Competition and the Use of Credit Lines

Modelling of Chinese Corporate Bond Default—A Machine Learning Approach

Market smart: How firms respond to the IPO P/E price-cap regulations in China

ZHOU LU

1 Acknowledgements

Undertaking the Ph.D. program at NHH has been one of the most challenging and rewarding experiences in my life. This dissertation is the result of five years of effort and passion as a Ph.D. Research Scholar at the Department of Finance at the Norwegian School of Economics. Throughout the process of writing this dissertation, I have received a great deal of support and help.

First of all, I would like to express my sincere gratitude to my two supervisors Xunhua Su and Kyeong Hun Lee, for their guidance during my Ph.D. program. Their enthusiasm and knowledge about financial research offer great inspiration to me. Without them, I could never become the researcher I am today.

I am also grateful to Xunhua Su, Daniel Chi, Zhuyao Zhuo, Xiaoyu Zhang for being good co-author and friends. Throughout my Ph.D. study, they have been great friends, and I am lucky to have them on board. I have learned a lot from them, and our discussion helps me become a better researcher.

I would also like to thank Espen Eckbo, Jonathan Karpoff, Darya Yuferova, Jørgen Haug, Tommy Stamland, Tore Leite, Eric De Bodt, Lars Lochstoer for their well-structured course. I would also like to thank the well-structured brown bag and seminar section, which offer a lot of inspiration for my research and grant the opportunity to discuss the research idea with elite researchers.

I would also like to thank all the faculty at the Department of Finance at NHH. Among others, José Albuquerque de Sousa, Eric De Bodt, Carsten Bienz, Espen Eckbo, Nils Friewald, Nataliya Gerasimova, Jørgen Haug, Thore Johnsen, Tore Leite, Jøril Mæland, Aksel Mjøs, Walt Pohl, Konrad Raff, Francisco Santos, Karin Thorburn, and Darya Yuferova.

I am also grateful to my Ph.D. colleagues and friends in NHH. Michael Axenrod, Damiano Maggie, Jing Lan, Markus Lithell, Andre Lot, Loreta Rapushi, Negar Ghanbari, Xiaoyu Zhang, Johan Karlsen Hengxiang Yu, Diego Bonelli, Trang Vu. They all have been good colleagues and friends during these years at NHH. I also want to thank the administration at the Department of Finance. Tonje Foss, Kjersti Hafastad, Olga Pushkash, and Linn Raknes James helped me smoothly navigate the Ph.D. program at NHH.

Lastly, I would also thank the support from my family. Without their support, I would not have made it during these years. My parents offer a lot of support and understanding during these years. As a finance professor, my father shares a lot of insight and vision of financial market development. I am highly indebted to his support and help. This dissertation is dedicated to him.

Zhou Lu

Bergen, January 2022

# 2 Introduction

This doctoral thesis consists of three essays on empirical corporate finance and is submitted to the Department of Finance at the Norwegian School of Economics in partial fulfillment of the requirements for the completion of the degree of Doctor of Philosophy at NHH.

These three essays explore three important areas in empirical corporate finance. The first paper investigates how product market competition affects firm liquidity choice. The second paper studied the corporate bond default in China in the machine learning approach. The third paper studied how government regulation on the IPO market (Pricing Cap) affects the firm information and investment flow.

While the topic may differ among themselves, these three papers use multiple state-of-art statistical methods. The first paper may be linked to liquidity risk and its hedging. The second paper highlights the importance of firm liquidity risk management and focuses on the Chinese corporate bond market. My third paper tries to connect firm behavior with the regulation in financial market and shed light on the unique Chinese market. The last two papers focus on Chinese economics and shed light on the financial market and the stability of the second-largest economy in the world.

## 2.1 Competition and the Use of Credit Lines

Competition in the product market continuously shapes corporate financial decisions, among which liquidity management is of particular relevance. The general finding of the literature is that cash offers greater strategic value as competition intensifies, and therefore firms facing more intense competition increase their cash holdings. But cash is only one source of corporate liquidity. The other major source of corporate liquidity credit lines provided by banks are used by over 65% of U.S. public firms. For these firms, the amount of available credit lines is as large as the amount of cash holdings. According to Federal Reserve data, in the last decade (2010 to 2019) the aggregate amount of committed commercial credit lines by U.S. banks is even larger than the amount of outstanding commercial loans. Thus, credit lines are a significant component of corporate liquidity and represent more than 50 % of bank commercial lending. Yet we know little about how competition affects firms' use of credit lines or how competition affects the relative importance of cash and credit lines as liquidity sources. We attempt to fill this void in this study. Competition increases the strategic benefits of having a credit line but also increases the difficulty of obtaining a line. What is the net effect of competition on credit line usage? Using a comprehensive sample of U.S. public firms from 2002 to 2019, we find that competition reduces the use of credit lines. Competition not only reduces the absolute usage of credit lines but also the relative usage when compared with cash holdings. The economic significance of competition is comparable to that of known determinants of credit line usage, such as cash flow level and asset tangibility.

To mitigate this endogeneity concern, we use import tariff rates as an instrument and apply a Two-Stage Least Squares (2SLS) regression. We use extra import tariffs imposed by Donald Trump and his administration as exogenous shocks and apply a propensity score matching (PSM) difference-in-difference framework for analysis to ascertain causality. Both methods are state-of-art statistical techniques for identification and are widely used in finance. In addition, Our paper may be the first paper using Trump tariff retaliation as the exogenous shock for product market competition.

This study contributes to the literature on how product market competition affects corporate liquidity management. Previous works in this literature mostly focus on cash holdings and largely overlook another equally important source of liquidity – credit lines. Empirical evidence shows that competition increases corporate cash holdings because cash provides greater strategic value in a more competitive environment (e.g., Bolton and Scharfstein, 1990; Haushalter, Klasa, and Maxwell, 2007; Frésard, 2010; Hoberg, Phillips, and Prabhala, 2014). Extant theory also suggests a positive effect of competition on credit line usage because credit lines offer obvious competitive advantages (e.g., Maksimovic, 1990; Martin and Santomero, 1997), but our evidence shows the opposite. This seeming inconsistency is not driven by firms' reduced demand for credit lines, but rather by banks' restricted supply of credit lines to firms facing intense competition. Our findings point to the need for more theoretical study on the effect of competition on the supply side of credit lines. The results also seem to be inconsistent with the traditional wisdom that treats cash and credit lines as perfect substitutes (e.g., Holmström and Tirole, 1998). Our findings suggest that on the demand side, cash and credit lines could be substitutes; but on the supply side, they are certainly not. Cash as unconditional liquidity is under the full control of firms, while the equilibrium holding of credit lines is conditional also on banks' supply. This finding is consistent with the notion that credit lines are contingent liquidity (e.g., Sufi, 2009; Nikolov, Schmid, and Steri, 2019).

Our findings on credit lines also have implications for the relation between competition and cash holdings. As competition intensifies, firms *want to* hold more cash because the strategic advantages of cash increase. But firms also *have to* rely more on cash as a liquidity source because credit line supply becomes more restricted. So, the accessability of alternative liquidity sources may also be a determinant of corporate cash holdings. This has the potential to help us better understand the determinants of cash holdings (e.g., Kim, Mauer, and Sherman, 1998; Opler, Pinkowitz, Stulz, and Williamson, 1999). By studying alternative corporate liquidity sources in a unified framework, we also extend Almeida, Campello, Cunha, and Weisbach (2014).

Furthermore, our findings point to a further dimension that differentiates credit lines and cash holdings as a corporate liquidity source. Sufi (2009) argues that firms prefer credit lines over cash when they are less likely to violate cash-flow-based financial covenants. Lins, Servaes, and Tufano (2010) find that cash and credit lines are held for different purposes: credit lines to fund future investment opportunities, whereas cash to buffer future cash flow shortfalls. Acharya, Almeida, and Campello (2012) show that firms with high asset betas hold more cash relative to credit lines because it is more costly for them to obtain credit lines from banks. We add to this literature by showing that competition reduces the use of credit lines relative to cash as a corporate liquidity source.

Finally, our findings add to the extensive literature on the interactions between product markets and various corporate financial policies or outcomes, including cost of debt (Valta, 2012), capital structure (e.g., Brander and Lewis, 1986; Phillips, 1995; Campello, 2003; MacKay and Phillips, 2005), and product strategies (e.g., Chevalier and Scharfstein, 1996; Dasgupta and Titman, 1998; Khanna and Tice, 2005; Phillips and Sertsios, 2013). Specifically, we add to this literature by showing how competition affects corporate liquidity management.

## 2.2 Modelling of Chinese Corporate Bond Default —A Machine Learning Approach

China's corporate bond market has grown exponentially over the past decade, increasing from 1.89 trillion RMB in 2009 to 32.9 trillion RMB in 2019. At the end of 2019, the corporate bond market accounted for 33.2% of China's bond market. According to statistics from the Wind Economic Database, 58 major issuers in China's corporate bond market defaulted before 2017, involving 113 defaulted bonds, with a balance of 37.992 billion RMB. From 2017 to 2019, the number of defaulted issuers increased to 103. This involved 425 defaulted bonds, and the balance of newly defaulted bonds reached 343.083 billion RMB. The reasons for the surge in defaults are as follows. After 2017, because of the impact of intense regulation and a deleveraging policy, Chinese enterprises' financing channels narrowed, the credit risk of firms relying on rolling financing continued to increase, and default events emerged at an accelerating pace. It is a challenge for highly leveraged firms to repay their principal and interest amid deterioration of the external finance environment, which has triggered the largest default wave of corporate bonds in China.

Our primary focus is predicting corporate bond defaults in China. Corporate bond defaults have been widely studied. Research can be sorted into three main categories: structure models, reduced-form models, and macroeconomic and/or accounting-based models. The structure credit model emerged from the options pricing theory of Black and Scholes (1973) and Merton (1973) and has been extended[1]. The structure model framework identifies key factors driving corporate bond value. The reduced-form model assumes that corporate bond defaults satisfy a Poisson distribution, and uses the default strength index to measure the default risk. (Jarrow and Turnbull, 1995; Duffie and Singleton, 1997, 1999). Another reduced-form model assumes that the probability of credit rating conversion follows the Markov process and constructs a credit rating transfer matrix to predict the default rate of debt (Jarrow, Lando, and Turnbull, 1997). Both the structure and reduced-form models predict the probability of future credit defaults under a specific theoretical framework. The advantage of these models lies in their minimal dependence on historical default data and their good foresight. However, these models have two disadvantages. First, they rely heavily on the validity of the assumptions. Second, the structure model relies on asset price data, which might not be available for non-listed firms. Other studies have focused on macroeconomic and accounting-based models. One stream of literature[2] evaluates how credit risk represents systematic risk in the

[1]See Black and Cox (1976); Geske (1977); Leland (1994, 2004); Longstaff and Schwartz (1995); Collin-Dufresne and Goldstein (2001), and Goldstein, Ju, and Leland (2001)

[2]The research includes Fama and French (1993); Dichev (1998); Chava and Jarrow (2004); Vassalou and Xing (2004); Campbell, Hilscher, and Szilagyi (2008).

market. Beaver (1966); Altman (1968); Ohlson (1980), and Beaver, McNichols, and Rhie (2005) evaluate how earnings and accounting ratios (reflecting the sales, expenses, growth, and liquidity of a firm, which indicate the macroeconomic and business environments of the firm) predict the default of corporate bonds.

The primary objective of most models, such as structural, reduced-form, or macroeconomic and accounting models, is to explain corporate bond defaults *within* sample, and they often emphasize causal inference. Our objective is different: we aim to develop a model that can accurately predict corporate bond defaults *out of sample* (i.e., a prediction problem). Satchidananda and Simha (2006) show that the problems of causal inference and prediction, although related, are fundamentally different. Specifically, causal inference modeling aims to minimize the bias resulting from model misspecification to obtain the most accurate representation of the underlying theory. In contrast, the objective of predictive modeling seeks to minimize the out-of-sample prediction error, that is, the combination of the bias and estimation variance resulting from using a sample to estimate model parameters. Although causal inference represents mainstream social science research, Kleinberg, Ludwig, Mullainathan, and Obermeyer (2015) show that many interesting prediction problems are neglected in the extant business and economics literature. Our models can effectively predict rare events with a small number of input variables and have better performance than traditional default risk models, such as structure or accounting models. Machine learning (ML) has been widely used in risk modeling, risk assessment, and risk prediction in recent decades. According to Altman, Marco, and Varetto (1994), combining traditional models, such as logit and probit regression models, significantly improves default predictability. In addition, ML model is powerful to detect non-linear relationship among variables. The ML technique is useful for improving model accuracy, especially when the credit market's complexity increases. Although ensemble learning has been successfully applied in many other fields ((Zhou, 2012)), ours is the first study to apply the method to a finance setting with a severe class imbalance problem, namely, rarity of corporate bond default . Whether ensembling models outperform traditional models is an empirical question. Our results suggest that ensemble learning, if properly used, is more powerful than SVM or KSVM for the purpose of corporate bond default prediction.

Second, we adopt models first using multiple ML models for feature engineering, which pick up the most important features (the so-called “importance index”) to predict the default risk and then perform logistic regression. The benefit of this type of model is that it perfectly combines the prediction power of complicated machine learning models (e.g., bagging or boosting models) with the economic intuition of the logistic model, which allows us to infer from the models.

Finally, we further compared our result with the traditional credit risk model. Because of data availability, we analyze the corporate default based on Merton (1974)’s and KMV model and show that our method outperforms traditional structure models.

### 2.3 Market smart: How firms respond to the IPO P/E price-cap regulations in China

Since the 1978 economic reform, China has gradually liberalized its economy. In 1992, the Chinese government started to develop a "socialist market economy." Unlike countries such as Poland, former Czechoslovakia, and Russia, which underwent rapid reforms according to comprehensive plans, China has enacted a step-by-step evolutionary reform by providing state-imposed market-like incentives, with different sectors having different reform speeds. As the manufacturing sector is more resilient to external shocks and has less government intervention, its reform was completed early. The large-scale privatization of state-owned enterprises (SOEs) began in the mid-1990s, as part of the broad economic reforms outlined in the ninth and tenth five-year plans. Gan (2009) estimate that between 1995 and 2005, firms with an aggregate 11.4 trillion RMB in assets were privatized in China, comprising two-thirds of China’s SOEs and state assets. Further, owing to a high unemployment rate, most firms in China have been operating based on market-like incentives. On the other hand, the financial sector, which is more fragile and reformed more conservatively, is heavily regulated in China. Thus, the coexistence of market forces and government interventions in China allows researchers to investigate how the market helps firms minimize the impact of government intervention and how firms respond to government regulation.

In this study, I examine this question by focusing on price caps. The China Securities Regulatory Commission (CSRC) has set regulatory guidelines on the price-to-earnings (P/E) ratio of companies that plan to go public. Before 2009 and after 2012, the CSRC applied a guideline that made it unlikely for companies with a P/E ratio greater than 23 to be approved by the regulator, with all industries being subject to a homogenous price-cap limitation. This regulation provides an exogenous shock to firms that intend to go public.

A key feature of China's approval-based public listing system that enables causal identification is that firms have little ability to time the initial public offering (IPO) market. Normally, IPO approval takes two to three years. Once approved, firms take several months to complete the final steps. In addition, unlike the US, withdrawal from the IPO market is associated with heavy sunk costs in the approval process. Thus, public listing in China serves as a strong signal to investors that the firm is eligible to pass multi-layer government regulation and has good profitability and earnings quality for at least three years. Therefore, instead of withdrawing from an IPO, firms are more likely to issue IPOs with losses to be compensated from the seasoned equity offering (SEO) market. Firms are unlikely to acknowledge the date the government imposed the price cap and jumps the queue of listing in advance to escape price-cap regulations. In addition, the implementation of the price cap is directly decided by the CSRC based on the IPO market conditions; thus, firms have little ability to influence price caps. Public equity is an especially important source of financing in China because alternative financing is limited, and public markets provide liquidity for early investors and entrepreneurs. Moreover, bank credit in China tends to typically favor SOEs or mature firms with good credit records. Although Chinese venture capital and private equity (VC/PE) are growing quickly, they remain less mature than their counterparts in the United States during the period investigated in this study. Financing under regulations became the most important channel of external financing for the majority of listed firms.

Although firm financing is inevitably associated with higher costs, firms may act strategically to try to minimize the negative impact of pricing cap regulation in the IPO market. There are two direct negative impacts of pricing cap regulation. Firstly, a pricing cap reduces the required rate of return in the IPO market. Compared to a non-pricing cap period, firms are more likely to use alternative financing with lower costs or higher returns during a pricing cap period. As a result, in this context, firms may try to minimize ownership dilution during their IPOs while seeking for alternative, less costly financing options such as seasoned equity offering (SEO) and corporate bond/loan financing. Firms also adjust their financial reporting; firms affected by the pricing cap anticipate that their IPOs will be influenced by the pricing cap, which reduces their incentive to manage earnings before the IPO and increases their incentive to manage earnings before the SEO.

Second, regulatory literature indicates that price regulation in the capital markets may postpone price discovery. Firms affected by the pricing cap are more likely to be associated with higher information asymmetry. As a counteraction, firms may increase self-disclosure during their IPOs to reduce the negative impact of the regulations.

I find consistent evidence that firms are likely to realize a positive abnormal return in the post-IPO market indication that its price is understated ; additionally, I find that this phenomenon disappears in the SEO market, indicating that the pricing cap distorts the stock price and hinders pricing efficiency in the short-term. Firms retain shares in IPOs and accelerating the speed and increase its amount of its SEO issuance or seeking alternative financing such as bank loan or corporate bond. The pricing cap also impacts whether affected firms are incentivized to manage their earnings well; firms actively engaged in information disclosure will increase their transparency.

Does this policy have long-term impacts on the cost of borrowing, the attractiveness of IPOs, and investment activity? Surprisingly, I find that the pricing cap policy, overall, does not affect borrowing costs, the willingness to go public, or firm investment . The probability of a price cascade in the first month after an IPO is also reduced. Overall, the results indicate that these regulations can be circumvented by the firm or the manager's strategic actions and that regulation can effectively reduce the short-term probability of a price cascading effect for an IPO.

This paper sheds light on financial market regulation in the IPO market and its impact on firm investment information flow and further extends government regulation research. Traditionally, it is difficult to isolate the impact of government regulation and firm behavior. My paper may be the first study using the exogenous shock "pricing cap regulation" to identify the regulation and its impact on firms. I apply a Dif-in-Dif analysis and further supplement a robustness test using the PSM Dif-in-Dif framework.

## References


Almeida, H., M. Campello, I. Cunha, M. S. Weisbach, 2014. Corporate liquidity management: A conceptual framework and survey. Annual Reviews of Finance and Economics 6, 135–162.

Altman, E. I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance 23(4), 589–609.

Altman, E. I., G. Marco, F. Varetto, 1994. Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). Journal of Banking & Finance 18(3), 505–529.

Beaver, W. H., 1966. Financial ratios as predictors of failure. Journal of Accounting Research pp. 71–111.

Beaver, W. H., M. F. McNichols, J.-W. Rhie, 2005. Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. Review of Accounting Studies 10(1), 93–122.

Black, F., J. C. Cox, 1976. Valuing corporate securities: Some effects of bond indenture provisions. The Journal of Finance 31(2), 351–367.

Black, F., M. Scholes, 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81(3), 637–654.

Bolton, P., D. S. Scharfstein, 1990. A Theory of Predation Based on Agency Problems in Financial Contracting. American Economic Review pp. 93–106.

Brander, J., T. Lewis, 1986. Oligopoly and Financial Structure: The Limited Liability Effect. American Economic Review 76, 956–970.

Campbell, J. Y., J. Hilscher, J. Szilagyi, 2008. In search of distress risk. The Journal of Finance 63(6), 2899–2939.

Campello, M., 2003. Capital structure and product market interactions: Evidence from business cycles. Journal of Financial Economics 68, 353–378.

Chava, S., R. A. Jarrow, 2004. Bankruptcy prediction with industry effects. Review of Finance 8(4), 537–569.

Chevalier, J. A., D. S. Scharfstein, 1996. Capital Market Imperfections and Countercyclical Markups: Theory and Evidence. American Economic Review 86, 703–725.

Collin-Dufresne, P., R. S. Goldstein, 2001. Do credit spreads reflect stationary leverage ratios?. The Journal of Finance 56(5), 1929–1957.

Dasgupta, S., S. Titman, 1998. Pricing strategy and financial policy. Review of Financial Studies 11, 705–737.

Dichev, I. D., 1998. Is the risk of bankruptcy a systematic risk?. the Journal of Finance 53(3), 1131–1147.

Duffie, D., K. J. Singleton, 1997. An econometric model of the term structure of interest-rate swap yields. The Journal of Finance 52(4), 1287–1321.

——— , 1999. Modeling term structures of defaultable bonds. The Review of Financial Studies 12(4), 687–720.

Fama, E. F., K. R. French, 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Frésard, L., 2010. Financial Strength and Product Market Behavior: The Real Effects of Corporate Cash Holdings. Journal of Finance 65(3), 1097–1122.

Gan, J., 2009. Privatization in China: experiences and lessons. in *China's Emerging Financial Markets*. Springer pp. 581–592.

Geske, R., 1977. The valuation of corporate liabilities as compound options. Journal of Financial and Quantitative Analysis pp. 541–552.

Goldstein, R., N. Ju, H. Leland, 2001. An EBIT-based model of dynamic capital structure. The Journal of Business 74(4), 483–512.

Haushalter, D., S. Klasa, W. F. Maxwell, 2007. The influence of product market dynamics on a firm's cash holdings and hedging behavior. Journal of Financial Economics 84, 797–825.

Hoberg, G., G. M. Phillips, N. R. Prabhala, 2014. Product Market Threats, Payouts, and Financial Flexibility. Journal of Finance 69(1), 8.

Holmström, B., J. Tirole, 1998. Private and public supply of liquidity. Journal of political Economy 106, 1–40.

Jarrow, R. A., D. Lando, S. M. Turnbull, 1997. A Markov model for the term structure of credit risk spreads. The Review of Financial Studies 10(2), 481–523.

Jarrow, R. A., S. M. Turnbull, 1995. Pricing derivatives on financial securities subject to credit risk. The

Journal of Finance 50(1), 53–85.

Khanna, N., S. Tice, 2005. Pricing, exit, and location decisions of firms: Evidence on the role of debt and operating efficiency. Journal of Financial Economics 75, 397–427.

Kim, C.-S., D. C. Mauer, A. E. Sherman, 1998. The determinants of corporate liquidity: Theory and evidence. Journal of financial and quantitative analysis 33(3), 335–359.

Kleinberg, J., J. Ludwig, S. Mullainathan, Z. Obermeyer, 2015. Prediction policy problems. American Economic Review 105(5), 491–95.

Leland, H. E., 1994. Corporate debt value, bond covenants, and optimal capital structure. The Journal of Finance 49(4), 1213–1252.

——— , 2004. Predictions of default probabilities in structural models of debt. Journal of Investment Management 2(2).

Longstaff, F. A., E. S. Schwartz, 1995. A simple approach to valuing risky fixed and floating rate debt. The Journal of Finance 50(3), 789–819.

MacKay, P., G. M. Phillips, 2005. How Does Industry Affect Firm Financial Structure?. Review of Financial Studies 18, 1433–1466.

Maksimovic, V., 1990. Product Market Imperfections and Loan Commitments. The journal of finance 45(5), 1641–1653.

Martin, J. S., A. M. Santomero, 1997. Investment Opportunities and Corporate Demand for Lines of Credit. Journal of Banking & Finance 21(10), 1331–1350.

Merton, R. C., 1973. Theory of rational option pricing. The Bell Journal of Economics and Management Science pp. 141–183.

——— , 1974. On the pricing of corporate debt: The risk structure of interest rates. The Journal of finance 29(2), 449–470.

Nikolov, B., L. Schmid, R. Steri, 2019. Dynamic corporate liquidity. Journal of Financial Economics 132, 76–102.

Ohlson, J. A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research pp. 109–131.

Opler, T., L. Pinkowitz, R. Stulz, R. Williamson, 1999. The Determinants and Implications of Corporate

Cash Holdings. Journal of Financial Economics 52, 3–46.

Phillips, G., G. Sertsios, 2013. How do firm financial conditions affect product quality and pricing?. Management Science 59, 1764–1782.

Phillips, G. M., 1995. Increased Debt and Industry Product Markets: An Empirical Analysis. Journal of Financial Economics 37, 189–238.

Satchidananda, S., J. B. Simha, 2006. Comparing decision trees with logistic regression for credit risk analysis. in *SUGI Asia Conference, Mumbai, India*.

Sufi, A., 2009. Bank lines of credit in corporate finance: An empirical analysis. The Review of Financial Studies 22(3), 1057–1088.

Valta, P., 2012. Competition and The Cost of Debt. Journal of Financial Economics 105(3), 661–682.

Vassalou, M., Y. Xing, 2004. Default risk in equity returns. The Journal of Finance 59(2), 831–868.

Zhou, Z.-H., 2012. Ensemble methods: foundations and algorithms. CRC press, .

# Competition and the Use of Credit Lines*

J. Daniel Chi† Zhou Lu‡ Xunhua Su§ Xiaoyu Zhang ¶

**Abstract**

Credit lines and cash holdings are the two main sources of corporate liquidity. Theory predicts that when facing more intense competition, a firm should demand more liquidity to strengthen its competitive position against rivals. We find that in contrast to the widely documented evidence that competition increases cash holdings, competition significantly reduces firms' use of credit lines. This finding is robust to alternative measures of competition as well as exogenous variation in competition. Further analysis suggests that the negative effect of competition on credit line usage is mainly driven by banks' restricted credit supply, rather than firms' reduced demand. Competition induces negative pressures on firm performance, making it more difficult to obtain credit lines from banks.

**Keywords:** *Competition, Corporate liquidity, Credit lines, Cash holdings*
**JEL Classification**: G31, G32, L10, L21

*This draft: June 26, 2020. For helpful comments and suggestions on previous drafts, we thank Heitor Almeida, Ramin Baghai, Jerry Hoberg, Gordon Phillips, Nagpurnanand Prabhala, Han Xia, and Lisa Yang. All errors are our own.

†Department of Finance, University of Nevada, Las Vegas (UNLV). daniel.chi@unlv.edu
‡Department of Finance, Norwegian School of Economics (NHH). zhou.lu@nhh.no
§Department of Finance, Norwegian School of Economics (NHH). xunhua.su@nhh.no
¶Department of Finance, Vrije Universiteit Amsterdam. xiaoyu.zhang@vu.nl

# Contents

# 1 Introduction

Competition in the product market continuously shapes corporate financial decisions, among which liquidity management is of particular relevance. After all, cash is king, and a firm goes bankrupt when running out of cash. The general finding of the literature is that cash offers greater strategic value as competition intensifies, and therefore firms facing more intense competition increase their cash holdings.[1] But cash is only one source of corporate liquidity. The other major source of corporate liquidity – credit lines provided by banks – are used by over 65% of U.S. public firms. For these firms, the amount of available credit lines is as large as the amount of cash holdings. According to Federal Reserve data, in the last decade (2010 to 2019) the aggregate amount of committed commercial credit lines by U.S. banks is even larger than the amount of outstanding commercial loans.[2] Thus, credit lines are a significant component of corporate liquidity and represent more than 50% of bank commercial lending. Yet we know little about how competition affects firms' use of credit lines or how competition affects the relative importance of cash and credit lines as liquidity sources. We attempt to fill this void in this study.

Theoretical and empirical evidence almost unequivocally shows that more intense competition leads to higher cash holdings. Similar intuition should also apply to credit lines, especially considering that credit lines arguably offer more efficient liquidity insurance than cash holdings (e.g., Holmström and Tirole, 1998). Studying the specific context of competition, Maksimovic (1990) and Martin and Santomero (1997) offer theoretical predictions that firms facing more intense competition should *increase* the use of credit lines. However, a fundamental difference between cash holdings and credit lines is that cash is internal liquidity that is unconditionally available to the firm. In contrast, the availability of credit lines is conditional on the supply of credit lines by banks (e.g., Sufi, 2009; Acharya, Almeida, and Campello, 2013). Because a credit line exposes a bank to significant risks of adverse selection and moral hazard, the bank will engage in extensive screening of potential borrowers and employ sophisticated contracts to protect itself. Therefore financially

[1] For example, see Bolton and Scharfstein (1990), Haushalter, Klasa, and Maxwell (2007), Frésard (2010), and Hoberg, Phillips, and Prabhala (2014).

[2] The Federal Reserve publishes data on U.S.-charted depository institutions' off-balance-sheet items, which include committed but unused C&I loan amount. The Fed also publishes outstanding C&I loan amount in its Table H.8. Between 2010 and 2019, the amount of "unused C&I loan commitments" was 15% higher than the amount of outstanding C&I loans.

weak firms, for example those with low operating cash flow, are less likely to secure a credit line commitment. Since a main effect of competition is reduced profit margin, it is plausible that more intense competition will *decrease the supply* of credit lines and hence reduce the use of credit lines.

Thus, competition increases the strategic benefits of having a credit line but also increases the difficulty of obtaining a line. What is the net effect of competition on credit line usage? Using a comprehensive sample of U.S. public firms from 2002 to 2019, we find that competition reduces the use of credit lines. A one-standard-deviation increase in competition intensity reduces the amount of undrawn credit relative to assets by 1.8 percentage points (ppt) or 23% of the sample mean, and reduces undrawn credit as a proportion of total corporate liquidity by 6.5 ppt or 19% of the sample mean. That is, competition not only reduces the absolute usage of credit lines but also the relative usage when compared with cash holdings. The economic significance of competition is comparable to that of known determinants of credit line usage, such as cash flow level and asset tangibility. This baseline result is robust to alternative measures of competition, different data sources of credit line usage, and different samples and sample periods. In Figure 1, we plot the dynamics of credit line usage as competition changes. For firms in the lowest competition-intensity quartile, the amount of credit lines represents over 60% of total liquidity (the sum of cash holdings and undrawn lines). As competition intensifies, the amount of credit lines decreases and eventually drops below 50% of total liquidity.

Can we draw a causal inference from the negative correlation between competition intensity and credit line usage? The answer hinges on whether the variation in our competition measures are exogenous. Our first competition measure is the Herfindahl–Hirschman Index (HHI). HHI is measured at the industry level and is largely exogenous to individual firms. A limitation of HHI is that it is static and backward-looking. Our second measure of competition is product market *Fluidity*, constructed by Hoberg, Phillips, and Prabhala (2014). Based on textual analysis of business descriptions in 10-K filings, *Fluidity* measures the intensity of change in a firm's product space. Higher *Fluidity* means greater threats in the product market. *Fluidity* is defined at the firm level, so it more closely captures inter-firm competition dynamics than HHI does; but *Fluidity* is partly influenced by a firm's own actions and could be partly endogenous. To mitigate this endogeneity concern, we use import tariff rates as an instrument for *Fluidity* (Li and Zhan, 2018).

Lower tariff rates reduce barriers of entry for foreign competitors and hence raise competition intensity. At the same time, it is difficult to conceive how tariff rates can affect corporate liquidity management other than through the competition channel. Therefore, tariff rates as an instrumental variable satisfy both the relevance condition and the exclusion-restriction condition. Two-Stage Least Squares (2SLS) analysis confirms the baseline results. We use extra import tariffs imposed by Donald Trump and his administration as exogenous shocks to ascertain causality . During Trump's presidency, the government policy does not have a predictable pattern and consider as a surprise to the company. On January 23, 2018, Trump had imposed tariffs on solar panels produced outside the United States, and the Office of the U.S. Trade Representative announced tariffs on washing machines. On March 1, 2018, Trump announced his intention to impose a 25% tariff on steel and a 10% tariff on aluminum imports. The import tariff shock served as a positive demand shock for producing solar panels, washing machines, steel, and the aluminum industry. The increased demand for those industries, at least in the short run, reduces competition intensity. This reduced competition intensity should increase the use of credit lines. We test this prediction using a difference-in-differences (DiD) regression. The results show that treated firms have more credit lines after tariff shock and rely more on credit lines as a liquidity source.

Taken together, the empirical findings suggest that when facing more intense competition, firms adjust their liquidity management strategy: they rely less on credit lines and more on cash holdings. To our knowledge, we are the first to provide evidence on the effect of competition on credit line usage. Unlike cash holdings that can be deployed at a firm's sole discretion, credit line usage reflects the equilibrium of demand by firms and supply by banks. Our baseline evidence implies that competition decreases banks' supply of credit lines more than it increases firms' demand, resulting in a negative net effect of competition on credit line usage.

We perform several tests to directly examine how competition affects the supply and demand channels of credit lines. First, it is well documented in the literature that asset tangibility increases the supply of credit by banks (e.g., Almeida and Campello, 2007). But there is no obvious reason that tangibility should affect a firm's demand for credit. Therefore, if the negative relation between competition and credit line usage is attenuated by asset tangibility, it is evidence that competition affects the supply channel of credit lines. Second, growth opportunities increases a firm's demand for credit lines (e.g., Lins, Servaes, and Tufano, 2010), but should not significantly affect the supply because banks are mostly care with the stability of cash flow, not the growth of cash flow. Therefore, if the negative relation between competition and credit line usage is attenuated by growth opportunities, it is evidence that competition affects the demand channel of credit lines. Third, lower levels of cash flow and higher levels of cash flow volatility should increase a firm's demand for credit lines but make banks more wary of supplying credit lines (Sufi, 2009). Thus, how the negative relation between competition and credit line usage varies with cash flow level and cash flow volatility shall tell us whether the demand channel or the supply channel dominates.

We implement a conditional-test framework by adding an interaction term between competition and one of the conditioning variables mentioned above. The conditional tests deliver several interesting findings. The negative effect of competition on credit line usage is less pronounced for firms with higher asset tangibility, consistent with a supply channel effect. In contrast, higher growth opportunities does not attenuate the negative effect of competition on credit line usage, suggesting that the demand channel does not exert a significant effect. In addition, the negative effect of competition on credit line usage is more negative for firms with lower cash flow and higher cash flow volatility, suggesting that the supply channel effect dominates the demand channel effect. Taken together, these findings suggest that competition reduces credit line usage not because firms demand lower credit lines when facing more intense competition, but rather that banks restrict credit line supply to these firms.

This study contributes to the literature on how product market competition affects corporate liquidity management. Previous works in this literature mostly focus on cash holdings and largely overlook another equally important source of liquidity – credit lines. Empirical evidence shows that competition increases corporate cash holdings because cash provides greater strategic value in a more competitive environment (e.g., Bolton and Scharfstein, 1990; Haushalter, Klasa, and Maxwell, 2007; Frésard, 2010; Hoberg, Phillips, and Prabhala, 2014). Extant theory also suggests a positive effect of competition on credit line usage because credit lines offer obvious competitive advantages (e.g., Maksimovic, 1990; Martin and Santomero, 1997), but our evidence shows the opposite. This seeming inconsistency is not driven by firms' reduced demand for credit lines, but rather by banks' restricted supply of credit lines to firms facing intense competition. Our findings point to the need for more theoretical study on the effect of competition on the supply side of credit lines. The results also seem to be inconsistent with the traditional wisdom that treats cash and credit lines as perfect substitutes (e.g., Holmström and Tirole, 1998). Our findings suggest that on the demand side, cash and credit lines could be substitutes; but on the supply side, they are certainly not. Cash as unconditional liquidity is under the full control of firms, while the equilibrium holding of credit lines is conditional also on banks' supply. This finding is consistent with the notion that credit lines are contingent liquidity (e.g., Sufi, 2009; Nikolov, Schmid, and Steri, 2019).

Our findings on credit lines also have implications for the relation between competition and cash holdings. As competition intensifies, firms *want to* hold more cash because the strategic advantages of cash increase. But firms also *have to* rely more on cash as a liquidity source because credit line supply becomes more restricted. So, the accessability of alternative liquidity sources may also be a determinant of corporate cash holdings. This has the potential to help us better understand the determinants of cash holdings (e.g., Kim, Mauer, and Sherman, 1998; Opler, Pinkowitz, Stulz, and Williamson, 1999). By studying alternative corporate liquidity sources in a unified framework, we also extend Almeida, Campello, Cunha, and Weisbach (2014).

Furthermore, our findings point to a further dimension that differentiates credit lines and cash holdings as a corporate liquidity source. Sufi (2009) argues that firms prefer credit lines over cash when they are less likely to violate cash-flow-based financial covenants. Lins, Servaes, and Tufano (2010) find that cash and credit lines are held for different purposes: credit lines to fund future investment opportunities, whereas cash to buffer future cash flow shortfalls. Acharya, Almeida, and Campello (2012) show that firms with high asset betas hold more cash relative to credit lines because it is more costly for them to obtain credit lines from banks. We add to this literature by showing that competition reduces the use of credit lines relative to cash as a corporate liquidity source.

Finally, our findings add to the extensive literature on the interactions between product markets and various corporate financial policies or outcomes, including cost of debt (Valta, 2012), capital structure (e.g., Brander and Lewis, 1986; Phillips, 1995; Campello, 2003; MacKay and Phillips, 2005), and product strategies (e.g., Chevalier and Scharfstein, 1996; Dasgupta and Titman, 1998; Khanna and Tice, 2005; Phillips and Sertsios, 2013). Specifically, we add to this literature by showing how competition affects corporate liquidity management.

The rest of the paper is organized as follows. Section 2 outlines the research question, the methodology, and the identification strategy. Section 3 describes the data, sample, and variables. Section 4 presents our baseline results, robustness tests, and endogeneity checks. Section 5 examines whether the baseline effect is the result of competition reducing the demand for credit lines or restricting the supply of credit lines. Finally, section 6 concludes.

# 2 Research Question, Methodology, and Identification

## 2.1 The Relation between Competition and Credit Line Usage

In the past several decades, cash holdings of U.S. publicly listed firms have increased to as much as 22% of their book assets. Much attention has been paid to why firms hold so much cash, and the precautionary motive is argued as one of the most important drivers (e.g., Opler, Pinkowitz, Stulz, and Williamson, 1999; Bates, Kahle, and Stulz, 2009; Duchin, Gilbert, Harford, and Hrdlicka, 2017). Cash holdings represent a safeguard against the inability to obtain financing when valuable investment opportunities arise, in particular when there are significant frictions in financial markets (e.g., Almeida, Campello, and Weisbach, 2004). A strand of the literature further shows that product market competition can be one reason for precautionary savings because competition induces negative pressures on firm performance and increases financially constraints, e.g., see the theoretical study by Bolton and Scharfstein (1990) and empirical studies by Haushalter, Klasa, and Maxwell (2007), Frésard (2010), and Hoberg, Phillips, and Prabhala (2014).[3]

But cash is only one source of corporate liquidity. The other important source of liquidity is credit lines provided by banks. How does competition affect the use of credit lines? First, competition affects the demand for credit lines. It is widely accepted that competition increases the demand for cash holdings. Similar arguments underlie the theory of credit lines, which views credit lines as option-like cash equivalents (e.g., Arnoud, Thakor, and Udell, 1987; Shockley and Thakor, 1997; Arnoud, Thakor, and Udell, 1987). Maksimovic (1990) outlines in a Cournot competition model

[3]Chi and Su (2016) provide theoretical and empirical evidence that competition increases the strategic value of cash. Also see Bates, Chang, and Chi (2018) for evidence in the 1990s.

that the availability of a credit line lowers expansion costs and increases the the severity of a firm's threat against rivals. Thus, credit lines enhance a firm's strategic position against rivals, and therefore competition increases the use of credit lines. Martin and Santomero (1997) show that credit lines allow a firm to capture valuable investment opportunities with speed and secrecy before rivals do. Competition makes investment opportunities more short-lived and therefore should increase firms' demand for credit lines. In sum, these theories argue that competition increases the demand for credit lines. We call this the *increased demand channel*.

Second, the literature has largely ignored the supply-side effect of competition on the use of credit lines. While cash as internal or unconditional liquidity is under the full discretion of firm managers, lines of credit as external or contingent liquidity are conditional on banks' supply. Firms draw down credit lines if and only if they think credit lines are the least costly financing source, for example, during the 2007-2009 financial crisis when other sources of financing were largely exhausted (Ivashina and Scharfstein, 2010). Therefore, credit lines expose a bank to significant risks of adverse selection and moral hazard (Avery and Berger, 1991; Shockley and Thakor, 1997), as well as liquidity risk (Acharya, Almeida, and Campello, 2013; Ippolito, Peydró, Polo, and Sette, 2016; Ippolito, Almeida, Orive, and Acharya, 2019). To mitigate these risks, the bank will engage in extensive screening of potential borrowers and employ sophisticated contracts to protect itself. Therefore financially weak firms, for example those with low operating cash flow, are less likely to secure a credit line commitment (Sufi, 2009). Since a main effect of competition is reduced profit margins, more intense competition should reduce the supply of credit lines by banks and hence reduce the equilibrium use of credit lines. We call this the *restricted supply channel*.

Collectively, competition can have both positive and negative effects on the use of credit lines through different economic channels. It may increase credit line usage through the *increased demand channel*, but may decrease credit line usage through the *restricted supply channel*. The net effect will depend on the relative strength of these channels and is ultimately an empirical question, which we examine next.

## 2.2 Baseline Model

### 2.2.1 Specification

To examine how competition affects firms' credit line usage, we use the following specification:

$$\textit{Credit Line Usage}_{i,t} = \alpha + \beta \cdot \textit{Competition}_{i,t} + \Gamma \cdot \mathbf{X}_{i,t} + \theta_i + \eta_t + \epsilon_{i,t} \tag{1}$$

The dependent variable, *Credit Line Usage*, is one of the variables measuring the use of credit lines. Note that the portion of a credit line that has been drawn has already been converted to either internal liquidity or other assets, and hence is not external liquidity any more. Therefore, we focus on the amount of undrawn credit lines. The independent variable of interest, *Competition*, is one of the measures of product market competition. We have a list of firm and industry characteristics as control variables ($\mathbf{X}$) as well as industry and year fixed effects $\theta$ and $\eta$. We describe variable construction in the following sections and summarize them in Table 2. Depending on the measure of credit line usage, we employ different estimation methods. For example, we estimate a Probit model when *Credit Line Usage* is a binary variable indicating whether the firm has undrawn credit lines.

### 2.2.2 Measuring Credit Line Usage

Following the literature (e.g., Sufi, 2009; Acharya, Almeida, and Campello, 2013), our main dependent variables include: (1) *HasLine*, (2) *Undrawn_Assets*, and (3) *Undrawn_TotLiq*. *HasLine* is a dummy that equals one if a firm has undrawn credit at the end of the year, and zero otherwise.[4] *Undrawn_Assets* is the amount of undrawn credit, scaled by the value of book assets. *Undrawn_TotLiq* is the amount of undrawn credit scaled by total corporate liquidity (*TotLiq*), defined as the sum of undrawn credit and cash holdings. This last variable captures the relative importance of credit lines vs. cash as a source of corporate liquidity.

