8.8000	9.7024
CAR	63	12.9778	1.1287	10.9400	15.9500
NPL	63	1.1787	0.3357	0.5900	2.3500
LR	63	6.0350	0.6227	5.2625	6.9700

and the loan rate (LR). The unobserved heterogeneity across banks is represented by u_i. Table 1 summarizes the results of descriptive statistics of the variables.

$$ROE_{it} = b_0 + b_1 BD_t + b_2 CAR_{it} + b_3 NPL_{it} + b_4 LR_t + u_i + +_{it} \quad (1)$$

3.2 Regression Results

The OLS regression result is shown in column 1 of Table 2. As a comparison, the fixed-effects [9] regression, and the random-effects (RE) regression are summarized in

Table 2. The summary of regression results (OLS, FE and RE).

	Dependent variable: ROE				
	(1)	(2)	(3)	(4)	(5)
	OLS	FE	FE_Robust	RE	RE_Robust
BD	-5.8618^{***}	-4.6516^{***}	-4.6516^{***}	-4.7330^{***}	-4.7330^{***}
	(1.6605)	(0.8660)	(1.3644)	(0.8657)	(1.3758)
CAR	0.8347^{**}	-0.0062	-0.0062	0.0533	0.0533
	(0.3555)	(0.2361)	(0.2142)	(0.2331)	(0.2117)
NPL	-2.0090^{*}	-3.5606^{***}	-3.5606^{***}	-3.4097^{***}	-3.4097^{***}
	(1.1857)	(0.9071)	(0.5945)	(0.8858)	(0.5986)
LR	1.4728^{*}	1.0225^{**}	1.0225^{***}	1.0658^{**}	1.0658^{***}
	(0.8569)	(0.4685)	(0.2815)	(0.4662)	(0.2977)
Cons.	52.6221^{***}	56.8670^{***}	56.8670^{***}	56.4098^{***}	56.4098^{***}
	(18.5954)	(9.6317)	(12.4382)	(9.6640)	(12.8749)
N	63	63	63	63	63
R^2	0.485	0.800	0.800	0.800	0.800
Standard errors in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)					

Table 3. Diagnostic tests of the econometric model.

Test method	Results
F test (FE vs OLS)	F(8,50) = 22.88; *p-value* = 0.0000;
LM-test (RE vs FE)	chibar2(01) = 86.44; *p-value* = 0.0000;
Pesaran CD test	Off-diagonal = 0.474; *p-value* = 0.2121;
Modified Wald test	chi2(9) = 102.35; *p-value* = 0.0000
Wooldridge test	F(1,8) = 26.754; *p-value* = 0.0009
Modified Hausman test	Hansen statistic = 543.466; *p-value* = 0.0000

column 2 and column 4 respectively. The use of clustered-robust standard errors [16] is also adopted in the analysis. As suggested by results in column 3 and column 5, it is evident that clustered-robust standard errors (SE) are larger than conventional standard errors. The use of clustered-robust SE allows for correlation between observations and would increase confidence intervals [17, 18]. Accordingly, the results in each column of Table 2 all suggest that BD and NPL negatively affect ROE while LR has a positive influence. This is consistent with the hypotheses proposed in Sect. 2. Meanwhile, as can be seen from Table 3, the results of the F test and LM test both indicate that either the FE estimator or the RE estimator is prior to the OLS estimator. The Hausman test [19] can further specify which model (FE or RE) is more appropriate and it is used to test whether the unobserved heterogeneity is correlated with independent variables.

It has been argued that the result from the standard Hausman test may not be reliable when the violations of static panel model assumptions are present [20]. Firstly, the present study examines the cross-sectional independence using a Pesaran's test and checks the heteroskedasticity with the modified Wald test. Secondly, the Wooldridge test for serial correlation in panel data is performed to check first-order autocorrelation and the result is shown in Table 3. It can be concluded from the diagnostic tests that a modified Hausman test based on Arellano's work [21] may be statistically reliable in this case. The result suggests that the FE model is more appropriate. Furthermore, as shown in Table 4, the results from the feasible generalized least square (FGLS) estimation, Prais-Winsten regression with panel-corrected standard error (PCSE), and the FE regression with Driscoll-Kraay (DK) standard errors [22] are also compared to the robustness of

Table 4. Regression results using different estimators.

	Dependent variable: ROE				
	(1)	(2)	(3)	(4)	(5)
	FE	FE_Robust	FGLS	PCSE	DK
BD	-4.6516^{***}	-4.6516^{***}	-4.4699^{***}	-4.4963^{***}	-4.6516^{***}
	(0.8660)	(1.3644)	(0.3430)	(0.8130)	(0.4561)
CAR	−0.0062	−0.0062	0.0712	0.1315	−0.0062
	(0.2361)	(0.2142)	(0.1040)	(0.2993)	(0.1977)
NPL	-3.5606^{***}	-3.5606^{***}	-2.6474^{***}	-1.9658^{**}	-3.5606^{**}
	(0.9071)	(0.5945)	(0.2624)	(0.8367)	(1.3427)
LR	1.0225^{**}	1.0225^{***}	1.2026^{***}	1.3714^{***}	1.0225^{**}
	(0.4685)	(0.2815)	(0.1227)	(0.3338)	(0.2946)
Cons.	56.8670^{***}	56.8670^{***}	51.9472^{***}	49.5771^{***}	56.8670^{***}
	(9.6317)	(12.4382)	(3.0503)	(7.4100)	(4.5956)
N	63	63	63	63	63
R^2	0.800	0.800	/	0.937	0.800
Standard errors in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)					

the relationships. It is indicated that the relationships between the three key determinants and ROE are still statistically significant.