[4] Our definition of *HasLine* is slightly different from Sufi (2009), who defines *HasLine* as a dummy that equals one if a firm has lines of credit, no matter whether the line has been drawn or not. We change the definition in order to capture the use of firms' external liquidity – *undrawn* credit. As argued, drawn credit has been converted to internal liquidity or other assets, and is hence not external liquidity any more. It is worth emphasizing that when we follow Sufi's definition, we obtain very similar results.

### 2.2.3 Measuring Competition

We employ several measures of competition that have been used in the literature. The first measure is the Herfindahl–Hirschman Index (HHI), which is defined as the sum of squared market shares of each firm and has long been used in economics, finance, and management. The lower the HHI, the less concentrated the industry, and hence the more intense the competition. We construct sales-based HHI for each 3-digit-SIC industry as follows:

$$HHI = \sum_{i=1}^{N} \left[ \frac{Sale_i}{\sum_{i=1}^{N} Sale_i} \right]^2 = \sum_{i=1}^{N} s_i^2 \tag{2}$$

where $Sale_i$ is the sales of firm $i$, and hence $s_i = \frac{Sale_i}{\sum_{i=1}^{n} Sale_i}$ is the market share of firm $i$ in the 3-digit SIC industry. HHI is defined at the industry level, so it assumes that every firm in the industry faces the same level of competition intensity. HHI is an inverse measure of competition. To ease interpretation, we define an inverse HHI: *HHIInv* $= -$ *HHI*. Higher *HHIInv* means more intense competition.

Our second measure of competition is product market fluidity (*Fluidity*), developed by Hoberg, Phillips, and Prabhala (2014). Based on textual analysis of business descriptions in 10-K filings, *Fluidity* is the dot product between the words used in a firm's business description and the change in the words used in its rivals' business descriptions. When rivals change their business descriptions to be more similar to the firm's, the overlap in word usage increases, and *Fluidity* increases. Because *Fluidity* captures the "change" in rivals' word usage relative to the firm's word usage, it is a dynamic measure of competition intensity at the firm level and is relatively exogenous to firm's own choices. The literature shows that higher *Fluidity* leads to lower dividend payout and higher cash holdings (Hoberg, Phillips, and Prabhala, 2014), higher value of cash (Chi and Su, 2016), as well as higher stock market crash risk (Li and Zhan, 2018). As product description likely precedes real market actions, *Fluidity* is forward-looking and is hence called product market "threats" in these previous studies. The *Fluidity* measure is defined at the firm level and more closely captures inter-firm competition dynamics.

We use two more competition measures in robustness checks. The first is an HHI measure where industries are classified not by SIC but by the text-based network industry classification (TNIC) (Hoberg and Phillips, 2016). TNIC groups firms into various industries based on textual analysis of each firm's product description. Relative to SIC-based HHI where industry classification is fairly static, TNIC-based HHI is dynamic because industry grouping changes over time. The second measure is *PctComp*, a measure of competition based on how extensively the management discusses competition in the 10-K filings (Li, Lundholm, and Minnis, 2013). The more the management discusses competition in 10-K filings, the higher the *PctComp*, and the greater the competition intensity.

#### 2.2.4 Control Variables

We construct a list of control variables that are widely used in the empirical literature on credit line usage (e.g., Sufi, 2009; Acharya, Almeida, and Campello, 2013). Firm size is measured by the natural logarithm of book value of non-cash assets (log*Assets*). Following Sufi (2009), we use non-cash assets, instead of total assets, when scaling the control variables. Firm age is measured by the natural logarithm of the number of years since IPO (log*FirmAge*). Arguably, larger and older firms build deeper relationship with banks and are more likely to use credit lines. As a profitability measure, *CashFlow* is defined as operating income before depreciation divided by non-cash assets. We capture firms' growth opportunities by the market-to-book ratio (*M2B*), calculated as the sum of the market value of equity and book value of total debt, divided by non-cash assets. Research and development investment captures innovation, which helps a firm to escape from competition. Similar to Chan, Lakonishok, and Sougiannis (2001), we define R&D Capital (*RDC*) as the accumulated R&D expenses in the past four years, using a 25% annual depreciation rate, scaled by non-cash assets. *Networth* is a measure of firm leverage calculated as cash-adjusted net equity, scaled by total non-cash assets. *Tangibility* is the ratio of net PP&E to non-cash assets and is a proxy for the size of assets that can be pledged as collateral. All else equal, a firm with higher tangibility is more likely to obtain a line of credit from banks. We control for two volatility measures. Cash flow volatility, *CFVol*, is the standard deviation of firm EBITDA in the past four quarters, scaled by non-cash assets. Industry sales volatility, *IndSaleVol*, is the (3-digit SIC) industry median value of the standard deviation of firm sales in the past four quarters, scaled by non-cash assets. Finally, we include *NonSP500*, a dummy indicator that equals one for firms not in the S&P500 Index.

### 2.3 Identification – Instrumental Variable Approach

Various papers have used import tariff rates to capture competition intensity (e.g., Frésard, 2010; Valta, 2012; Xu, 2012). Xu (2012) uses tariff rates as an instrument for import penetration, and Li and Zhan (2018) use tariff rates as an instrument for *Fluidity*.[5] Lower tariff rates reduce barriers of entry for foreign competitors and hence raise competition intensity. At the same time, it is difficult to conceive how tariff rates can affect corporate liquidity management other than through the competition channel. Therefore, tariff rates as an instrumental variable for *Fluidity* satisfy both the relevance condition and the exclusion-restriction condition.[6]

## 3 Data, Sample, and Summary Statistics

### 3.1 Data and Sample Selection

We obtain credit line data from the S&P Capital IQ - Capital Structure database. Capital IQ's credit line data start as early as 1995, but the data before 2002 are sporadic. For example, before 2001, only about 2% firms are shown to have a credit line. The number jumps to 32% in 2001, 65% in 2002, and stays fairly stable in the following years. To ensure data quality, we start our sample period from 2002. The sample ends in 2017 because our key measure of product market competition, *Fluidity*, is available up to 2017.

We merge Capital IQ with CRSP-Compustat Merged (CCM) database, and construct our full sample with the following steps. First, we retain only US-based firms, and remove all regulated (Standard Industrial classification (SIC) 4900-4949) and financial industries (SIC 6000-6999). Second, we drop observations if total assets are missing or less than US$ 1 million, if cash holdings are negative, or if total assets are smaller than the amount of cash holdings. With this step, we re-

[5]They also use exchange rates as another instrumental variable. Constrained by the availability of credit line data, our sample period is from 2002 to 2017, which is a more recent period than their sample period. For our sample period, exchange rates are a weak instrument for *Fluidity*.

[6]Alternatively, the literature uses large reductions in import tariff rates as exogenous shocks of competition (e.g., Frésard, 2010; Valta, 2012; Bharath and Hertzel, 2019). Most large reductions in U.S. import tariff rates occurred in 1980s and 1990s. During our sample period from 2002 to 2017, there are few significant reductions that we can employ as exogenous shocks.

move tiny firms as well as firms with erroneous cash holdings data relative to total assets. Finally, we drop observations with missing information on the variables of interest or the control variables. The full sample has 47,765 firm-year observations for 6,511 unique firms between 2002 and 2019. We are also interested in a subsample of firms with undrawn credit lines (where *HasLine*=1), which consists of 36,350 firm-year observations for 4,917 unique firms. All variables are winsorized at the 1st and 99th percentiles to reduce the effect from extreme outliers.[7]

### 3.2 Credit Line Usage by Year

To provide an overview of the key corporate-liquidity variables, we first report in Table 3 the annual means of these variables. The far-right column reports the time series averages. Panel A is for the full sample, and Panel B is for the subsample of firms that have a credit line.

Panel A shows that throughout the sample period, about 76% of firms have a credit line. This percentage is fairly stable over time, with a peak of 81% in 2013 and a trough of 70% in 2002. On average, the size of credit lines is about 14% the size of firm assets, of which 4% is drawn credit and 10% is undrawn credit. That is, undrawn credit is almost three times as large as the amount of drawn credit. This undrawn credit is about 42% of firms' total liquidity and 74% of firms' total debt.

Moving on to Panel B where we consider only firms with a credit line, it becomes clear that credit lines are an even more important liquidity source for these firms. On average, the size of undrawn credit is 13% the size of total assets, versus cash at 13%. Undrawn credit lines represent 55% of total liquidity. Relative to the amount of outstanding debt, the amount of undrawn credit lines is two and a half times as large. The literature has paid much attention to corporate cash holdings, but these statistics suggest that credit lines are an equally important source of corporate liquidity.

[7] We further winsorize *Networth* and *CashFlow* at the 5th and 99th percentiles, and M2B and RDC at 1st and 95th percentiles, because of their extremely skewness.

### 3.3 Summary Statistics

Table 4 reports detailed summary statistics on liquidity variables, competition measures, and control variables. All unscaled variables, e.g., total assets and credit line amounts, are inflation-adjusted to 2010 USD using the Headline CPI published by the Bureau of Labor Statistics. Again, in Table 4, Panel A is for the full sample, and Panel B is for the subsample of firms that have a credit line. The *NonSP500* indicator in Panel A shows that 87% of observations are not in the S&P500 index, suggesting that our sample consists of a wide spectrum of firms in the economy.

We will focus our discussion on Panel B which contains firms that have credit lines. First, the amount of undrawn credit (*Undrawn*) has a mean of US $325.48 million and a median of 83.56 million. Although the mean of *Undrawn* is smaller than that of cash holdings, i.e., *Cash* (359), the median of *Undrawn* is slightly higher than that of *Cash* (54.68). In the data, a disproportionally large amount of aggregate cash is held by a small group of large firms, such as Apple and Alphabet, and therefore cash is more positively skewed than undrawn credit. Second, the ratios of cash and undrawn credit to total assets show a similar pattern: for the mean, *Cash_Assets* and *Undrawn_Assets* are paramount (13% vs. 13%); for the median, *Undrawn_Assets* is larger (10% vs. 8%). The ratio of undrawn credit to total liquidity, *Undrawn_TotLiq*, is fairly normally distributed, with a mean of 55% and a median of 57%. These figures again show the importance of credit lines as a source of corporate liquidity. Third, this subsample of observations have lower *Fluidity* and higher *HHI*, indicating that firms with credit lines face lower competition. This is consistent with our main hypothesis that competition decreases the use of credit lines as a liquidity source. Finally, firms with credit lines are relatively larger (higher *Assets*), older (higher *FirmAge*), and more profitable (higher *CashFlow*). Compared with the full sample, these firms also have lower market-to-book ratio (*M2B*, 1.98 vs. 2.74) and lower R&D capital (*RDC*, 0.08 vs. 0.16). To the extent that *M2B* and *RDC* measures investment opportunities, both figures seem inconsistent with the survey findings of corporate CFOs (Lins, Servaes, and Tufano, 2010) that credit lines are mainly used to finance future investment opportunities.

# 4 Competition and Credit Line Usage

## 4.1 Univariate Results

Our first main research question is how competition affects the use of (undrawn) credit lines in corporate liquidity management. As argued in Section 2, the effect of competition on credit line usage depends on two opposite drivers and is unclear ex ante. Before employing more sophisticated methodology, we first take a look at the univariate relationship between competition and credit line usage through a simple figure.

We plot in Figure 1 the average ratios of undrawn credit to total liquidity (*Undrawn_TotLiq*) and book assets (*Undrawn_Assets*) across the four competition quartiles. We consider only the sub-sample of firms that have undrawn credit lines (i.e., *HasLine*=1). In Panel A, we use inverse HHI (*HHIInv*) to measure competition. As *HHIInv* or competition intensity increases from the lowest quartile to the highest quartile, the undrawn-credit-to-total-liquidity ratio decreases monotonically from about 60.2% to about 45.0% (the left scale). The difference (15.2 percentage points) is one fifth of the sample average undrawn-credit-to-total-liquidity ratio (0.42). Similarly, the undrawn-credit-to-assets ratio also monotonically decreases from 13.5% to 11.2% (the right scale). The difference (2.3 percentage points) is about 23% of the sample average (0.10).

In Panel B, we use *Fluidity* to measure competition. Unlike *HHI* that is defined at the industry level and therefore all firms in an SIC-year cohort have the same *HHI* value, *Fluidity* is defined at the firm-level and has greater cross-sectional variation. The figure shows that as Fluidity increases, credit line usage monotonically decrease. Specifically, the undrawn credit-to-total liquidity ratio decreases from 61.2% for the highest *Fluidity* quartile to 52.0% for the lowest quartile, while the ratio of undrawn credit-to-assets drops from 14.2% to 11.5%. The magnitudes of the decreases seem economically large. To summarize, the univariate results show that competition decreases the use of credit lines in corporate liquidity management.

## 4.2 Baseline Tests

### 4.2.1 HHI and Credit Line Usage

We next employ more rigorous techniques to examine our research question. A spoiler warning is that the results from fancier econometrics confirm the findings of Figure 1. To start, we use sales-based HHI at the 3-digit-SIC level as a measure of product market competition and run Equation (1). Table 5 reports the marginal effects of the regressions.

We first run a Probit regression in Column (1) to examine whether changes in HHI affect the probability of having a credit line. The dependent variable is *HasLine*, and the explanatory variable of interest is inverse HHI (*HHIInv*). We control a large set of firm characteristics as well as industry and year fixed effects. The result shows a significant decrease in credit line usage when *HHIInv* increases, i.e., when competition intensifies. This decrease is economically large. For a one-standard-deviation increase in *HHIInv*, the probability of having a credit line decreases by 2.0 percentage points. Putting this number into context, a one-standard-deviation increase in *CashFlow* raises the probability of having a line by 22 percentage points, and a one-standard-deviation increase in *Tangibility* increase the probability by 1 percentage point. Cash flow and tangibility are among the most important determinants of corporate borrowing capacity (e.g., Almeida and Campello, 2007; Sufi, 2009). The economic significance of competition suggests that competition is another important determinant of credit line usage.

In Columns (2) and (3), we employ Tobit models to examine whether HHI affects the amount of undrawn credit lines. We run Tobit regressions because more than one third of observations have zero undrawn credit, mostly because they do not have a credit line.[8] The dependent variable in Column (2) is the ratio of undrawn-credit-to-assets, and in Column (3) the ratio of undrawn-credit-to-total-liquidity. The results in both columns suggest that as competition increases, the amount of undrawn credit decreases. The result in Column (3) shows that as competition intensifies, credit line becomes a less prominent liquidity source relative to cash holdings. The regression coefficients exhibit high statistical significance and are economically large. For example, a one-standard deviation increase in *HHIInv* raises the undrawn-credit-to-total-liquidity ratio by 1.8 percentage

[8]OLS regressions produce qualitatively similar results.

points, which is approximately a 4% increase relative to the sample mean (42%).

Columns (4) to (6) of Table 5 repeat the regressions in the first three columns but replace the continuous *HHIInv* with a *HighHHIInv* dummy that equals one if *HHIInv* is above sample median and zero otherwise. The results are consistent with the first three columns. As competition intensifies, firms are less likely to use credit lines, and the amount of undrawn credit significantly declines as a proportion of assets or total liquidity. Using a dummy variable also eases interpretation of the economic magnitude. If a firm transitions from a below-median competition environment to an above-median competition environment, the probability of having a line is lower by 5 percentage points, the undrawn-credit-to-assets ratio is lower by 2 percentage points, and the undrawn-credit-to-total-liquidity ratio is lower by 9 percentage points. These magnitudes seem economically significant in the context of the respective sample means of 76%, 10%, and 42%.

Overall, the results in Table 5 consistently show that credit line usage decreases as competition intensifies. As argued earlier, competition can increase the demand for credit lines but can also reduce the supply. The empirical evidence shows that the net effect of competition is negative, implying that restricted supply dominates increased demand.

Across different specifications, the coefficient estimates for the control variables are relatively stable in terms of signs and significance levels, and are generally consistent with extant literature. Similar to Sufi (2009) and Acharya, Almeida, and Campello (2013), we find that firms' credit line usage is positively associated with firm size (log*Assets*), age (logFirmAge), and profitability (*CashFlow*), but negatively associated with market-to-book (*M2B*). Asset tangibility also has a positive coefficient, consistent with asset tangibility increasing firms' borrowing capacity. Cash flow volatility and industry sales volatility mostly have negative coefficients, consistent with the explanation that volatile operating performance makes it more difficult for firms to obtain credit lines. Finally, firms not in the S&P 500 Index (*NonSP500*) have higher credit line usage – likely because these firms have less access to the capital market and therefore need to rely more on bank financing.

### 4.2.2 Fluidity and Credit Line Usage

To capture competition dynamics at the firm level, we employ *Fluidity* as the competition measure and re-run the regressions in Table 5. The results are in Table 6 and are very consistent with those in Table 5. As competition intensifies (*Fluidity* increases), firms are less likely to have a credit line, and the amount of undrawn credit decreases as a proportion of assets or total liquidity. Across all columns, *Fluidity* or *HighFluidity* is highly significant with *t*-stats above 8. The economic magnitudes are remarkably large. According to Column (1), a one-standard-deviation increase in *Fluidity* decreases the probability of having a credit line by 3.9 percentage points. The economic magnitudes are 4.4 for *CashFlow* and 1 for *Tangibility*. Similar calculation for the other five specifications in the table also confirms that *Fluidity* is an important determinant of credit line usage. The coefficient estimates in Columns (2) and (3) imply that a one-standard-deviation increase in *Fluidity* decreases the *Undrawn_Assets* ratio by 1.4 percentage points and the *Undrawn_TotLiq* ratio by 5.6 percentage points.

The findings on *Undrawn_TotLiq* in Columns (3) and (6) warrant further elaboration. The effect of *Fluidity* in both columns is highly significant with *t*-stat around 10. Based on Column (6), firms with above-median *Fluidity* on average have an undrawn-credit-to-total-liquidity ratio that is 5 percentage points higher than firms with below-median *Fluidity*. The difference is about 7.9% of the sample mean. While competition reduces the likelihood of having a credit line and the amount of undrawn line relative to assets, the *Undrawn_TotLiq* regressions tell us that competition reduces the relative importance of credit line vs. cash in corporate liquidity management. It is not obvious ex ante why this is the case. In extant theory, competition increases both cash and credit lines because both offer competitive advantages. Conversely, if competition makes it more difficult for a firm to obtain a line, it is not clear how competition makes it easier to accumulate more cash. We will explore this intriguing question in later sections.

## 4.3 Robustness Checks

We perform an extensive set of robustness checks on our baseline findings and summarize the results in Table 8. Specifically, we check whether our baseline findings are robust to alternative measures of competition, different sub-samples, and different sub-periods. These robustness checks include the same set of control variables and fixed effects as earlier. But for brevity, we tabulate only the coefficients on the competition variables.

First, we use two alternative competition measures in Panels A and B of Table 8. The first measure is *TNIC HHI*, a Herfindahl–Hirschman Index for industries defined using the text-based network industry classification (TNIC), developed by Hoberg and Phillips (2016). The SIC-based industry classification is static, and most firms stay in the same industry over time. TNIC is a dynamic classification, based on product descriptions in annual 10-K filings. The second measure is *PctComp*, a measure of competition based on management's discussion of competition in their 10-K filings, developed by Li, Lundholm, and Minnis (2013). They show that the measure is related to diminishing marginal returns in the future. The two alternative measures generate consistent results that competition reduces firms' credit line usage, and the effect is highly significant both statistically and economically.

Second, we replicate our baseline tests using data from Sufi (2009). The credit line data for our baseline tests are from the Capital IQ database, with a long sample period of 2002-2017. While Capital IQ contains comprehensive information about firms' capital structure, the literature reports that the credit line information in Capital IQ is not always accurate (e.g., Mathers and Giacomini, 2016). Sufi (2009) hand-collects a sample of credit line usage with 1,380 firm-year observations for 294 unique firms between 1997 and 2003. His sample mostly precedes our sample period and thus provides the opportunity of an out-of-sample test. On average, 72% of his sample firms have a credit line, and for these firms, the undrawn credit averages at 10% of total assets and 44% of total liquidity. These summary statistics are slightly higher than those of our sample. We use Sufi's sample to re-run our regressions of Table 6 and report the results in Panel C of Table 8. Despite the small sample size, we obtain highly significant results that are consistent with our main findings. It is worth noting that the coefficient estimates of *Fluidity* are much larger than those in Table 6, with all coefficients doubling in size.

Third, Ivashina and Scharfstein (2010) show that an extraordinarily high level of credit-line drawndowns occurred during the 2007-2009 financial crisis. To test whether our baseline findings are significantly affected by this crisis period, we exclude observations of 2007-2009 and re-run the regressions. The results in Panel D shows that the results are nearly identical to those in Table 6. In un-tabulated tests, we check the robustness of our baseline results for various subperiods between 2002 and 2017. It turns out that our baseline results are robust to any subperiod of 2002-2017 so long as the number of observations in the subperiod is sufficiently large.

Fourth, product market competition is arguably more relevant for manufacturing industries than other industries. We thus focus only on manufacturing industries and redo our baseline regressions in Panel E. Across all columns, competition remains highly significant with larger coefficient estimates.

Fifth, in Capital IQ, some observations that showing missing credit line information may actually have a line, but the information is not recorded in Capital IQ. To test whether our baseline results are robust to this potential database error, we keep only observations that have non-missing credit line information (i.e., HasLine = 1) and rerun the baseline tests in Panel F. Again, we find consistent results.

Finally, Mathers and Giacomini (2016) report that the credit line data in Capital IQ have better quality for firms that are (1) smaller, (2) with non-investment grade credit ratings, (3) more profitable, or (4) have lower market-to-book ratios. We run our baseline tests for each of these four subsets of firms and obtain consistent results. For brevity, we report in Panel G only the regression results using non-investment grade firms. Although the numbers of observation are much smaller than our baseline tests, the negative coefficients on *Fluidity* remain highly significant.

### 4.4 Instrumental Variable Regression

We have documented a robust and negative association between competition and credit line usage, but this association could be endogenous. The causal inference from competition to credit line usage hinges on whether the variation in our competition measures is exogenous. HHI is measured at industry level and is mostly exogenous to individual firms. Fluidity is influenced by changes in rivals' product description and changes in a firm's own product description, so it may be partly endogenous. Before delving into the solution for endogeneity, it is worth contemplating the potential cause of endogeneity and the expected direction of bias.

Endogeneity can be caused by reverse causality or omitted correlated variables. As for reverse causality, it is possible that greater liquidity provided by credit lines allows a firm to quickly capture investment opportunities and to differentiate their products from rivals' products, i.e., making the product market more "fluid." In this scenario, the baseline coefficient estimate on *Fluidity* would have been biased towards a positive sign (from the currently reported negative sign). Therefore, if we mitigate this reverse-causality-induced bias, the coefficient estimate on *Fluidity* will be more negative. Note, however, while this reverse causality may bias the estimate on whether a firm has a line or the line size relative to asset size, it is not clear how it will bias the estimate on the relative size between line and cash (i.e., the undrawn-to-total-liquidity regression). That is, it is not clear how the relative size between line and cash should reverse-cause product market fluidity. This note also applies to potential endogeneity caused by omitted correlated variables, which we discussed next.

The most plausible omitted variable is that a firm, anticipating competition pressure and the negative effect on performance, obtains more credit lines to shore up its liquidity and simultaneously differentiates its products from rivals, making the product market more "fluid." If the already-included control variables on performance cannot adequately account for this effect, the baseline coefficient estimate on *Fluidity* would have been biased towards a positive sign (from the currently reported negative sign). Therefore, if we mitigate this omitted variable bias, the coefficient estimate on *Fluidity* will be more negative. In sum, if we adequately mitigate potential endogeneity, whether it is caused by reserve causality or omitted variables, the coefficient estimate on *Fluidity* will be more negative.

As discussed in Section 2.3, we use import tariff rates as an instrument for *Fluidity*. We calculate tariff rates following Frésard and Valta (2016) and Schott (2010).[9] Our tariff rates data are at the 4-digit SIC level for manufacturing industries (SIC 2000 to 3999) from 2002 to 2017.

We report the results of two-stage least squares (2SLS) regressions in Table 9. Specifically, we redo all regressions of Table 6 using tariff rates as an instrument for *Fluidity*. Columns (4) and (8) report the first-stage regressions, and Columns (1)-(3) and (5)-(7) of report the second-stage results.

[9] Data on the MFN duty rates are from The United States International Trade Commission (USITC) database, while U.S. imports data are from Schott (2008). We merge the two data sets and calculate import-weighted tariff rates for every 8-digits harmonized (HS) industry.

As expected, the first-stage regressions show that tariff rates are significantly negatively related to *Fluidity* or *HighFluidity*. Diagnostic tests confirm that tariff rates are a strong instrument. Across all second-stage regressions, *Fluidity* and *HighFluidity* have a significantly negative coefficient, consistent with the baseline results.

Compared with the coefficient estimates on *Fluidity* in the baseline regressions, the estimates in 2SLS regressions are more negative, confirming the expected direction of potential endogeneity biases. It is reassuring that 2SLS mitigates the endogeneity bias and confirms the findings of the baseline regressions. But we readily acknowledge that without a truly exogenous experiment that randomly changes competition intensity, we cannot definitively infer causality running from competition to credit line usage.

### 4.5 Nature Experiment

To further ascertain causality, we use extra import tariff imposed by Donald Trump and his administrative as an exogenous shocks. During Trump's presidency, the government policy do not have a predictable pattern and consider as a surprise to the company. On January 23, 2018, Trump had imposed tariffs on solar panels produced outside the United States, and the Office of the U.S. Trade Representative announced tariffs on washing machines. On March 1, 2018, Trump announced his intention to impose a 25% tariff on steel and a 10% tariff on aluminum imports. The import tariff shock served as a positive demand shock for industries producing solar panels, washing machines, steel and aluminum. The increased demand for those industry, at least in the short run, reduces competition intensity. This reduced competition intensity should increase the use of credit lines. We test this prediction using a difference-in-differences (DiD) regression. For each year from 2002 to 2019, we match each firm producing solar panels, washing machines, steel and aluminum (the treatment group, SIC 3633, 3674, 1011, 3312-3317, 3321-3325, 3334, 3341, 3351-3357, 3363, 3365) with a firm in another service-oriented industry (the "control" group, SIC 3312-3399, 1011-1099, 3631-3639, and 3671-3679 excluding the SIC code in treatment group). We perform a one-to-one propensity-score matching using the control variables in our regression. We conduct the match without replacement, and the maximum difference in the propensity score allowed for a match is 1%. This results in a sample of 173 firm-years, in which 93 firm-years treated firms are matched with 80 firm-years with controlled firms.

Our matched sample satisfies three important validity criteria (see e.g., Fang, Tian, and Tice (2014)). First, in Panel A of Table 11, we run a Probit model using the original sample and the matched sample. The post-match regression in Column (2) shows that majority of the control variables are statistically insignificant after matching. Second, as shown in Panel B, there is no difference in the propensity score between observations in treatment group and those in control group. Third, as shown in the lower part of Panel C, the differences in firm controls are mostly insignificant.

We designate 2016-2017 as pre-event years and 2018-2019 as post-event years. The final sample has 87 firm-year pairs and 173 firm-year observations. For various credit-line measures, we compared the average of treated group and controlled group. Figure 2 plots the treatment-control differences by year. The figure show that on average firm in the control group are more likely to have credit line before Trump imposed extra import tariff. Control group maintained a stable credit line level before and after Donald Trump imposed extra import tariff. Treated firms (firm producing solar panels, washing machines, steel and aluminum) have similar credit line tendency before Donald Trump imposed extra import tariff and significantly increase its undrawn credit line level after Trump imposed extra import tariff. The figure show that our result fulfilled the parallel trend assumption.

Table 7 reports the Dif-in-Dif regression result. Treat is an indicator variable equal to one if firm produce solar panels, washing machines, steel and aluminum, and zero for control firms. Post is an indicator variable equal to one for post-2018, and zero for pre-2018. The coefficient estimates on Treat$\times$Post measure the effects of tariff shock imposed by Trump administrative. Consistent with the plots in Figure 2, the regression results show that post tariff shock, treated firms have more credit lines, and rely more on credit lines as a liquidity source. To ensure that the Dif-in-Dif results are not because of chance, we conduct placebo tests for years before 2018. That is, we randomly pick a year outside the Trump tariff shock window, and implement the same Dif-in-Dif test. These placebo tests do not show any statistically significant result.

Collectively, the empirical findings suggest a causal link between product market competition and corporate liquidity management. Confirming extant evidence in the literature that competition increases cash holdings, we present new evidence that competition reduces firms' use of credit lines as a liquidity source.

# 5 Cross-sectional Heterogeneity

We have documented empirically that competition reduces credit line usage, suggesting that overall, restricted supply dominates increased demand. The supply and demand effects are intertwined and difficult to disentangle.[10] But as challenging as it is, in this section we attempt to shed some light on the supply vs. demand effects. For example, one question we ask is that while extant theory predicts a positive demand effect, is this positive demand effect empirically significant? To explore these questions, we identify several firm characteristics that should ex ante affect the supply of or the demand for credit lines. We employ these firm characteristics as conditioning variables, that is, we interact them with the competition measure in the credit-line regressions. Coefficient estimates on the interaction terms will tell us the strength of the supply effect or the demand effect.

## 5.1 Asset Tangibility

Previous studies (e.g., Almeida and Campello, 2007; Sufi, 2009) show that higher asset tangibility raises firms' borrowing capacity because tangible assets can be more easily pledged as collateral. But there is no obvious reason that tangibility should affect a firm's demand for credit. As competition intensifies, firms with higher borrowing capacity should face less restricted supply of credit lines. Therefore, if the negative relation between competition and credit line usage is attenuated by asset tangibility, it is evidence that competition affects the supply channel of credit lines.

To implement this test, we add to the specifications in Table 6 an interaction term between *Fluidity* and *HighTang*, an indicator variable that equals one for firms with above-median asset tangibility. Regression results are in Table 13. The coefficient estimate of the interaction term *Flu-*

[10] Similarly, the literature on the transmission of monetary policies faces difficulty in separating the demand channel and the supply channel. See for example Kashyap and Stein (2000) for more details.

*idity*×*HighTang* captures the difference of the competition effect between firms with above- and below-median tangibility. In all specifications, this estimate is significantly positive, indicating that asset tangibility indeed attenuates the negative effect of competition on credit line usage. The effect is also economically large. For example, according to Column (3), the effect of *Fluidity* for firms with below-median tangibility is twice as large as that for firms with above-median tangibility. As it is unlikely that asset tangibility affects a firm's demand for credit lines, the conditional tests in Tables 13 show the importance of the supply channel.

## 5.2 Growth Opportunities

Survey results of CFOs suggest that firms hold credit lines mainly to fund investment opportunities. Theory suggests that facing intense competition, firms' demand for credit lines should be even stronger because they need to capture the investment opportunities before their rivals do. At the same time, there is no convincing reason that growth opportunities should significantly affect the supply of credit lines because banks are mostly concerned with the stability, not the growth, of cash flow. Therefore, if the negative relation between competition and credit line usage is attenuated by growth opportunities, it is evidence that competition affects the demand channel of credit lines. To implement this test, we interact competition with growth opportunities, measured by firms' R&D capital and market-to-book ratio.

The regression results are in Table 14. In Panel A, we add to our baseline regressions an interaction term between *Fluidity* and *HighRDC*, which equals one for firms with above-median R&D capital (*RDC*), and zero otherwise. Coefficient estimates on the interaction terms are significantly negative. That is, rather than attenuating the negative effect of competition, higher R&D actually exacerbates the negative effect. The conditional effects are also economically large. For example, the result in Column (3) suggests that the marginal effect of competition among high R&D firms is almost four times as large as the effect among low R&D firms. In Panel B, we replace *HighRDC* by *HighM2B*, an indicator variable for whether a firm has above–median market-to-book ratio in the previous year. Coefficient estimates on the interaction terms are small and not statistically significant. As a whole, results in this table is inconsistent with the prediction that competition should increase the demand for credit lines.

### 5.3 Cash Flow Level and Cash Flow Volatility

Sufi (2009) shows that the level and volatility of operating cash flow are important determinants of credit line usage. A higher level of cash flow *reduces a firm's demand* for external liquidity and hence the demand for credit lines. At the same time, a higher level of operating cash flow makes the firm a safer borrower and *increases banks' supply* of credit lines. If we interact competition with cash flow, the coefficient estimate on the interaction term will tell whether competition mostly affects the demand channel or the supply channel: a *positive* interaction term will suggest that competition mostly affects the supply channel; a *negative* interaction term the demand channel.

To implement this test, we add to the specifications in Table 6 an interaction term between *Fluidity* and *HighCF*, a dummy that equals one for firms with above-median *CashFlow* in the previous year, and zero otherwise. The results are in Panel A of Table 15. The coefficient estimate on *Fluidity*×*HighCF* is significantly positive in Column (1) and (3), and marginally positive in Column (2) (*t*-stat = 3.04). As discussed above, the positive interaction terms suggest that competition affects credit line usage mostly through the supply channel, consistent with the findings from asset tangibility.

Compared with cash flow level, cash flow volatility should have the opposite effect on credit line usage. A firm with higher cash flow volatility should have a greater demand for credit lines, but will be less likely to obtain a line. Therefore, if we introduce an interaction term with competition, the coefficient estimate will tell whether competition mostly affects the demand channel or the supply channel. Here, the interpretation of the interaction term is exactly the opposite of when interacting competition with cash flow level: a *negative* interaction term will suggest that competition mostly affects the supply channel; a *positive* interaction term the demand channel. The results are in Panel B of of Table 15. *HighCFV* in a dummy indicator that equals one for firms with above-median cash flow volatility in the previous year, and zero otherwise. Coefficient estimates on the interaction term all have a negative sign and are statistically significant in Columns (2) and (3), again suggesting that competition affects credit line usage mostly through the supply channel.

One concern with the above test is that competition and cash flow may be endogenously determined because the first-order effect of competition is to reduce profit margin and operating cash flow. Competition may also increase cash flow volatility. We note that *Fluidity* is a forward-looking measure while cash flow level and cash flow volatility are lagged by one year, which to some extent mitigates this concern.

# 6 Conclusion

The literature on product market competition and corporate liquidity has mostly focused on one source of liquidity, cash holdings, while largely overlooked another important source of liquidity, credit lines provided by banks. We extend this literature by studying how competition affects the usage of credit lines. Similar to the prediction about cash holdings, extant theory predicts that competition increases the use of credit lines, which will help firms with constrained internal liquidity to mitigate competition-induced under-investment (e.g., Maksimovic, 1990; Martin and Santomero, 1997; Lins, Servaes, and Tufano, 2010). Our empirical findings, however, show that competition decreases, rather than increases, the use of credit lines. This finding is robust to alternative measures of competition and exogenous variation in competition. Further cross-sectional analysis suggests that the reduction in credit line usage is driven by restricted supply by banks, rather than reduced demand by firms. Our results support the notion that competition induces negative pressure on firm performance, making it more difficult for firms to obtain credit lines. In this respect, we add to the studies that highlight the contingency nature of credit lines as a source of corporate liquidity (e.g., Sufi, 2009; Nikolov, Schmid, and Steri, 2019).

## References

Acharya, V. V., H. Almeida, M. Campello, 2013. Aggregate Risk and The Choice Between Cash and Lines of Credit. Journal of Finance 68(5), 2059–2116.

Almeida, H., M. Campello, 2007. Financial constraints, asset tangibility, and corporate investment. The Review of Financial Studies 20, 1429–1460.

Almeida, H., M. Campello, I. Cunha, M. S. Weisbach, 2014. Corporate liquidity management: A conceptual framework and survey. Annual Reviews of Finance and Economics 6, 135–162.

Almeida, H., M. Campello, M. S. Weisbach, 2004. The cash flow sensitivity of cash. The Journal of Finance 59(4), 1777–1804.

Arnoud, B., A. V. Thakor, G. F. Udell, 1987. Competition, Risk Neutrality and Loan Commitments. Journal of Banking and Finance 11, 449–471.

Avery, R. B., A. N. Berger, 1991. Risk-based capital and deposit insurance reform. Journal of Banking & Finance 15(4-5), 847–874.

Bates, T. W., C.-H. Chang, J. D. Chi, 2018. Why has the value of cash increased over time?. Journal of Financial and Quantitative Analysis 53(2), 749–787.

Bates, T. W., K. M. Kahle, R. M. Stulz, 2009. Why Do US Firms Hold So Much More Cash than They Used To?. Journal of Finance 64, 1985–2021.

Bharath, S. T., M. Hertzel, 2019. External governance and debt structure. The Review of Financial Studies 32(9), 3335–3365.

Bolton, P., D. S. Scharfstein, 1990. A Theory of Predation Based on Agency Problems in Financial Contracting. American Economic Review pp. 93–106.

Brander, J., T. Lewis, 1986. Oligopoly and Financial Structure: The Limited Liability Effect. American Economic Review 76, 956–970.

Campello, M., 2003. Capital structure and product market interactions: Evidence from business cycles. Journal of Financial Economics 68, 353–378.

Chan, L. K., J. Lakonishok, T. Sougiannis, 2001. The stock market valuation of research and development expenditures. The Journal of Finance 56, 2431–2456.

Chevalier, J. A., D. S. Scharfstein, 1996. Capital Market Imperfections and Countercyclical Markups: Theory and Evidence. American Economic Review 86, 703–725.

Chi, J. D., X. Su, 2016. Product market threats and the value of corporate cash holdings. Financial Management 45(3), 705–735.

Dasgupta, S., S. Titman, 1998. Pricing strategy and financial policy. Review of Financial Studies 11, 705–737.

Duchin, R., T. Gilbert, J. Harford, C. Hrdlicka, 2017. Precautionary savings with risky assets: When cash is not cash. The Journal of Finance 72(2), 793–852.

Fang, V. W., X. Tian, S. Tice, 2014. Does stock liquidity enhance or impede firm innovation?. The journal of Finance 69(5), 2085–2125.

Frésard, L., 2010. Financial Strength and Product Market Behavior: The Real Effects of Corporate Cash Holdings. Journal of Finance 65(3), 1097–1122.

Frésard, L., P. Valta, 2016. How does corporate investment respond to increased entry threat?. The Review of Corporate Finance Studies 5, 1–35.

Haushalter, D., S. Klasa, W. F. Maxwell, 2007. The influence of product market dynamics on a firm's cash holdings and hedging behavior. Journal of Financial Economics 84, 797–825.

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

Hoberg, G., G. M. Phillips, N. R. Prabhala, 2014. Product Market Threats, Payouts, and Financial Flexibility. Journal of Finance 69(1), 8.

Holmström, B., J. Tirole, 1998. Private and public supply of liquidity. Journal of political Economy 106, 1–40.

Ippolito, F., H. Almeida, A. P. Orive, V. Acharya, 2019. Bank lines of credit as contingent liquidity: Covenant violations and their implications. Journal of Financial Intermediation pp. 1008–17.

Ippolito, F., J.-L. Peydró, A. Polo, E. Sette, 2016. Double bank runs and liquidity risk management. Journal of Financial Economics 122(1), 135–154.

Ivashina, V., D. Scharfstein, 2010. Bank lending during the financial crisis of 2008. Journal of Financial economics 97, 319–338.

Kashyap, A. K., J. C. Stein, 2000. What do a million observations on banks say about the transmission of monetary policy?. American Economic Review 90, 407–428.

Khanna, N., S. Tice, 2005. Pricing, exit, and location decisions of firms: Evidence on the role of debt and operating efficiency. Journal of Financial Economics 75, 397–427.

Kim, C.-S., D. C. Mauer, A. E. Sherman, 1998. The determinants of corporate liquidity: Theory and evidence. Journal of financial and quantitative analysis 33(3), 335–359.

Li, F., R. Lundholm, M. Minnis, 2013. A measure of competition based on 10-K filings. Journal of Accounting Research 51, 399–436.

Li, S., X. Zhan, 2018. Product Market Threats and Stock Crash Risk. Management Science 65(9), 4011–4031.

Lins, K. V., H. Servaes, P. Tufano, 2010. What Drives Corporate Liquidity? An International Survey of Cash Holdings and Lines of Credit. Journal of Financial Economics 98(1), 160–176.

MacKay, P., G. M. Phillips, 2005. How Does Industry Affect Firm Financial Structure?. Review of Financial Studies 18, 1433–1466.

Maksimovic, V., 1990. Product Market Imperfections and Loan Commitments. The journal of finance 45(5), 1641–1653.

Martin, J. S., A. M. Santomero, 1997. Investment Opportunities and Corporate Demand for Lines of Credit. Journal of Banking & Finance 21(10), 1331–1350.

Mathers, A. M., E. Giacomini, 2016. A Note on Capital IQ's Credit Line Data. Financial Review 51, 435–461.

Nikolov, B., L. Schmid, R. Steri, 2019. Dynamic corporate liquidity. Journal of Financial Economics 132, 76–102.

Opler, T., L. Pinkowitz, R. Stulz, R. Williamson, 1999. The Determinants and Implications of Corporate Cash Holdings. Journal of Financial Economics 52, 3–46.

Phillips, G., G. Sertsios, 2013. How do firm financial conditions affect product quality and pricing?. Management Science 59, 1764–1782.

Phillips, G. M., 1995. Increased Debt and Industry Product Markets: An Empirical Analysis. Journal of Financial Economics 37, 189–238.

Schott, P., 2010. US manufacturing exports and imports by SIC or NAICS category and partner country, 1972 to 2005. Working paper.

Schott, P. K., 2008. The relative sophistication of Chinese exports. Economic policy 23, 6–49.

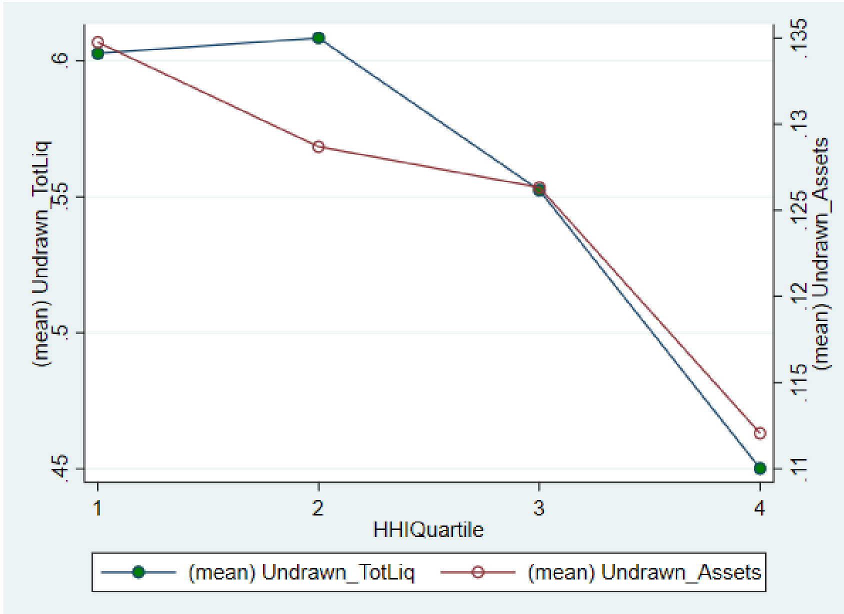
Shockley, R. L., A. V. Thakor, 1997. Bank Loan Commitment Contracts: Data, Theory, and Tests. Journal of Money, Credit, and Banking pp. 517–534.

Sufi, A., 2009. Bank lines of credit in corporate finance: An empirical analysis. The Review of Financial Studies 22(3), 1057–1088.

Valta, P., 2012. Competition and The Cost of Debt. Journal of Financial Economics 105(3), 661–682.

Xu, J., 2012. Profitability and Capital Structure: Evidence from Import Penetration. Journal of Financial Economics 106(2), 427–446.

Panel A: Inverse HHI as a Measure of Competition




Panel B: Fluidity as a Measure of Competition

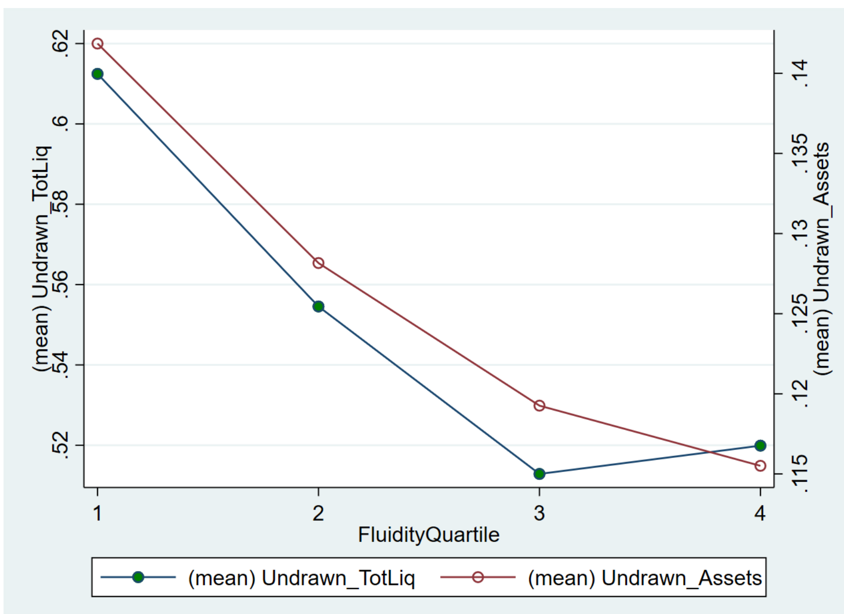
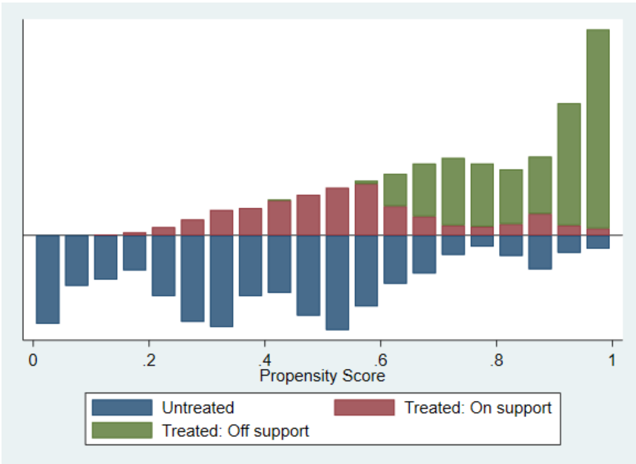


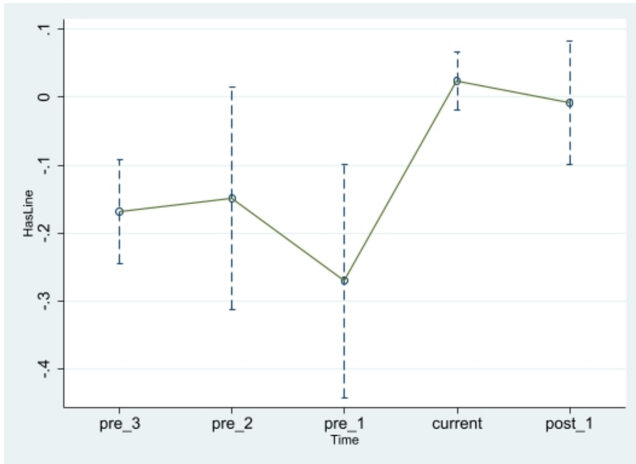

Figure 1: **Univariate Results - Undrawn Credit across Competition Quartiles**

This figure plots the ratio of undrawn credit lines to total liquidity (*Undrawn_TotLiq*, left scale) or total assets (*Undrawn_Assets*, right scale) across the four quartiles of competition, measured by Inverse HHI (*HHIInv*) in Panel A and *Fluidity* in Panel B. The figure shows that as competition increases, both *Undrawn_Assets* and *Undrawn_TotLiq* decrease. The sample includes only firms that have undrawn credit lines (i.e., *HasLine*=1). There are 30,024 firm-year observations for 4,277 publicly traded U.S. firms during 2002-2019.

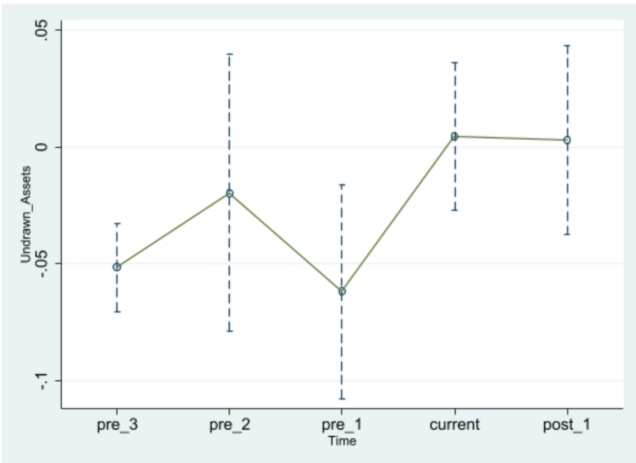
(a) Propensity Score




(b) Trump Tariff Policy and Having Credit Line




(c) Trump Tariff Policy and Undrawn_Asset




(d) Trump Tariff Policy and Undrawn_TotLiq

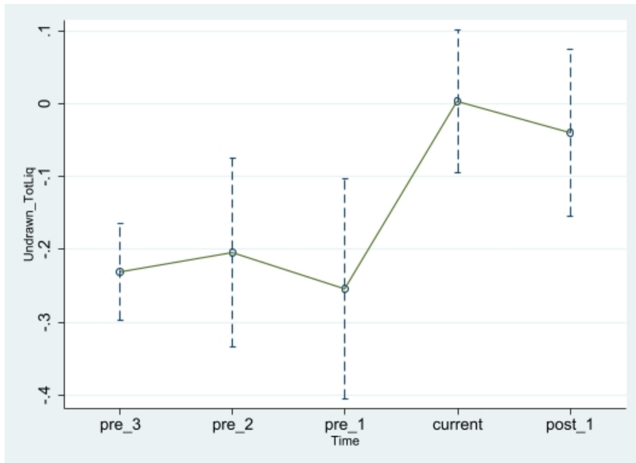



Figure 2: **DID Results Credit Line in Trump Tariff Policy**

This figure shows the treatment-control group differences of key variables (HasLine, Undrawn_Assets, and Undrawn_TotLiq) and the propensity score in our Dif-in-Dif sample from 2015 to 2019. The event year is 2018. Trump passed memorandum and imposing extra imposed tariffs on solar panels, washing machines, steel and aluminum on January 23 and March 1.

Table 1: **Variable Definitions**

The table defines the variables used in our analyses and lists their data sources: CIQ = S&P Capital IQ, CCM = Compustat-CRSP merged, HP = Hoberg-Phillips Data Library. Tariff rates are constructed following Frésard and Valta (2016) and Schott (2010). All dollar-denominated variables are inflation-adjusted to 2010 value.

| Variable | Definition | Source |
|---|---|---|
| ***A. Variables of Interest*** | | |
| HasLine | Indicator that the firm has undrawn credit lines | CIQ |
| Undrawn | Undrawn amount of credit lines in millions $US | CIQ |
| Undrawn_Assets | *Undrawn* scaled by assets, $[Undrawn/at]$ | CIQ, CCM |
| Drawn | Drawn amount of credit line in millions $US | CIQ |
| Drawn_Assets | *Drawn* scaled by assets, $[Drawn/at]$ | CIQ, CCM |
| TotLiq | Cash + Undrawn credit line, $[Undrawn/(che + Undrawn)]$ | CIQ, CCM |
| Undrawn_TotLiq | *Undrawn* scaled by total liquidity, $[Undrawn/\ (che + Undrawn)]$ | CIQ, CCM |
| TotLine_TotLiq | $(Drawn + Undrawn)/(Undrawn + Cash)$ | CIQ, CCM |
| Undrawn_Debt | *Undrawn* scaled by total debt $Undrawn/(dltt + dlc)$ | CIQ, CCM |
| Cash | Cash and short-term investments in millions $US, $che$. | CIQ |
| Cash_Assets | *Cash* scaled by assets, $[che/at]$ | CCM |
| Fluidity | Product market fluidity from Hoberg, Phillips, and Prabhala (2014) | HP |
| HighFluidity | Indicator that the firm has above median *Fluidity* in current year | HP |
| HHI | Sales-based Herfindahl Index for each 3-digit-SIC industry | CCM |
| HHIInv | Inverse HHI, or -1× *HHI* | CCM |
| HighHHIInv | Indicator that the firm has above median *HHIInv* in current year | CCM |
| TariffRate | Natural logarithm of import tariff rates for each 4-digit-SIC manufacturing industry | |
| ***B. Control Variables*** | | |
| Assets | The book value of non-cash assets in millions $US, $[at - che]$ | CCM |
| logAssets | Natural logarithm of *Assets* | CCM |
| FirmAge | Year difference between IPO and the current year | |
| logFirmAge | Natural logarithm of *FirmAge* | CCM |
| CashFlow | EBITDA scaled by non-cash assets, $[ebitda/(at - che)]$ | CCM |
| CFVol | The standard deviation of $oibdpq/(atq - cheq)$ in the past 4 quarters | CCM |
| M2B | (Total debt + market value of equity) scaled by non-cash assets, $[(dltt + dlc + prccq * csho)/(at - che)]$ | CCM |
| NetWorth | (Total non-cash assets – total liability) scaled by non-cash assets, $[(at - che - lt)/(at - che)]$ | CCM |
| RD | R&D expenses in millions $US, $xrd$ | CCM |
| RDC | Accumulated R&D expenses in the past four year using a 25% annual depreciation rate, scaled by non-cash asset $[(RD_{t-1} + 0.75RD_{t-2} + 0.5RD_{t-3} + 0.25RD_{t-4})/(at - che)]$ | CCM |
| Tangibility | Net PP&E scaled by non-cash assets, $[ppent/(at - che)]$ | CCM |
| IndSaleVol | The 3-digit-SIC industry median value of standard deviation of $saleq/(atq - cheq)$ in the past 4 quarters | CCM |
| NonSP500 | Indicator equal to 1 for firms not in the S&P500 Index | CCM |

(...*continued from the previous page*)

| Variable | Definition | Source |
|---|---|---|
| ***C. Other Variables*** | | |
| HighCF | Indicator that the firm has above median *CashFlow* | CCM |
| HighCFV | Indicator that the firm has above median *CFVol* | CCM |
| HighM2B | Indicator that the firm has above median *M2B* | CCM |
| HighRDC | Indicator that the firm has above median *RDC* | CCM |
| HighTang | Indicator that the firm has above median *Tangibility* | CCM |
| Rated | Indicator that the firm has a S&P long-term debt rating | CCM |
| Governance | Indicator that the firm has above median governance index | CCM |

Table 2: **Treatment and Control group of Trump Tariff Shock**

The table shows the 4 digit SIC code that affected by the Trump tax retaliation policy and its categories

| 4-digit SIC | Detailed Information | Treatment | Control1 | Control2 | Control3 |
|---|---|---|---|---|---|
| 3633 | Household Laundry Equipment | Yes | | | |
| 3674 | Semiconductors and Related Devices (solar cells) | Yes | | | |
| 3334 | Primary Production of Aluminum | Yes | | | |
| 3341 | Secondary Smelting and Refining of Nonferrous Metals | Yes | | | |
| 3363 | Aluminum Die-Castings | Yes | | | |
| 3365 | Aluminum Founries | Yes | | | |
| 1010-1019 | Iron Ores | Yes | | | |
| 3310-3319 | Steel Works, Blast Furnaces, and Rolling and Finishing Mills | Yes | | | |
| 3320-3329 | Iron and Steel Foundries | Yes | | | |
| 3350-3359 | Rolling, Drawing, and Extruding of Nonferrous | Yes | | | |
| 3631-3632 & 3634-3639 | Household Appliances | | Yes | Yes | Yes |
| 3670-3673 & 3675-3679 | Electronic Components And Accessories | | Yes | Yes | Yes |
| 3320-3399 | Primary Metal Industries | | Yes | Yes | Yes |
| 1020-1099 | Metal Mining | | Yes | Yes | Yes |
| 3600-3630 & 3640-3699 | Electronic And Other Electrical Equipment And Components, Except Computer Equipment | | | Yes | Yes |
| 2000-3299 & 3400-3599 & 3700-3999 | Manufacturing | | | | Yes |

Table 3: **Summary Statistics for Key Dependent Variables by Year**

This table reports annual means of the corporate liquidity variables. Panel A is for the full sample of 49,566 firm-year observations of 6,516 publicly traded firms. Panel B is for the subsample of firms with undrawn credit lines (HasLine = 1), including 31,583 firm-year observations of 4,260 firms. All variable definitions are in Table 2.

| Year | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Panel A: The Full Sample (Obs: 47,765)*** | | | | | | | | | | | | | | | | | | | |
| HasLine | 0.70 | 0.71 | 0.72 | 0.74 | 0.75 | 0.76 | 0.78 | 0.77 | 0.78 | 0.80 | 0.81 | 0.81 | 0.80 | 0.78 | 0.79 | 0.79 | 0.76 | 0.69 | 0.76 |
| Drawn_Assets | 0.05 | 0.04 | 0.04 | 0.04 | 0.04 | 0.05 | 0.06 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.05 | 0.05 | 0.05 | 0.04 | 0.04 | 0.04 |
| Undrawn_Assets | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.10 | 0.10 | 0.10 | 0.10 | 0.11 | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 | 0.09 | 0.10 |
| Cash_Assets | 0.18 | 0.19 | 0.19 | 0.19 | 0.18 | 0.18 | 0.17 | 0.19 | 0.19 | 0.18 | 0.17 | 0.19 | 0.19 | 0.19 | 0.19 | 0.19 | 0.20 | 0.23 | 0.19 |
| Undrawn_TotLiq | 0.41 | 0.40 | 0.40 | 0.41 | 0.43 | 0.44 | 0.44 | 0.40 | 0.40 | 0.42 | 0.44 | 0.43 | 0.44 | 0.43 | 0.43 | 0.44 | 0.44 | 0.40 | 0.42 |
| UndrawnDebt | 0.85 | 0.96 | 0.99 | 1.03 | 0.94 | 0.89 | 0.81 | 0.87 | 0.92 | 0.76 | 0.68 | 0.67 | 0.60 | 0.51 | 0.47 | 0.44 | 0.43 | 0.31 | 0.74 |
| Observations | 2950 | 3109 | 2954 | 2858 | 2799 | 2744 | 2713 | 2682 | 2619 | 2524 | 2445 | 2459 | 2508 | 2544 | 2453 | 2409 | 2398 | 2597 | |
| ***Panel B: The Sub-sample with HasLine=1 (Obs: 36,350)*** | | | | | | | | | | | | | | | | | | | |
| HasLine | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Drawn_Assets | 0.07 | 0.05 | 0.05 | 0.05 | 0.05 | 0.06 | 0.07 | 0.05 | 0.04 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |
| Undrawn_Assets | 0.13 | 0.12 | 0.12 | 0.13 | 0.13 | 0.12 | 0.12 | 0.12 | 0.13 | 0.12 | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 | 0.12 | 0.13 |
| Cash_Assets | 0.12 | 0.13 | 0.13 | 0.13 | 0.12 | 0.12 | 0.12 | 0.14 | 0.14 | 0.14 | 0.13 | 0.14 | 0.14 | 0.13 | 0.13 | 0.13 | 0.12 | 0.11 | 0.13 |
| Undrawn_TotLiq | 0.59 | 0.57 | 0.56 | 0.56 | 0.58 | 0.58 | 0.56 | 0.52 | 0.52 | 0.53 | 0.54 | 0.54 | 0.54 | 0.55 | 0.55 | 0.56 | 0.57 | 0.58 | 0.55 |
| UndrawnDebt | 1.19 | 1.32 | 1.35 | 1.37 | 1.24 | 1.16 | 1.03 | 1.12 | 1.18 | 0.95 | 0.84 | 0.83 | 0.74 | 0.64 | 0.60 | 0.55 | 0.55 | 0.45 | 0.97 |
| Observations | 2073 | 2216 | 2124 | 2117 | 2098 | 2078 | 2104 | 2064 | 2038 | 2011 | 1985 | 1989 | 2014 | 1996 | 1929 | 1893 | 1829 | 1792 | |

Table 4: **Summary Statistics for All Variables**

This table reports descriptive statistics for the variables. Panel A is for the full sample. Panel B is for the subsample of firms that have undrawn credit lines (i.e., *HasLine* = 1). All variables are winsorized at the 1% and 99%. All variable definitions are in Table 2.

| **Variable** | **N** | **mean** | **sd** | **p1** | **p25** | **p50** | **p75** | **p99** |
|---|---|---|---|---|---|---|---|---|
| ***Panel A: The Full Sample*** | | | | | | | | |
| HasLine | 47765 | 0.76 | 0.43 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Undrawn | 47765 | 247.7 | 521.9 | 0.00 | 0.38 | 31.56 | 213.1 | 2651 |
| Drawn | 47765 | 47.81 | 123.2 | 0.00 | 0.00 | 0.00 | 23.50 | 666 |
| TotLine | 47765 | 317.4 | 610.5 | 0.00 | 2.66 | 50.90 | 310.8 | 3002 |
| Cash | 47765 | 348.9 | 1020 | 0.05 | 9.30 | 48.84 | 206.3 | 7039 |
| Undrawn_Assets | 47765 | 0.10 | 0.10 | 0.00 | 0.00 | 0.07 | 0.14 | 0.45 |
| Drawn_Assets | 47765 | 0.04 | 0.09 | 0.00 | 0.00 | 0.00 | 0.05 | 0.43 |
| Cash_Assets | 47765 | 0.19 | 0.22 | 0.00 | 0.03 | 0.11 | 0.26 | 0.93 |
| TotLine_Assets | 47765 | 0.14 | 0.15 | 0.00 | 0.02 | 0.10 | 0.21 | 0.67 |
| Undrawn_TotLiq | 47765 | 0.42 | 0.35 | 0.00 | 0.01 | 0.41 | 0.74 | 0.99 |
| TotLine_TotLiq | 47765 | 0.47 | 0.36 | 0.00 | 0.09 | 0.49 | 0.82 | 1.00 |
| UndrawnDebt | 44101 | 0.74 | 1.20 | 0.00 | 0.02 | 0.28 | 0.79 | 5.59 |
| Fluidity | 47765 | 0.66 | 0.35 | 0.14 | 0.40 | 0.58 | 0.84 | 1.75 |
| HHI | 47765 | 0.17 | 0.15 | 0.03 | 0.06 | 0.12 | 0.21 | 0.84 |
| HHIInv | 47765 | -0.17 | 0.15 | -0.84 | -0.21 | -0.12 | -0.06 | -0.03 |
| TNICHHI | 42724 | 0.31 | 0.28 | 0.03 | 0.10 | 0.20 | 0.44 | 1.00 |
| InvTNICHHI | 42724 | -0.31 | 0.28 | -1.00 | -0.44 | -0.20 | -0.10 | -0.03 |
| pctcomp | 4802 | 0.36 | 0.30 | 0.03 | 0.16 | 0.27 | 0.46 | 1.53 |
| Assets | 47765 | 3069 | 9185 | 1.83 | 65.94 | 384.2 | 1736 | 55028 |
| FirmAge | 47765 | 19.93 | 15.79 | 2.00 | 8.00 | 15.00 | 27.00 | 62.00 |
| CashFlow | 47765 | -0.01 | 0.49 | -1.86 | 0.03 | 0.12 | 0.18 | 0.62 |
| CFVol | 47765 | 0.04 | 0.13 | 0.00 | 0.00 | 0.01 | 0.02 | 0.86 |
| M2B | 47765 | 2.74 | 3.26 | 0.34 | 0.96 | 1.48 | 2.77 | 13.89 |
| NetWorth | 47765 | 0.29 | 0.45 | -1.31 | 0.18 | 0.39 | 0.56 | 0.87 |
| Tangibility | 47765 | 0.65 | 0.50 | 0.02 | 0.27 | 0.54 | 0.93 | 2.65 |
| RDC | 43627 | 0.16 | 0.34 | 0.00 | 0.00 | 0.00 | 0.12 | 1.30 |
| NonSP500 | 47765 | 0.87 | 0.33 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| IndSaleVol | 47765 | 0.02 | 0.02 | 0.00 | 0.02 | 0.02 | 0.03 | 0.09 |
| logAssets | 47765 | 5.84 | 2.31 | 0.61 | 4.19 | 5.95 | 7.46 | 10.92 |

(*continuing on the next page...*)

(...*continued from the previous page*)

| Variable | N | mean | sd | p1 | p25 | p50 | p75 | p99 |
|---|---|---|---|---|---|---|---|---|
| ***Panel B: The Sub-sample with HasLine=1*** | | | | | | | | |
| HasLine | 36350 | 1.00 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Undrawn | 36350 | 325.4 | 576.7 | 0.24 | 16.21 | 83.56 | 327.0 | 2651 |
| Drawn | 36350 | 60.12 | 134.9 | 0.00 | 0.00 | 1.23 | 45.47 | 666.9 |
| TotLine | 36350 | 413.1 | 668.7 | 0.75 | 25.28 | 129.0 | 474.0 | 3002 |
| Cash | 36350 | 381.1 | 1063 | 0.05 | 10.40 | 54.68 | 233.1 | 7315 |
| Undrawn_Assets | 36350 | 0.13 | 0.10 | 0.00 | 0.06 | 0.10 | 0.17 | 0.45 |
| Drawn_Assets | 36350 | 0.05 | 0.09 | 0.00 | 0.00 | 0.00 | 0.07 | 0.43 |
| Cash_Assets | 36350 | 0.13 | 0.14 | 0.00 | 0.03 | 0.08 | 0.18 | 0.66 |
| TotLine_Assets | 36350 | 0.18 | 0.14 | 0.00 | 0.08 | 0.14 | 0.24 | 0.67 |
| Undrawn_TotLiq | 36350 | 0.55 | 0.29 | 0.01 | 0.31 | 0.57 | 0.82 | 0.99 |
| TotLine_TotLiq | 36350 | 0.60 | 0.29 | 0.02 | 0.37 | 0.64 | 0.88 | 1.00 |
| UndrawnDebt | 33868 | 0.97 | 1.29 | 0.01 | 0.19 | 0.45 | 1.10 | 5.60 |
| Fluidity | 36350 | 0.60 | 0.32 | 0.14 | 0.37 | 0.54 | 0.76 | 1.61 |
| HHI | 36350 | 0.18 | 0.16 | 0.03 | 0.08 | 0.14 | 0.23 | 0.84 |
| HHIInv | 36350 | -0.18 | 0.16 | -0.84 | -0.23 | -0.14 | -0.08 | -0.03 |
| TNICHHI | 32704 | 0.33 | 0.29 | 0.03 | 0.11 | 0.22 | 0.46 | 1.00 |
| InvTNICHHI | 32704 | -0.33 | 0.29 | -1.00 | -0.46 | -0.22 | -0.11 | -0.03 |
| pctcomp | 3972 | 0.34 | 0.29 | 0.03 | 0.15 | 0.25 | 0.42 | 1.45 |
| Assets | 36350 | 3682 | 9858 | 6.75 | 150.8 | 641.8 | 2364 | 61435 |
| logAssets | 36350 | 6.41 | 2.03 | 1.91 | 5.02 | 6.46 | 7.77 | 11.03 |
| FirmAge | 36350 | 22.08 | 16.42 | 2.00 | 9.00 | 17.00 | 32.00 | 62.00 |
| CashFlow | 36350 | 0.11 | 0.23 | -0.94 | 0.08 | 0.13 | 0.19 | 0.58 |
| CFVol | 36350 | 0.02 | 0.05 | 0.00 | 0.00 | 0.01 | 0.02 | 0.14 |
| IndSaleVol | 36350 | 0.03 | 0.02 | 0.00 | 0.02 | 0.02 | 0.03 | 0.09 |
| M2B | 36350 | 1.98 | 2.08 | 0.36 | 0.91 | 1.33 | 2.14 | 13.89 |
| NetWorth | 36350 | 0.36 | 0.32 | -1.05 | 0.23 | 0.40 | 0.55 | 0.86 |
| Tangibility | 36350 | 0.64 | 0.47 | 0.03 | 0.27 | 0.54 | 0.91 | 2.19 |
| RDC | 33686 | 0.08 | 0.20 | 0.00 | 0.00 | 0.00 | 0.05 | 1.30 |
| NonSP500 | 36350 | 0.85 | 0.36 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |

Table 5: **Baseline Results: HHI and Credit Lines**

This table reports the marginal effects from regressions of credit line usage on competition. Competition is measured by the sales-based Herfindahl Index (*HHI*) at the 3-digit SIC level. We run a Probit regression in Columns (1) and (4), where the dependent variable is *HasLine*, a dummy that equals one if the firm has undrawn lines of credit, and zero otherwise. The rest columns report the results of Tobit regressions, where the dependent variable in Columns (2) and (5) is the ratio of undrawn credit to assets (*Undrawn_Assets*), and in Columns (3) and (6) is the ratio of undrawn credit to firm total liquidity (*Undrawn_TotLiq*), defined as the sum of undrawn credit and cash. *HHIInv* is the inverse HHI and equals to -1×HHI. *HighHHIInv* is a dummy equal to one if *HHIInv* is above the sample median. Both higher *HHIInv* and *HighHHIInv* indicate more intense competition. All regressions control for year and industry fixed effects. In parentheses are *t*-stats based on robust standard errors adjusted for heteroskedasticity and firm-level clustering. ***,**, and * denote statistical significance at the 1%, 5%, and 10% levels. All variable definitions are in Table 2. Variables of interest are in boldface.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | HasLine | Undrawn_Assets | Undrawn_TotLiq | HasLine | Undrawn_Assets | Undrawn_TotLiq |
| **HHIInv** | **-0.13*** ** | **-0.03*** ** | **-0.12*** ** | | | |
| | **(-4.73)** | **(-3.65)** | **(-4.23)** | | | |
| **HighHHIInv** | | | | **-0.05*** ** | **-0.02*** ** | **-0.09*** ** |
| | | | | **(-4.73)** | **(-6.69)** | **(-9.41)** |
| CashFlow | 0.11*** | 0.09*** | 0.21*** | 0.11*** | 0.09*** | 0.20*** |
| | (13.13) | (18.24) | (15.73) | (12.65) | (18.04) | (15.31) |
| Tangibility | 0.02** | 0.01*** | 0.01 | 0.02** | 0.01*** | 0.01 |
| | (2.28) | (2.82) | (1.23) | (2.31) | (2.91) | (1.39) |
| logAssets | 0.04*** | 0.01*** | 0.05*** | 0.04*** | 0.01*** | 0.05*** |
| | (20.61) | (6.03) | (16.58) | (19.44) | (6.08) | (16.76) |
| NetWorth | 0.02** | 0.01*** | 0.05*** | 0.02** | 0.01*** | 0.05*** |
| | (2.01) | (3.38) | (4.24) | (1.97) | (3.41) | (4.29) |
| M2B | -0.01*** | -0.00*** | -0.04*** | -0.01*** | -0.00*** | -0.04*** |
| | (-9.45) | (-8.22) | (-21.55) | (-9.15) | (-7.77) | (-20.95) |
| IndSaleVol | -0.34 | 0.23** | -1.06*** | -0.28 | 0.24** | -1.05*** |
| | (-1.18) | (2.26) | (-3.17) | (-0.98) | (2.34) | (-3.15) |
| CFVol | -0.01 | -0.01 | -0.06 | -0.01 | -0.01 | -0.05 |
| | (-0.41) | (-0.52) | (-1.27) | (-0.26) | (-0.41) | (-1.11) |
| logFirmAge | 0.02*** | 0.01*** | 0.02*** | 0.03*** | 0.01*** | 0.02*** |
| | (7.66) | (5.25) | (4.85) | (7.52) | (5.20) | (4.76) |
| NonSP500 | 0.06*** | 0.03*** | 0.15*** | 0.07*** | 0.03*** | 0.14*** |
| | (4.93) | (8.26) | (10.54) | (4.90) | (8.25) | (10.59) |
| Year FEs | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FEs | Yes | Yes | Yes | Yes | Yes | Yes |
| *N* | 47690 | 47765 | 47765 | 47690 | 47765 | 47765 |

Table 6: **Baseline Results: Fluidity and Credit Lines**

This table reports the marginal effects from regressions of credit line usage on competition. Competition is measured by *Fluidity*. We run a Probit regression in Columns (1) and (4), where the dependent variable is *HasLine*, a dummy that equals one if the firm has undrawn lines of credit, and zero otherwise. The rest columns report the results of Tobit regressions, where the dependent variable in Columns (2) and (5) is the ratio of undrawn credit to assets (*Undrawn_Assets*), and in Columns (3) and (6) is the ratio of undrawn credit to firm total liquidity (*Undrawn_TotLiq*), defined as the sum of undrawn credit and cash. *HighFluidity* is a dummy equal to one if *Fluidity* is above the sample median. All regressions control for year and industry fixed effects. In parentheses are *t*-stats based on robust standard errors adjusted for heteroskedasticity and firm-level clustering. ***,**, and * denote statistical significance at the 1%, 5%, and 10% levels. All variable definitions are in Table 2. Variables of interest are in boldface.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | HasLine | Undrawn_Assets | Undrawn_TotLiq | HasLine | Undrawn_Assets | Undrawn_TotLiq |
| **Fluidity** | **-0.11*** ** | **-0.04*** ** | **-0.16*** ** | | | |
| | **(-10.68)** | **(-10.06)** | **(-11.17)** | | | |
| **High Fluidity** | | | | **-0.05*** ** | **-0.02*** ** | **-0.08*** ** |
| | | | | **(-8.93)** | **(-8.84)** | **(-11.00)** |
| CashFlow | 0.09*** | 0.08*** | 0.18*** | 0.10*** | 0.09*** | 0.20*** |
| | (11.07) | (16.96) | (13.88) | (12.39) | (17.75) | (14.93) |
| Tangibility | 0.02** | 0.01*** | 0.01 | 0.02** | 0.01*** | 0.01 |
| | (2.29) | (2.93) | (1.36) | (2.13) | (2.76) | (1.16) |
| logAssets | 0.05*** | 0.01*** | 0.05*** | 0.05*** | 0.01*** | 0.05*** |
| | (22.45) | (7.38) | (18.27) | (21.61) | (6.74) | (17.60) |
| NetWorth | 0.01 | 0.01*** | 0.04*** | 0.01* | 0.01*** | 0.04*** |
| | (1.61) | (3.14) | (3.98) | (1.95) | (3.31) | (4.17) |
| M2B | -0.01*** | -0.00*** | -0.03*** | -0.01*** | -0.00*** | -0.03*** |
| | (-7.92) | (-6.89) | (-19.90) | (-8.68) | (-7.46) | (-20.57) |
| IndSaleVol | -0.23 | 0.23** | -1.07*** | -0.18 | 0.24** | -1.02*** |
| | (-0.82) | (2.22) | (-3.24) | (-0.64) | (2.39) | (-3.07) |
| CFVol | 0.00 | -0.00 | -0.04 | -0.01 | -0.01 | -0.06 |
| | (0.09) | (-0.14) | (-0.82) | (-0.44) | (-0.56) | (-1.33) |
| logFirmAge | 0.02*** | 0.01*** | 0.01*** | 0.02*** | 0.01*** | 0.02*** |
| | ( 5.52) | (3.33) | (2.67) | (6.38) | (4.13) | (3.45) |
| NonSP500 | 0.08*** | 0.04*** | 0.15*** | 0.07*** | 0.03*** | 0.15*** |
| | (5.48) | (8.77) | (11.16) | (5.19) | (8.53) | (10.91) |
| Year FEs | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FEs | Yes | Yes | Yes | Yes | Yes | Yes |
| *N* | 47690 | 47765 | 47765 | 47690 | 47765 | 47765 |

Table 7: **Baseline Results: Total Line**

This table reports the marginal effects from regressions of credit line usage on competition. Competition is measured by *Fluidity* and inverse *HHI*. We run a Tobit regression, where the dependent variable in odd columns is the ratio of total credit line (Undrawn+ Drawn) to assets (*TotLine_Assets*), and in even columns is the ratio of total credit line (Undrawn+ Drawn) to firm total Liquidity (*Undrawn+ Drawn+Cash*). *High-Fluidity* and *LowHHI* is a dummy equal to one if *Fluidity* is above the sample median or *HHI* is below the sample median. All regressions control for year and industry fixed effects. In parentheses are *t*-stats based on robust standard errors adjusted for heteroskedasticity and firm-level clustering. ***,**, and * denote statistical significance at the 1%, 5%, and 10% levels. All variable definitions are in Table 2. Variables of interest are in boldface.

| | (1) Assets | (2) TotLiq | (3) Assets | (4) TotLiq | (5) Assets | (6) TotLiq | (7) Assets | (8) TotLiq |
|---|---|---|---|---|---|---|---|---|
| **HHIInv** | **-0.04*** ** | **-0.12*** ** | | | | | | |
| | **(-3.08)** | **(-4.14)** | | | | | | |
| **Low_HHI** | | | **-0.03*** ** | **-0.09*** ** | | | | |
| | | | **(-6.42)** | **(-9.25)** | | | | |
| **Fluidity** | | | | | **-0.05*** ** | **-0.17*** ** | | |
| | | | | | **(-7.88)** | **(-11.47)** | | |
| **High Fluidity** | | | | | | | **-0.03*** ** | **-0.09*** ** |
| | | | | | | | **(-8.06)** | **(-11.25)** |
| CashFlow | 0.11*** | 0.19*** | 0.11*** | 0.19*** | 0.10*** | 0.17*** | 0.11*** | 0.18*** |
| | (16.83) | (14.75) | (16.60) | (14.33) | (15.65) | (12.75) | (16.30) | (13.87) |
| Tangibility | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 |
| | (1.07) | (0.07) | (1.16) | (0.22) | (1.17) | (0.22) | (1.03) | (0.00) |
| LogAssets | 0.00 | 0.04*** | 0.00 | 0.04*** | 0.00 | 0.04*** | 0.00 | 0.04*** |
| | (0.28) | (13.27) | (0.32) | (13.43) | (1.40) | (15.01) | (0.93) | (14.31) |
| NetWorth | -0.01** | 0.01 | -0.01** | 0.01 | -0.01*** | 0.01 | -0.01*** | 0.01 |
| | (-2.56) | (1.13) | (-2.57) | (1.15) | (-2.80) | (0.83) | (-2.66) | (1.03) |
| M2B | -0.01*** | -0.04*** | -0.01*** | -0.04*** | -0.01*** | -0.04*** | -0.01*** | -0.04*** |
| | (-13.41) | (-24.66) | (-13.01) | (-24.11) | (-12.26) | (-23.02) | (-12.70) | (-23.70) |
| IndSaleVol | -0.01 | -1.30*** | -0.00 | -1.29*** | -0.01 | -1.32*** | 0.00 | -1.27*** |
| | (-0.04) | (-3.96) | (-0.01) | (-3.95) | (-0.07) | (-4.06) | (0.02) | (-3.89) |
| CFVol | -0.04* | -0.05 | -0.03 | -0.05 | -0.03 | -0.03 | -0.04* | -0.06 |
| | (-1.70) | (-1.11) | (-1.61) | (-0.96) | (-1.46) | (-0.70) | (-1.77) | (-1.19) |
| LogFirmAge | 0.01** | 0.02*** | 0.00** | 0.02*** | 0.00 | 0.01 | 0.00 | 0.01** |
| | (2.30) | (3.85) | (2.22) | (3.74) | (0.75) | (1.53) | (1.24) | (2.33) |
| NonSP500 | 0.06*** | 0.15*** | 0.05*** | 0.15*** | 0.06*** | 0.15*** | 0.06*** | 0.15*** |
| | (9.69) | (10.72) | (9.67) | (10.78) | (10.09) | (11.36) | (9.93) | (11.10) |
| Year FEs | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FEs | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| *N* | 47765 | 47765 | 47765 | 47765 | 47765 | 47765 | 47765 | 47765 |

Table 8: **Baseline Results: Robustness Check**

This table reports the marginal effects from regressions replicating those in Tables 5 and 6 while employing alternative competition measures, sub-samples, and sub-periods. In Panels A and B, we measure competition by the TNIC HHI and PctComp. Panel C uses the data set of Sufi (2009). Panel D excludes the 2007-2009 financial crisis period. Panel E includes only manufacturing firms. Panel F keeps only those firms that have undrawn credit lines (i.e., those with *HasLine* = 1). All regressions include the set of control variables as in Tables 5 and 6, as well as year and industry fixed effects. In parentheses are *t*-stats based on robust standard errors adjusted for heteroskedasticity and firm-level clustering. ***,**, and * denote statistical significance at the 1%, 5%, and 10% levels. All variable definitions are in Table 2.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | HasLine | Undrawn_Assets | Undrrawn_TotLiq | HasLine | Undrawn_Assets | Undr._TotLiq |
| *Panel A: Using TNIC HHI as an Alternative Measure of Competition* | | | | | | |
| InvTNICHHI | -0.05*** | -0.02*** | -0.11*** | | | |
| | (-4.44) | (-4.30) | (-7.20) | | | |
| LowTNICHHI | | | | -0.04*** | -0.01*** | -0.06*** |
| | | | | (-6.03) | (-5.53) | (-8.11) |
| *N* | 42657 | 42724 | 42724 | 42657 | 42724 | 42724 |
| *Panel B: Using PctComp as an Alternative Measure of Competition* | | | | | | |
| PctComp | -0.06*** | -0.02** | -0.08*** | | | |
| | (-2.64) | (-2.43) | (-2.63) | | | |
| HighPctComp | | | | -0.04*** | -0.01 | -0.05*** |
| | | | | (-2.87) | (-1.61) | (-3.63) |
| *N* | 4715 | 4802 | 4802 | 4715 | 4802 | 4802 |
| *Panel C: Using Sufi (2009)'s Data* | | | | | | |
| Fluidity | -0.50 | -0.11** | -0.31*** | | | |
| | (-1.14) | (-2.53) | (-2.65) | | | |
| High Fluidity | | | | -0.24 | -0.03 | -0.07 |
| | | | | (-0.87) | (-0.76) | (-0.90) |
| *N* | 225 | 225 | 224 | 225 | 225 | 224 |
| *Panel D: Excluding the 2007-2009 Financial Crisis Period* | | | | | | |
| Fluidity | -0.11*** | -0.04*** | -0.16*** | | | |
| | (-10.56) | (-10.11) | (-11.08) | | | |
| High Fluidity | | | | -0.05*** | -0.02*** | -0.08*** |
| | | | | (-8.55) | (-8.65) | (-10.84) |
| *N* | 39562 | 39626 | 39626 | 39562 | 39626 | 39626 |

(*continuing on the next page...*)

(*...continued from the previous page*)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | HasLine | Undrawn_Assets | Undrawn_TotLiq | HasLine | Undrawn_Assets | Undrawn_TotLiq |
| *Panel E: Manufacturing Firms* | | | | | | |
| Fluidity | -0.18*** | -0.08*** | -0.31*** | | | |
| | (-12.75) | (-12.61) | (-15.52) | | | |
| High Fluidity | | | | -0.08*** | -0.04*** | -0.15*** |
| | | | | (-9.75) | (-10.34) | (-12.80) |
| *N* | 23964 | 24019 | 24019 | 23964 | 24019 | 24019 |
| *Panel F: Keeping Firms that Have Credit Lines (HasLine=1)* | | | | | | |
| Fluidity | | -0.02*** | -0.09*** | | | |
| | | (-5.47) | (-7.41) | | | |
| High Fluidity | | | | | -0.01*** | -0.05*** |
| | | | | | (-5.11) | (-8.17) |
| *N* | | 36350 | 36350 | | 36350 | 36350 |
| *Panel G: Non-Investment Grade Firms* | | | | | | |
| Fluidity | -0.65*** | -0.02*** | -0.13*** | | | |
| | (-5.41) | (-2.84) | (-5.14) | | | |
| High Fluidity | | | | -0.33*** | -0.02*** | -0.07*** |
| | | | | (-4.73) | (-4.03) | (-5.35) |
| *N* | 8204 | 8562 | 8562 | 8204 | 8562 | 8562 |

Table 9: **Addressing Endogeneity: Tariff Rates as An Instrumental Variable (MLE)**

This table reports the 2SLS result of using tariff rates as an instrument for product market fluidity. *TariffRate* is import tariff rates for manufacturing firms (SIC 2000-3999) at the 4-digit-SIC level during 2002-2017. All regressions control for year and industry fixed effects. In parentheses are *t*-stats based on robust standard errors adjusted for heteroskedasticity and firm-level clustering. ***,**, and * denote statistical significance at the 1%, 5%, and 10% levels. All variable definitions are in Table 2. Variables of interest are in boldface.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | 2nd Stage | | | 1st Stage | 2nd Stage | | | 1st Stage |
| | HasLine | Undrawn_Assets | Undrawn_TotLiq | Fluidity | HasLine | Undrawn_Assets | Undrawn_TotLiq | HighFluidity |
| Fluidity | -1.93*** | -0.15*** | -0.54*** | | | | | |
| | (-3.79) | (-2.98) | (-4.20) | | | | | |
| High Fluidity | | | | | -1.28*** | -0.08** | -0.35*** | |
| | | | | | (-4.62) | (-2.50) | (-4.36) | |
| TariffRates | | | | -6.55*** | | | | -9.71*** |
| | | | | (-8.13) | | | | (-8.82) |
| ExchangeRate | | | | 1.24*** | | | | 3.93*** |
| | | | | (2.82) | | | | (4.83) |
| CashFlow | 0.10 | 0.05*** | 0.10*** | -0.16*** | 0.24*** | 0.06*** | 0.14*** | -0.11*** |
| | (0.87) | (4.47) | (3.38) | (-11.64) | (3.11) | (7.62) | (6.58) | (-7.40) |
| Tangibility | -0.02 | 0.00 | -0.03 | -0.02 | -0.00 | 0.00 | -0.02 | -0.02 |
| | (-0.28) | (0.33) | (-1.35) | (-1.51) | (-0.03) | (0.62) | (-1.00) | (-1.18) |
| logAssets | 0.20*** | 0.01*** | 0.04*** | 0.03*** | 0.17*** | 0.00** | 0.03*** | 0.03*** |
| | (9.11) | (2.59) | (5.47) | (8.37) | (8.90) | (2.10) | (5.30) | (5.14) |
| NetWorth | -0.04 | 0.00 | -0.01 | -0.04*** | 0.06 | 0.01 | 0.01 | 0.01 |
| | (-0.56) | (0.03) | (-0.66) | (-2.95) | (1.02) | (1.05) | (0.52) | (0.55) |
| M2B | -0.02 | -0.00** | -0.03*** | 0.02*** | -0.02* | -0.00*** | -0.03*** | 0.02*** |
| | (-1.00) | (-2.16) | (-6.70) | (11.11) | (-1.65) | (-3.29) | (-8.29) | (10.31) |
| IndSaleVol | -2.23 | -0.25 | -3.62*** | 0.07 | -2.44 | -0.25 | -3.68*** | -0.31 |
| | (-0.79) | (-0.98) | (-4.71) | (0.14) | (-0.89) | (-0.94) | (-4.72) | (-0.35) |
| CFVol | -0.08 | -0.03 | -0.02 | -0.00 | -0.15 | -0.03 | -0.06 | -0.04 |
| | (-0.42) | (-1.18) | (-0.33) | (-0.10) | (-0.80) | (-1.45) | (-0.76) | (-1.44) |
| logFirmAge | 0.04 | 0.00 | 0.01 | -0.08*** | 0.05 | 0.00 | 0.02 | -0.09*** |
| | (0.63) | (0.31) | (1.05) | (-10.30) | (0.99) | (1.00) | (1.62) | (-8.93) |
| NonSP500 | 0.17 | 0.02** | 0.10*** | 0.04* | 0.11 | 0.02** | 0.09*** | 0.01 |
| | (1.41) | (2.47) | (3.65) | (1.73) | (0.95) | (2.17) | (3.26) | (0.42) |
| Year FEs | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FEs | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| *N* | 11380 | 11507 | 11507 | 11380 | 11380 | 11507 | 11507 | 11380 |
| Hensen J | 14.39(0) | 2.50 (0.11) | 9.65 (0) | | 2.76(0.09) | 8.82(0.00) | 1.32(0.25) | |
| Wald Chi$^2$ | 34.69(0) | 39.59 (0) | 58.30 (0) | | 44.23(0) | 31.18 (0) | 60.77(0) | |

Table 10: **Propensity Score Matching: Trump Tariff Policy**

This table provides quasi-natural experiment evidence for U.S. firms over the sample period 2002-2019 using Tariff increasing by Trump administrative as exogenous shock. For each observation of treated firm, we match an observation with a controlled firm in the same year, employing a propensity-score-matching approach. In Panel A, we report results from Probit regressions used to calculate the propensity scores for the matching procedure, where the dependent variable is a dummy equals to 1 if firm is in the treated group and 0 if firm is in the control group 1. Columns (1) and (2) respectively show marginal effects for the full sample before matching and the sub-sample with matched observations. Panel B displays the distribution of propensity scores from the regression in Column (2) of Panel A. Panel C compares credit line usage between treated firm vs. controlled firms. ***, **, and * indicate significance at the 1are defined in Appendix II.