However, it should be noted that the static panel model in the present work is based on the strict assumption that the variable BD is conditionally exogenous with the inclusion of properly controlled covariates, and the unobserved individual heterogeneity is time-invariant. Since the development of big data technology may be viewed as a macro-level exogenous shock to the banks that are at the micro-level, the estimation of the key coefficient is likely to be asymptotically consistent if other endogeneity issues are absent as well. It is suggested that further examination may be needed to check the potential violations of the causal-inference assumptions associated with this static panel model.

4 Conclusions

This study is performed to understand the big data applications in banking via a CAP framework and an empirical analysis is conducted to examine the influence of big data technology on the bank profitability of city commercial banks in China during the 2013–2019 period. It is shown that big data technology currently has a negative influence on bank profitability. Panel regressions with various adjustments have been carried out to ensure the validity of the regression results. It can be seen from the empirical results that CCBs have not yet benefited from technological innovations. Therefore, it is suggested that the role of big data in driving business growth and transformation should be strengthened within CCBs and they should continuously invest in big data, which could help provide innovative approaches to improve their banking products and services. The integration of big data technology into daily business activities could significantly enhance their bank values in the future. It should be admitted that the current study may be subject to some limitations. Firstly, the results are only based on the analysis of the data from a limited period and a limited number of banks. Secondly, accounting financial ratios are extensively used in this study and the financial data may be subject to potential measurement errors, which could influence the robustness of statistical results to an unknown extent.

Acknowledgment. The author would like to acknowledge the fund support from Sanda University (No.2021JSZX0101 and No.2021ZD03).

References

1. U. Srivastava and S. J. P. C. S. Gopalkrishnan, "Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks," vol. 50, pp. 643-652, 2015.
2. J. Zhang, W. Peng, and B. J. C. E. R. Qu, "Bank risk taking, efficiency, and law enforcement: Evidence from Chinese city commercial banks," vol. 23, no. 2, pp. 284-295, 2012.
3. S. J. A. Hsu and t. P. P. Studies, "China's Banking Sector as the Foundation of Financial Reform," vol. 3, no. 2, 2016.
4. S. Agarwal and Y. H. J. C. F. R. I. Chua, "FinTech and household finance: a review of the empirical literature," vol. ahead-of-print, no. ahead-of-print, 2020.

5. Y. Shim and D.-H. Shin, "Analyzing China's Fintech Industry from the Perspective of Actor-Network Theory," Telecommunications Policy, vol. 40, no. 2, pp. 168-181, 2016/03/01/ 2016.
6. M. H. Tumin and R. M. J. S. E. J. Said, "Performance and Financial Ratios of Commercial Banks in Malaysia and China," vol. 7, no. 2, 2010.
7. F. Sufian, "Determinants of Bank Profitability in a Developing Economy: Empirical Evidence from the China Banking Sector," Journal of Asia-Pacific Business, vol. 10, no. 4, pp. 281-307, 2009/11/30 2009.
8. A. C. H. Lei and Z. Song, "Liquidity creation and bank capital structure in China," Global Finance Journal, vol. 24, no. 3, pp. 188-202, 2013/01/01/ 2013.
9. P. M. J. S. E. J. Schueffel, "Taming the Beast: A Scientific Definition of Fintech," vol. 4, no. 4, pp. 32-54, 2016.
10. B. Prasad Bhattarai, "Effects of Non-performing Loan on Profitability of Commercial Banks in Nepal," 2020.
11. B. J. A. E. Duraj and F. Review, "Factors Influencing the Bank Profitability - Empirical Evidence from Albania," vol. 5, no. 3, pp. 483-494, 2015.
12. M. M. J. E. M. Ahamed, "Asset quality, non-interest income, and bank profitability: Evidence from Indian banks," vol. 63, no. JUN., pp. 1-14, 2017.
13. H. Dong, H. Wang, and X. J. S. S. E. P. Yu, "Interest Rate Determination in China: Past, Present, and Future," vol. 11, no. 4, pp. 1355-61, 2015.
14. M. M. Rahman, M. K. Hamid, and M. A. M. J. I. J. o. B. Khan, "Determinants of Bank Profitability: Empirical Evidence from Bangladesh," vol. 10, p. 135, 2015.
15. M. A. J. I. J. o. A. R. i. A. F. Almumani and M. Sciences, "Impact of Managerial Factors on Commercial Bank Profitability: Empirical Evidence from Jordan," vol. 3, no. 3, pp. 298-310, 2013.
16. Wooldridge J M 2010 Econometric analysis of cross section and panel data (Cambridge, MA: MIT press)
17. Baltagi B 2008 Econometric analysis of panel data (Chichester: John Wiley & Sons, Inc.)
18. Wooldridge J M 2015 Introductory econometrics: A modern approach (Boston, MA: Cengage Learning)
19. J. A. J. E. Hausman, "Specification Tests in Econometrics," vol. 46, no. 6, pp. 1251-1271, 1978.
20. D. J. S. J. P. C. o. S. Hoechle and Stata, "Robust Standard Errors for Panel Regressions with Cross-Sectional Dependence," vol. 7, no. 3, pp. 281-312, 2007.
21. M. Arellano, "On the testing of correlated effects with panel data," Journal of Econometrics, vol. 59, no. 1, pp. 87-97, 1993/09/01/ 1993.
22. J. C. Driscoll, A. C. J. T. R. o. E. Kraay, and Statistics, "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data," vol. 80, no. 4, pp. 549-560, 1998.