**Panel A: Pre-Match and Post-Match Probit Regression**

| | (1) | (2) |
|---|---|---|
| | Pre-Match Regression | Post-Match Regression |
| CashFlow | -0.16 | -0.85* |
| | (-0.58) | (-1.68) |
| Tangibility | 0.15 | -0.77** |
| | (0.72) | (-2.25) |
| LogAssets | 0.13 | 0.05 |
| | (1.62) | (0.50) |
| NetWorth | 0.14 | -0.07 |
| | (0.54) | (-0.15) |
| M2B | 0.11*** | -0.10 |
| | (3.56) | (-1.37) |
| IndSaleVol | -90.05*** | -23.38*** |
| | (-4.36) | (-2.64) |
| CFVol | -0.73 | 0.06 |
| | (-1.05) | (0.08) |
| logFirmAge | -0.34** | 0.36* |
| | (-2.29) | (1.67) |
| NonSP500 | -0.27 | 0.62 |
| | (-0.81) | (1.24) |
| Year FEs | Yes | Yes |
| Industry FEs | Yes | Yes |
| *N* | 812 | 173 |

**Panel B: Estimated Propensity Score Distributions**

| **Propensity Score** | **N** | **mean** | **sd** | **p1** | **p25** | **p50** | **p75** | **p99** |
|---|---|---|---|---|---|---|---|---|
| Control | 80 | 0.53 | 0.20 | 0.17 | 0.40 | 0.51 | 0.65 | 1.00 |
| Treated | 93 | 0.48 | 0.17 | 0.18 | 0.36 | 0.45 | 0.58 | 0.88 |
| Total | 173 | 0.50 | 0.19 | 0.18 | 0.37 | 0.49 | 0.59 | 0.89 |

(*continuing on the next page...*)

(...*continued from the previous page*)

**Panel C: Difference in Credit Line Usage and Other Firm Characteristics**

| | Treatment | | | Control | | | Difference |
|---|---|---|---|---|---|---|---|
| | N | Mean | SD | N | Mean | SD | |
| HasLine | 80 | 0.838 | 0.37 | 93 | 0.763 | 0.43 | 0.074 |
| Undrawn_Assets | 80 | 0.110 | 0.10 | 93 | 0.105 | 0.10 | 0.005 |
| Undrawn_TotLiq | 80 | 0.375 | 0.29 | 93 | 0.403 | 0.32 | -0.028 |
| | | | | | | | |
| *Firm controls:* | | | | | | | |
| CashFlow | 80 | 0.076 | 0.36 | 93 | 0.050 | 0.26 | 0.026 |
| Tangibility | 80 | 0.742 | 0.49 | 93 | 0.697 | 0.44 | 0.045 |
| LogAssets | 80 | 6.295 | 2.36 | 93 | 6.624 | 1.67 | -0.329 |
| NetWorth | 80 | 0.345 | 0.34 | 93 | 0.414 | 0.32 | -0.070 |
| M2B | 80 | 3.460 | 3.76 | 93 | 1.749 | 1.27 | 1.712*** |
| IndSaleVol | 80 | 0.020 | 0.01 | 93 | 0.018 | 0.00 | 0.002* |
| CFVol | 80 | 0.028 | 0.13 | 93 | 0.014 | 0.02 | 0.014 |
| LogFirmAge | 80 | 2.848 | 1.01 | 93 | 3.260 | 0.67 | -0.412*** |
| NonSP500 | 80 | 0.813 | 0.39 | 93 | 0.935 | 0.25 | -0.123** |

Table 11: **Quasi-natural Experiment: Trump Tariff Policy**

This table examines how Trump's tariff policy affects competition's effect on credit line usage. *Treat* is a dummy that equals one if the firm is in treatment group after applying the propensity score matching. *Post* is a dummy equals one if current fiscal year is 2018 or 2019, and 0 if current fiscal year is 2016 or 2017. All regressions control for year and industry fixed effects. In parentheses are *t*-stats based on robust standard errors adjusted for heteroskedasticity and firm-level clustering. ***,**, and * denote statistical significance at the 1%, 5%, and 10% levels. All variable definitions are in Table 2. Variables of interest are in boldface.

| | (1)<br>HasLine | (2)<br>Undrawn_Assets | (3)<br>Undrawn_TotLiq |
|---|---|---|---|
| Treat× Post | 0.44*** | 0.15*** | 0.40** |
| | (3.40) | (3.10) | (2.45) |
| Treat | -0.30*** | -0.08*** | -0.27*** |
| | (-4.07) | (-2.85) | (-3.90) |
| CashFlow | 0.05 | 0.05 | 0.09 |
| | (0.62) | (1.07) | (0.71) |
| Tangibility | -0.07 | -0.06** | -0.17*** |
| | (-1.18) | (-2.31) | (-3.06) |
| LogAssets | 0.06*** | 0.01 | 0.04* |
| | (3.11) | (1.40) | (1.79) |
| NetWorth | -0.12 | -0.02 | -0.06 |
| | (-1.59) | (-0.64) | (-0.68) |
| M2B | -0.02 | 0.00 | -0.04** |
| | (-1.05) | (0.70) | (-2.48) |
| IndSaleVol | 6.58 | 2.19*** | 4.23** |
| | (0.50) | (3.47) | (2.22) |
| CFVol | -0.69 | -1.03* | -2.74* |
| | (-0.57) | (-1.81) | (-1.74) |
| LogFirmAge | -0.02 | -0.02 | -0.03 |
| | (-0.41) | (-1.29) | (-0.88) |
| NonSP500 | 0.00 | 0.07*** | 0.07 |
| | (0.01) | (2.62) | (0.92) |
| Year FEs | Yes | Yes | Yes |
| Industry FEs | Yes | Yes | Yes |
| *N* | 173 | 173 | 173 |

Table 12: **Placebo Test: Trump Tariff Policy**

This table examines how Trump's tariff policy affects competition's effect on credit line usage using random assigned shock. *Treat* is a dummy that equals one if the firm is in treatment group after applying the propensity score matching. *Placebo1* is a dummy equals one if fiscal year is 2004 or 2005, and 0 if current fiscal year is 2002 or 2003. *Placebo2* is a dummy equals one if fiscal year is 2012 or 2013, and 0 if current fiscal year is 2014 or 2015. All regressions control for year and industry fixed effects. In parentheses are *t*-stats based on robust standard errors adjusted for heteroskedasticity and firm-level clustering. ***,**, and * denote statistical significance at the 1%, 5%, and 10% levels. All variable definitions are in Table 2. Variables of interest are in boldface.

| | (1) | (2) | (3) |
|---|---|---|---|
| | HasLine | Undrawn_Assets | Undrawn_TotLiq |
| **Treat× Placebo1** | **-0.12** | **-0.03** | **-0.04** |
| | **(-0.36)** | **(-0.98)** | **(-0.49)** |
| Treat | -0.92*** | -0.06*** | -0.25*** |
| | (-3.59) | (-2.59) | (-3.39) |
| Controls | Yes | Yes | Yes |
| Year FEs | Yes | Yes | Yes |
| Industry FEs | Yes | Yes | Yes |
| *N* | 173 | 173 | 173 |
| **Treat× Placebo2** | **-0.03** | **-0.03** | **-0.04** |
| | **(-0.36)** | **(-0.98)** | **(-0.49)** |
| Treat | -0.24*** | -0.06*** | -0.25*** |
| | (-3.59) | (-2.59) | (-3.39) |
| Controls | Yes | Yes | Yes |
| Year FEs | Yes | Yes | Yes |
| Industry FEs | Yes | Yes | Yes |
| *N* | 332 | 332 | 332 |

Table 13: **Cross-sectional Heterogeneity: Asset Tangibility**

This table examines how asset tangibility affects competition's effect on credit line usage. *HighTang* is a dummy that equals one if the firm has above-median *Tangibility*, and zero otherwise. All regressions control for year and industry fixed effects. In parentheses are *t*-stats based on robust standard errors adjusted for heteroskedasticity and firm-level clustering. ***,**, and * denote statistical significance at the 1%, 5%, and 10% levels. All variable definitions are in Table 2. Variables of interest are in boldface.

| | (1) | (2) | (3) |
|---|---|---|---|
| | HasLine | Undrawn_Assets | Undrawn_TotLiq |
| **Fluidity×HighTang** | **0.08*** ** | **0.04*** ** | **0.18*** ** |
| | **(5.25)** | **(6.10)** | **(8.40)** |
| Fluidity | -0.15*** | -0.06*** | -0.24*** |
| | (-11.77) | (-11.65) | (-14.27) |
| HighTang | -0.03*** | -0.02*** | -0.09*** |
| | (-3.20) | (-3.63) | (-6.11) |
| CashFlow | 0.09*** | 0.08*** | 0.19*** |
| | (11.41) | (17.36) | (14.61) |
| logAssets | 0.04*** | 0.01*** | 0.05*** |
| | (21.07) | (7.15) | (18.31) |
| NetWorth | 0.01* | 0.01*** | 0.05*** |
| | (1.88) | (3.34) | (4.53) |
| M2B | -0.01*** | -0.00*** | -0.03*** |
| | (-7.92) | (-6.82) | (-19.86) |
| IndSaleVol | -0.26 | 0.22** | -1.09*** |
| | (-0.91) | (2.16) | (-3.28) |
| CFVol | 0.01 | 0.00 | -0.03 |
| | (0.31) | (0.03) | (-0.63) |
| logFirmAge | 0.02*** | 0.01*** | 0.01*** |
| | (5.74) | (3.71) | (2.87) |
| NonSP500 | 0.07*** | 0.03*** | 0.15*** |
| | (5.34) | (8.62) | (11.01) |
| Year FEs | Yes | Yes | Yes |
| Industry FEs | Yes | Yes | Yes |
| $N$ | 47690 | 47765 | 47765 |

Table 14: **Cross-sectional Heterogeneity: Investment Opportunities**

This table examines how growth opportunities affect competition's effect on credit line usage. Growth opportunities are measured by R&D capital (*RDC*) in Panel A and market-to-book ratio (*M2*B) in Panel B. *HighRDC* is a dummy that equals one if the firm has above-median R&D capital in the previous year, and zero otherwise. HighM2B is a dummy that equals one if the firm has above-median M2B in the previous year, and zero otherwise. All regressions control for year and industry fixed effects. In parentheses are *t*-stats based on robust standard errors adjusted for heteroskedasticity and firm-level clustering. ***,**, and * denote statistical significance at the 1%, 5%, and 10% levels. All variable definitions are in Table 2. Variables of interest are in boldface.

| | (1) | (2) | (3) |
|---|---|---|---|
| | HasLine | Undrawn_Assets | Undrawn_TotLiq |
| *Panel A: R&D Intensity* | | | |
| **Fluidity× HighRD** | **-0.65*** ** | **-0.07*** ** | **-0.27*** ** |
| | **(-6.99)** | **(-9.38)** | **(-10.96)** |
| Fluidity | -0.13 | -0.01 | -0.01 |
| | (-1.63) | (-1.03) | (-0.74) |
| HighRD | 0.14** | 0.02*** | 0.03 |
| | (2.01) | (3.64) | (1.54) |
| *N* | 43414 | 43627 | 43627 |
| *Panel B: Market-to-book* | | | |
| **Fluidity× HighM2B** | **0.00** | **-0.01*** | **-0.02** |
| | **(0.02)** | **(-1.85)** | **(-1.25)** |
| Fluidity | -0.11*** | -0.04*** | -0.17*** |
| | (-9.03) | (-8.63) | (-9.67) |
| HighM2B | -0.04*** | 0.00 | -0.12*** |
| | (-4.25) | (0.24) | (-8.49) |
| *N* | 47690 | 47765 | 47765 |

Table 15: **Cross-sectional Heterogeneity: Cash Flow and Cash Flow Volatility**

This table examines how cash flow and cash flow volatility affect competition's effect on credit line usage. *HighCF* is a dummy that equals one if the firm has above-median *CashFlow* in the previous year, and zero otherwise. *HighCFV* is a dummy that equals one if the firm has above-median cash flow volatility in the previous year and zero otherwise. All regressions control for year and industry fixed effects. In parentheses are *t*-stats based on robust standard errors adjusted for heteroskedasticity and firm-level clustering. ***,**, and * denote statistical significance at the 1%, 5%, and 10% levels. All variable definitions are in Table 2. Variables of interest are in boldface.

| | (1) | (2) | (3) |
|---|---|---|---|
| | HasLine | Undrawn_Assets | Undrawn_TotLiq |
| *Panel A: Cash Flow* | | | |
| **HighCF×Fluidity** | **0.03*****| **0.02***** | **0.11***** |
| | **(2.62)** | **(3.04)** | **(6.03)** |
| Fluidity | -0.13*** | -0.05*** | -0.22*** |
| | (-11.31) | (-10.99) | (-12.84) |
| HighCF | 0.06*** | 0.04*** | 0.05*** |
| | (5.24) | (9.94) | (3.76) |
| $N$ | 47690 | 47765 | 47765 |
| *Panel B: Cash Flow Volatility* | | | |
| **HighCFV×Fluidity** | **-0.01** | **-0.02***** | **-0.04**** |
| | **(-1.06)** | **(-3.33)** | **(-2.26)** |
| Fluidity | -0.10*** | -0.03*** | -0.14*** |
| | (-7.84) | (-7.04) | (-8.26) |
| HighCFV | -0.01* | 0.01 | -0.04*** |
| | (-1.87) | (1.30) | (-3.55) |
| $N$ | 47690 | 47765 | 47765 |

ACCOUNTING & FINANCE

afaanz

Accounting & Finance

# Modelling of Chinese corporate bond default – A machine learning approach

Zhou Lu[a], Zhuyao Zhuo[b,c]

[a]*Department of Finance, Norwegian School of Economics (NHH), Bergen, Norway*
[b]*Department of Finance, Lingnan (University) College, Sun Yat-sen University (SYSU), Guangzhou, China*
[c]*China Resource Financial Holdings Co., Ltd. (CR Capital), Hong Kong, China*

---

**Abstract**

We apply machine learning techniques to construct a series of models of corporate bond defaults. By combining Chinese accounting information and corporate bond data from January 2012 to December 2019, we construct an out-of-sample forecast that significantly improves the identification rate of corporate bond defaults, with an area under the receiver operating characteristics curve greater than 90 percent. Our models are robust to different machine learning models, including stacking, boosting, and bagging ensembling models. Our models consider cross-sectional heterogeneity, such as different ownership structures, accessibility to external finance, industry heterogeneity, different sample periods, and government policy impact.

*Key words*: Corporate bond; Ensembling models; Machine learning; Prediction of default probability

*JEL classification*: G31, G32, C50, C87

*doi*: 10.1111/acfi.12846

---

We would like to thank the Editor for handling the manuscript and to thank the referees for their insightful comments, which have helped to improve the manuscript. We thank Yun Dai, Evgeny Lyandres, Yanchu Liu, Xunhua Su, Qianwei Ying, Tong Zhou, and seminar participants at the Lingnan University College of Sun Yat-sen University (SYSU), Norwegian School of Economics, and the 2021 CAFC for their helpful comments and suggestions on previous drafts. All errors are our own.
This work was supported by the National Natural Science Foundation of China (No. 71873152) and China Postdoctoral Science Foundation (No. 2020M672910).

Please address correspondence to Zhou Lu via email: zhou.lu@nhh.no

## 1. Introduction

China's corporate bond market has grown exponentially over the past decade, increasing from 1.89 trillion RMB in 2009 to 32.9 trillion RMB in 2019. At the end of 2019, the corporate bond market accounted for 33.2 percent of China's bond market. According to statistics from the Wind Economic Database, 58 major issuers in China's corporate bond market defaulted before 2017, involving 113 defaulted bonds, with a balance of 37.992 billion RMB. From 2017 to 2019, the number of defaulted issuers increased to 103. This involved 425 defaulted bonds, and the balance of newly defaulted bonds reached 343.083 billion RMB. The reasons for the surge in defaults are as follows. After 2017, because of the impact of intense regulation and a deleveraging policy, Chinese enterprises' financing channels narrowed, the credit risk of firms relying on rolling financing continued to increase, and default events emerged at an accelerating pace. It is a challenge for highly leveraged firms to repay their principal and interest amid deterioration of the external finance environment, which has triggered the largest default wave of corporate bonds in China.

Our primary focus is predicting corporate bond defaults in China. Corporate bond defaults have been widely studied. Research can be sorted into three main categories: structure models, reduced-form models and macroeconomic and/or accounting-based models. The structure credit model emerged from the options pricing theory of Black and Scholes (1973) and Merton (1973) and has been extended.[1] The structure model framework identifies key factors driving corporate bond value. The reduced-form model assumes that corporate bond defaults satisfy a Poisson distribution, and uses the default strength index to measure the default risk (Jarrow and Turnbull, 1995; Duffie and Singleton, 1997, 1999). Another reduced-form model assumes that the probability of credit rating conversion follows the Markov process and constructs a credit rating transfer matrix to predict the default rate of debt (Jarrow *et al.*, 1997). Both the structure and reduced-form models predict the probability of future credit defaults under a specific theoretical framework. The advantage of these models lies in their minimal dependence on historical default data and their good foresight. However, these models have two disadvantages. First, they rely heavily on the validity of the assumptions. Second, the structure model relies on asset price data, which might not be available for non-listed firms. Other studies have focused on macroeconomic and accounting-based models. One stream of literature[2] evaluates how credit risk represents systematic risk in the market. Beaver (1966), Altman (1968), Ohlson (1980) and Beaver *et al.* (2005) evaluate

[1]See Black and Cox (1976), Geske (1977), Leland (1994, 2012), Longstaff and Schwartz (1995), Collin-Dufresne and Goldstein (2001) and Goldstein *et al.* (2001).

[2]The research includes Fama and French (1993), Dichev (1998), Chava and Jarrow (2004), Vassalou and Xing (2004) and Campbell *et al.* (2008).

how earnings and accounting ratios (reflecting the sales, expenses, growth and liquidity of a firm, which indicate its macroeconomic and business environments) predict the default of corporate bonds.

The primary objective of most models, such as structural, reduced-form or macroeconomic and accounting models, is to explain corporate bond defaults *within* sample, and they often emphasise causal inference. Our objective is different: we aim to develop a model that can accurately predict corporate bond defaults *out of sample* (i.e., a prediction problem). Satchidananda and Simha (2006) show that the problems of causal inference and prediction, although related, are fundamentally different. Specifically, causal inference modelling aims to minimise the bias resulting from model misspecification to obtain the most accurate representation of the underlying theory. In contrast, the objective of predictive modelling seeks to minimise the out-of-sample prediction error, that is, the combination of the bias and estimation variance resulting from using a sample to estimate model parameters. Although causal inference represents mainstream social science research, Kleinberg *et al.* (2015) show that many interesting prediction problems are neglected in the extant business and economics literature. As shown in Table 1, the actual rate of corporate bond defaults in 2 years is 0.51 percent, which is a rare event, and 70 percent of firms are not publicly listed, suggesting that it usually does not have reliable public information. Our models can effectively predict rare events with a small number of input variables and have better performance than traditional default risk models, such as structure or accounting models.

Machine learning (ML) has been widely used in risk modelling, risk assessment and risk prediction in recent decades. According to Altman *et al.* (1994), combining traditional models, such as logit and probit regression models, significantly improves default predictability. Figini *et al.* (2017) propose an ML technique for credit risk estimation. Khandani *et al.* (2010) construct an ML nonlinear nonparametric forecasting model of consumer credit risk, which significantly improves the model's accuracy. Kim and Sohn (2004) apply a neural network model and improve the classical credit scoring models. Wang and Huang (2009) and Xu *et al.* (2014) use a neural network model for risk assessment and risk forecasting. Yu (2011) shows that ML technology helps assess the risk of complex projects. The ML technique is useful for improving model accuracy, especially when the credit market's complexity increases.

This study uses two typical default risk ML models as benchmarks. The linear support vector machine (SVM) model and kernel support vector machine (KSVM) model with a financial kernel that maps raw financial data into a broader set of ratios within the same year and changes in ratios across different years. Our proposed default prediction model differs from both these benchmark models in two key ways. First, we use ensemble learning, a state-of-the-art ML paradigm, to predict corporate bond default. Although ensemble learning has been successfully applied in many other fields (Zhou, 2012), ours is

Table 1
Default status categorised by listed status, ownership structure and Wind primary industry classifier

| | Default | Non-default | Default rate |
|---|---|---|---|
| **Panel A: Default status within period** | | | |
| 0.5 years | 20 | 39,538 | 0.05% |
| 1 year | 59 | 39,499 | 0.15% |
| 1.5 years | 115 | 39,443 | 0.29% |
| 2 years | 202 | 39,356 | 0.51% |
| **Panel B: Categorised by listed status** | | | |
| Non-listed | 138 | 27,470 | 0.50% |
| Listed | 64 | 11,886 | 0.54% |
| **Panel C: Categorised by ownership structure** | | | |
| SOEs (Central) | 8 | 4,339 | 0.18% |
| SOEs (Local) | 22 | 24,559 | 0.09% |
| Public companies | 6 | 580 | 1.02% |
| Collective enterprises | 4 | 186 | 2.10% |
| Private enterprises | 135 | 8,127 | 1.63% |
| Foreign-owned enterprises | 8 | 573 | 1.37% |
| Foreign companies | 4 | 318 | 1.24% |
| Chinese–Foreign equity joint ventures | 15 | 485 | 3.09% |
| Other enterprises | 0 | 189 | 0.00% |
| **Panel D: Categorised by Wind primary industry classifier** | | | |
| Energy | 23 | 1,415 | 1.60% |
| Information technology | 21 | 1,550 | 1.33% |
| Consumption staples | 16 | 1,316 | 1.20% |
| Material | 38 | 4,492 | 0.84% |
| Consumer discretionary | 19 | 3,401 | 0.55% |
| Utilities | 9 | 2,257 | 0.39% |
| Industrials | 67 | 17,854 | 0.37% |
| Financial | 6 | 2,521 | 0.24% |
| Health care | 1 | 913 | 0.10% |
| Real estate | 2 | 3,590 | 0.06% |
| Telecommunication services | 0 | 47 | 0.00% |

The table shows the number of realised default and non-default cases, and the portion of default cases over the sum of default and non-default cases after the announcement of semi-annual accounting reports categorised by listed status, ownership structure, firm property, and Wind primary industry classifier.

the first study to apply the method to a finance setting with a severe class imbalance problem, namely, rarity of corporate bond default (see Table 1). Whether ensembling models outperform traditional models is an empirical question. Our results suggest that ensemble learning, if properly used, is more powerful than SVM or KSVM for the purpose of corporate bond default prediction.

Second, we adopt models first using multiple ML models for feature engineering, which pick up the most important features (the so-called 'importance index') to predict the default risk and then perform logistic regression. The benefit of this type of model is that it perfectly combines the prediction power of complicated ML models (e.g., bagging or boosting models) with the economic intuition of the logistic model, which allows us to draw inferences from the models. Prediction power is defined as an area under the receiver operating characteristics curve (AUC) of the receiver operating characteristics (ROC) curve greater than 90 percent.

To compare the out-of-sample performance of different default prediction models, we adopt two performance evaluation metrics. First, since default status is a binary variable, we can measure the model performance by constructing a confusion matrix and evaluate the models by its *Accuracy*, *Precision*, *Recall* and *F_Score*. A detailed explanation of the criteria is provided in Section 5. Second, we use the AUC of the ROC curve as a performance evaluation metric. The AUC is equivalent to the probability that a randomly chosen fraud observation will be ranked higher by a classifier than a randomly chosen non-default observation will be ranked. The AUC for random guesses is 50 percent. Therefore, any reasonable default prediction model must have an AUC higher than 50 percent, and any model with an AUC greater than 80 percent has good prediction power.

To put the performance evaluation of all fraud prediction models on an equal footing, we require all models to start with a common set of raw financial data. We used 27 financial ratios or normalised financial variables and 31 dummy variables as the starting point to evaluate the performance of all default prediction models.

This study uses bond data in China, namely, 2012–2017 semi-annual and annual accounting information, for listed and non-listed firms, and predicts the default risk 2 years after the publication of the accounting report. We first report the out-of-sample performance results for the two benchmark models. Using AUC as the performance evaluation metric, we find that the out-of-sample performance of the two benchmark models is significantly better than the performance of random guesses. The SVM yields an AUC of 88.72 percent, and the KSVM model yields an AUC of 91.75 percent. Our models outperform those of Barboza *et al.* (2017), whose SVM and KSVM models have an AUC of 66 and 90 percent and perform better than the logistic regression and GTBoost models of Bracke *et al.* (2019).

The performance of the prediction models increases after we adopt ensembling models. All ensembling models yield an AUC greater than 90 percent. We first show the result of boosting models, including AdaBoost (ABC), gradient boosting classifier (GBC) and histogram gradient boost (HGBC) models. The AUC of the ABC, GBC and HGBC models is 97.53, 94.83 and 94.96 percent, respectively, outperforming the benchmark models. We then show the performance of the bagging models by combining a simple learner, such as SVM, logistic regression and neural network with the bagging classifier. All models have good prediction power. We apply a random forest (RF), an extra tree (ET) and two classical bagging models, and the results show that by using the bagging model, the model's prediction power improves. The bagging model combined with a basic learner, such as SVM, KSVM, multilayer perceptron (MLP) classifier and logistic regression (LR), yields an AUC greater than 92 percent, which is higher than the AUC of both benchmark models. The AUC of RF and ET is 97.65 and 96.69 percent, respectively, which is higher than the AUC of both benchmark models. We apply multiple stacking models that combine the basic learner (SVM, KSVM, LR and MLP). The model outperforms the baseline model. We further investigate its performance by combining a stacking model with boosting and bagging models. The final model that combines the stacking, boosting and bagging models yields an AUC of 97.89 percent, indicating its good prediction property.

We find that after using the boosting and bagging methods for feature engineering, our models outperform the benchmark models. For instance, the AUC obtained by combining ABC + logistic regression is 92.16 percent, that by combining the GBC + logistic regression is 91.68 percent, that by combining RF + logistic regression is 92 percent, and that by combining ET+ logistic regression is 92.15 percent. Feature engineering provides an intuitive prediction of what drives corporate bond defaults.

We further investigate the most important features that predict corporate bond default using the importance index (for details on the method, see Appendix I). We further investigate the most important features predicting corporate bond defaults for publicly listed firms, non-listed firms, state-owned enterprises (SOEs), foreign enterprises and firms before and after China's structural reform.

We further compared our result with the traditional credit risk model. Because of data availability, we analyse the corporate default based on the model of Merton (1974) and the KMV model (for details on the method, see Appendix II) and show that our method outperforms traditional structure models. The remainder of this study is organised as follows. Section 2 presents the data, samples and variables. Section 3 presents the baseline results and robustness tests. Section 4 concludes.

## 2. Empirical analysis

### 2.1. The measure of default risk

Default risk can be measured as a conditional expectation. Assuming that there are $n$ observed individuals, given a non-random vector $x_i$, the probability of a default event of the $i^{\text{th}}$ individual is

$$p(x_i) = P(y_i = 1|x_i) = E(y_i = 1|x_i), \tag{1}$$

where $y_i$ is the state variable of the $i^{\text{th}}$ individual concerning event and $x_i = \{x_{i1}, x_{i2}, \ldots, x_{ik}\}$ is a $k$-dimensional non-random observation vector that may affect the occurrence of events.

ML algorithms for default rate prediction include SVM, KSVM, boosting, bagging, stacking, and models combining boosting, bagging, and LR are used to predict China's credit bond issuers' default rate and compare prediction accuracy. We introduce the mechanism of these models in Appendix I.

### 2.2. Data and sample selection

We obtained 2012–2017 semi-annual and annual accounting information on Chinese firms' debt issuance to construct our explanatory variables. We downloaded the information on listing status, ownership structure, industry classifier and credit rating downgrade and constructed a data set of semi-annual firms. Data were collected from the Wind Economic Database. A total of 39,558 observations were obtained.

The explanatory variable is the default status of corporate bonds within a specific period after the observation (financial report). The sample period runs from July 2012 to December 2019 for the extrapolation prediction. We calculated the default status of firms at 0.5, 1, 1.5 and 2 years. The default status is a dummy that equals one if a firm cannot meet interest or principal payments within the period, and zero otherwise. Panel A of Table 1 shows the number of realised default and non-default cases and the proportion of default cases over the sum of default and non-default cases after the announcement of the semi-annual accounting report for 0.5, 1, 1.5 and 2 years. As the table shows, the 'within 2 years' sample has the highest probability of default.

In this study, credit status in the next 2 years was selected as the prediction variable. With the improvement of Chinese firms' financial database and the increasing number of default events, the default rate within 0.5–1 year can be calculated to enrich the study's structure. To capture the variables that affect the change in firm credit status, this study calculated the default distribution from three dimensions: listing status, ownership structure, and Wind primary industry classifier (see Table 1).

Panel B of Table 1 shows the default status of listed and non-listed companies. The sample number of non-listed companies is about double that of listed companies, and the default rate of the former is almost the same as that of the latter. Listed firms can obtain funds through stock issuance or stock-pledge financing. Their liquidity and debt-paying ability are stronger than those of non-listed firms, with a lower probability of credit default and sample data in line with expectations. Therefore, we use 'Listed' as a proxy for firms' access to external finance. Panel C of Table 1 shows the sample default status by firm type or ownership structure. Foreign-funded firms, private firms, collective firms and other non-SOEs have higher default rates. Compared with SOEs, non-SOEs have insufficient collateral to pledge, have poor financing capacity, and have higher probability of default. Panel D of Table 1 shows the industry's default status by industry at the first level (arranged by the default rate from large to small). The default rate of firms in the energy, information technology, consumer staples and materials industries is higher than that of firms in other industries.

### *2.3. Input variables, summary statistics and samples*

There are multiple reasons for corporate bond defaults. Goldstein and Pauzner (2005), Wagner (2007), Gatev *et al.* (2009), Acharya and Viswanathan (2011), Acharya *et al.* (2011), Gorton and Metrick (2012), He and Xiong (2012) and Acharya and Mora (2015) show that credit risk and liquidity risk influence each other and affect a bank's stability. Liquidity risk includes market liquidity risk and funding liquidity risk. Market liquidity risk refers to the possibility that an asset cannot be repaid in full as scheduled when it matures. In this case, a firm may be unable to meet the demand for repaying mature liabilities or for new investments, which reflects the asset's liquidity. Funding liquidity risk refers to the possibility that a firm cannot obtain enough cash flow in the future to repay the maturing debt, which reflects the firm's financing ability. On the one hand, when the future liquidity of assets worsens, or financing channels are blocked, firms have difficulty refinancing, repayment of unmatured debts is at risk, and liquidity risk is transmitted to credit risk. On the other hand, the increasing credit risk causes financial institutions to reduce the asset valuation, increases the pressure of financial institutions to redeem their liability, and increases firms' liquidity risk. For these reasons, we added five factors (*Current Ratio*, *Acid Ratio*, *Cash Ratio*, *Cash Ratio2* and *WCL*) that represent liquidity risk in our prediction models. The definitions of the factors are presented in Table 2.

The trade-off theory of capital structure suggests that firms balance the costs and benefits of debt to determine an optimal leverage ratio (Gorton and Metrick, 2012). Leland (1994, 1998) Leland and Toft (1996) and Titman and Tsyplakov (2007) show that the consequences of leverage ratios for default risk and prices of defaultable bonds are significant. The trade-off theory of capital structure shows that fluctuations in firms' capital structure affect their value in two ways. Increasing the ratio of debt capital is conducive to increasing firm value. Financial constraints and agency costs limit the amount of debt financing, and an excessive debt capital ratio increases the burden of future interest payments. Weakening solvency causes firms to face higher credit risks. This means that excessive financial leverage could jeopardise a firm's solvency. Therefore, firms have a higher probability of default. For these reasons, we added 16 factors (*Leverage*, *IBLiability*, *BSBTA*, *#IB*, *#SB*, *SR*) related to firm solvency risk and capital structure. The definitions of the factors are presented in Table 2.

Merton (1974), Rendleman (1978), Acharya *et al.* (2006) and Asvanunt *et al.* (2011) show that cash flow and value drive corporate bond defaults, which

Table 2
Variable definitions

| Primary categories | Variables | Definition |
|---|---|---|
| Liquidity risk | *Current Ratio* | $\frac{\text{Current Assets}}{\text{Current Liabilities}}$ |
| | *Acid Ratio* | $\frac{\text{Current Assets}-\text{Inventory}}{\text{Current Liabilities}}$ |
| | *Cash Ratio* | $\frac{\text{Current Liabilities}}{\text{Cash and Cash Equivalents}}$ |
| | *Cash Ratio2* | $\frac{\text{Total Liabilities}}{\text{Cash and Cash Equivalents}}$ |
| | *WCL* | $\frac{\text{Current Liabilities}}{\text{Total Liabilities}}$ |
| Solvency risk (Capital Structure) | *Leverage* | $\frac{\text{Total Assets}}{\text{Total Liabilities}}$ |
| | *IBLiability* | $\frac{\text{InterestBearing Debt}}{\text{Total Invested Capital}}$ |
| | *BSBTA* | The total balance of secured bond/total assets |
| | *#IB* | The number of bonds issued by companies |
| | *#SB* | The number of secured bonds |
| | *SR* | Secured ratio (The balance of secured bond/total equity) |
| | *SB3TE* | Secured bonds under 3 years/total equity |
| | *SB1TE* | Secured bonds under 1 year /total equity |
| | *SBTL* | The balance of secured bond/ total liability |
| | *External Guarantee* | Whether target firm provide a guarantee to other firms to issue bond or not |
| | *Guarantee* | Whether target firm has been granted a guaranty |

(continued)

Table 2 (continued)

| Primary categories | Variables | Definition |
|---|---|---|
| | | for its bond issuance or not |
| | *SB13* | Secured bond matures within 1–3 years |
| | *SB35* | Secured bond matures within 3–5 years |
| | *SB57* | Secured bond matures within 5–7 years |
| | *SB710* | Secured bond matures within 7–10 years |
| | *SB10+* | Secured bond matures above 10 years |
| Cash flow management | *OCF Ratio* | $\frac{\text{Net Cash Outflow (Operating)}}{\text{Prime Operating Revenue}}$ |
| | *ICF Ratio* | $\frac{\text{Net Cash Outflow (Investment)}}{\text{Total Liabilities}}$ |
| | *FCF Ratio* | $\frac{\text{Net Cash Outflow (Financing)}}{\text{Total Liabilities}}$ |
| | *OCF Outflow* | Net cash outflow (Operating) |
| | *ICF Outflow* | Net cash outflow (Investment) |
| | *FCF Outflow* | Net cash outflow (Financing) |
| 4*Profitability | *OPE* | $\frac{\text{Operating Income}}{\text{Prime Operating Revenue}}$ |
| | *GrowthOR* | $\frac{\text{Prime Operating Revenue}_t - \text{Prime Operating Revenue}_{t-1}}{\text{Prime Operating Revenue}_{t-1}}$ |
| | *ROA* | $\frac{\text{Net Income}}{\text{Total Assets}}$ |
| | *ORoA* | $\frac{\text{Prime Operating Revenue}}{\text{Total Assets}}$ |
| Profitability for shareholder | *ROE* | $\frac{\text{Net Income}}{\text{Total Equities}}$ |
| Operational ability | *InvTO* | $\frac{\text{2*Sale}}{\text{(Beginning Inventory+Ending Inventory)}}$ |
| Regional development | *GDPPC* | GDP per capita in state level |
| | *City* | City construction investment bond/local government bond |
| | *Local* | Local government bond amount |
| Credit rating | *Downgrade* | *Downgrade* |
| | *Credit rating* | $= max\{CreditRating_t - CreditRating_{t-1}, 0\}$’ |

(continued)

Table 2 (continued)

| Primary categories | Variables | Definition |
|---|---|---|
| | | $CreditRating_t$ is an ordered integer, which ranges from 0 to 18 for firms with highest credit rating AAA to lowest credit rating *C* |
| Accessibility of external finance | *Listed* | *Listed* = 1 if the firm is publicly listed, 0 otherwise |
| Ownership structure | *SOE (Central), SOE (Local), Public Company, Collective Enterprise, Private Enterprise, Foreign-Owned Enterprise, Foreign Company, Chinese–Foreign Equity Joint Venture, and Other Firms* | Indicator variables for ownership structures |
| Wind primary industry classifier | *Energy, Information Technology, Consumer Staples, Material, Consumer Discretionary, Industrials, Utilities, Financial, Health Care, Real Estate, and Telecommunication Services* | Indicator variables for industries |
| Accounting report date dummy | *2012-06-30, 2012-12-31, 2013-06-30, 2013-12-31, 2014-06-30, 2014-12-31, 2015-06-30, 2015-12-31, 2016-06-30, 2016-12-31, 2017-06-30, 2017-12-31* | Indicator variables for accounting report date |

The table defines the variables used in our analyses. All dollar-denominated variables are inflation adjusted to 2010 values.

implies that business risk can affect the credit risk of firms. Business risk refers to the possibility that firms' future operating cash flow fluctuates because of changes in the market environment or in production and operation activities, thereby affecting their market value. When future profitability is unlikely to generate sufficient operating cash flow, firms need to obtain funds to support their daily operations and production by cashing in assets or financing. When business risk is high, we would expect firms to prioritise their daily operations to maintain liquidity. Firms may have higher liquidity pressure, which increases their solvency pressure. For these reasons, we added 12 factors (*OCF ratio*, *ICF ratio*, *FCF Ratio*, *OCF Outflow*, *ICF Outflow*, *FCF Outflow*, *OPE*, *GrowthOR*, *ROA*, *ORoA*, *ROE* and *InvTO*) related to firm profitability, cash flow management, and operational ability. The definitions of the factors are presented in Table 2.

Credit ratings reflect firms' future solvency. When a firm's business situation deteriorates, the rating company considers downgrading its credit ratings. Therefore, *Downgrade* can be used as a forecasting indicator. This study used credit rating information provided by the Wind database. We first converted the credit rating from AAA, AA+, AA, AA−, A+, A, A−, BBB+, BBB, BBB−, BB+, BB, BB−, B+, B, B−, CCC, CC and C into ordered integers ranging from 0 to 18; with greater numbers meaning worse credit rating. We then observed whether the firm has a credit rating downgrading compared to the previous year. For firms without a change in credit rating or that experience a credit rating upgrade, *Downgrade* was set to 0. *Downgrade* was calculated as follows:

$$Downgrade = \max\{CreditRating_t - CreditRating_{t-1}, 0\}. \tag{2}$$

Figure 1 exhibits the histogram between credit rating downgrading and default risk. The figure shows that the higher the order of credit rating downgrades in the past year, the more likely the firm will default in the next 2 years. Therefore, this variable should be included in the prediction model.

Corporate bond defaults in China may depend largely on the financial status of the local government. Some local SOEs have inferior performance and rely heavily on local government financial support to repay their debt. Therefore, it is important to consider macroeconomic factors, such as regional per capita GDP, city construction investment bonds, and local government bonds.

Essentially, our models used 36 financial factors combined with multiple ordered or dummy variables (e.g., downgrade, listed, ownership structure, Wind primary industry classifier, and accounting report date dummy). The total number of input variables is 69. Table 2 lists the factors and their definitions, and Table 3 presents the summary statistics.

Firms with bond defaults behave differently from firms that satisfy the continuing operations assumption. For example, the asset–liability ratio of defaulting firms is larger than that of firms that continue to operate, and the growth rate of their main business income can be negative. This study took the

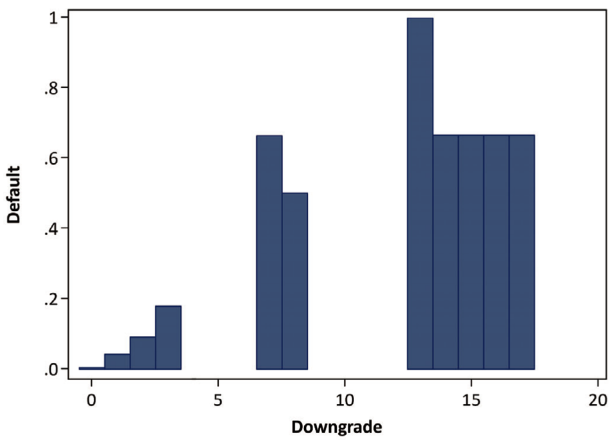



Figure 1 Default risk and credit downgrading.This figure shows a histogram of default risk and credit downgrading. The *y*-axis indicates the probability of default. The *x*-axis is the level of credit rating downgrading.

reciprocal of the current ratio and the acid ratio to highlight the short-term solvency of defaulting firms. To avoid omitting any sample of non-defaulting firms, we excluded only the extreme value sample of non-defaulting firms. After removing the outliers, the numerical distribution of all the variables was close to normal. The effective sample size is 39,558 firms, accounting for 89.89 percent of the original sample. The number of defaulting firms remains unchanged. We further scaled all the variables to ensure that they fall in the [0,1] interval using the following transformation:

$$X_{scaled} = \frac{(X - \min(X))}{(\max(X) - \min(X))}. \tag{3}$$

To verify the prediction effect, 39,558 firm-year observations were randomly sampled in the ratio of 7:3 to ensure that the within-group default rates of the training and testing groups were similar. The sampling distribution is shown in Table 4, Panel A.

## 3. Default risk prediction

### *3.1. Prediction accuracy of models*

According to the model's actual status and predicted status, we can show its performance using a confusion matrix for a dichotomy problem. Table 5 shows the format of the confusion matrix. TN denotes the number of firms correctly classified as non-default firms. TP denotes the number of firms correctly

Table 3
Summary statistics

| | Count | Mean | SD | Min | P25 | P50 | p75 | Max |
|---|---|---|---|---|---|---|---|---|
| *OPE* | 39,558 | 0.98 | 0.00 | 0.00 | 0.98 | 0.98 | 0.98 | 1.00 |
| *GrowthOR* | 39,558 | 0.22 | 0.09 | 0.00 | 0.18 | 0.20 | 0.24 | 1.00 |
| *ORoA* | 39,558 | 0.65 | 0.06 | 0.00 | 0.62 | 0.63 | 0.65 | 1.00 |
| *ROA* | 39,558 | 0.67 | 0.07 | 0.00 | 0.63 | 0.65 | 0.69 | 1.00 |
| *ROE* | 39,558 | 0.80 | 0.06 | 0.00 | 0.78 | 0.79 | 0.83 | 1.00 |
| *InvCurrentRatio* | 39,558 | 0.09 | 0.08 | 0.00 | 0.04 | 0.08 | 0.12 | 1.00 |
| *InvAcidRatio* | 39,558 | 0.08 | 0.07 | 0.00 | 0.03 | 0.06 | 0.09 | 1.00 |
| *InvTO* | 39,558 | 0.03 | 0.07 | 0.00 | 0.00 | 0.01 | 0.02 | 1.00 |
| *Leverage* | 39,558 | 0.55 | 0.17 | 0.00 | 0.44 | 0.57 | 0.68 | 1.00 |
| *WCL* | 39,558 | 0.58 | 0.24 | 0.00 | 0.39 | 0.59 | 0.78 | 1.00 |
| *IBLiability* | 39,558 | 0.52 | 0.22 | 0.00 | 0.36 | 0.52 | 0.69 | 1.00 |
| *Cash Ratio* | 39,558 | 0.03 | 0.04 | 0.00 | 0.01 | 0.02 | 0.03 | 1.00 |
| *Cash Ratio2* | 39,558 | 0.04 | 0.06 | 0.00 | 0.02 | 0.03 | 0.05 | 1.00 |
| *Listed* | 39,558 | 0.30 | 0.46 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 |
| *OCF Ratio* | 39,558 | 0.19 | 0.01 | 0.00 | 0.19 | 0.19 | 0.19 | 1.00 |
| *FCF Ratio* | 39,558 | 0.75 | 0.02 | 0.00 | 0.74 | 0.75 | 0.76 | 1.00 |
| *ICF Ratio* | 39,558 | 0.16 | 0.01 | 0.00 | 0.15 | 0.16 | 0.16 | 1.00 |
| *Downgrade* | 39,558 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *Local* | 39,558 | 0.19 | 0.19 | 0.00 | 0.08 | 0.14 | 0.23 | 1.00 |
| *GDPPC* | 39,558 | 0.42 | 0.23 | 0.00 | 0.21 | 0.40 | 0.59 | 1.00 |
| *City* | 39,558 | 0.32 | 0.20 | 0.00 | 0.16 | 0.26 | 0.43 | 1.00 |
| *SB13* | 39,558 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *SB35* | 39,558 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *SB57* | 39,558 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *SB710* | 39,558 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *SB10+* | 39,558 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *BSBTA* | 39,558 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *#IB* | 39,558 | 0.01 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *#SB* | 39,558 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *SR* | 39,558 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *SB3TE* | 39,558 | 0.01 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *SB1TE* | 39,558 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *#SBTL* | 39,558 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *FCF Outflow* | 39,558 | 0.32 | 0.47 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 |
| *OCF Outflow* | 39,558 | 0.36 | 0.48 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 |
| *ICF Outflow* | 39,558 | 0.88 | 0.33 | 0.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| *External Guarantee* | 39,558 | 0.13 | 0.33 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *Guarantee* | 39,558 | 0.02 | 0.13 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *Chinese–Foreign Equity Joint Ventures* | 39,558 | 0.01 | 0.11 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *SOEs (Central)* | 39,558 | 0.11 | 0.31 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *SOEs (Local)* | 39,558 | 0.61 | 0.49 | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 |
| *Public ENTERPRISES* | 39,558 | 0.02 | 0.12 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *Foreign-Owned Enterprises* | 39,558 | 0.01 | 0.12 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *Foreign Enterprises* | 39,558 | 0.01 | 0.09 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *Private Enterprises* | 39,558 | 0.22 | 0.41 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |

(continued)

Table 3 (continued)

| | Count | Mean | SD | Min | P25 | P50 | p75 | Max |
|---|---|---|---|---|---|---|---|---|
| *Collective Enterprises* | 39,558 | 0.22 | 0.41 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *IT* | 39,558 | 0.04 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *Utilities* | 39,558 | 0.06 | 0.23 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *Health Care* | 39,558 | 0.03 | 0.16 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *Optional Consumption* | 39,558 | 0.09 | 0.28 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *Industrials* | 39,558 | 0.45 | 0.50 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 |
| *Real Estate* | 39,558 | 0.09 | 0.29 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *Consumer Staples* | 39,558 | 0.03 | 0.18 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *Material* | 39,558 | 0.11 | 0.32 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *Energy* | 39,558 | 0.04 | 0.19 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| *Finance* | 39,558 | 0.06 | 0.24 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |

The table shows the summary statistics for all the variables used in our analysis. All variables are winsorised at 1 and 99 percent. The variables in our prediction model and their definitions are shown in Table 2.

classified as default firms. FN denotes the number of firms misclassified as non-default firms. FP denotes the number of firms misclassified as default firms.

Using the information in the confusion matrix, we calculate the *Accuracy*, *Precision*, *Recall*, and *F_Score* of the model. *Accuracy* is a measure of the proportion of the total sample that correctly predicts the outcomes, calculated as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}. \tag{4}$$

*Precision* measures the proportion of real defaults in a sample that predicts defaults, calculated as:

$$Precision = \frac{TP}{TP + FP}. \tag{5}$$

*Recall* measures the proportion of real defaults that are predicted defaults in the sample, calculated as:

$$Recall = \frac{TP}{TP + FN}. \tag{6}$$

*F_score* is the harmonic mean of the *Recall* and the *Precision* rates, which can be used to compare the ML model's prediction accuracy comprehensively; the higher the value, the higher the prediction accuracy. The *F_score* formula is as follows:

Table 4
Default status categorised by training and testing sample

| | Default | Sample size | Default rate |
|---|---|---|---|
| **Panel A: Training and testing sample (Total)** | | | |
| Training | 131 | 27,690 | 0.47% |
| Testing | 71 | 11,868 | 0.59% |
| Testing group proportion | 30.00% | 35.15% | |
| **Panel B: Training and testing sample (Pre-reform)** | | | |
| Training | 14 | 7,714 | 0.18% |
| Testing | 2 | 3,306 | 0.06% |
| Testing group proportion | 12.50% | 30.00% | |
| **Panel C: Training and testing sample (Post-reform)** | | | |
| Training | 141 | 19,976 | 0.70% |
| Testing | 45 | 8,562 | 0.52% |
| Testing group proportion | 24.19% | 30.00% | |
| **Panel D: Training and testing sample (SOE)** | | | |
| Training | 22 | 20,249 | 0.11% |
| Testing | 8 | 8,679 | 0.09% |
| Testing group proportion | 26.67% | 30.00% | |
| **Panel E: Training and testing sample (Non-SOE)** | | | |
| Training | 106 | 7,441 | 1.40% |
| Testing | 66 | 3,189 | 2.02% |
| Testing group proportion | 38.37% | 30.00% | |
| **Panel F: Training and testing sample (Listed)** | | | |
| Training | 44 | 8,365 | 0.52% |
| Testing | 22 | 3,585 | 0.61% |
| Testing group proportion | 33.33% | 30.00% | |
| **Panel G: Training and testing sample (Non-listed)** | | | |
| Training | 95 | 19,325 | 0.49% |
| Testing | 43 | 8,283 | 0.52% |
| Testing group proportion | 31.16% | 30.00% | |
| **Panel H: Training and testing sample (Foreign)** | | | |
| Training | 19 | 982 | 1.90% |
| Testing | 8 | 421 | 1.86% |
| Testing Group Proportion | 29.63% | 30.00% | |
| **Panel I: Training and testing sample (Non-foreign)** | | | |
| Training | 125 | 26,708 | 0.47% |
| Testing | 50 | 11,447 | 0.43% |
| Testing group proportion | 28.57% | 30.00% | |

The table shows the number of realised default and non-default cases, the default rate, and tested group proportion within each sub-sample, training and testing groups.

$$F_{Score} = \frac{2Precision*Recall}{Precision + Recall}. \tag{7}$$

We depict the ROC curve whose $y$-axis is $TPR\ (Recall)$ and $x$-axis is $FPR$, calculated as:

Table 5
Confusion matrix

| | Non-default in prediction | Default in prediction |
|---|---|---|
| Non-default in reality | TN | FP |
| Default in reality | FN | TP |

The table shows the format of the confusion matrix. TN denotes the number of firms correctly classified as non-default firms. TP denotes the number of firms correctly classified as default firms. FN denotes the number of firms misclassified as non-default firms. FP denotes the number of firms misclassified as default firms.

$$TPR(Recall) = \frac{TP}{TP + FN}. \tag{8}$$

$$FPR = \frac{FP}{FP + TN}. \tag{9}$$

We then calculate the area under the ROC curve (denoted as $AUC$). For a given sample, we aim to maximise the proportion of firms that are correctly predicted to default and to minimise the proportion of firms that are mistakenly predicted to default but do not in reality. The higher the $AUC$, the better the predicted performance of the model. The $AUC$ shows the probability of the model ranking a random positive category sample (default) on top of a random negative category sample (non-default). The optimal ML model is selected by combining $AUC$ and *Recall*.

### *3.2. Ensembling model*

The traditional model usually selects one of the basic ML models to fit the data. This strategy may lead to problems if the best model is unknown. The model may lose some of the information in the data through an arbitrary selection of the models. To address this problem, we use the ensembling model method. For example, buying a car entails multiple decision-making processes. The purchaser can listen to friends' suggestions, discuss the matter with an expert, or use a search engine, and, by learning from different processes, finally selects a car. The ensembling process is similar to this learning process. By correctly assembling weak learning (friends' suggestions, expert suggestions and search engine recommendations) and learning from this process, the purchaser can finally make a decision. By considering multiple models, the ML model usually performs better than the traditional weak learning process, because it considers the model uncertainty without losing information.

Table 6 shows the performance of the ensembling model and our benchmark models. We first use the boosting and stacking models for analysis. The process

Table 6
Performance of ensembling models

| Ensembling method | Benchmark | | Boosting | | | Stacking | | | |
|---|---|---|---|---|---|---|---|---|---|
| Prediction performance criteria | SVM | KSVM | ABC | GBC | HGBC | Benchmark | Benchmark +ABC+GBC | Benchmark +RF+ET | Benchmark +ABC+GBC+RF+ET |
| **Panel A: Boosting and stacking models** | | | | | | | | | |
| *Accuracy* | 0.82 | 0.76 | 0.99 | 0.94 | 0.89 | 0.87 | 0.92 | 0.80 | 0.85 |
| *Precision* | 0.03 | 0.02 | 0.76 | 0.07 | 0.05 | 0.04 | 0.06 | 0.03 | 0.04 |
| *Recall* | 0.79 | 0.90 | 0.18 | 0.75 | 0.89 | 0.90 | 0.87 | 1.00 | 1.00 |
| *F_Score* | 0.05 | 0.04 | 0.30 | 0.14 | 0.09 | 0.07 | 0.12 | 0.06 | 0.07 |
| *AUC* | 88.72% | 91.75% | 97.53% | 94.83% | 94.96% | 94.63% | 96.36% | 97.67% | 97.89% |

| Ensembling method | Benchmark | | Bagging | | | | Special bagging models | |
|---|---|---|---|---|---|---|---|---|
| Prediction performance criteria | SVM | KSVM | SVM | KSVM | LR | MLP | RF(Bagging+DT) | ET (Bagging+DT) |
| **Panel B: Bagging models** | | | | | | | | |
| *Accuracy* | 0.82 | 0.76 | 0.05 | 0.61 | 0.82 | 0.78 | 0.77 | 0.84 |
| *Precision* | 0.03 | 0.02 | 0.01 | 0.01 | 0.03 | 0.03 | 0.03 | 0.04 |
| *Recall* | 0.79 | 0.90 | 1.00 | 0.97 | 0.87 | 0.94 | 1.00 | 0.96 |
| *F_Score* | 0.05 | 0.04 | 0.01 | 0.03 | 0.05 | 0.05 | 0.05 | 0.07 |
| *AUC* | 88.72% | 91.75% | 92.54% | 93.58% | 92.26% | 92.94% | 97.65% | 96.69% |

The table shows the predictive performance of multiple ensembling models. Panel A shows the benchmark, boosting and stacking models. The boosting model includes the ABC, GBC and HGBC models. The stacking model uses the SVM, MLPClassifier (MLP) with 1 hidden layer, LR and KSVM models as the benchmark stacking model, and further adds ABC, GBC, RF and ET to the stacking model. Panel B exhibits the benchmark models, SVM, KSVM, MLP, ABC and GBC models combined with the bagging classifier, and special bagging models, such as RF and ET.

of AdaBoost assigns the same initial weight to each sample. After each round of learner training, each sample's weight is adjusted according to its performance to increase the weight of the misclassified sample. In this way, more attention can be paid to samples with mistakes in the past. According to this process, learners are repeatedly trained and, finally, the weighted average of the learner is computed. The model needs to be weighted. Therefore, the SVM and neural network cannot be used as boosting models. Stacking regression is an integrated learning technique that combines multiple regression models with a meta-regressor. Moreover, each base regression model must use the complete training set when training, and the output of each base regression model in the integrated learning process becomes the input of the meta-regressor as the meta-feature. The meta-regressor ensembles the meta-features and uses them to predict the outcome. Panel A of Table 6 shows the model performance of the boosting and stacking models and Figures 2 and 3 shows their ROC curves.

All boosting models yield an AUC greater than 94 percent, which is far greater than the AUC of 88.72 and 91.75 percent for the benchmark models. The AUC of the ABC model is greater than 97 percent, indicating its good prediction power. All stacking models yield an AUC greater than 94 percent, which is greater than the AUC of 88.72 and 91.75 percent for the benchmark models. We also observe that the higher the number of models added to the stacking models, the higher the prediction power of the models. The final stacking model yields an AUC of 97.89 percent by combining the SVM, MLP, LR, KSVM, ABC, GBC, RF and ET models with a max voting classifier.

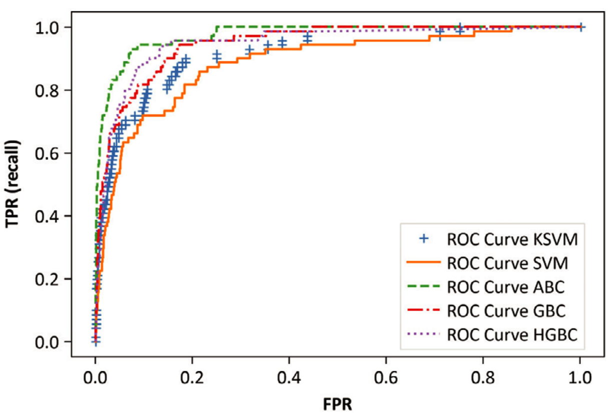



Figure 2 ROC curve of boosting models.This figure shows the ROC curve of the boosting models. This figure shows the ROC curve of ABC, GBC, HGBC, and the benchmark SVM and KSVM models. The *y*-axis indicates TPR (recall rate), which is the proportion of real defaults that are predicted to default. The *x*-axis is the FPR, which is the proportion of non-default firms that are falsely predicted as default firms. AUC is the area under the curve.

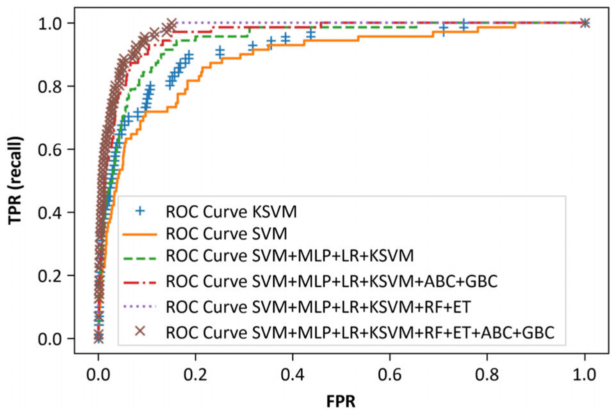



Figure 3 ROC curve of max voting classifier (stacking) models.This figure shows the ROC curve of the stacking model whose final prediction model is a max voting classifier. We use a max voting classifier combined with SVM, KSVM, LR and MLP classifiers as our benchmark for stacking models and further combined with the ABC, GBC, RF and ET models. The *y*-axis indicates TPR (recall rate), which is the proportion of real defaults that are predicted to default. The *x*-axis is the FPR, which is the proportion of non-default firms that are falsely predicted as default firms. AUC is the area under the curve.

We further consider using the bagging method to ensemble the models. Typical bagging methods are the RF and ET, random sampling from the training data set, training a different model separately, and ensembling all the training outcomes when predicting the outcomes. Panel B of Table 6 shows the model performance of bagging models and Figure 4 shows their ROC curves.

All bagging models have an AUC greater than 92 percent, which is greater than the AUC of 88.72 and 91.75 percent for the benchmark models. The table shows that in combination with the bagging classifier, the prediction power of our benchmark models increases. The model that combines SVM with the bagging classifier yields an AUC of 92.54 percent, and the model that combines the KSVM with the bagging classifier yields an AUC of 93.58 percent. We further test two bagging methods, RF and ET, which are based on the decision tree (DT) and bagging models. The RF model yields an AUC of 97.65 percent, and the ET model yielded an AUC of 96.69, indicating that it has good predictive power for corporate bond defaults.

### *3.3. Feature engineering*

Feature engineering allows us to extract important features from the original data and to use them as inputs for multiple ML models. We use the ABC, GBC, RF and ET models to extract important features and combine them with an LR model. Our model is similar to a stacking model. The difference is that a stacking model allows us to combine the output of multiple models. This model

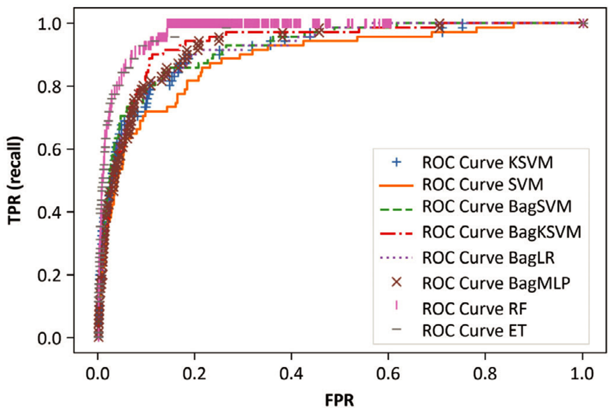



Figure 4 ROC curve of bagging models.This figure shows the ROC curve of the bagging and benchmark models. This figure shows the ROC curve of the bagging classifier combined with SVM, KSVM, LR, MLP and DT (RF and ET models) and the benchmark SVM and KSVM models. The *y*-axis indicates TPR (recall rate), which is the proportion of real defaults that are predicted to default. The *x*-axis is the FPR, which is the proportion of non-default firms that are falsely predicted as default firms. AUC is the area under the curve.

can combine only one model with an LR model. However, our model allows us to extract important features from the importance index, which provides further insight into what predicts the default. Figure 5 shows the ROC curve, and Table 7 shows the model performance.

The model combining ABC, GBC, RF and ET with the LR model performs better than our benchmark SVM and KSVM models. All models yield an AUC greater than 91 percent, indicating good prediction quality. Table 8 shows the top 10 important features of ML models for default prediction. The ABC model places a heavy weight on secured bonds under 3 years/total equity, while other influencing factors are not paramount with these factors. The remaining three models rank *InvTO*, *IBLiability*, *Leverage*, *WCL* and *ROA* in the top 10 important corporate bond default prediction factors covering liquidity risk, capital structure and firm profitability. Two of the models rank *ROE*, *InvAcid Ratio* and *GDPPA* in the top 10 important factors of corporate bond default prediction, showing that macroeconomic factors have strong power for default prediction.

### 3.4. *Cross-sectional heterogeneity*

#### 3.4.1. *Structural reform*

In 2015, considering the over-leveraged situation for most industrial firms, the Chinese government decided to cut excessive industrial capacity, reduce the leverage of firms, lower corporate costs, improve the weaknesses of enterprises,

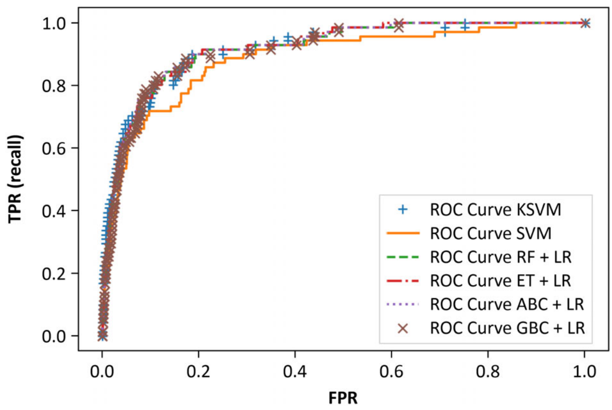



Figure 5 ROC curve of boosting and bagging models combined with logistic regression.This figure presents the ROC curve of the logistic regression model using the random forest for data selection, the RF model, and KSVM model. The *y*-axis indicates TPR (recall rate), which is the proportion of real defaults that are predicted to default. The *x*-axis is the FPR, which is the proportion of non-default firms that are falsely predicted as default firms. AUC is the area under the curve.

Table 7
Performance of feature engineering models

| Prediction performance criteria | SVM | KSVM | ABC+LR | GBC+LR | RF+LR | ET+LR |
|---|---|---|---|---|---|---|
| *Accuracy* | 0.82 | 0.76 | 0.82 | 0.81 | 0.82 | 0.82 |
| *Precision* | 0.03 | 0.02 | 0.03 | 0.03 | 0.03 | 0.03 |
| *Recall* | 0.79 | 0.90 | 0.86 | 0.89 | 0.86 | 0.87 |
| *F_Score* | 0.05 | 0.04 | 0.05 | 0.05 | 0.05 | 0.06 |
| *AUC* | 88.72% | 91.75% | 92.16% | 91.68% | 92.00% | 92.15% |

The table shows the predictive performance of logistic regression, after feature engineering by the ABC, GBC, RF and ET models.

and start a supply-side structural reform. Our study investigates the structural changes in corporate bond defaults before and after the supply-side structural reform. Panel C of Table 4 shows the default status for the training and testing groups of pre-reform firms.

We observe that the within-sample default rates are different between training and testing samples. The testing group proportion of the default group differs from that of the non-default group, and the testing group has only two default observations, imposing a huge challenge for our ML model's prediction power. Regardless of the challenge, most of our ML models maintain good performance. As shown in Panel A of Table 9 and Figure 6, the AUCs of all our benchmark models and bagging models are greater than 91 percent.

Table 8
Importance index of ABC, GBC, RF and ET models

| Models | ABC | | GBC | | RF | | ET | |
|---|---|---|---|---|---|---|---|---|
| Ranking | Variables | Importance | Variables | Importance | Variables | Importance | Variables | Importance |
| 1 | *SB3TE* | Inf | *InvTO* | 0.106 | *IBLiability* | 0.052 | *InvTO* | 0.045 |
| 2 | *Listed* | 0.008 | *IBLiability* | 0.096 | *WCL* | 0.051 | *Cash Ratio2* | 0.043 |
| 3 | *#IB* | 0.008 | *Leverage* | 0.093 | *InvTO* | 0.051 | *Cash Ratio* | 0.042 |
| 4 | *OCF Ratio* | 0.006 | *WCL* | 0.077 | *InvAcid Ratio* | 0.049 | *WCL* | 0.040 |
| 5 | *Chinese–Foreign Joint Venture* | 0.003 | *GDPPA* | 0.057 | *Cash Ratio2* | 0.047 | *GDPPA* | 0.039 |
| 6 | *Utilities* | 0.003 | *ROA* | 0.057 | *ROA* | 0.046 | *IBLiability* | 0.039 |
| 7 | *IT* | 0.002 | *GrowthOR* | 0.056 | *OROA* | 0.046 | *InvCurrent Ratio* | 0.038 |
| 8 | *Foreign-Owned Enterprise* | 0.001 | *ICF Ratio* | 0.055 | *ROE* | 0.046 | *Leverage* | 0.037 |
| 9 | *Consumer Staples* | 0.001 | *InvAcid Ratio* | 0.040 | *ICF Ratio* | 0.045 | *ROE* | 0.037 |
| 10 | *Health Care* | 0.001 | *OROA* | 0.038 | *Leverage* | 0.044 | *ROA* | 0.037 |

The table shows the ranking of the top 10 variables with the highest importance using multiple bagging and boosting models.

Table 9
Performance of ensembling models for corporate bond default before and after structural reform

| Ensembling method | Benchmark | | Boosting | | Bagging | |
|---|---|---|---|---|---|---|
| Prediction performance criteria | SVM | KSVM | ABC | GBC | RF | ET |
| **Panel A: Before structural reform** | | | | | | |
| *Accuracy* | 0.05 | 0.70 | 1.00 | 0.99 | 0.79 | 0.75 |
| *Precision* | 0.00 | 0.00 | 0.25 | 0.00 | 0.00 | 0.00 |
| *Recall* | 1.00 | 1.00 | 0.50 | 0.00 | 1.00 | 1.00 |
| *F_Score* | 0.00 | 0.00 | 0.33 | 0.00 | 0.01 | 0.00 |
| *AUC* | 93.30% | 98.71% | 74.95% | 24.89% | 91.21% | 99.50% |
| **Panel B: After structural reform** | | | | | | |
| *Accuracy* | 0.52 | 0.76 | 0.99 | 0.73 | 0.71 | 0.80 |
| *Precision* | 0.01 | 0.02 | 0.75 | 0.02 | 0.02 | 0.03 |
| *Recall* | 0.93 | 0.93 | 0.07 | 0.89 | 1.00 | 1.00 |
| *F_Score* | 0.02 | 0.04 | 0.12 | 0.03 | 0.03 | 0.05 |
| *AUC* | 84.46% | 90.52% | 97.53% | 90.74% | 96.97% | 97.74% |
| **Panel C: State-owned enterprises (SOEs)** | | | | | | |
| *Accuracy* | 0.00 | 0.68 | 1.00 | 0.90 | 0.78 | 0.73 |
| *Precision* | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| *Recall* | 1.00 | 0.88 | 0.00 | 0.50 | 1.00 | 0.88 |
| *F_Score* | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 |
| *AUC* | 91.99% | 92.35% | 49.92% | 71.46% | 96.11% | 87.74% |
| **Panel D: Non-state-owned enterprises (Non-SOEs)** | | | | | | |
| *Accuracy* | 0.02 | 0.73 | 0.98 | 0.60 | 0.69 | 0.74 |
| *Precision* | 0.02 | 0.06 | 0.60 | 0.04 | 0.06 | 0.06 |
| *Recall* | 1.00 | 0.83 | 0.14 | 0.77 | 0.89 | 0.85 |
| *F_Score* | 0.04 | 0.11 | 0.22 | 0.07 | 0.11 | 0.12 |
| *AUC* | 81.15% | 83.23% | 91.16% | 76.71% | 88.40% | 89.59% |
| **Panel E: Listed** | | | | | | |
| *Accuracy* | 0.42 | 0.77 | 0.99 | 0.79 | 0.69 | 0.70 |
| *Precision* | 0.01 | 0.02 | 0.08 | 0.02 | 0.02 | 0.02 |
| *Recall* | 0.95 | 1.00 | 0.05 | 0.80 | 0.05 | 1.00 |
| *F_Score* | 0.02 | 0.05 | 0.06 | 0.04 | 0.03 | 0.04 |
| *AUC* | 93.68% | 96.14% | 52.33% | 87.81% | 94.18% | 97.37% |
| **Panel F: Non-listed** | | | | | | |
| *Accuracy* | 0.26 | 0.55 | 0.99 | 0.80 | 0.71 | 0.80 |
| *Precision* | 0.01 | 0.01 | 0.33 | 0.02 | 0.02 | 0.02 |
| *Recall* | 0.98 | 0.98 | 0.02 | 0.88 | 1.00 | 1.00 |
| *F_Score* | 0.01 | 0.02 | 0.04 | 0.04 | 0.03 | 0.05 |
| *AUC* | 88.85% | 94.87% | 97.15% | 92.61% | 97.65% | 98.26% |
| **Panel G: With foreign investment** | | | | | | |
| *Accuracy* | 0.29 | 0.79 | 0.98 | 0.85 | 0.72 | 0.78 |
| *Precision* | 0.03 | 0.08 | 0.46 | 0.11 | 0.06 | 0.08 |
| *Recall* | 1.00 | 1.00 | 0.75 | 1.00 | 1.00 | 1.00 |
| *F_Score* | 0.05 | 0.15 | 0.57 | 0.21 | 0.12 | 0.15 |
| *AUC* | 97.12% | 99.30% | 86.65% | 97.44% | 99.94% | 100% |
| **Panel H: Without foreign investment** | | | | | | |

(continued)

Table 9 (continued)

| Ensembling method<br>Prediction performance criteria | Benchmark | | Boosting | | Bagging | |
|---|---|---|---|---|---|---|
| | SVM | KSVM | ABC | GBC | RF | ET |
| *Accuracy* | 0.58 | 0.78 | 0.97 | 0.76 | 0.78 | 0.84 |
| *Precision* | 0.01 | 0.02 | 0.10 | 0.02 | 0.02 | 0.02 |
| *Recall* | 0.98 | 0.86 | 0.72 | 0.88 | 0.92 | 0.94 |
| *F_Score* | 0.02 | 0.03 | 0.17 | 0.03 | 0.04 | 0.05 |
| *AUC* | 89.61% | 90.31% | 94.26% | 90.77% | 95.38% | 96.36% |

The table shows the predictive performance of multiple ensembling models. Panel A exhibits the benchmark, boosting and stacking models. The boosting model includes the ABC, GBC and HGBC models. The stacking model uses the SVM, MLPClassifier (MLP) with 1 hidden layer, LR and KSVM models as the benchmark stacking model, and further adds ABC, GBC, RF and ET to the stacking model. We combine the bagging classifier with the SVM, KSVM, MLP, ABC and GBC models.

The boosting model performance is less satisfactory, but the AUC of the ABC model is greater than 70 percent. Considering that the intragroup default rate and testing group proportion are imbalanced, the outcome is acceptable. Panel A of Table 10 shows the importance index for our models for pre-reform firms' default prediction.

Four models ranked the *OROA* and *OCF ratio*, and three models ranked the *FCF ratio* and *ROE* as important factors of corporate bond default. This shows that operational risk and firm cash flow management were the key factors in corporate bond default prediction before the structural reform. The two models ranked the *ICF ratio*, *ROA*, *IBLiability*, *Leverage*, *Downgrade*, *Cash Ratio*, *SOE (Local)* and *City* in the top 10 important factors of corporate bond default. The results show that cash flow management, capital structure, credit rating, liquidity risk, firm ownership and macroeconomic factors have strong power to predict corporate bond default.

For post-reform firms, Panel D of Table 4 shows the default status for the training and testing groups of pre-reform firms. The testing group proportion and intragroup default status are similar. Additionally, as shown in Panel B of Table 9 and Figure 6, all of our models have an AUC greater than 84 percent, showing their good prediction power for corporate bond defaults.

We observe that the SVM and KSVM models perform well in the pre-reform sample but less so in the post-reform period. On the contrary, the boosting models perform well in the post-reform period, but less so in the pre-reform period. This finding indicates that our benchmark performs well in tackling the imbalance sample. Furthermore, in both the pre-reform and post-reform samples, our bagging models RF and ET yield good results with an AUC greater than 90 percent. This indicates that the bagging method may

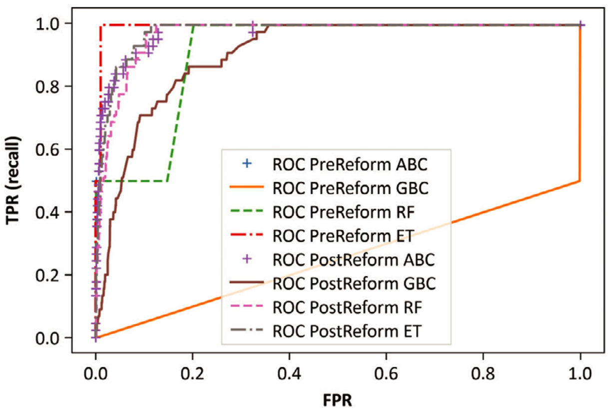



Figure 6 ROC curve of firms before and after structural reform.This figure shows the ROC curve of the ABC, GBC, RF and ET models for firms before and after structural reform. The *y*-axis indicates TPR (recall rate), which is the proportion of real defaults that are predicted to default. The *x*-axis is the FPR, which is the proportion of non-default firms that are falsely predicted as default firms. AUC is the area under the curve.

have good universality and robustness for different samples. Panel B of Table 10 shows the importance index for our models for post-reform firm default prediction.

The three models rank *IBLiability*, *InvTO* and *WCL* among the top 10 important factors for corporate bond default. The results show that capital structure, operational risk and liquidity risk are strong predictors of corporate bond defaults in the post-reform period in China. The two models rank *InvAcid Ratio*, *Cash Ratio2*, *OROA*, *ICF Ratio*, *Leverage*, *Downgrade*, *GDPPA*, *Local* and *Private Enterprises* as the top 10 important factors of corporate bond default prediction. The results show that liquidity risk, operational risk, cash flow management, capital structure, credit rating, macroeconomic factors and firm ownership structure have strong predictive power for corporate bond defaults.

### *3.4.2. State-owned enterprises*

China has the largest group of SOEs globally. As Table 1 shows, SOEs comprise 66 percent of our sample. What is the difference between SOEs and non-SOEs in terms of predicting default risk? Panel D of Table 4 shows the default status for the training and testing groups of SOEs. The testing group proportion and intragroup default status are similar. As shown in Panel C of Table 9 and Figure 7, our benchmark models and bagging models all have an AUC greater than 80 percent.

Table 10
Importance of sub-sample analysis

| Models | ABC | | GBC | | RF | | ET | |
|---|---|---|---|---|---|---|---|---|
| Ranking | Variables | Importance | Variables | Importance | Variables | Importance | Variables | Importance |
| **Panel A: Before structural reform** | | | | | | | | |
| 1 | *ICF Ratio* | 0.198 | *OROA* | 0.434 | *OROA* | 0.116 | *ROE* | 0.092 |
| 2 | *OROA* | 0.166 | *IBLiability* | 0.216 | *ROA* | 0.073 | *Downgrade* | 0.075 |
| 3 | *GrowthOR* | 0.119 | *Leverage* | 0.178 | *ICF Ratio* | 0.059 | *OROA* | 0.062 |
| 4 | *OCF Ratio* | 0.093 | *Downgrade* | 0.026 | *ROE* | 0.058 | *OCF Ratio* | 0.047 |
| 5 | *Foreign-Owned Enterprise* | 0.068 | *OCF Ratio* | 0.020 | *OPE* | 0.051 | *SOE (Local)* | 0.040 |
| 6 | *ROA* | 0.056 | *FCF Ratio* | 0.013 | *IBLiability* | 0.050 | *City* | 0.039 |
| 7 | *InvAcid Ratio* | 0.048 | *City* | 0.013 | *Cash Ratio2* | 0.044 | *WCL* | 0.036 |
| 8 | *Local* | 0.048 | *GDPPA* | 0.013 | *Leverage* | 0.041 | *Private Enterprises* | 0.036 |
| 9 | *IT* | 0.048 | *ROE* | 0.013 | *OCF Ratio* | 0.039 | *Cash Ratio* | 0.034 |
| 10 | *SOE (Local)* | 0.029 | *Cash Ratio* | 0.013 | *FCF Ratio* | 0.039 | *FCF Ratio* | 0.034 |
| **Panel B: After structural reform** | | | | | | | | |
| 1 | *Material* | 0.012 | *Downgrade* | 0.332 | *IBLeverage* | 0.077 | *InvTO* | 0.046 |
| 2 | *FCF Ratio* | 0.005 | *Local* | 0.157 | *InvTO* | 0.057 | *IBLeverage* | 0.044 |
| 3 | *Optional Consumption* | 0.005 | *IBLiability* | 0.136 | *WCL* | 0.053 | *WCL* | 0.038 |
| 4 | *ICF Ratio* | 0.005 | *Private Enterprises* | 0.079 | *OROA* | 0.049 | *GDPPA* | 0.035 |
| 5 | *Consumption Staples* | 0.003 | *Chinese–Foreign Joint Venture* | 0.077 | *Cash Ratio2* | 0.047 | *Cash Ratio2* | 0.034 |
| 6 | *Foreign-Owned Enterprise* | 0.003 | *InvTO* | 0.050 | *Leverage* | 0.043 | *Leverage* | 0.033 |
| 7 | *Listed* | 0.003 | *GDPPA* | 0.033 | *ICF Ratio* | 0.043 | *City* | 0.032 |
| 8 | *Health Care* | 0.002 | *WCL* | 0.033 | *ROE* | 0.041 | *Downgrade* | 0.031 |
| 9 | *#SB* | 0.001 | *InvAcid Ratio* | 0.025 | *InvAcid Ratio* | 0.041 | *Local* | 0.031 |
| 10 | *IT* | 0.001 | *SOE (Local)* | 0.021 | *Private Enterprises* | 0.041 | *OROA* | 0.031 |

(continued)

Table 10 (continued)

| Models | ABC | | GBC | | RF | | ET | |
|---|---|---|---|---|---|---|---|---|
| Ranking | Variables | Importance | Variables | Importance | Variables | Importance | Variables | Importance |
| **Panel C: State-owned enterprises (SOEs)** | | | | | | | | |
| 1 | *ICF Ratio* | 0.147 | *OROA* | 0.229 | *IBLeverage* | 0.128 | *Energy* | 0.099 |
| 2 | *ROA* | 0.144 | *Downgrade* | 0.163 | *Leverage* | 0.099 | *Leverage* | 0.098 |
| 3 | *GrowthOR* | 0.105 | *Leverage* | 0.134 | *ICF Ratio* | 0.081 | *IBLeverage* | 0.069 |
| 5 | *OPE* | 0.087 | *IBLeverage* | 0.085 | *ROE* | 0.056 | *ROE* | 0.055 |
| 6 | *Cash Ratio* | 0.056 | *ROE* | 0.076 | *GDPPA* | 0.053 | *WCL* | 0.051 |
| 7 | *OROA* | 0.055 | *ICF Ratio* | 0.034 | *OCF Ratio* | 0.047 | *External Guarantee* | 0.046 |
| 8 | *IBLeverage* | 0.055 | *BSBTA* | 0.028 | *InvCurrent Ratio* | 0.045 | *GDPPA* | 0.041 |
| 9 | *BSBTA* | 0.053 | *Cash Ratio* | 0.023 | *FCF Ratio* | 0.045 | *Local* | 0.031 |
| 10 | *Leverage* | 0.046 | *FCF Ratio* | 0.021 | *OROA* | 0.045 | *FCF Ratio* | 0.031 |
| **Panel D: Non-state-owned enterprises (Non-SOEs)** | | | | | | | | |
| 1 | *Optional consumption* | 0.078 | *Downgrade* | 0.239 | *IBLeverage* | 0.056 | *InvTO* | 0.048 |
| 2 | *Private Enterprises* | 0.027 | *OROA* | 0.077 | *InvTO* | 0.051 | *Downgrade* | 0.043 |
| 3 | *Cash Ratio2* | 0.019 | *SR* | 0.065 | *WCL* | 0.049 | *Cash Ratio2* | 0.041 |
| 4 | *GR* | 0.008 | *WCL* | 0.053 | *OROA* | 0.048 | *WCL* | 0.040 |
| 5 | *#SB* | 0.007 | *GDPPA* | 0.052 | *InvAcid Ratio* | 0.047 | *IBLeverage* | 0.040 |
| 6 | *Foreign-Owned Enterprise* | 0.004 | *InvCurrent Ratio* | 0.046 | *ROE* | 0.046 | *GDPPA* | 0.038 |
| 7 | *#IB* | 0.003 | *ROE* | 0.044 | *InvCurrent Ratio* | 0.046 | *InvCurrent Ratio* | 0.037 |
| 8 | *Energy* | 0.001 | *OCF Ratio* | 0.042 | *Cash Ratio2* | 0.045 | *Cash Ratio* | 0.037 |
| 9 | *Finance* | 0.000 | *InvTO* | 0.042 | *Cash Ratio* | 0.044 | *ROE* | 0.037 |
| 10 | *Downgrade* | 0.000 | *Cash Ratio* | 0.038 | *GrowthOR* | 0.043 | *GrowthOR* | 0.035 |
| **Panel E: Listed** | | | | | | | | |
| 1 | *Leverage* | 0.127 | *OROA* | 0.221 | *GrowthOR* | 0.062 | *WCL* | 0.065 |
| 2 | *OROA* | 0.098 | *IBLeverage* | 0.131 | *OROA* | 0.060 | *Private Enterprises* | 0.057 |
| 3 | *InvAcid Ratio* | 0.096 | *Downgrade* | 0.130 | *OCF Ratio* | 0.056 | *IT* | 0.045 |
| 4 | *Cash Ratio2* | 0.090 | *GrowthOR* | 0.103 | *WCL* | 0.056 | *GDPPA* | 0.044 |

(continued)

Table 10 (continued)

| Models | ABC | | GBC | | RF | | ET | |
|---|---|---|---|---|---|---|---|---|
| Ranking | Variables | Importance | Variables | Importance | Variables | Importance | Variables | Importance |
| 5 | *FCF Ratio* | 0.084 | *Leverage* | 0.101 | *Cash Ratio2* | 0.055 | *ROE* | 0.038 |
| 6 | *GrowthOR* | 0.074 | *InvAcid Ratio* | 0.067 | *ROE* | 0.054 | *OCF Outflow* | 0.036 |
| 7 | *IBLeverage* | 0.060 | *WCL* | 0.048 | *FCF Ratio* | 0.050 | *Energy* | 0.035 |
| 8 | *ICF Ratio* | 0.057 | *ICF Ratio* | 0.022 | *Cash Ratio* | 0.048 | *SOE (Local)* | 0.031 |
| 9 | *Cash RATIO* | 0.050 | *Local* | 0.019 | *ROA* | 0.047 | *IBLeverage* | 0.031 |
| 10 | *OCF Ratio* | 0.048 | *OCF Ratio* | 0.013 | *GDPPA* | 0.046 | *Optional Consumption* | 0.030 |
| **Panel F: Non-listed** | | | | | | | | |
| 1 | *WCL* | 0.038 | *Cash Ratio* | 0.095 | *InvTO* | 0.053 | *Cash Ratio2* | 0.047 |
| 2 | *Private Enterprises* | 0.017 | *IBLeverage* | 0.094 | *Cash Ratio2* | 0.051 | *InvTO* | 0.046 |
| 3 | *Foreign-Owned Enterprise* | 0.012 | *Downgrade* | 0.079 | *IBLeverage* | 0.051 | *Cash Ratio* | 0.045 |
| 4 | *SOE (Local)* | 0.011 | *Private Enterprises* | 0.079 | *Leverage* | 0.051 | *Leverage* | 0.040 |
| 5 | *Foreign-Owned Enterprise* | 0.004 | *ICF Ratio* | 0.060 | *ROA* | 0.049 | *GDPPA* | 0.040 |
| 6 | *SOE (Central)* | 0.004 | *#SB* | 0.056 | *Cash Ratio* | 0.047 | *WCL* | 0.038 |
| 7 | *OCF Outflow* | 0.004 | *Leverage* | 0.055 | *InvCurrent Ratio* | 0.047 | *IBLeverage* | 0.037 |
| 8 | *Energy* | 0.002 | *Cash Ratio2* | 0.052 | *ROE* | 0.045 | *ROA* | 0.036 |
| 9 | *Chinese–Foreign Joint Venture* | 0.002 | *InvAcid Ratio* | 0.050 | *WCL* | 0.044 | *ROE* | 0.036 |
| 10 | *Optional Consumption* | 0.002 | *GDPPA* | 0.045 | *InvAcid Ratio* | 0.042 | *InvCurrent Ratio* | 0.034 |
| **Panel G: With foreign investment** | | | | | | | | |
| 1 | *InvTO* | 0.119 | *GR* | 0.217 | *Cash Ratio2* | 0.078 | *Cash Ratio2* | 0.049 |
| 2 | *Industrial* | 0.101 | *Cash Ratio2* | 0.102 | *InvTO* | 0.055 | *City* | 0.044 |
| 3 | *Cash RATIO* | 0.098 | *OROA* | 0.091 | *ROA* | 0.055 | *Cash Ratio* | 0.044 |
| 4 | *GrowthOR* | 0.086 | *InvAcid Ratio* | 0.082 | *InvAcid Ratio* | 0.054 | *WCL* | 0.042 |

(continued)

Table 10 (continued)

| Models Ranking | ABC Variables | ABC Importance | GBC Variables | GBC Importance | RF Variables | RF Importance | ET Variables | ET Importance |
|---|---|---|---|---|---|---|---|---|
| 5 | *WCL* | 0.077 | *ROA* | 0.079 | *WCL* | 0.052 | *Leverage* | 0.041 |
| 6 | *Leverage* | 0.067 | *InvTO* | 0.055 | *OCF Ratio* | 0.049 | *Local* | 0.037 |
| 7 | *SB3TE* | 0.062 | *City* | 0.055 | *InvCurrent Ratio* | 0.049 | *ROE* | 0.036 |
| 8 | *Cash Ratio2* | 0.055 | *SB3TE* | 0.052 | *Leverage* | 0.048 | *OROA* | 0.032 |
| 9 | *FCF Ratio* | 0.054 | *IBLerverage* | 0.042 | *ROE* | 0.045 | *GDPPA* | 0.032 |
| 10 | *InvAcid ratio* | 0.051 | *InvCurrent Ratio* | 0.040 | *GR* | 0.043 | *ICF Ratio* | 0.032 |
| **Panel H: Without foreign investment** | | | | | | | | |
| 1 | *#SB* | Inf | *OROA* | 0.148 | *IBLeverage* | 0.071 | *InvTO* | 0.049 |
| 2 | *Optional Consumption* | 0.006 | *IBLeverage* | 0.145 | *WCL* | 0.054 | *IBLeverage* | 0.040 |
| 3 | *SOE (Local)* | 0.006 | *Private enterprises* | 0.123 | *InvTO* | 0.052 | *Cash Ratio2* | 0.040 |
| 4 | *GR* | 0.005 | *Downgrade* | 0.123 | *ROA* | 0.050 | *ROE* | 0.039 |
| 5 | *SB1TE* | 0.005 | *Leverage* | 0.055 | *InvAcid Ratio* | 0.048 | *OROA* | 0.039 |
| 6 | *Consumption Staples* | 0.005 | *InvCurrent ratio* | 0.049 | *ROE* | 0.048 | *Cash Ratio* | 0.037 |
| 7 | *Utility* | 0.003 | *WCL* | 0.044 | *Cash Ratio2* | 0.048 | *OPE* | 0.037 |
| 8 | *Private Enterprises* | 0.003 | *GDPPA* | 0.042 | *OROA* | 0.046 | *InvCurrent Ratio* | 0.036 |
| 9 | *OCF Outflow* | 0.003 | *Local* | 0.031 | *Leverage* | 0.045 | *GDPPA* | 0.036 |
| 10 | *ICF Outflow* | 0.003 | *ROE* | 0.029 | *ICF Ratio* | 0.043 | *ROA* | 0.036 |

The table shows the ranking of the top 10 variables with the highest importance using multiple bagging and boosting models within different sample specifications.

The boosting model performance is less satisfactory, but the AUC of the GBC model is greater than 70 percent. Considering that we had few sample observations, the outcome is acceptable. Panel C of Table 10 shows the importance index for our models for SOE default prediction.

Four models ranked *IBLiability* and *Leverage* among the top 10 important factors for corporate bond default. The results show that capital structure is a strong predictor of corporate bond defaults for SOEs in China. Three models ranked *ICF Ratio*, *OROA*, *ROE* and *FCF Ratio* among the top 10 important corporate bond default prediction factors. The results show that operational risk and cash flow management have strong predictive power for corporate bond defaults. The two models ranked *BGBTA*, *Cash Ratio* and *GDPPA* among the top 10 important corporate bond default prediction factors. The results show that liquidity risk, capital structure and regional development are strong predictors of corporate bond defaults by SOEs in China.

For non-SOEs, Panel E of Table 4 shows the default status for the training and testing groups of pre-reform firms. The testing group proportion and intragroup default status are similar. Additionally, as shown in Panel D of Table 9 and Figure 7, all our models have an AUC greater than 80 percent.

For both SOEs and non-SOEs, our bagging models RF and ET yield good results with an AUC greater than 90 percent. This indicates that the bagging method may have good universality and robustness for different samples. Panel D of Table 10 shows the importance index for our models for post-reform firm default prediction.

Three models ranked *Cash Ratio2*, *Cash Ratio*, *WCL*, *ROE*, *InvTO* and *Downgrade* among the top 10 important corporate bond default prediction factors for non-SOEs. The results show that non-SOE corporate bond defaults may largely be explained by a lack of liquidity and poor firm performance. The two models ranked *SR*, *OROA*, *GrowthOR*, *InvCurrent Ratio* and *GDPPA* among the top 10 important corporate bond default prediction factors. The results further strengthen the argument in this section, and show that regional development and capital structure are strong predictors of corporate bond default for non-SOEs in China.

### 3.4.3. *Listed and non-listed firms*

Our study investigates the default status of listed and non-listed firms. Panel F of Table 4 shows the default status for the training and testing groups of listed firms. The testing group proportion and intragroup default status are similar. As shown in Panel E of Table 9 and Figure 8, our benchmark models and bagging models all have an AUC greater than 90 percent.

The boosting model performance is less satisfactory, but the AUC of the GBC model is 87.81 percent. Considering that there were few observations in our sample, the outcome is acceptable. Panel E of Table 10 shows the importance index for our models for predicting listed firms' defaults.

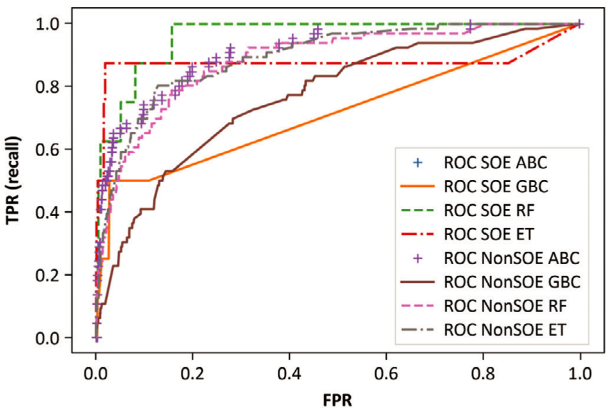



Figure 7 ROC curve of enterprises with SOEs and non-SOEs.This figure shows the ROC curve of the ABC, GBC, RF and ET models for SOEs and non-SOEs. The *y*-axis indicates TPR (recall rate), which is the proportion of real defaults that are predicted to default. The *x*-axis is the FPR, which is the proportion of non-default firms that are falsely predicted as default firms. AUC is the area under the curve.

The three models ranked *OCF Ratio*, *OROA*, *GrowthOR*, *IBLeverage* and *WCL* among the top 10 important corporate bond default prediction factors. The results show that operational risk and cash flow management have strong predictive power for listed firm defaults. The two models ranked *Leverage*, *InvAcid Ratio*, *Cash Ratio*, *Cash Ratio2*, *FCF ratio*, *ICF Ratio*, *ROE* and *GDPPA* among the top 10 important corporate bond default prediction factors. The results show that liquidity risk, cash flow management, profitability and regional development are strong predictors of corporate bond default for listed firms in China.

For non-listed firms, Panel G of Table 4 shows the default status for the training and testing groups of non-listed firms. The testing group proportion and intragroup default status are similar. Additionally, as shown in Panel F of Table 9 and Figure 8, all our models have an AUC greater than 80 percent.

For both the listed and non-listed samples, our bagging models RF and ET yield good results with an AUC greater than 90 percent. The results indicate that the bagging method may have good universality and robustness for different samples. Panel F of Table 10 shows the importance index for our models for non-listed firm default prediction.

The three models ranked *Cash Ratio2*, *Cash Ratio*, *WCL*, *IBLeverage* and *Leverage* among the top 10 important corporate bond default prediction factors for non-listed firms. The results show that non-listed firms' corporate bond defaults may largely be explained by a lack of liquidity and over-leveraging. The two models ranked *InvCurrent Ratio*, *InvAcid Ratio*, *GDPPA*, *InvTO*, *ROA*, *ROE* and *GDPPA* among the top 10 important corporate bond

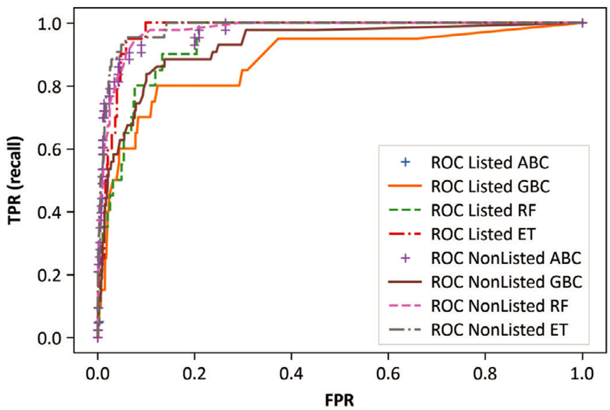



Figure 8 ROC curve of listed enterprises and non-listed enterprises.This figure shows the ROC curve of the ABC, GBC, RF and ET models for listed and non-listed firms. The $y$-axis indicates TPR (recall rate), which is the proportion of real defaults that are predicted to default. The $x$-axis is the FPR, which is the proportion of non-default firms that are falsely predicted as default firms. AUC is the area under the curve.

default prediction factors. The results further strengthen the argument in this section and show that regional development, firm profitability and operating risk are strong predictors of corporate bond default for non-listed firms in China.

*3.4.4. Foreign and non-foreign firms*

Our study investigates the default status of foreign and non-foreign firms. Panel H of Table 4 shows the default status for the training and testing groups of foreign firms. The testing group proportion and intragroup default status are similar. As shown in Panel G of Table 9 and Figure 9, our benchmark models, boosting and bagging models all have an AUC greater than 85 percent. Additionally, our ET model yields a 100 percent AUC, suggesting that it perfectly predicts corporate bond defaults of foreign firms in China.

Panel G of Table 10 shows the importance index for our models for foreign firms' default prediction. Four models ranked *Cash Ratio2* among the top 10 important factors for corporate bond default prediction. The results show that firm liquidity strongly predicts listed firms' default. Three models ranked *InvTO*, *WCL* and *InvAcid Ratio* among the top 10 important corporate bond default prediction factors. The results show that liquidity risk and operating risk are strong predictors of corporate bond defaults for listed firms in China. The two models ranked *InvCurrent Ratio*, *Cash Ratio*, *Leverage*, *GB3TE*, *SR*, *OROA*, *ROA*, *ROE* and *City* among the top 10 important factors of corporate bond default prediction. The results show that liquidity risk, capital structure, profitability and local government financial status are strong predictors of corporate bond defaults for foreign firms in China.

For non-foreign firms, Panel I of Table 4 shows the default status for the training and testing groups of non-foreign firms. The testing group proportion and intragroup default status are similar. Additionally, as shown in Panel H of Table 9 and Figure 9, all our models have an AUC greater than 80 percent.

For both listed and non-foreign samples, our bagging models RF and ET yield good results with an AUC greater than 90 percent. This indicates that the bagging method may have good universality and robustness for different samples.

Panel H of Table 10 shows the importance index for our models for non-foreign firm default prediction. Three models ranked *OROA*, *ROE* and *IBLeverage* among the top 10 important corporate bond default prediction factors for non-foreign firms. The results show that non-foreign firms' corporate bond defaults may largely be explained by operating risk and capital structure. The two models ranked *Leverage*, *InvCurrent Ratio*, *Cash Ratio2*, *WCL*, *InvTO*, *ROA* and *GDPPA* among the top 10 important corporate bond default prediction factors. The results further strengthen the argument in this section and show that regional development and liquidity risk are strong predictors of corporate bond defaults for non-foreign firms in China.

### 3.4.5. Industry heterogeneity

Regarding industry heterogeneity, we provide a model for representative industries (information technology, materials, energy and industrial).

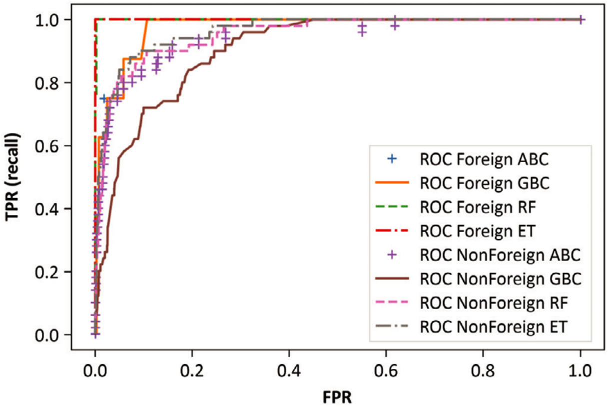



Figure 9 ROC curve of enterprises with and without foreign investment.This figure shows the ROC curve of the ABC, GBC, RF and ET models for firms with and without foreign investment. The *y*-axis indicates TPR (recall rate), which is the proportion of real defaults that are predicted to default. The *x*-axis is the FPR, which is the proportion of non-default firms that are falsely predicted as default firms. AUC is the area under the curve.

The solvency and liquidity risk and credit rating downgrading for each type of industry are found to have a significant positive prediction power on the default rate, and this conclusion is consistent with the results of the main sample. However, there are differences in the adverse prediction criteria for the default rates in different industries. Among them, the accessibility of external finance and *FCF* in the material industry and the energy industry's profitability are effective negative criteria for default rate prediction.

### 3.4.6. Traditional credit risk model

We further analyse the performance of traditional credit risk models. Reduced-form models are usually used for analysing the default status of asset portfolios. They assume that the default status of individual firms is unpredictable and each default event is independent and exogenous. They assume that the default probability satisfies a certain distribution, and it can be used to calculate the default probability of a portfolio. Our paper focuses on analysing the default status of the individual entity or the default event. Thus, it is not suitable to use a reduced-form model and compare its performance with our models.

In contrast, structure models are more suitable as the benchmark of our models, as they assume that the probability of default is endogenously determined by firm performance. The widely used structure models include CreditMatrics, IRB models, Merton distance-to-default, and the KMV model. The input to the CreditMatrics and IRB models require loss-given default data, which is not public disclosed. We only provide the result for the KMV model as it extends Merton's model by considering the debt structure of the firms. Details of the KMV model are presented in Appendix II.

The KMV model is established on the availability of stock market data, and we have 6,684 firm-year observations that have corresponding distance-to-default measures. The default status and the model performance matrix are shown in Table 11.

If the EDF of DD is greater than 0.5, we consider it to have a high probability of default. The distance-to-default model estimates that there are 461 annual default events during our sample period. However, the realised default event number is 0. The recall rate of the KMV model is 0.00 with AUC also at 0, indicating its poor prediction ability. We further increase the shed hold of default EDF to 0.95. The predicted number of defaults based on the KMV model remains unchanged. The KMV model overestimates the default status of Chinese firms. Multiple explanations exist for this phenomenon. Firstly, in China, getting publicly listed is difficult. Once the firm goes public, it is unlikely for them to be financially constrained. Access to the public market is a strong signal for the corporate bond investor, who will continue offering money unless the firm has defaulted. The publicly listed firm can quickly leverage up to a high level, even if it cannot pay the principal of the debt. Firms can continue borrowing new short-term debt to finance their corporate bond. Secondly, large listed firms, no matter whether an SOE or a non-SOE, usually create various job opportunities for local government. Sometimes the firm can even hijack the local government to provide subsidiaries for maintaining the stability of the financial market and local economic growth. The default of the listed non-SOE Evergrande is the typical example. On 24 September 2020, Evergrande urged the Guangdong government to support major asset restructuring. The firm threatened the government that if the restructuring could not complete on schedule, the firm would be bankrupt and threatened that its bankruptcy could bring in huge systemic risks to the financial market. Because of its mature bankruptcy law and the political nature, this situation is unlikely to occur in the United States. Firms may have a hidden buffer from local government financial support for corporate bond default.

### *3.4.7. Robustness test*

We analyse multiple training and testing groups using different random seeds to generate different random training and testing groups. The results show that solvency risk and credit downgrading are effective positive indicators of corporate default rate forecasts, and the profitability and accessibility of external finance are effective negative indicators of corporate default rate prediction. Operating ability and cash flow management ability are not strong predictors of the default rate for our main sample. The variables selected in the model may vary within the primary categories. For instance, some of the random samples choose *IBLiability* as the variable that enters the model, while others choose *Leverage*. The outcome is that some of the variables are strongly correlated with other variables with similar information content and importance for credit risk prediction. This variation does not affect the accuracy and interpretability of the prediction. Moreover, independent of the drawn training set, the recall rate and AUC are above 85 and 90 percent, respectively. The robustness results indicate the model's strong predictive power.

## 4. Conclusion

In 2018, China's corporate bond market experienced an unexpected historical wave of corporate bond default, which was difficult to predict. Hence, an important area of finance research is developing effective models for corporate bond default prediction, which limits the damage of default. This study aims to develop a new out-of-sample default prediction model based on a sample of public and private firms over the period January 2012 to December 2019.

In adherence to existing research, we use only readily available financial data as input in fraud prediction. However, we depart from most existing research in

Table 11
Performance of KMV model

| Default point | W0.5 Predicteddefault | W0.5 Predictednon-default | W0.7 Predicteddefault | W0.7 Predictednon-default | W0 Predicted default | W0 Predictednon-default |
|---|---|---|---|---|---|---|
| **Panel A: Confusion Matrix Default = 1 if Prob(Default) ≥ 0.5** | | | | | | |
| ActuralDefault | 0 | 0 | 0 | 0 | 0 | 0 |
| ActuralNon-Default | 461 | 6,223 | 756 | 5,928 | 552 | 6,132 |
| **Panel B: Confusion Matrix Default = 1 if Prob(Default) ≥ 0.95** | | | | | | |
| ActuralDefault | 0 | 0 | 0 | 0 | 0 | 0 |
| ActuralNon-Default | 374 | 6,310 | 566 | 6,118 | 464 | 6,220 |

| Default point / Prediction performance criteria | W0.5 Prob(Default) ≥ 0.5 | W0.5 Prob(Default) ≥ 0.95 | W0.7 Prob(Default) ≥ 0.5 | W0.7 Prob(Default) ≥ 0.95 | W0 Prob(Default) ≥ 0.5 | W0 Prob(Default) ≥ 0.95 |
|---|---|---|---|---|---|---|
| **Panel C: KMV Model Performance Default = 1 if Prob(Default) ≥ 0.5** | | | | | | |
| *Accuracy* | 0.93 | 0.94 | 0.89 | 0.92 | 0.92 | 0.96 |
| *Precision* | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| *Recall* | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| *F_Score* | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| *AUC* | – | – | – | – | – | – |

The table shows the predictive performance of KMV models with different default points and different default criteria. Panel A is the confusion matrix, and firms are considered to default if their probability to default is greater than 50 percent in 1 year. The default point is Current Liability+0.5*non-Current Liability(W0.5), Current Liability+0.7*non-Current Liability(W0.7), and Current Liability (W0). Panel B is the confusion matrix, and firms are considered to default if their probability to default is greater than 95 percent in 1 year. Panel C is the performance matrix, and firms are considered to default if their probability to default is greater than 50 or 95 percent in 1 year.

finance in several important ways. First, we predict fraud out-of-sample rather than explain fraud determinants within the sample. Second, we use ensemble learning, one of the state-of-the-art paradigms in machine learning, for fraud prediction rather than the commonly used logit regression. Finally, we attempt to extract the most influential factors that predict the corporate bond default.

Our study explained different sample periods and different enterprises (listed, non-listed, SOEs, non-SOEs, foreign and non-foreign firms), the majority of which have good prediction power. Our models can effectively predict rare events with a small number of input variables and have better performance than traditional default risk models, such as structure models or accounting models.

Because of the data availability of the private firms, we limit our empirical analyses to 58 factors. As the 58 data items represent only a small fraction of the hundreds of possible raw financial data items emerging from the accounting system, our study could be further extended using other available accounting information or using international data to predict corporate bond default. We do not rule out the possibility that better default prediction models could be developed by performing a more systematic and theory-driven selection of model input from hundreds of readily available financial data items.

## References

Acharya, V., J.-Z. Huang, M. Subrahmanyam, and R. K. Sundaram, 2006, When does strategic debt-service matter? *Economic Theory* 29, 363–378.

Acharya, V. V., and N. Mora, 2015, A crisis of banks as liquidity providers, *The Journal of Finance* 70, 1–43.

Acharya, V. V., H. S. Shin, and T. Yorulmazer, 2011, Crisis resolution and bank liquidity, *The Review of Financial Studies* 24, 2166–2205.

Acharya, V. V., and S. Viswanathan, 2011, Leverage, moral hazard, and liquidity, *The Journal of Finance* 66, 99–138.

Altman, E. I., 1968, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *The Journal of Finance* 23, 589–609.

Altman, E. I., G. Marco, and F. Varetto, 1994, Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks (the Italian experience), *Journal of Banking and Finance* 18, 505–529.

Ao, Y., H. Li, L. Zhu, S. Ali, and Z. Yang, 2019, The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling, *Journal of Petroleum Science and Engineering* 174, 776–789.

Asvanunt, A., M. Broadie, and S. Sundaresan, 2011, Managing corporate liquidity: strategies and pricing implications, *International Journal of Theoretical and Applied Finance* 14, 369–406.

Barboza, F., H. Kimura, and E. Altman, 2017, Machine learning models and bankruptcy prediction, *Expert Systems with Applications* 83, 405–417.

Beaver, W. H., 1966, Financial ratios as predictors of failure, *Journal of Accounting Research* 4, 71–111.

Beaver, W. H., M. F. McNichols, and J.-W. Rhie, 2005, Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy, *Review of Accounting Studies* 10, 93–122.

Black, F., and J. C. Cox, 1976, Valuing corporate securities: some effects of bond indenture provisions, *The Journal of Finance* 31, 351–367.
Black, F., and M. Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637–654.
Boser, B. E., I. M. Guyon, and V. N. Vapnik, 1992, A training algorithm for optimal margin classifiers, in: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 144–152.
Bracke, P., A. Datta, C. Jung, and S. Sen, 2019, Machine learning explainability in finance: an application to default risk analysis. Bank of England Staff Working Paper No. 816. Available at: http://dx.doi.org/https://doi.org/10.2139/ssrn.3435104.
Breiman, L., 2001, Random forests, *Machine Learning* 45, 5–32.
Cadenas, J. M., M. C. Garrido, R. Martínez, and P. P. Bonissone, 2012, Extending information processing in a Fuzzy Random Forest ensemble, *Soft Computing* 16, 845–861.
Campbell, J. Y., J. Hilscher, and J. Szilagyi, 2008, In search of distress risk, *The Journal of Finance* 63, 2899–2939.
Chava, S., and R. A. Jarrow, 2004, Bankruptcy prediction with industry effects, *Review of Finance* 8, 537–569.
Chesser, D. L., 1974, Predicting loan noncompliance, *The Journal of Commercial Bank Lending* 56, 28–38.
Collin-Dufresne, P., and R. S. Goldstein, 2001, Do credit spreads reflect stationary leverage ratios? *The Journal of Finance* 56, 1929–1957.
Cugnata, F., and S. Salini, 2014, Model-based approach for importance–performance analysis, *Quality and Quantity* 48, 3053–3064.
Danenas, P., G. Garsva, and S. Gudas, 2011, Credit risk evaluation model development using support vector based classifiers, *Procedia Computer Science* 4, 1699–1707.
Dichev, I. D., 1998, Is the risk of bankruptcy a systematic risk? *The Journal of Finance* 53, 1131–1147.
Duffie, D., and K. J. Singleton, 1997, An econometric model of the term structure of interest-rate swap yields, *The Journal of Finance* 52, 1287–1321.
Duffie, D., and K. J. Singleton, 1999, Modeling term structures of defaultable bonds, *The Review of Financial Studies* 12, 687–720.
Fama, E. F., and K. R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.
Figini, S., F. Bonelli, and E. Giovannini, 2017, Solvency prediction for small and medium enterprises in banking, *Decision Support Systems* 102, 91–97.
Gatev, E., T. Schuermann, and P. E. Strahan, 2009, Managing bank liquidity risk: how deposit-loan synergies vary with market conditions, *The Review of Financial Studies* 22, 995–1020.
Geske, R., 1977, The valuation of corporate liabilities as compound options, *Journal of Financial and Quantitative Analysis* 12, 541–552.
Girosi, F., 1998, An equivalence between sparse approximation and support vector machines, *Neural Computation* 10, 1455–1480.
Goldstein, I., and A. Pauzner, 2005, Demand–deposit contracts and the probability of bank runs, *The Journal of Finance* 60, 1293–1327.
Goldstein, R., N. Ju, and H. f, 2001, An EBIT-based model of dynamic capital structure, *The Journal of Business* 74, 483–512.
Gorton, G., and A. Metrick, 2012, Securitized banking and the run on repo, *Journal of Financial Economics* 104, 425–451.
Grunert, J., L. Norden, and M. Weber, 2005, The role of non-financial factors in internal credit ratings, *Journal of Banking and Finance* 29, 509–531.

He, Z., and W. Xiong, 2012, Rollover risk and credit risk, *The Journal of Finance* 67, 391–430.
Huang, Z., H. Chen, C.-J. Hsu, W.-H. Chen, and S. Wu, 2004, Credit rating analysis with support vector machines and neural networks: a market comparative study, *Decision Support Systems* 37, 543–558.
Janitza, S., E. Celik, and A.-L. Boulesteix, 2018, A computationally fast variable importance test for random forests for high-dimensional data, *Advances in Data Analysis and Classification* 12, 885–915.
Jarrow, R. A., D. Lando, and S. M. Turnbull, 1997, A Markov model for the term structure of credit risk spreads, *The Review of Financial Studies* 10, 481–523.
Jarrow, R. A., and S. M. Turnbull, 1995, Pricing derivatives on financial securities subject to credit risk, *The Journal of Finance* 50, 53–85.
Khandani, A. E., A. J. Kim, and A. W. Lo, 2010, Consumer credit-risk models via machine-learning algorithms, *Journal of Banking and Finance* 34, 2767–2787.
Kim, Y. S., and S. Y. Sohn, 2004, Managing loan customers using misclassification patterns of credit scoring model, *Expert Systems with Applications* 26, 567–573.
Kleinberg, J., J. Ludwig, S. Mullainathan, and Z. Obermeyer, 2015, Prediction policy problems, *American Economic Review* 105, 491–495.
Leland, H. E., 1994, Corporate debt value, bond covenants, and optimal capital structure, *The Journal of Finance* 49, 1213–1252.
Leland, H. E., 1998, Agency costs, risk management, and capital structure, *The Journal of Finance* 53, 1213–1243.
Leland, H. E. 2012, Predictions of default probabilities in structural models of debt, in: H. G. Fong, ed., *The Credit Market Handbook: Advanced Modeling Issues* (John Wiley & Sons Inc., New York, NY), 39–64.
Leland, H. E., and K. B. Toft, 1996, Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads, *The Journal of Finance* 51, 987–1019.
Lennox, C., 1999, Identifying failing companies: a re-evaluation of the logit, probit and DA approaches, *Journal of Economics and Business* 51, 347–364.
Liu, J., J. Li, W. Xu, and Y. Shi, 2011, A weighted $L_q$ adaptive least squares support vector machine classifiers – robust and sparse approximation, *Expert Systems with Applications* 38, 2253–2259.
Longstaff, F. A., and E. S. Schwartz, 1995, A simple approach to valuing risky fixed and floating rate debt, *The Journal of Finance* 50, 789–819.
Lunetta, K. L., L. Brooke Hayward, J. Segal, and P. Van Eerdewegh, 2004, Screening large-scale association study data: exploiting interactions using random forests, *BMC Genetics* 5, 32.
Martin, D., 1977, Early warning of bank failure: a logit regression approach, *Journal of Banking and Finance* 1, 249–276.
Merton, R. C., 1973, Theory of rational option pricing, *The Bell Journal of Economics and Management Science* 4, 141–183.
Merton, R. C., 1974, On the pricing of corporate debt: the risk structure of interest rates, *The Journal of Finance* 29, 449–470.
Ohlson, J. A., 1980, Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research* 18, 109–131.
Rendleman, R. J., 1978, The effects of default risk on the firm's investment and financing decisions, *Financial Management* 7, 45–53.
Satchidananda, S. S., and J. B. Simha, 2006, Comparing decision trees with logistic regression for credit risk analysis, in: *SUGI Asia Conference, Mumbai, India*.
Schwarz, D. F., I. R. König, and A. Ziegler, 2010, On safari to random jungle: a fast implementation of random forests for high-dimensional data, *Bioinformatics* 26, 1752–1758.

Suykens, J. A. K., J. De Brabanter, L. Lukas, and J. Vandewalle, 2002, Weighted least squares support vector machines: robustness and sparse approximation, *Neurocomputing* 48, 85–105.

Tanveer, M., 2015, Robust and sparse linear programming twin support vector machines, *Cognitive Computation* 7, 137–149.

Titman, S., and S. Tsyplakov, 2007, A dynamic model of optimal capital structure, *Review of Finance* 11, 401–451.

Vapnik, V., and A. Y. Lerner, 1963, Recognition of patterns with help of generalized portraits, *Avtomatika i Telemekhanika* 24, 774–780.

Vassalou, M., and Y. Xing, 2004, Default risk in equity returns, *The Journal of Finance* 59, 831–868.

Wagner, W., 2007, The liquidity of bank assets and banking stability, *Journal of Banking and Finance* 31, 121–139.

Wang, G., and J. Ma, 2012, A hybrid ensemble approach for enterprise credit risk assessment based on support vector machine, *Expert Systems with Applications* 39, 5325–5331.

Wang, Y., and L. Huang, 2009, Risk assessment of supply chain based on BP neural network, in: *2009 Second International Symposium on Knowledge Acquisition and Modeling*, vol. 2 (IEEE, Piscataway, NJ), 186–188.

Wei, L., J. Li, and Z. Chen, 2007, Credit risk evaluation using support vector machine with mixture of kernel, in: *International Conference on Computational Science* (Springer, New York), 431–438.

Westgaard, S., and N. Van der Wijst, 2001, Default probabilities in a corporate bank portfolio: a logistic model approach, *European Journal of Operational Research* 135, 338–349.

Wright, M. N., and A. Ziegler, 2017, ranger: a fast implementation of random forests for high dimensional data in C++ and R, *Journal of Statistical Software* 77, 1–17.

Xu, B., J. Z. Huang, G. Williams, M. J. Li, and Y. Ye, 2012, Hybrid random forests: advantages of mixed trees in classifying text data, in *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (Springer, New York), 147–158.

Xu, W., Z. Zhang, D. Gong, and X. Guan, 2014, Neural network model for the risk prediction in cold chain logistics, *International Journal of Multimedia and Ubiquitous Engineering* 9, 111–124.

Yu, Y., 2011, Risk management game method of the weapons project based on BP neural network, in: *2011 International Conference of Information Technology, Computer Engineering and Management Sciences*, vol. 1 (IEEE, Piscataway, NJ), 113–117.

Zavgren, C. V., 1985, Assessing the vulnerability to failure of American industrial firms: a logistic analysis, *Journal of Business Finance and Accounting* 12, 19–45.

Zhou, Z.-H., 2012, *Ensemble Methods: Foundations and Algorithms* (CRC Press, Boca Raton, FL).

## Appendix I

### Default risk model based on machine learning

*Logistic regression model*

The logistic regression model is the most common generalised linear probability classifier. It uses the sigmoid function to map the linear model with a range of $(-\infty, \infty)$ to (0, 1), and the output result is the probability value of the predicted event:

$$ln\frac{P\{y=1\}}{P\{y=0\}}=\beta_0+\beta_1 x+\ldots+\beta_k x_k, \tag{A1}$$

$$P\{y_i=1|x_i\}=\frac{1}{1+e^{-(\beta_0+\beta_1 x+\ldots+\beta_k x_k)}}. \tag{A2}$$

The advantage of logistic regression lies in coefficients' explanatory ability and better robustness under the uncertain distribution hypothesis. Therefore, the logistic regression model is widely used in research to predict the credit default rate. Unlike the structure model, the logistic model can incorporate both financial and non-financial factors for predicting default. While financial ratios capture firm-specific information, the non-financial factors help evaluate firms' links with macroeconomic factors. Logit or probit models have been widely used to analyse firm failure (Chesser, 1974; Martin, 1977; Ohlson, 1980; Zavgren, 1985; Lennox, 1999; Westgaard and Van der Wijst, 2001; Grunert *et al.*, 2005). However, they are country specific; thus, they might not consider the heterogeneity of country differences. Additionally, different industries have different risk characteristics that must be considered.

*Kernel support vector machine*

Vapnik and Lerner (1963) state that SVM has been widely used in pattern recognition, such as portrait recognition and text classification. SVM is a generalised linear classifier with a binary classification of data according to supervised learning. It uses the hinge loss function to calculate risk and adds regularisation to the solving system to optimise structural risk. Therefore, the SVM classifier is sparse and robust (Girosi, 1998; Suykens *et al.*, 2002; Liu *et al.*, 2011; Tanveer, 2015). In empirical application, Wang and Ma (2012) study Chinese firms' credit risk; however, their sample is limited and does not include non-listed firms. Huang *et al.* (2004) study default risk in Taiwan and the US using SVM and find that SVM has high prediction accuracy. Danenas *et al.* (2011) calculate corporate credit risk using multiple linear SVM estimation methods based on US service companies' financial data during 2005–2007. When faced with linear inseparability, traditional SVM cannot effectively distinguish positive and negative classes by a hyperplane. Boser *et al.* (1992) apply the kernel method to the SVM model for the first time and obtain a nonlinear KSVM. The basic idea of KSVM is based on the kernel function. The nonlinear fractional data can be mapped from the original variable space to a higher-dimensional Hilbert space, and can then take the inner product to achieve linear separability. Kernel functions operate in high-dimensional and implicit variable space without computing the coordinates of data in the space, and thus, they have the advantages of low computational complexity and high computational efficiency. Wei *et al.* (2007) study the application of a kernel mixed SVM model in credit risk assessment using credit data of US commercial banks and prove that the SVM model has higher efficiency and better predictability than traditional SVM models.

### *Decision tree*

A DT is a commonly used classification and regression method. In the DT model, the variable value is evaluated at each level. The information gain (or Gini index) is used to select the classification attribute, and the classification process is completed by bifurcating layer by layer until the end of the leaf node. The advantages of DTs are improved model intuition and speed of classification. Satchidananda and Simha (2006) use a DT model for analysing credit risk and show that DT classifiers produce good results with parsimonious models. However, DT has the following obvious disadvantage: the process is based on a greedy algorithm. A DT derived from a training set allows excessive fitting to obtain higher prediction accuracy, with too many nodes and poor generalisation ability. To solve the overfitting problem, Breiman (2001) and Lunetta *et al.* (2004) propose a tree-based integrated ML method, the RF. The basic principle of stochastic forest modelling is to construct multiple DT using random resampling and node random splitting techniques and to obtain the prediction results using a tree classifier. RFs can deal with high-dimensional variable input samples (without dimension reduction). Additionally, RFs have good robustness and specialise in handling noisy and missing data. They maintain a high learning speed when proceeding with a large database and offer a so-called variable importance index, which can be used to rank variables according to their predictive abilities (Breiman, 2001; Lunetta *et al.*, 2004; Xu *et al.*, 2012; Ao *et al.*, 2019). RFs have been widely used in various problems of classification, prediction, variable selection and outlier detection in recent years based on their advantages.

### *Importance index*

A boosting or bagging algorithm can improve the predictive robustness of models and rank the importance of variables and assist other models in variable selection. Schwarz *et al.* (2010), Wright and Ziegler (2017) and Janitza *et al.* (2018) discuss the application of the RF variable selection method in high-dimensional data. Cadenas *et al.* (2012) and Cugnata and Salini (2014) show that the RF variable selection method is better than the step-wise elimination method when the variables interact, and the results of screening variables are stable and have a good predictive outcome. The variable selection process is as follows:

1 Assuming that there are $N$ trees in the forest, for each DT, estimate the model's prediction error, namely, the out-of-bag (OOB) error, denoted as $err\_OOB_1$.
2 Add random noise interference to variable $X$ for all sample observations in OOB and recalculate the OOB error, denoted as $err\_OOB_2$.
3 The weight or importance of variable $j$ is denoted as $FI_j = \sum_{n=i}^{N} \frac{(err\_OOB_2 - err\_OOB_1)}{N}$. After adding a random noise term to the variable, if the accuracy of OOB prediction significantly decreases ($err\_OOB_2$ increases), then the variable has a significant impact on the sample's prediction results.

## Appendix II

### Default risk model based on structure model

The structure model can be traced back to the studies of Black and Scholes (1973) and Merton (1974). Merton's model assumes that the liability or the equity of a firm is the contingent claim of firm assets and can be priced by an option price model. The model is based on following assumptions:

**Assumption 1:** No market friction, no transaction cost and tax. The asset can be fully divisible and can be traded continuously. No limitation of short selling. No bid and ask spread, and the rate of borrowing is equal to that of lending.

**Assumption 2:** The financial market has a sufficient number of investors and does not have arbitrage opportunity.

**Assumption 3:** If a risk-free asset exists, the return of the asset is fixed and known.

Firms only have two kinds of security: the equity and the zero coupon bond. The firm will default when the asset value is lower than that of the bond value. The bond investor acquires all of the residual value of the firm, while the equity holder gains 0. The liability can be considered similar to buying a European call option, with the execution and the market value of the firm asset as underlying, and with 1 year maturity. The movement of firm asset value satisfies a log-normal distribution, and the volatility of the asset return is stable. The following equation can be used to calculate the value of asset ($V$) and its volatility $\sigma_v$:

$$\begin{aligned} E &= VN(d_1) - Be^{-rt}N(d_2) \\ \sigma_e &= N(d_1)V\sigma_v/E \\ d_1 &= \left[\ln\left(\frac{V}{B}\right) + \left(r + \frac{1}{2}\sigma_v^2\right)t\right]/\sigma_v\sqrt{t} \\ d_2 &= \left[\ln\left(\frac{V}{B}\right) + \left(r - \frac{1}{2}\sigma_v^2\right)t\right]/\sigma_v\sqrt{t} \\ N(x) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{x^2}{2}}dx, \end{aligned} \tag{A3}$$

where $E$ is the value of firm equity, $B$ is the par value of firm leverage, $\sigma_E$ is the volatility of firm stock price, $r$ is the risk-free rate of the market (the risk-free rate we used is the 1-year fixed deposit rate in China), and $T$ is the maturity of the liability. The distance to default ($DD$) is:

$$DD = \frac{V - B}{V\sigma_v} \tag{A4}$$

where $DD$ is the default rate of the firm that satisfies a normal distribution. The expected firm default probability (EDF) shall be:

$$P(T) = 1 - N(DD). \tag{A5}$$

The KMV model further extends Merton's model by considering the maturity and capital structure of the firm. Firm long-term liability can relax solvency pressure. Merton's model uses the par value of firm liability as a default point, whereas the KMV model places more weight on the current liability. The Equation (A4) can be restructured to

$$DD = \frac{V - DP}{V\sigma_v}$$

$$DP = CurrentLiability + W^{*}LongtermLiability, \tag{A6}$$

where $W$ is the weight of long-term liability, and it can be any number between 0 and 1. In this study, we use 0.5 as the weight of the long-term liability. Changing the value of the weight does not increase the performance of the KMV models.

# Market smart: How firms respond to the IPO P/E price-cap regulations in China *

Zhou Lu†

**Abstract**

China represents a special case in which market dynamics and government intervention coexist to provide insights into (1) how firms respond to government regulation, and (2) how the market helps firms minimize the impact of government intervention. Using the difference-in-differences (DID) method, this study investigates how firms respond to the stock price-to-earnings (P/E) price-cap regulation in China. The results reveal that firms subject to price caps have higher cumulative abnormal returns (CARs) in initial public offerings (IPOs), retain ownership in IPOs, shorten their time for their first seasoned equity offering (SEO) issuance, increase earnings management before SEO issuance, and use more leverage and loan financing when they cannot issue SEOs. Additionally, since price caps delay price discovery in IPOs, firms reveal more information as countermeasures. Therefore, although the financial market is regulated, firms act strategically to evade regulations. Nevertheless, although firms experience a negative impact of regulations in the short term, regulations do not affect borrowing cost, investment, and willingness to go public in the long term.

**Keywords:** *P/E Ratio, price-cap regulations, Government Intervention, IPO , Ownership Retention, Earning Management*

**JEL Classification**: G31, G32, L10, L21

*This draft: November 23, 2021. For their helpful comments and suggestions on previous drafts, I would like to thank Yun Dai, Hui Chen, Kyeong Hun (Kyle) Lee, Xunhua Su, Danielle Zhang, and the audience who attended the NHH brown bag seminar. All errors are my own.

†Department of Finance, Norwegian School of Economics (NHH). zhou.lu@nhh.no

# Contents

# 1 Introduction

Since the 1978 economic reform, China has gradually liberalized its economy. In 1992, the Chinese government started to develop a "socialist market economy." Unlike countries such as Poland, former Czechoslovakia, and Russia, which underwent rapid reforms according to comprehensive plans, China has enacted a step-by-step evolutionary reform by providing state-imposed market-like incentives, with different sectors having different reform speeds. As the manufacturing sector is more resilient to external shocks and has less government intervention, its reform was completed early. The large-scale privatization of state-owned enterprises (SOEs) began in the mid-1990s, as part of the broad economic reforms outlined in the ninth and tenth five-year plans. Gan [2009] estimate that between 1995 and 2005, firms with an aggregate 11.4 trillion RMB in assets were privatized in China, comprising two-thirds of China's SOEs and state assets. Further, owing to a high unemployment rate, most firms in China have been operating based on market-like incentives. On the other hand, the financial sector, which is more fragile and reformed more conservatively, is heavily regulated in China. Thus, the coexistence of market forces and government interventions in China allows researchers to investigate how the market helps firms minimize the impact of government intervention and how firms respond to government regulation.

In this study, I examine this question by focusing on price caps. The China Securities Regulatory Commission (CSRC) has set regulatory guidelines on the price-to-earnings (P/E) ratio of companies that plan to go public. Before 2009 and after 2012, the CSRC applied a guideline that made it unlikely for companies with a P/E ratio greater than 23 to be approved by the regulator, with all industries being subject to a homogenous price-cap limitation. This regulation provides an exogenous shock to firms that intend to go public.

A key feature of China's approval-based public listing system that enables causal identification is that firms have little ability to time the initial public offering (IPO) market. Normally, IPO approval takes two to three years. Once approved, firms take several months to complete the final steps. In addition, unlike the US, withdrawal from the IPO market is associated with heavy sunk costs in the approval process. Thus, public listing in China serves as a strong signal to investors that the firm is eligible to pass multi-layer government regulation and has good profitability and earnings quality for at least three years. Therefore, instead of withdrawing from an IPO, firms are more likely to issue IPOs with losses to be compensated from the seasoned equity offering (SEO) market. Firms are unlikely to acknowledge the date the government imposed the price cap and jumps the queue of listing in advance to escape price-cap regulations. In addition, the implementation of the price cap is directly decided by the CSRC based on the IPO market conditions; thus, firms have little ability to influence price caps. Public equity is an especially important source of financing in China because alternative financing is limited, and public markets provide liquidity for early investors and entrepreneurs. Moreover, bank credit in China tends to typically favor SOEs or mature firms with good credit records. Although Chinese venture capital and private equity (VC/PE) are growing quickly, they remain less mature than their counterparts in the United States during the period investigated in this study. Financing under regulations became the most important channel of external financing for the majority of listed firms.

Although firm financing is inevitably associated with higher costs, firms may act strategically to try to minimize the negative impact of pricing cap regulation in the IPO market. There are two direct negative impacts of pricing cap regulation. Firstly, a pricing cap reduces the required rate of return in the IPO market. Compared to a non-pricing cap period, firms are more likely to use alternative financing with lower costs or higher returns during a pricing cap period. As a result, in this context, firms may try to minimize ownership dilution during their IPOs while seeking for alternative, less costly financing options such as seasoned equity offering (SEO) and corporate bond/loan financing. Firms also adjust their financial reporting; firms affected by the pricing cap anticipate that their IPOs will be influenced by the pricing cap, which reduces their incentive to manage earnings before the IPO and increases their incentive to manage earnings before the SEO. Second, regulatory literature indicates that price regulation in the capital markets may postpone price discovery. Firms affected by the pricing cap are more likely to be associated with higher information asymmetry. As a counteraction, firms may increase self-disclosure during their IPOs to reduce the negative impact of the regulations.

I find consistent evidence that firms are likely to realize a positive abnormal return in the post-IPO market indication that its price is understated ; additionally, I find that this phenomenon disappears in the SEO market, indicating that the pricing cap distorts the stock price and hinders pricing efficiency in the short-term. Firms retain shares in IPOs and accelerating the speed and increase its amount of its SEO issuance or seeking alternative financing such as bank loan or corporate bond. The pricing cap also impacts whether affected firms are incentivized to manage their earnings well;

firms actively engaged in information disclosure will increase their transparency.

Does this policy have long-term impacts on the cost of borrowing, the attractiveness of IPOs, and investment activity? Surprisingly, I find that the pricing cap policy, overall, does not affect borrowing costs, the willingness to go public, or firm investment . The probability of a price cascade in the first month after an IPO is also reduced. Overall, the results indicate that these regulations can be circumvented by the firm or the manager's strategic actions and that regulation can effectively reduce the short-term probability of a price cascading effect for an IPO.

The remainder of this paper is organized as follows. Section 2 reviews the related literature. Section 3 provides an institutional background about China and develops the hypotheses. Section 4 describes the data and outlines the methodology. Section 5 presents the results and the robustness test. Section 6 summarizes the main conclusions.

## 2 Literature Review

There is a growing body of literature on IPOs in China, most of which highlight the extraordinarily high underpricing of Chinese IPOs and subsequently attempt to interpret it [Chan et al., 2004, Chang et al., 2008, Mok and Hui, 1998, Su and Fleisher, 1999]. Tian [2011] was the first paper to highlight how government regulation distorts the market and results in high stock underpricing. Liu et al. [2011] show that IPO pricing efficiency significantly improved after the first-stage book-building policy reform and the CSRC's window guidance abolition. However, institutional investors shifted investment risk to retail investors, resulting in a winner's curse problem. Liu et al. [2021] find that after removing the P/E price cap, the underpricing phenomenon significantly decreased, associated with lower initial returns and improved resource allocation effectiveness. As the issuer and underwriter cannot determine the issuance price based on market demand and supply, this weakens the motivation and ability of information discovery for underwriters under the price cap, which distorts the market allocation and pricing efficiency. Bekaert et al. [2005] show that equity market liberalizations on average lead to a 1% annual real economic growth.

Moreover, several studies show that retaining ownership can signal issue quality and prospects [Demers and Joos, 2007, Leland and Pyle, 1977]. Jain and Kini [1994] and Kim et al. [2004] find a positive relationship between performance changes and the portion of shares retained by pre-offering shareholders.

According to the signaling theory, IPO underpricing is positively related to the speed and probability of conducting SEOs [Jegadeesh et al., 1993, Slovin et al., 1994, Welch, 1989]. Good-quality companies issue SEOs when the market has realized the company's true value [Welch, 1996]. In other words, good-quality firms are associated with faster SEO issuance and higher SEO probability.

Teoh et al. [1998a] and Teoh et al. [1998b] document the earnings management phenomenon before IPO and associate it with poor post-IPO performance. Earnings management has also been observed in global markets such as Hong Kong [Mathew, 2002], Korea [Cheon et al., 2011], China [Aharony et al., 2000], the UK [Levis, 1995], and the Netherlands [Roosenboom et al., 2003]. Liu et al. [2021] find that firms increased earnings management after removing the P/E price cap.

Further, the extant literature shows that, as the stock pricing efficiency of affected firms during the price cap period is low, such firms do not have an incentive for earnings management before the IPO. Meanwhile, firms not affected by price caps tend to window dress their performance before the IPO. After the IPO, firms are more transparent, with more standard accounting reports and legal requirements. Thereby, IPOs may be the only chance for unaffected firms to utilize the information asymmetry between the firm and investors to increase equity financing. Conversely, in the SEO market, which does not have a price cap and has a better price discovery, the firm is more likely to have an incentive for earnings management to reduce the cost of capital and increase the capital raised in SEOs. In particular, the underwriter's commission fee is based on the portion of raised capital in SEOs. Underwriters have an incentive to increase the issuing prices. Thus, earnings management serves as an important channel to signal a firm's quality to the market and investors.

According to McConnell and Servaes [1995] and Stulz [1990], high-valued small growth firms avoid debt usage, as for these companies, the cost of borrowing is higher than that of large companies, which has a negative effect on their earnings. This phenomenon also exists in the Chinese stock market. As documented by Huang and Zhang [2001], this phenomenon is driven by 1) the low cost of equity financing; 2) the quota control and approval system for stock issuance partially shifting the risk of stock issuance to the local governments or the central government; 3) firms having a low dividend payout ratio, with the payment of the dividend being voluntarily based; 4) the assessment system of listed companies, which makes the company pay no attention to optimizing the financing structure; and 5) the majority of the shares being held by the state, making firms in the stock market subject to serious internal control problems. This study investigates an extreme case under McConnell and Servaes [1995] and Stulz [1990]'s framework and analyzes firms' financial strategies under market friction caused by government regulation.

In a perfect world without market friction, firm investment would be determined by investment opportunities [Modigliani and Miller, 1958, Stein, 2003]. However, in the real world, it has been documented that information asymmetry, moral hazard, or agency problems can distort investment behavior [Fazzari et al., 1988, Jensen, 1986, Myers and Majluf, 1984]. Recently, a new source of distortion, government intervention, has gradually attracted attention in academia [Fang et al., 2017]. Bekaert and Harvey [2000] show that financial liberalization decreases the cost of equity and increases investment. Brown et al. [2009, 2012], Hall et al. [2010], Kim and Weisbach [2008] study the relationship between firm innovation input and firm financial constraints. Fazzari et al. [1988] show that cash flow and external equity is the leading factor predicting firm innovation input for young firms. Brown et al. [2012] show that due to information problems and lack of collateral value, innovation activity is influenced by firm accessibility to external finance and cash storage, and highlights the importance of stock market development liberalization in promoting economic growth and increasing firm-level innovative activity. Kim and Weisbach [2008] highlight the importance of external equity financing for firm innovation. Schumpeter [1982] highlights the impact of government regulation on innovation activity. Hombert and Matray [2018] show that large R&D stocks can help firms escape from the product market competition by increasing product differentiation. In this study, I examine firm innovation activity under government regulations. Firms affected by price caps are financially distressed; thus, it is unlikely for them to increase their innovation input. In the long term, when the market realizes the value of the affected firm, compared to firms that are not subjected to the price cap, the affected firms are more likely to have better quality and better access to external finance. Thereby, firms may increase their innovation input to differentiate themselves from their peers.

## 3 Institutional Background and Hypotheses development

China's IPO pricing system has experienced many transformations: fixed prices, price control, and market pricing. The CSRC regulates the issue price by setting the price earnings ratio or issuing a ceiling. At the end of 2004, the CSRC introduced the book-building process and imposed less price regulations in the market. In June 2009, the CSRC implemented the Guiding Opinions on Further Reforming and Improving the Issuance System of New Share, which removed the price-cap regulation on the P/E ratio and handed over the right of pricing to the market. However, firms listed during this period were associated with high offer prices, P/E ratios, and excessive raising problems. Book builders incurred commission fees based on the total IPO funding. Additionally, they collaborated with analysts and firms, and sold the overvalued stock to investors. In the post-IPO market, stock prices are usually associated with negative price adjustments, resulting in a price cascade. To address this problem, in April 2012, the CSRC issued supplementary regulations that issuance pricing should be linked to the P/E ratio of the same industry. For newly listed enterprises whose P/E ratio exceeds 25% of the median P/E ratio of the same industry, it is necessary to explain the reasons and indicate the risk to investors. However, this policy did not stop the price cascade in the stock market, and the CSRC suspended the IPO listing from October 2012 to December 31, 2013, with 48 firms approved to listing in the following two months. The CSRC then paused firm approval for listing until June 9, 2014. All firms had the same P/E ratio in IPOs. Although there was no compulsory regulation or guiding opinion to be enacted, there was a tacit agreement for book builders to set up the offer price forcing firms to have an IPO P/E ratio less than 23.

Before 2006, the main forms of equity refinancing of listed companies in China were public issuance and stock allotment. Thereby, public issuance and stock allotment were associated with restricted regulations on return on equity (ROE) and dividend payout. After 2006, the Administrative Measures for the Issuance of Securities by Listed Companies introduced private placements to firms that allow a limited number of investors to issue stocks with a certain lock-up period to raise funds or acquire assets. Firms that wished to issue private placements also needed to be approved by the CSRC, but compared to the IPOs, it had a much simpler procedure and a higher rate of approval. It replaced public issuance and stock allotment as the major source of equity refinancing channels. To issue private placements, firms needed to host general meetings of shareholders and host board meetings twice. Additionally, they needed to be approved by the CSRC, which usually takes three to six months to be approved. Firms usually take about one year to prepare for SEOs.

The price of new shares in China largely depends on earnings per share (EPS). Therefore, the P/E ratio and P/E ratio ceiling increase the listing cost of enterprises with higher fair value and reduce the number of shares they are willing to sell when listing. After the implementation of the issue price control, the CSRC set the upper limit of the P/E ratio of the enterprises to be listed. The higher the fair value, the greater was the cost of issuing shares in the price-cap period. Using a case study of Italian companies, Pagano et al. [1998] show that the median market-to-book ratio of publicly traded firms in the same industry is an important determinant of when Italian firms go public, and that enterprises with high growth in the future will delay the listing time and wait until the firm has stable profits. However, this is unlikely to hold in China for multiple reasons. First, owing to the IPO suspension, the listing procedure is associated with policy uncertainty. That is, no one can anticipate whether the regulator will change its current policy . To be on the safe side, a firm may accelerate its speed for an IPO. Second, unlike other markets, it is time consuming and costly to obtain the qualification of listing in China. Firms usually need to stay three to four years in queue to obtain approval for listing from the CSRC. If the firm withdraws from the queue, the firm needs to line up for listing for an additional three to four years. Therefore, most firms choose to stay in the queue instead of withdrawing from the IPO, unless they are not qualified for listing. Third, going public has multiple advantages. Celikyurt et al. [2010] show that companies listed earlier in the industry tend to initiate more acquisitions and have competitive advantages. Lee et al. [2019] highlight the regulatory costs associated with IPOs in China. Qualification is a strong signal to investors and a scarce resource in China, which provides a valuation premium for the firm. Owing to information asymmetry, private firms have a natural disadvantage in expanding client and external financing. Thus, Chinese firms heavily rely on bank loans as their major source of financing, and investment banks tend to favor publicly listed firms. To increase the accessibility of finance, expand client base, and increase market competitiveness, instead of withdrawing from the IPO market, enterprises will strive to be listed as soon as possible. It is impossible for a firm to postpone or have an earlier listing to escape the regulatory impact. First, even assuming that a firm anticipates its approval by the CSRC in one year, it is impossible for the firm to identify the exact time. Second, even assuming that the firm already anticipates the exact time it will be approved, they are unable to predict when the P/E regulation is implemented. Third, even assuming that the firm anticipates the exact time of approval and the time regulator imposes the P/E price cap, it is difficult for the firm to time the IPOs. The approval will automatically expire if the firm fail to list within six months after it receives the qualification from the CSRC. In other words, within six months, the firm needs to decide the offer price, devise a road map, advertise the company, thus making it difficult to time the IPO market. Additionally, it is also difficult to escape from the regulation by manipulating the P/E ratio. The P/E ratio is equal to the price per share scaled by the EPS, which is directly determined by the market. The firm can manipulate the EPS, as the firm's target is to maximize the stock price. Owing to price-cap regulations, the firm needs to decrease the EPS to boost the stock price. However, the price will adjust downwards if the firm decreases EPS. In other words, it is impossible to escape the regulations by manipulating the EPS.

### 3.1 Hypotheses development

For an industry with an average P/E ratio greater than 23, the underwriter cannot determine the issue price. Therefore, it is difficult for the IPO price to reflect the value of the company. The price-cap limit creates an upper limit on firm IPO returns, resulting in a price upward adjustment in the IPO period, with most firms exceedingly underpriced due to price-cap regulations. Therefore, I postulate Hypothesis I as follows:

*Hypothesis I: During the price-cap period, firms affected by the regulation are more likely to experience an upward price adjustment in the post-IPO market. Additionally, firms affected by the price cap are more likely to have higher initial returns and higher underpricing.*

For the SEO market, since it does not have price-cap regulations, it can be used as falsification test. The stock price is fully adjusted during the post-IPO period to prevent the firm from experiencing an upward price adjustment in the post-SEO issuance period and to prevent a systematic price anomaly between affected and non-affected firms. Hence, Hypothesis II is as follows:

*Hypothesis II: During the price-cap period, no significant difference in the cumulative return in the post-SEO period between affected and non-affected firms can be observed.*

Owing to the price cap, firms with higher fair values have lower IPO profit margins. Thus, firms are less willing to sell their stocks in an IPO because IPO issuance is associated with ownership dilution. Therefore, firms will retain their shares in IPOs. In SEOs, the profit margin between the affected and non-affected firms does not exhibit systematic differences. Consequently, there will be no systematic difference between the affected and non-affected firms. The securities law of the People's Republic of China stipulates the minimum proportion of outstanding shares of listed enterprises. Specifically, if the total share capital after listing is less than 400 million RMB, the proportion of public shareholding shall not be less than 25%; if the total share capital exceeds 400 million RMB, the proportion of public shareholding shall not be less than 10%. To minimize the negative impact of price-cap regulation, firms may manipulate the stock capital, increasing it by more than 400 million RMB to dilute less ownership in IPOs. Hence, we propose Hypothesis III as follows:

*Hypothesis III: During the price-cap period, firms retain shares in IPOs. The affected firms will have a higher probability of increasing their stock capital to 400 million RMB, making themselves eligible to issue less shares (10%) to public shareholders. This phenomenon is unlikely to be observed in the SEO market.*

Ownership retention in IPOs triggers two problems. First, both the price and volume of issuance decline in an IPO, and the affected firm is poorly funded in the IPO. Therefore, the affected firm must seek alternative financing. The affected firms will have to choose to increase bank loans, increase corporate bond financing, accelerate the speed for an SEO, or increase the SEO issuance amount to minimize the negative impacts of low-funded IPOs. Therefore, Hypothesis IV is as follows:

*Hypothesis IV: During the price-cap period, the treated firms will increase corporate bonds and loan financing, increase the amount and probability of SEO issuance, or accelerate the speed of SEO issuance.*

Kim and Rhee [1997] document that three negative impacts are associated with price regulation: volatility spillover, delayed price discovery, and trading interference. In this study, I focus on the negative impact of price discovery. Owing to the price-cap regulation, the affected firms have an upper limit on their IPO offer price. Book builders do not have an incentive to discover the value of firms. Compared to the non-price-cap period, firms and investors are associated with higher information asymmetry. To minimize the negative impact from the information side, firms are more likely to disclose information. The greater the information asymmetry between investors and firms, the higher is the probability of self-disclosure. However, the disclosure of business secrets is costly, and we would expect that firms reduce their self-disclosure when they are financially less constrained. Thus, I postulate Hypothesis V as follows:

*Hypothesis V: During the price-cap period, treated firms are more likely to be associated with information asymmetry due to the pricing cap regulation. Accordingly, firms are more likely to disclose more information to minimize the negative impact of price-cap regulations. This effect* *will gradually diminish when a firm has an access to alternative financing.*

I want to answer four questions associated with price caps. First, can the affected firm successfully circumvent price-cap regulations in China? Second, is the affected firm associated with a higher financing cost? Third, does it have a negative impact on a firm?s willingness to conduct an IPO? Finally, is this policy effective against the price cascade in the post-IPO period? Lemmon and Roberts [2010] highlight that when the investment scope is limited, the company's capital expenditure will be reduced. Kahle and Stulz [2013] show how problems in bank capital supply negatively affect firm investment during financial crises. If the firm cannot minimize the impact of price-cap regulation, it may have a negative impact on its willingness to issue an IPO, decrease investment, and be associated with a higher cost of capital. If a firm successfully circumvents the regulation, the price-cap regulation will not have a significant impact on investment, cost of capital, or issuance willingness.

Earnings management is widely observed in the IPO and SEO markets, and price-cap regulation may distort firms' incentives to manage earnings. On the one hand, firms affected by the price cap anticipate that their IPOs will be poorly funded; thus, they do not have an incentive to manage their earnings before their IPO. Instead, the SEO market has a higher profit margin, and the affected firm is more likely to manage earnings before SEO issuance. On the other hand, non-affected firms are more likely to manipulate earnings before an IPO, as it is the only time at which a firm can utilize information asymmetry to boost its stock funding. However, this opportunity will wane after a stock trade in the secondary market. Thus, I put forward Hypothesis VI as follows:

*Hypothesis VI: During the price-cap period, affected firms are less motivated to manage earnings for their IPOs. Instead, they are more likely to manipulate their earnings before SEOs.*

# 4 Materials and Methods

## 4.1 Data and Sample Selection

I obtained the accounting information besides information on IPOs, seasoned offerings, corporate debt issuance, and bank loan usage from the Wind and CSMAR databases. These commercial databases are similar to the CRSP and COMPUSTAT databases and provide IPO prospectus data (sometimes called "pre-disclosure" data) as well as listing and financial statement data. These databases have been widely used in studies published in leading journals.

I downloaded the data for all mainboard firms, limiting my sample to firms listed after 2009. I dropped observations with missing information on the variables of interest or the control variables, and dropped firms listed from 2012 to 2014 during the IPO suspension period. The full sample had 6961 firm-year observations for 1296 firms listed from 2009 to 2012 and 2014 to 2020. All variables were winsorized at the 1st and 99th percentiles to reduce the effect of extreme outliers.

Table 1 reports detailed summary statistics on the variables of interest and control variables. Again, in Table 1, Panel A includes the full sample, and Panel B includes the subsample of firms in industries with a trading P/E ratio greater than 23. I used the CSRC guidelines for the Industry Classification of Listed Companies (2012 Revision) as the industry identifier (equivalent to the 2 digits SIC code).

[Insert Table 1 here]

Table 1 Panel A exhibits the summary statistics for all variables of interest. On average, the CAR for SEOs was around -2%, and the 30 days average CAR for IPO was 53%. The probability of the stock dropping below the offer price in 30 day was 18%. On average, firms had 161% of cumulative initial returns after IPO issuance. For firms with debt and loan issuance, firms on average held 13% of debt and loan as a ratio of their total assets. In my sample, only 10% of firms had SEO issuance. For firms that had SEO issuance, firms on average took 1371 days to issue SEO after going public. The average proportion of publicly issued shares was 23%, and 22% had stock capital greater than 400 million CNY before the IPO issuance. The firm issued SEO increased their stock capital by 12% on average in SEOs. On average, 2On average, 13 firms queued up for listing each year. The firm-weighted average cost of capital was 6The firm leverage was 34%, with a profitability of 9%, a sale growth rate of 18%, 1.50 market-to-book ratio, and ROA of 8.45. The shareholder concentration for the top 10 shareholders was approxiamtely 38% and, on average, the top 1On average, 12% of firms were backed by private equity or venture capitals, wehereas 15% were SOEs.

Panel B compares the t-statistics between firms in industries with an average P/E greater than 23, which were subject to the price-cap policy (treatment), and firms with an average P/E less than 23, which were not subject to the policy (control). Compared to the control group, firms in industries with an average P/E greater than 23 were associated with a 4% higher probability of having an SEO issuance and a 222-day earlier SEO issuance. Additionally, firms in the treatment group had a 2% lower ownership retention rate in IPOs. Further, treated firms had a 28.5 percentage point lower probability of having a stock capital greater than 400 million CNY before the IPO. The treated group had a 1% higher RD input.

Moreover, firms in the treated group had nine firms that queued up for listing each year. The control group had 4.6 less firms queue up for listing. The affected firms had a 0.7% higher WACC. The treated group had 8% lower leverage, 1% lower cash flow, 2.8% higher sale growth, 52.9% higher market-to-book ratio, and 89% lower ROA. Moreover, firm shareholder concentration and shares held by the top 1% of the shareholders for the treatment firm were 10% lower than those of firms in the control group. The affected firms older by an average of 11.6 percent , 15% less likely to be backed by private equity and venture capital companies, and were 23% less likely to be SOEs.

Figure 1a shows the P/E ratio of firms listed on the main board after 2009. The IPO average P/E ratio was in the range (40-50) before 2012 and experienced a sharp drop after 2010. The IPO P/E ratio was steadily below 23 after 2014. Figure 1c shows the P/E ratio of firms listed for at least 3 years for both the treatment and control groups at the end of the listing year. Figure 1d shows the P/E ratio of new listed firms. We can observe that the average difference in the P/E ratios between the treatment and control groups at the end of the listing year fluctuated near 80, experienced a sharp drop after 2012, and dropped to 30 in 2020. Figure 1b indicates that the price-cap regulation helped reduce speculation, resulting in a decline in the industry P/E overall, which significantly decreased the probability of a price cascade driven by high offer price, P/E ratio, and stock over-raise.

[Insert Figure 1 here]

Figure 1c shows the P/E ratio of firms listed for at least 3 years in the treatment group and control group. The results show that pricing cap effectively reduced the P/E ratio difference between the treatment and control group and that firms in the treatment group have a more reasonable P/E ratio than those in the non-pricing cap period. Figure 1d shows the P/E ratio of new listed firms. The results show that before the enaction of the pricing cap, the P/E ratio difference between the newly listed firms in the treatment group and control group was as high as 45 and during 2011 the difference decreased to 10; moreover, in the pricing cap period the difference between the two groups reached the highest historical value—60 in 2017—and significantly decreased to 20 or 30 and maintained a stable level in the subsequent years.

### 4.2 Baseline Model

I performed a difference in differences (DID) analysis and estimated how the price cap affects firm investment and finance using the following model:

$$Y_{i,j,t} = \alpha + \beta Treated_j \times Post_t + \gamma X_{i,j,t} + \eta_j + \theta_t + \epsilon_{i,j,t} \tag{1}$$

X$_{i,j,t}$ are the firm controls, industry, or year characteristics, respectively, of firm i in industry j in year t. $\eta_j$ is the industry fixed effect, $\theta_t$ is the time fixed effect, and $\epsilon_{i,j,t}$ is the error term. The coefficient estimate of the interaction term Treated×Post captures the price-cap effect between firms with an average P/E ratio greater than 23 (affected by the price cap) and below 23 (not affected by the price cap).

## 5 Empirical Results

### 5.1 Impact

[Insert Table 2 here]

Table 2 Panel A shows the CARs and initial returns in the post-IPO period, using the fuzzy DID method. It reveals that firms affected by the price cap (treatment group) had a 10% higher CAR in the first day after the IPO. The CAR difference between the treatment and control groups increased steadily after 20 days. Additionally, firms affected by the price cap on average had 154% higher CARs after being listed for 20 days. They were also associated with 212% higher initial returns. These results indicate that in the post-IPO period, firms affected by the price cap were associated with higher stock returns compared to firms in the control group. This effect was even sustainable after one month. Moreover, affected firms were associated with deep underpricing, indicating that the stock prices were distorted in the primary market. Table 2 Panel B shows the CAR in the post-SEO period, using the fuzzy DID method. The findings show that, in the post-SEO period, the price anomaly for firms affected by the price cap disappeared. After multiple stock trading, the firms' information were fully revealed to the public. If the market were efficient, regulation would not have sustainable impact on firm stock performance, and there would not be systematic difference between firms in the treatment and control groups. Thus, these results verify Hypotheses I and II.

## 5.2 Firm Counteract Measures

This subsection investigates how firms respond to price-cap regulations. Here, I provide two possible channels for a firm's reaction.

### 5.2.1 Alternative External Financing

Owing to price-cap regulations, firms are associated with lower returns in IPOs. Considering the cost of funding, firms are less willing to increase funding in IPOs. Instead, since there is no price-cap regulation in the SEO market, firms may stop funding until the stock price fully reflects the value of the firm. Firms may try to access alternative markets with lower financing costs, such as corporate bonds and loan markets.

[Insert Tables 3 here]

Table 3 show the results. Table 3 Panel A shows that firms are more likely to retain shares in IPOs to reduce the negative impact of price-cap regulations. On average, firms in the treatment group had a 3% lower stock capital release to the public in an IPO. The affected firms attempted to increase their stock capital to 400 million RMB to be eligible to issue less stock to the public. Further, firms affected by the price cap had a 15% higher probability of having 400 million RMB stock capital before the IPO. Table 3 Panel B shows that firms were more likely to conduct SEOs, shorten the time of their first SEO issuance, and have higher SEO issuance amount. The results show that firms affected by the price-cap regulation increased the likelihood of SEO issuance by 2% and reduced the time of its first SEO issuance by 75 percentage points. In addition, firms increased the SEO issuance amount by 27 percentage points in the second year and 48 percentage points in the third year after the IPO. Table 3 Panel C shows the corporate bond and loan usage after IPOs. Firms affected by the price cap increased the total amount raised in corporate bonds and bank loans by 4%, 3%, and 2%, respectively, for one, two, or three years. This positive effect gradually diminishes when firms had more access to external financing. These results are consistent with Hypotheses III and IV, thus confirming them.

### 5.2.2 Self Disclosure

Firms affected by the price-cap regulation had distorted IPO offer prices. The book builder did not have an incentive to discover the intrinsic value of the firm, since regardless of how much effort they devoted to price discovery, the offer price was fixed and pre-determined by the P/E regulations. Consequently, compared to the non-price-cap period, firms affected by the price cap were more likely to be associated with higher information asymmetry.

As a countermeasure, firms had to self-disclose information, especially when they had less access to finance or during the period when the firm was associated with higher information asymmetry.

[Insert Tables 4 and 5 here]

Table 4 and 5 exhibit the results. Table 4 Panel A shows that affected firms actively increased their firm transparency and the length of their field research reports. The firm transparency was measured by the transparency index powered by SSE and SZSE. SSE and SZSE construct this index based on information disclosure, relationships with investors, ESG, and cooperativeness with the stock exchange. Information disclosure is measured by the availability of publicly available information and the voluntary disclosure of information. Firms increased the length of their field research reports by 269% in the first year of their IPOs and experienced increases of 0.14 and 0.13 in firm transparency in the first and second years of their IPOs. Table 4 Panel B shows that firms affected by the price cap invited 72% more researchers to conduct the field research, with this increase being mainly from institutions. Although the effect gradually diminished, we can observe that the affected firms had more elaborate field research reports, even for firms that went public for 3 years, with 56% more words in field research reports on average. Table 5 Panel A shows that affected firms had 5% lower attention from analysts during the year of IPO issuance. This effect dropped to 1% in the subsequent years. The effect reversed after the second years of listing. After the third year of listing, affected firms attracted 3% more attention from analysts. Attention to research report showed a similar result, where affected firms received 7% lower attention in the research report in the first year of listing, and 5% more attention in the third year of listing. These results indicate that firms associated with greater information asymmetry and firm self-disclose information attempt to minimize the effect of price-cap regulation.

We use Standardized Unexpected Earnings (SUE) as a proxy for information asymmetry. Firms associated with higher SUE are more likely to be associated with greater information asymmetry. Table 5 Panel B shows the results. The table shows that firms are associated with a lower SUE in both time series and analyst forecast models if they are subjected to pricing cap regulation. The SUE decreased by 40 percentage points in the time series model and 155 percentage points in the analyst forecast model. This effect is only significant for the IPO year, indicating that firms improve their information environments when they are subjected to pricing cap regulation.

## 5.3 Policy Effectiveness and Real Cost

### 5.3.1 Willingness to go public

[Insert Table 6 here]

In the long term, if good-quality firms cannot minimize the negative impact of price-cap regulations, firms may be less willing to have an IPO. Table 6 Panel A Columns (1) and (2) show the results, indicating no evidence that the price-cap regulation affects the willingness of a firm to go public.

### 5.3.2 Cost of Borrowing (WACC)

[Insert Table 6 here]

In the long term, the market friction triggered by the price-cap regulation in the IPO market may potentially increase the firm's financial demand from other markets (e.g., corporate bonds and bank loans), which may increase the firm's cost of capital. I investigate the weighted average cost of capital in the post-IPO period. Columns (3) to (6) of Table 6 Panel A show the results, which indicate that the weighted average cost of capital for the affected firm, on average, decreased by 1% during the price-cap period.

### 5.3.3 Investment

[Insert Table 6 here]

In the long term, if a firm successfully minimizes the effect of the price cap, we shall not observe a significant impact on firm investment. Table 6 Panel B shows that compared to firms not affected by the regulation, the affected firms had slightly higher R&D investment and capital expenditures during the price period, with R&D increasing by 1% and CAPEX increasing by 4% in the first year of an IPO.

#### 5.3.4 Price Cascade, and Information Environment

[Insert Table 6 here]

Table 6 Panel C shows that firms affected by the price-cap regulation were less likely to result in a price cascade. The affected firms had a 43% lower probability of a price cascade after being listed for five days. Further, price-cap regulations were found to effectively reduce the probability of a price cascade even after only one month of the listing. On average, the affected firm had a 51% lower probability, resulting in a price cascade during the first month of the listing.

#### 5.3.5 Earning Management

[Insert Table 6 here]

Table 6 Panel D shows that firms affected by price-cap regulation changed their incentives for earnings management. Firms affected by price-cap regulations anticipate that their IPOs will be poorly funded; thus, they do not have an incentive to manipulate earnings before their IPOs. However, when the market becomes less distorted, the affected firms attempt to boost their earnings to increase the funding raised in the SEO market. I used the modified Jones model by Dechow et al. [1995] to capture earnings management. The detailed calculation is shown in Appendix A. Table 6 lists the results. It shows that firms affected by the price cap had 4% lower discretion accruals before their IPOs. Firms increased discretion accruals by 1% after being listed for two or three years. These results confirm Hypothesis VI.

### 5.4 Robustness Test

[Insert Table A2 here]

The direct estimation of the difference between the two groups could result in a biased result because a firm may arbitrarily decide whether it is treated. Although it is rather unlikely for firms to choose whether they are treated, I performed a robustness test using the propensity score matching (PSM) method. The PSM method requires a large sample size and common support, which may significantly reduce the economic magnitude of the coefficients and the sample size. As shown in Table A2, after applying the PSM-DID technique, all results held, indicating that the effect is unlikely to be driven by sample selection or self-selected treatment problems, thus confirming all hypotheses.

## 6 Discussion and Conclusions

This study shows that price caps distort the stock price, prevent information transmission between investors and firms, and reduce the motivation for earnings management in the short term. These results are in line with Bekaert et al. [2005]'s hypothesis, where the short-term impacts on CARs after IPOs and their disappearance in SEO show that price caps may distort the firm's stock market performance. In addition, price caps increase the cost of IPOs, prompting affected firms to shift their financing to SEOs, corporate bonds, and bank loan markets to strategically minimize the cost of borrowing. However, price caps may not have a homogenous impact on all firms. Firms affected by a price cap have different earnings management incentives, resulting in different behaviors before IPOs or SEOs. During the price-cap period, firms strongly prefer stock financing after having SEO issuance. A firm may switch to less expensive external finance (e.g., bond and loan financing) when they are not eligible for SEO issuance. In line with [Welch, 1996], the findings reveal that firms subject to the price cap are more likely to issue SEOs and have a faster speed of conducting their first SEO issuance. Further, firms actively self-disclose to address the intensified information asymmetry problem. In the long run, firms are more likely to improve earnings management using the SEO market to boost stock prices and alter the choice between CEOs' salaries and stock options. Nevertheless, the study does not find evidence that firm investment and the cost of capital increase by price-cap regulations. Moreover, the price cap does not affect the firm's willingness to go public. Furthermore, the study shows that ownership retention can be driven by government regulations and stock market distortion, with firms being less willing to dilute their ownership, especially when they are less funded by IPOs.

Regarding the theoretical contribution of the study, although the current literature focuses on how government regulation distorts market efficiency, by investigating the dynamic equilibrium of government regulation, firm finance, and investment behavior, my single-country focus and unique identification technique suggest that firms can act strategically to escape regulation, even when the market is heavily regulated.

Further, the study has significant practical implications for policy makers. The CSRC recently removed the price cap in the Sci-Tech innovation board (STAR) and the Growth Enterprise Market (GEM). My results shows that the homogenous price cap distorts resource allocation in industries with a P/E ratio greater than 23, thus affecting market efficiency. Therefore, although the market forces can avert distortion, policy makers have to improve price-cap regulations to achieve market efficiency and optimal resource allocation.

## A Discretion Accrual

Using Dechow et al. [1995]'s modified Jones model, I capture the "window dressing" phenomenon. I first estimate the predicted total accrual based on the following regression:

$$\frac{TA_{it}}{A_{it-1}} = \alpha_i[\frac{\Delta REV_{it}}{A_{it-1}}] + \beta_{1i}[\frac{1}{A_{it-1}}] + \beta_{2i}[\frac{PPE_{it}}{A_{it-1}}] + \epsilon_{it}, \tag{2}$$

where

$$TotalAccural_{it} = OperationRevenue - NetOperationCashFlow \tag{3}$$

$$DiscretionaryAccrual = \frac{TotalAccural_{it}}{A_{it-1}} - \widehat{\alpha}_{j,t}[\frac{\Delta REV_{it} - \Delta AR_{it}}{A_{it-1}}] - \widehat{\beta}_{1j,t}[\frac{1}{A_{it-1}}] - \widehat{\beta}_{2j,t}[\frac{PPE_{it}}{A_{it-1}}] \tag{4}$$

$A_{it-1}$ is the total assets in the previous year, $\Delta REV_{it}$ is the changes in operating income in the current year, and PPE is the net fixed assets. $\Delta AR_{it}$ is the change in account receivable in the current year.

I first perform a regression by industry and year based on equation (2) to obtain the regression coefficients. Then, they are inserted into Equation (4) to estimate the discretionary accrual.

## B Weighted Average Capital Cost (WACC)

I capture the cost of borrowing using WACC which is calculated as following:

Total capital = total owners' equity + provision for asset impairment
- provision for impairment of construction in progress - net amount of construction in progress
+ deferred income tax liabilities - deferred income tax assets + short-term loans
+ trading financial liabilities + non-current liabilities due within one year + long-term loans
+ bonds payable + long-term accounts payable
(5)

$$WACC = CBC \times (1 - Tax) \times W_B + CEC \times W_E$$

$$CEC = r_f + \beta \times r_E$$

(6)

CBC is the cost of bond capital. CEC is the cost of equity capital. Tax is the corporate tax rate. $W_B$ is the weight of the corporate bond over total capital. $W_E$ is the weight of equity over total capital. $r_f$ is the risk free rate, $\beta$ is the risk factor, and $r_E$ is the market risk premium.

The cost of bond capital is defined as the one-year bank loan interest rate. The risk-free rate of return is defined as the bank's one-year deposit interest rate. I uses the beta value weighted by the circulating market value of stocks in the Shanghai and Shenzhen stock markets (SSE and SZSE) for 250 trading days as the risk factors. Considering the excessive volatility of China's stock market, the market risk premium is defined as 4%.

## C Standardized Unexpected Earnings

We use SUE as a proxy for price informativeness; this had been widely used in previous studies [Foster et al., 1984]. Unexpected earnings (UE) measures is

$$UE_{i,t} = EPS_{i,t} - E(EPS_{i,t}) = EPS_{i,t} - EPS_{i,t-1} \tag{7}$$

where i indexes for stocks and t indexes for year, $EPS_{i,t}$ is earnings per share.

Next, we calculate standardized unexpected earnings (SUE) [Foster et al., 1984] as

$$SUE_{i,t} = \frac{UE_{i,t}}{|EPS_{i,t-1}|} \tag{8}$$

Traditionally, the per-share earnings are assumed to follow a random walk, and we can use the actual earnings for the current period as the forecast value of current earnings, when we want to estimate expected earnings. Another way to form earnings estimates is to use analyst forecasts, which tend to be more accurate than time series in predicting annual earnings. In this paper, I use both time series and analyst forecasts, and denote them as *SUETime* and *SUEAnalyst*. The time series model assumes that $E(EPS_{i,t}) = EPS_{i,t-1}$ while the analyst forecasts model assumes $E(EPS_{i,t}) = \widehat{EPS_{i,t-1}}$, where $\widehat{EPS_{i,t-1}}$ equals the analyst?s prediction of EPS from the previous year.

**Table A1: Variable Definitions**

The table defines the variables used in my analyses and lists their data sources.

| Variable | Definition | Source |
|---|---|---|
| ***A. Variables of Interest*** | | |
| CAR_IPO | Cumulative abnormal return of IPO. | Wind |
| CAR_SEO | Cumulative abnormal return of SEO. | Wind |
| Cascade | Dummy variable equals one if the stock price of the firm drops below the offering price | Wind |
| Initial_Return | Cumulative stock return between open-board day and IPO date $InitialReturn = (\frac{R_O}{R_L})$ | Wind |
| Debt_Loan | Total amount of corporate debt issuance and loan granted scaled by total assets | CSMAR |
| Have_SEO | Indicator that the firm has SEO issuance | Wind |
| First_SEO | The date difference between firm's first SEO and IPO | Wind |
| log_First_SEO | The natural log of *FirstSEO* | Wind |
| log$SEO | The natural log of the total funding raised in the SEO issuance. | Wind |
| Own_Dilute | Change in stock capital in the IPO scaled by stock capital after IPO | Wind |
| Issue400M | Indicator variable equal one if the firm has stock capital greater than 400 million CNY before the IPO issuance | Wind |
| Own_Dilute_SEO | Change in stock capital in SEOs scaled by stock capital after SEOs | Wind |
| DAccrual | Discretionary accruals, refer to Appendix A | Wind |
| Researcher | Natural log of the number of researcher conducting field research | CSMAR |
| Inst_Researchers | Natural log of the number of researcher from institution conducting field research | CSMAR |
| Length | Natural log of words in the field research report | CSMAR |
| Transparency (Voluntarily Disclosure) | Transparency index powered by CSRC (4-Excellent; 3-Good; 2-Passed; 1-Not Pass). The index is constructed based on the *Measures for the assessment of the information disclosures of companies listed on the Shenzhen Stock Exchange* and *Measures for the administration of information disclosure of listed companies* | CSMAR |
| Analyst | Natural log of the number of analysts (teams) have tracked and analyzed the company in a year | CSMAR |
| Report | Natural log of the number of research report have tracked and analyzed the company in a year | CSMAR |
| CAPX | Capital expenditure$[(\Delta FixedAsset + DepreciationExpense)/FixedAsset]$ | Wind |
| R&D | R&D expense scaled by total assets | Wind |
| inQueue | Number of firms that line up for listing | Wind |
| WACC | Weighted average cost of capital. Refer to Appendix B | CSMAR |
| SUE_time | Price informativeness, refer to Appendix C | CSMAR |
| SUE_Analyst | Price informativeness, refer to Appendix C | CSMAR |

| Variable | Definition | Source |
|---|---|---|
| ***B. Control Variables*** | | |
| Leverage | Book leverage ratio, total liability scaled by total assets | Wind |
| Size | Natural logarithm of total assets | Wind |
| EBITDA | Profitability, EBITDA scaled by total assets | Wind |
| PE_VC | Indicator variable equal to one of the firms is backed by private equity, or venture capital | Wind |
| SOE | Indicator variable equal one if the firm is an SOE | Wind |
| Age | Natural log of the number of years from the establishment of the firm | Wind |
| Growth | Growth rate of operation revenue | Wind |
| TobinQ | Total market value of the firm/ total asset value of the firm | Wind |
| ROA | Net income/ total asset | Wind |
| HHI_Top_Share | Herfindal index of the share holding ratio of the top ten shareholders before the IPO | CSMAR |
| Top_Share | Proportion of share hold by the largest shareholder before IPO | CSMAR |

**Table A2: Robustness: PSM-DID**

This table reports the results from the PSM-DID regressions of post-IPO CAR, initial return, financing activity, self-disclosure, investment activity, willingness to launch an IPO, cost of borrowing, and probability of price cascade. Column (2) in Panel B and Columns (7) to (10) report the marginal effect of the Probit regression. The rest of the columns report the results of the OLS regressions. For Panel A, the dependent variables are post-IPO CAR at 1 day, 5 days, 10 days, 20 days, and 30 days and the initial return of the IPO. The dependent variables of Panel B are the percentage change in ownership (*OwnDilute*) in the IPO; a dummy variable equal to 1 if the firm has a stock capital greater than 400 million RMB before IPO issuance and 0 otherwise (*Issue400M*); a percentage change in ownership (*OwnDiluteSEO*) in the SEO; and post-IPO corporate bond and bank loan usage in the IPO year and 1 year, 2 years, and 3 years after the IPO. For Panel C, the dependent variables are the natural log of the number of analysts tracking and analyzing the firm (*Analyst*) and its field report (*Report*) at 0 years, 1 year, 2 years, and 3 years after the firm goes public. For Panel D, the dependent variables are post-IPO R&D expenditure and capital expenditure for the current year, first year, second year, and third years. For Panel E, the dependent variables are the number of firms lined up for listing (*inQueue*); its natural logarithm (*loginQueue*); and WACC at 0 years, 1 year, 2 years, and 3 years after the firm goes public. The dummy variable equals 1 if the stock price of the firm drops below its offer price at 1 day, 5 days, 10 days, 20 days, and 30 days and 0 otherwise. All regressions control for year and industry-fixed effects. In parentheses are t-stats based on robust standard errors adjusted for heteroscedasticity and firm-level clustering. ***,**, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively . All variable definitions are in Table A1.

(*continuing on the next page...*)

*(...continued from the previous page)*

***Panel A: Price Anomaly***

| | (1) CAR_IPO in (0,1)day | (2) CAR_IPO in (0,5)days | (3) CAR_IPO in (0,10)days | (4) CAR_IPO in (0,20)days | (5) CAR_IPO in (0,30)days | (6) Initial_Return |
|---|---|---|---|---|---|---|
| Treated×Post | 0.09*** | 0.45*** | 0.87*** | 1.42*** | 1.26*** | 2.17*** |
| | (6.66) | (8.10) | (5.62) | (4.39) | (4.39) | (3.50) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 266 | 266 | 266 | 266 | 266 | 266 |
| adj. $R^2$ | 0.614 | 0.698 | 0.636 | 0.515 | 0.522 | 0.531 |

***Panel B: Firm Financing***

| | (1) Own_Dilute | (2) Issue400M | (3) Own_Dilute_SEO | (4) Debt_Loan in (0,1)Year | (5) Debt_Loan in (1,2)Year | (6) Debt_Loan in (2,3)Year | (7) Debt_Loan in (3,4)Year |
|---|---|---|---|---|---|---|---|
| Treated×Post | -0.06*** | 0.67*** | 0.42 | 0.06*** | 0.06** | 0.01 | -0.03 |
| | (-3.14) | (5.58) | (0.00) | (3.24) | (2.14) | (0.42) | (-0.99) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 258 | 120 | 18 | 266 | 261 | 238 | 233 |
| adj. $R^2$or Pseudo $R^2$ | 0.586 | 0.624 | 0.00 | 0.659 | 0.582 | 0.695 | 0.628 |

***Panel C: Firm Information Environment***

| | (1) Analyst in (0,1)Year | (2) Analyst in (1,2)Year | (3) Analyst in (2,3)Year | (4) Analyst in (3,4)Year | (5) Report in (0,1)Year | (6) Report in (1,2)Year | (7) Report in (2,3)Year | (8) Report in (3,4)Year |
|---|---|---|---|---|---|---|---|---|
| Treated×Post | -0.06*** | -0.04 | -0.03 | -0.07** | -0.09*** | -0.06 | -0.08 | -0.17** |
| | (-3.83) | (-1.10) | (-1.02) | (-2.19) | (-3.39) | (-0.78) | (-1.17) | (-2.16) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 256 | 224 | 215 | 218 | 256 | 224 | 215 | 218 |
| adj. $R^2$ | 0.651 | 0.525 | 0.651 | 0.617 | 0.661 | 0.498 | 0.611 | 0.695 |

***Panel D: Firm Investment***

| | (1) R&D in (0,1)Year | (2) R&D in (1,2)Year | (3) R&D in (2,3)Year | (4) R&D in (3,4)Year | (5) CAPX in (0,1)Year | (6) CAPX in (1,2)Year | (7) CAPX in (2,3)Year | (8) CAPX in (3,4)Year |
|---|---|---|---|---|---|---|---|---|
| Treated×Post | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 | -0.01 | 0.03 |
| | (0.99) | (1.17) | (0.67) | (0.29) | (0.90) | (0.73) | (-0.30) | (0.81) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 266 | 261 | 238 | 233 | 266 | 261 | 238 | 233 |
| adj. $R^2$ | 0.817 | 0.821 | 0.694 | 0.711 | 0.349 | 0.636 | 0.627 | 0.611 |

***Panel E: Policy Effectiveness***

| | (1) inQueue | (2) log(1+inQueue) | (3) WACC in (0,1)Year | (4) WACC in (1,2)Year | (5) WACC in (2,3)Year | (6) WACC in (3,4)Year | (7) Cascade in (0,5)days | (8) Cascade in (0,10)days | (9) Cascade in (0,20)days | (10) Cascade in (0,30)days |
|---|---|---|---|---|---|---|---|---|---|---|
| Treated×Post | 1.64 | -0.02 | -0.01*** | -0.01*** | -0.01*** | -0.01*** | 0.00 | -1.83** | -1.96*** | -2.20*** |
| | (0.60) | (-0.12) | (-6.35) | (-3.66) | (-3.89) | (-4.11) | (.) | (-2.17) | (-2.66) | (-6.28) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 158 | 158 | 266 | 261 | 238 | 233 | 116 | 139 | 162 | 162 |
| adj. $R^2$or Pseudo $R^2$ | 0.634 | 0.779 | 0.748 | 0.717 | 0.660 | 0.657 | 0.324 | 0.414 | 0.423 | 0.441 |

## References


Joseph Aharony, Chi-Wen Jevons Lee, and Tak Jun Wong. Financial packaging of ipo firms in china. *Journal of Accounting Research*, 38(1):103–126, 2000.

Geert Bekaert and Campbell R Harvey. Foreign speculators and emerging equity markets. *Journal of Finance*, 55(2):565–613, 2000.

Geert Bekaert, Campbell R Harvey, and Christian Lundblad. Does financial liberalization spur growth? *Journal of Financial Economics*, 77(1):3–55, 2005.

James R Brown, Steven M Fazzari, and Bruce C Petersen. Financing innovation and growth: cash flow, external equity, and the 1990s r&d boom. *Journal of Finance*, 64(1):151–185, 2009.

James R Brown, Gustav Martinsson, and Bruce C Petersen. Do financing constraints matter for r&d? *European Economic Review*, 56(8):1512–1529, 2012.

Ugur Celikyurt, Merih Sevilir, and Anil Shivdasani. Going public to acquire? the acquisition motive in ipos. *Journal of Financial Economics*, 96(3):345–363, 2010.

Kalok Chan, Junbo Wang, and KC John Wei. Underpricing and long-term performance of ipos in china. *Journal of Corporate Finance*, 10(3):409–430, 2004.

Eddy Chang, Chao Chen, Jing Chi, and Martin Young. Ipo underpricing in china: new evidence from the primary and secondary markets. *Emerging Markets Review*, 9(1):1–16, 2008.

Youngsoon S Cheon, Moonchul Kim, and Munho Hwang. Post-initial public offering earnings management driven by insider selling motives: using kosdaq initial public offerings. *Asia-Pacific Journal of Financial Studies*, 40(5):627–657, 2011.

Patricia M Dechow, Richard G Sloan, and Amy P Sweeney. Detecting earnings management. *Accounting Review*, pages 193–225, 1995.

Elizabeth Demers and Philip Joos. Ipo failure risk. *Journal of Accounting Research*, 45(2):333–371, 2007.

Lily H Fang, Josh Lerner, and Chaopeng Wu. Intellectual property rights protection, ownership, and innovation: evidence from china. *Review of Financial Studies*, 30(7):2446–2477, 2017.

Steven M. Fazzari, R. Glenn Hubbard, Bruce C. Petersen, Alan S. Blinder, and James M. Poterba. Financing constraints and corporate investment. *Brookings papers on Economic Activity*, (1):141–206, 1988.

George Foster, Chris Olsen, and Terry Shevlin. Earnings releases, anomalies, and the behavior of security returns. *Accounting Review*, pages 574–603, 1984.

Jie Gan. Privatization in china: experiences and lessons. In *China's Emerging Financial Markets*, pages 581–592. Springer, 2009.

Bronwyn H Hall, Jacques Mairesse, and Pierre Mohnen. Measuring the returns to r&d. In *Handbook of The Economics of Innovation*, volume 2, pages 1033–1082. Elsevier, 2010.

Johan Hombert and Adrien Matray. Can innovation help us manufacturing firms escape import competition from china? *Journal of Finance*, 73(5):2003–2039, 2018.

Shaoan Huang and Gang Zhang. Analysis on equity financing preference of chinese listed companies. *Journal of Economic Research*, 11(1):77–83, 2001.

Bharat A Jain and Omesh Kini. The post-issue operating performance of ipo firms. *Journal of Finance*, 49 (5):1699–1726, 1994.

Narasimhan Jegadeesh, Mark Weinstein, and Ivo Welch. An empirical investigation of ipo returns and subsequent equity offerings. *Journal of Financial Economics*, 34(2):153–175, 1993.

Michael C Jensen. Agency costs of free cash flow, corporate finance, and takeovers. *American Economic Review*, 76(2):323–329, 1986.

Kathleen M Kahle and René M Stulz. Access to capital, investment, and the financial crisis. *Journal of Financial Economics*, 110(2):280–299, 2013.

Kenneth A Kim and S Ghon Rhee. Price limit performance: evidence from the tokyo stock exchange. *Journal of Finance*, 52(2):885–901, 1997.

Kenneth A Kim, Pattanaporn Kitsabunnarat, and John R Nofsinger. Ownership and operating performance in an emerging market: evidence from thai ipo firms. *Journal of Corporate Finance*, 10(3):355–381, 2004.

Woojin Kim and Michael S Weisbach. Motivations for public equity offers: an international perspective. *Journal of Financial Economics*, 87(2):281–307, 2008.

Charles MC Lee, Yuanyu Qu, and Tao Shen. Going public in china: Reverse mergers versus ipos. *Journal of Corporate Finance*, 58:92–111, 2019.

Hayne E Leland and David H Pyle. Informational asymmetries, financial structure, and financial intermediation. *Journal of Finance*, 32(2):371–387, 1977.

Michael Lemmon and Michael R Roberts. The response of corporate financing and investment to changes in the supply of credit. *Journal of Financial and Quantitative Analysis*, 45(3):555–587, 2010.

Mario Levis. Seasoned equity offerings and the short-and long-run performance of initial public offerings in the uk. *European Financial Management*, 1(2):125–146, 1995.

YingChi Liu, Jiangchao Meng, and Hongquan Zhu. A dilemma? the price earning ratui cao if ipo pricing in chinese stock markets. *China Journal of Econometrics (in Chinese)*, 1(2):409, 2021.

Zhiyuan Liu, Kai Zheng, and Yanan He. Is the first-stage book-building reform effective. *Journal of Financial Research (in Chinese)*, (4):158–173, 2011.

Prem G Mathew. Long-horizon seasoned equity offerings performance in pacific rim markets. *Review of Financial Economics*, 11(4):317–333, 2002.

John J McConnell and Henri Servaes. Equity ownership and the two faces of debt. *Journal of Financial Economics*, 39(1):131–157, 1995.

Franco Modigliani and Merton H Miller. The cost of capital, corporation finance and the theory of investment. *American Economic Review*, 48(3):261–297, 1958.

Henry MK Mok and YV Hui. Underpricing and aftermarket performance of ipos in shanghai, china. *Pacific-Basin Finance Journal*, 6(5):453–474, 1998.

Stewart C. Myers and Nicholas S. Majluf. Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics*, 13(2):187–221, 1984.

Marco Pagano, Fabio Panetta, and Luigi Zingales. Why do companies go public? an empirical analysis. *Journal of Finance*, 53(1):27–64, 1998.

Peter Roosenboom, Tjalling van der Goot, and Gerard Mertens. Earnings management and initial public offerings: evidence from the netherlands. *The International Journal of Accounting*, 38(3):243–266, 2003.

Joseph A Schumpeter. The theory of economic development: an inquiry into profits, capital, credit, interest, and the business cycle. *Transaction Publishers.–1982.–January*, 1:244, 1982.

Myron B Slovin, Marie E Sushka, and Yvette M Bendeck. Seasoned common stock issuance following an ipo. *Journal of Banking & Finance*, 18(1):207–226, 1994.

Jeremy C. Stein. Chapter 2 - agency, information and corporate investment. Handbook of The Economics of Finance, pages 111–165. Elsevier, 2003.

RenéM Stulz. Managerial discretion and optimal financing policies. *Journal of Financial Economics*, 26 (1):3–27, 1990.

Dongwei Su and Belton M Fleisher. An empirical investigation of underpricing in chinese ipos. *Pacific-*

*Basin Finance Journal*, 7(2):173–202, 1999.

Siew Hong Teoh, Ivo Welch, and Tak J Wong. Earnings management and the long-run market performance of initial public offerings. *Journal of Finance*, 53(6):1935–1974, 1998a.

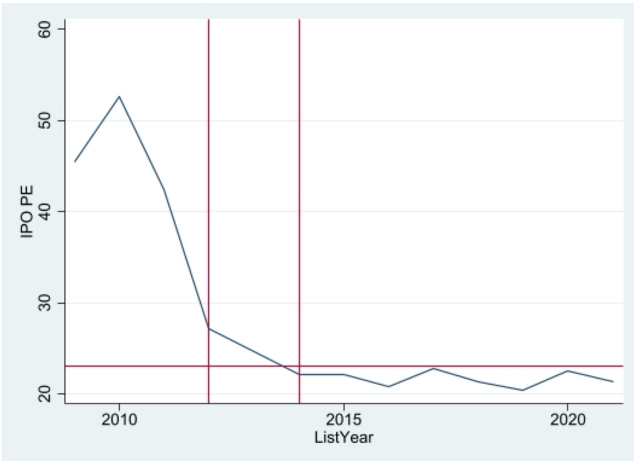
Siew Hong Teoh, Tak J Wong, and Gita R Rao. Are accruals during initial public offerings opportunistic? *Review of Accounting Studies*, 3(1):175–208, 1998b.

Lihui Tian. Regulatory underpricing: determinants of chinese extreme ipo returns. *Journal of Empirical Finance*, 18(1):78–90, 2011.

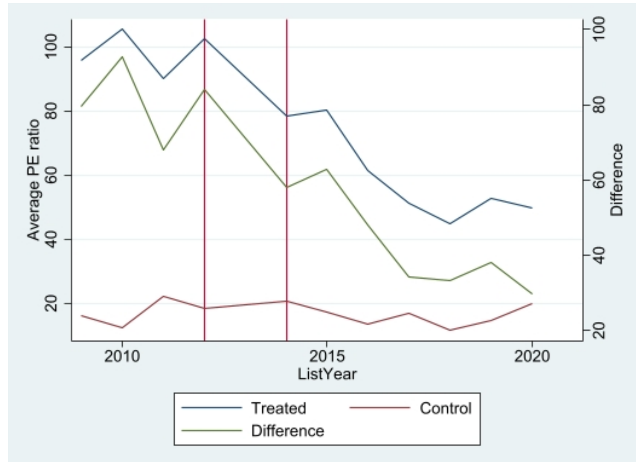
Ivo Welch. Seasoned offerings, imitation costs, and the underpricing of initial public offerings. *Journal of Finance*, 44(2):421–449, 1989.

Ivo Welch. Equity offerings following the ipo theory and evidence. *Journal of Corporate Finance*, 2(3): 227–259, 1996.

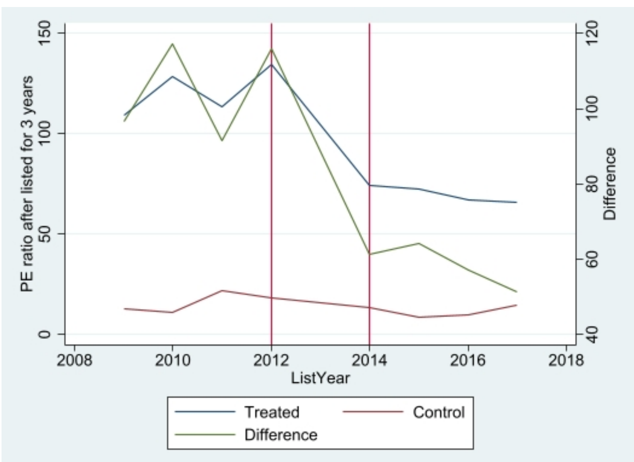
**(a)** IPO PE




**(b)** Average P/E ratio for each year




**(c)** P/E ratio for firms listed for at least 3 years




**(d)** Average P/E ratio for new listed firms

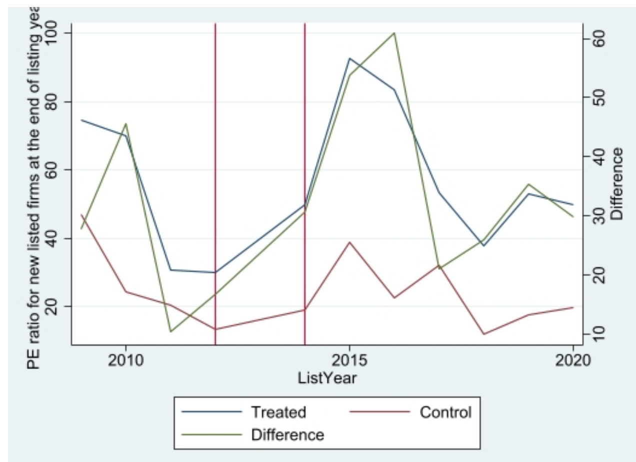



**Figure 1: Pricing Cap and P/E ratio**

This figure plots the pricing cap, IPO P/E ratio, P/E ratio for firms listed for at least 3 years, average P/E ratio, average P/E ratio for new listed firms at the end of listing year and for each year from 2009 to 2020. The sample only includes the firms listed on the main board. There are 6,961 firm-year observations for 1296 firms. The spaces between two vertical lines indicate the IPO suspension period, which I exclude from my sample. The horizontal line is the pricing cap limit equal to 23.

**Table 1: Summary Statistics for all Variables**

This table reports descriptive statistics for the variables of interest. Panel A is for the full sample. Panel B shows t-statistics for treatment and control firms. All variables are winsorized at 1% and 99%. All variable definitions are in Table A1.

| Variable | N | mean | sd | p1 | p25 | p50 | p75 | p99 |
|---|---|---|---|---|---|---|---|---|
| ***Panel A: The Full Sample*** | | | | | | | | |
| CAR_SEOin1Day | 710 | -0.00 | 0.03 | -0.08 | -0.02 | -0.00 | 0.01 | 0.10 |
| CAR_SEOin5Day | 710 | -0.00 | 0.05 | -0.12 | -0.03 | -0.01 | 0.02 | 0.14 |
| CAR_SEOin10Day | 710 | -0.01 | 0.06 | -0.16 | -0.05 | -0.01 | 0.03 | 0.21 |
| CAR_SEOin20Day | 710 | -0.01 | 0.09 | -0.23 | -0.06 | -0.02 | 0.03 | 0.27 |
| CAR_SEOin30Day | 710 | -0.02 | 0.11 | -0.30 | -0.08 | -0.03 | 0.04 | 0.37 |
| CARin1Day | 6961 | 0.03 | 0.07 | -0.11 | -0.02 | 0.05 | 0.10 | 0.14 |
| CARin5Day | 6961 | 0.20 | 0.31 | -0.25 | -0.06 | 0.06 | 0.60 | 0.67 |
| CARin10Day | 6961 | 0.41 | 0.65 | -0.28 | -0.07 | 0.07 | 0.89 | 1.68 |
| CARin20Day | 6961 | 0.56 | 1.12 | -0.32 | -0.08 | 0.06 | 0.79 | 5.69 |
| CARin30Day | 6961 | 0.53 | 1.06 | -0.32 | -0.10 | 0.07 | 0.75 | 5.86 |
| Cascadein1Day | 6961 | 0.11 | 0.31 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| Cascadein5Day | 6961 | 0.13 | 0.34 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| Cascadein10Day | 6961 | 0.16 | 0.36 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| Cascadein20Day | 6961 | 0.18 | 0.38 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| Cascadein30Day | 6961 | 0.18 | 0.39 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| Initial_Return | 6961 | 1.61 | 2.27 | -5.25 | 0.19 | 1.39 | 2.79 | 17.61 |
| Debt_Loan | 6961 | 0.13 | 1.18 | 0.00 | 0.00 | 0.04 | 0.16 | 59.88 |
| Have_SEO | 6961 | 0.10 | 0.30 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| First_SEO | 710 | 1371 | 631.7 | 414.0 | 910.0 | 1254 | 1771 | 3341 |
| Own_Dilute | 6936 | 0.23 | 0.05 | 0.08 | 0.25 | 0.25 | 0.25 | 0.30 |
| Issuer400M | 6961 | 0.22 | 0.41 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| Own_Dilute_SEO | 710 | 0.12 | 0.11 | 0.01 | 0.06 | 0.10 | 0.15 | 0.72 |
| Researcher | 2969 | 3.19 | 1.60 | 0.00 | 2.08 | 3.30 | 4.32 | 8.22 |
| Inst_Researcher | 2960 | 3.16 | 1.60 | 0.00 | 2.08 | 3.26 | 4.30 | 8.21 |
| Length | 3306 | 9.11 | 1.61 | 1.79 | 8.29 | 9.38 | 10.21 | 13.53 |
| Transparency | 4307 | 3.11 | 0.60 | 1.00 | 3.00 | 3.00 | 3.00 | 4.00 |
| Analyst | 5229 | 0.11 | 0.11 | 0.01 | 0.03 | 0.07 | 0.15 | 0.64 |
| Report | 5240 | 0.22 | 0.27 | 0.01 | 0.04 | 0.12 | 0.32 | 2.90 |
| DAccrual | 6961 | 0.03 | 0.12 | -6.68 | 0.00 | 0.02 | 0.07 | 4.81 |
| CAPX | 6961 | 0.03 | 1.44 | -89.17 | 0.00 | 0.00 | 0.05 | 4.81 |
| RD | 6961 | 0.02 | 0.02 | 0.00 | 0.01 | 0.02 | 0.03 | 0.25 |
| inQueue | 5070 | 13.31 | 21.57 | 1.00 | 2.00 | 7.00 | 11.00 | 104.0 |
| WACC | 6863 | 0.06 | 0.01 | 0.01 | 0.05 | 0.06 | 0.07 | 0.11 |
| SUE_Time | 945 | 0.04 | 0.71 | -2.55 | -0.28 | 0.01 | 0.25 | 4.55 |
| SUE_Analyst | 942 | -0.25 | 0.93 | -6.56 | -0.37 | -0.01 | 0.20 | 0.83 |
| Leverage | 6961 | 0.34 | 0.18 | 0.06 | 0.20 | 0.32 | 0.46 | 0.94 |
| Size | 6961 | 21.83 | 1.10 | 18.91 | 21.06 | 21.69 | 22.37 | 26.75 |
| EBITDA | 6961 | 0.09 | 0.05 | 0.00 | 0.06 | 0.08 | 0.11 | 0.37 |
| Growth | 6961 | 1.18 | 0.26 | 0.68 | 1.03 | 1.14 | 1.28 | 2.36 |
| TobinQ | 6961 | 1.50 | 0.94 | 0.32 | 0.89 | 1.21 | 1.81 | 5.67 |
| ROA | 6961 | 8.46 | 5.42 | 0.81 | 4.72 | 7.43 | 10.92 | 40.93 |
| HHI_Top_Share | 6961 | 0.38 | 0.20 | 0.02 | 0.23 | 0.35 | 0.49 | 0.93 |
| Top_Share | 6961 | 0.53 | 0.20 | 0.09 | 0.38 | 0.52 | 0.67 | 0.96 |
| Age | 6961 | 2.76 | 0.37 | 0.00 | 2.56 | 2.77 | 3.00 | 3.50 |
| PE_VC | 6961 | 0.12 | 0.33 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| SOE | 6961 | 0.15 | 0.36 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |

(*continuing on the next page...*)

(*...continued from the previous page*)

| ***Panel B: T-statistics for Treated and Control Groups*** | | | | | | |
|---|---|---|---|---|---|---|
| **Variable** | **Total** | **Mean (C)** | **Mean (T)** | **Control** | **Treated** | **t-stat** |
| CAR_SEOin1Day | 710 | 0.00 | -0.00 | 9 | 701 | 0.00 |
| CAR_SEOin5Day | 710 | 0.02 | -0.00 | 9 | 701 | 0.02 |
| CAR_SEOin10Day | 710 | 0.03 | -0.00 | 9 | 701 | 0.04 |
| CAR_SEOin20Day | 710 | 0.06 | -0.01 | 9 | 701 | 0.07** |
| CAR_SEOin30Day | 710 | 0.02 | -0.016 | 9 | 701 | 0.03 |
| CARin1Day | 6961 | -0.01 | 0.03 | 134 | 6827 | -0.04*** |
| CARin5Day | 6961 | -0.00 | 0.20 | 134 | 6827 | -0.21*** |
| CARin10Day | 6961 | 0.01 | 0.41 | 134 | 6827 | -0.40*** |
| CARin20Day | 6961 | -0.02 | 0.57 | 134 | 6827 | -0.60*** |
| CARin30Day | 6961 | -0.03 | 0.53 | 134 | 6827 | -0.57*** |
| Cascadein1Day | 6961 | 0.09 | 0.10 | 134 | 6827 | -0.01 |
| Cascadein5Day | 6961 | 0.29 | 0.12 | 134 | 6827 | 0.16*** |
| Cascadein10Day | 6961 | 0.29 | 0.15 | 134 | 6827 | 0.14*** |
| Cascadein20Day | 6961 | 0.36 | 0.17 | 134 | 6827 | 0.19*** |
| Cascadein30Day | 6961 | 0.36 | 0.17 | 134 | 6827 | 0.18*** |
| Initial_Return | 6961 | 1.68 | 1.60 | 816 | 6145 | 0.08 |
| Debt_Loan | 6961 | 0.09 | 0.13 | 816 | 6145 | -0.04 |
| Have_SEO | 6961 | 0.06 | 0.10 | 816 | 6145 | -0.04*** |
| First_SEO | 710 | 1165 | 1388 | 52 | 658 | -222.3* |
| Own_Dilute | 6936 | 0.21 | 0.23 | 791 | 6145 | -0.01*** |
| Issuer400M | 6961 | 0.46 | 0.18 | 816 | 6145 | 0.28*** |
| Own_Dilute_SEO | 710 | 0.11 | 0.12 | 52 | 658 | -0.01 |
| Researcher | 2969 | 3.02 | 3.19 | 122 | 2847 | -0.16 |
| Inst_Researcher | 2960 | 3.01 | 3.17 | 122 | 2838 | -0.15 |
| Length | 3306 | 8.90 | 9.11 | 148 | 3158 | -0.21 |
| Transparency | 4307 | 3.14 | 3.10 | 179 | 4128 | 0.03 |
| Analyst | 5229 | 0.10 | 0.10 | 573 | 4656 | -0.00 |
| Report | 5240 | 0.21 | 0.22 | 573 | 4667 | -0.01 |
| DAccrual | 6961 | 0.01 | 0.02 | 816 | 6145 | -0.00 |
| CAPX | 6961 | 0.05 | 0.03 | 816 | 6145 | 0.02 |
| RD | 6961 | 0.01 | 0.02 | 816 | 6145 | -0.00*** |
| inQueue | 5070 | 9.15 | 13.81 | 550 | 4520 | -4.66*** |
| WACC | 6863 | 0.05 | 0.06 | 770 | 6093 | -0.00*** |
| SUE_Time | 945 | 0.01 | 0.04 | 88 | 857 | -0.02 |
| SUE_Analyst | 942 | -0.10 | -0.26 | 88 | 854 | 0.15 |
| Leverage | 6961 | 0.41 | 0.33 | 816 | 6145 | 0.08*** |
| Size | 6961 | 22.52 | 21.74 | 816 | 6145 | 0.77*** |
| EBITDA | 6961 | 0.09 | 0.09 | 816 | 6145 | 0.00*** |
| Growth | 6961 | 1.15 | 1.18 | 816 | 6145 | -0.02** |
| TobinQ | 6961 | 1.03 | 1.56 | 816 | 6145 | -0.52*** |
| ROA | 6961 | 9.24 | 8.35 | 816 | 6145 | 0.89*** |
| HHI_Top_Share | 6961 | 0.46 | 0.37 | 816 | 6145 | 0.09*** |
| Top_Share | 6961 | 0.61 | 0.52 | 816 | 6145 | 0.09*** |
| Age | 6961 | 2.66 | 2.77 | 816 | 6145 | -0.11*** |
| PEVC | 6961 | 0.26 | 0.10 | 816 | 6145 | 0.15*** |
| SOE | 6961 | 0.35 | 0.12 | 816 | 6145 | 0.23*** |

(*continuing on the next page...*)

**Table 2: Natural Experiment: Cumulative Abnormal Return and Initial Return**

This table reports the results from DID regressions of post-IPO CAR, initial return, and post-SEO CAR. All columns report the results of OLS regressions. For Panel A, the dependent variables are post-IPO CAR at 1 day, 5 days, 10 days, 20 days, and 30 days and the initial return of the IPO. For Panel B, the dependent variables are post-SEO CAR at 1 day, 5 days, 10 days, 20 days, and 30 days. All regressions control for year and industry-fixed effects. In parentheses are t-stats based on robust standard errors adjusted for heteroscedasticity and firm-level clustering. ***,**, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. All variable definitions are in Table A1.

***Panel A: IPO Cumulative Abnormal Return (CAR_IPO) and Initial Return***

| | (1) CAR_IPO in (0,1)day | (2) CAR_IPO in (0,5)days | (3) CAR_IPO in (0,10)days | (4) CAR_IPO in (0,20)days | (5) CAR_IPO in (0,30)days | (6) Initial_Return |
|---|---|---|---|---|---|---|
| Treated×Post | 0.10*** | 0.55*** | 1.05*** | 1.54*** | 1.44*** | 2.12*** |
| | (24.28) | (40.50) | (32.19) | (19.64) | (20.47) | (12.04) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 6961 | 6961 | 6961 | 6961 | 6961 | 6961 |
| adj. $R^2$ | 0.539 | 0.830 | 0.736 | 0.495 | 0.493 | 0.254 |

***Panel B: SEO Cumulative Abnormal Return (CAR_SEO)***

| | (1) CAR_SEO in (0,1)day | (2) CAR_SEO in (0,5)days | (3) CAR_SEO in (0,10)days | (4) CAR_SEO in (0,20)days | (5) CAR_SEO in (0,30)days |
|---|---|---|---|---|---|
| Treated×Post | 0.00 | -0.00 | 0.01 | 0.01 | 0.01 |
| | (0.73) | (-0.15) | (0.86) | (0.53) | (0.87) |
| Controls | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| $N$ | 829 | 829 | 829 | 829 | 829 |
| adj. $R^2$ | -0.023 | -0.010 | -0.016 | -0.003 | 0.015 |

## Table 3: Natural Experiment: Financing

This table reports the results of DID regressions on firm financial activity. Column (2) in Panel A and Columns (1) and (2) in Panel B report the marginal effect of the Probit regression. The other columns report the OLS regression results. The dependent variables of Panel A are percentage change in ownership (*OwnDilute*) in the IPO, a dummy variable equal to 1 if the firm has stock capital greater than 400 million RMB before IPO issuance and 0 otherwise (*Issue400M*), and percentage change in ownership (*OwnDilute-SEO*) in the SEO. The dependent variables of Panel B are a dummy variable equal to 1 if the firm has an SEO issuance 1 year or 2 years after going public and 0 otherwise (*HaveSEO@2Year*, *HaveSEO@3Year*), the natural logarithm of the SEO issuance amount 1 year or 2 years after going public (*logSEOAmt@2Year*, *logSEOAmt@3Year*), and the natural logarithm of the date difference between the firm's first SEO and IPO (*logFirstSEO*). The dependent variables of Panel C are post-IPO corporate bond and bank loan usage in the IPO year and 1 year, 2 years, and 3 years after the IPO. All regressions control for year and industry-fixed effects. In parentheses are t-stats based on robust standard errors adjusted for heteroscedasticity and firm-level clustering. ***,**, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. All variable definitions are in Table A1.

***Panel A: Ownership Dilution***

| | (1) Own_Dilute | (2) Issue400M | (3) Own_Dilute_SEO |
|---|---|---|---|
| Treated×Post | -0.03*** | 0.15*** | -0.02 |
| | (-7.78) | (6.89) | (-1.23) |
| Controls | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes |
| $N$ | 6936 | 6743 | 710 |
| adj. $R^2$ or Pseudo $R^2$ | 0.292 | 0.433 | 0.107 |

***Panel B: SEO Issuance and Speed***

| | (1) Have_SEO in (1,2) Year | (2) Have_SEO in (2,3) Year | (3) log$SEO in (1,2) Year | (4) log$SEO in (2,3) Year | (5) logFirst_SEO |
|---|---|---|---|---|---|
| Treated×Post | 0.05*** | 0.02** | 0.23 | 0.56*** | -0.77*** |
| | (6.52) | (2.55) | (1.34) | (3.34) | (-14.45) |
| Controls | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| $N$ | 5641 | 5887 | 553 | 705 | 710 |
| adj. $R^2$ or Pseudo $R^2$ | 0.246 | 0.114 | 0.557 | 0.589 | 0.520 |

***Panel C: Corporate Bonds and Loans***

| | (1) Debt_Loan in (0,1)Year | (2) Debt_Loan in (1,2)Year | (3) Debt_Loan in (2,3)Year | (4) Debt_Loan in (3,4)Year |
|---|---|---|---|---|
| Treated×Post | 0.04*** | 0.03*** | 0.02*** | 0.01 |
| | (5.86) | (3.89) | (2.84) | (0.66) |
| Controls | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| $N$ | 6961 | 6907 | 6756 | 6539 |
| adj. $R^2$ | 0.420 | 0.43[illegible] | 0.417 | 0.375 |

**Table 4: Natural Experiment: Self Disclosure**

This table reports the results of DID regressions on self-disclosure. All columns report the results of OLS regressions. The dependent variables for Panel A are the firm transparency indices of the SZSE and SSE (*Transparent)* and natural log of the length of performance briefing (*Length*) in the IPO year and 1 year, 2 years, and 3 years after the IPO. The dependent variables for Panel B are the natural log of the number of researchers conducting field research (*NumResearcher)*, and the natural log of the number of researchers from institutions conducting field research (*Inst_Researcher*) in the IPO year and 1 year, 2 years, and 3 years after the IPO. All regressions control for year and industry-fixed effects. In parentheses are t-stats based on robust standard errors adjusted for heteroscedasticity and firm-level clustering. ***,**, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. All variable definitions are in Table A1.

***Panel A: Firm Transparency and Performance Briefing Informativeness***

| | (1) Transparency in (0,1)Year | (2) Transparency in (1,2)Year | (3) Transparency in (2,3)Year | (4) Transparency in (3,4)Year | (5) Length in (0,1)Year | (6) Length in (1,2)Year | (7) Length in (2,3)Year | (8) Length in (3,4)Year |
|---|---|---|---|---|---|---|---|---|
| Treated×Post | 0.14*** | 0.13** | 0.02 | -0.05 | 2.74*** | 2.13*** | 1.17*** | 0.40** |
| | (3.58) | (2.13) | (0.35) | (-0.70) | (11.41) | (11.27) | (6.06) | (2.03) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 4340 | 4320 | 4261 | 4180 | 2006 | 3181 | 3670 | 3586 |
| adj. $R^2$ | 0.110 | 0.131 | 0.159 | 0.219 | 0.535 | 0.425 | 0.241 | 0.185 |

***Panel B: Field Report Status***

| | (1) Researcher in (0,1)Year | (2) Researcher in (1,2)Year | (3) Researcher in (2,3)Year | (4) Researcher in (3,4)Year | (5) Inst_Researchers in (0,1)Year | (6) Inst_Researcher in (1,2)Year | (7) Inst_Researcher in (2,3)Year | (8) Inst_Researcher in (3,4)Year |
|---|---|---|---|---|---|---|---|---|
| Treated×Post | 0.01 | 0.75*** | 0.33 | 0.27 | -0.00 | 0.75*** | 0.32 | 0.27 |
| | (0.01) | (3.35) | (1.62) | (1.33) | (-0.01) | (3.39) | (1.57) | (1.34) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 708 | 1560 | 2799 | 3401 | 708 | 1560 | 2788 | 3398 |
| adj. $R^2$ | 0.316 | 0.317 | 0.263 | 0.271 | 0.314 | 0.317 | 0.257 | 0.268 |

**Table 5: Natural Experiment: Information Environment**

This table reports the results of DID regressions on information environment. All columns report the results of OLS regressions. For Panel A, the dependent variables are the natural log of the number of analysts tracking and analyzing the firm (*Analyst*) and the natural log of the number of analysts tracking and analyzing the firm?s field report (*Report*) at 0 year, 1 year, 2 years, and 3 years after the firm went public. For Panel B, the dependent variables are standardized unexpected earnings (SUE) calculated using time series and analyst prediction models for the current year, first year, second year, and third year. All regressions control for year and industry-fixed effects. In parentheses are t-stats based on robust standard errors adjusted for heteroscedasticity and firm-level clustering. ***,**, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. All variable definitions are in Table A1.

***Panel A: Analyst and Field Report Attention***

| | (1) Analyst in (0,1)Year | (2) Analyst in (1,2)Year | (3) Analyst in (2,3)Year | (4) Analyst in (3,4)Year | (5) Report in (0,1)Year | (6) Report in (1,2)Year | (7) Report in (2,3)Year | (8) Report in (3,4)Year |
|---|---|---|---|---|---|---|---|---|
| Treated×Post | -0.05*** | -0.01** | 0.00 | 0.03*** | -0.07*** | -0.01 | 0.01 | 0.05** |
| | (-9.79) | (-2.14) | (0.44) | (2.88) | (-8.63) | (-0.78) | (0.77) | (2.12) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 6551 | 5715 | 5388 | 5057 | 6551 | 5750 | 5426 | 5057 |
| adj. $R^2$ | 0.451 | 0.291 | 0.315 | 0.336 | 0.399 | 0.293 | 0.340 | 0.330 |

***Panel B: Standardized Unexpected Earnings (SUE)***

| | (1) SUE_Time in (0,1)Year | (2) SUE_Time in (1,2)Year | (3) SUE_Time in (2,3)Year | (4) SUE_Analyst in (0,1)Year | (5) SUE_Analyst in (1,2)Year | (6) SUE-Analyst in (2,3)Year |
|---|---|---|---|---|---|---|
| Treated×Post | -0.40*** | 0.08 | 0.33 | -1.55*** | -0.11 | 0.20* |
| | (-1.93e+13) | (1.23) | (0.97) | (-4.32e+13) | (-0.82) | (1.98) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 114 | 1232 | 973 | 114 | 1232 | 973 |
| adj. $R^2$ | 1.000 | 0.394 | 0.247 | 1.000 | 0.342 | 0.406 |

### Table 6: Natural Experiment: Long-term Impact and Policy Effectiveness

This table reports the results of DID regressions on the firm?s long-term impact and policy effectiveness. All columns report the results of OLS regressions for Panel A and B. All columns report the marginal effect of the Probit regression in Panel C. For Panel A, the dependent variables are the number of firms lining up for listing (N_inQueue), its natural logarithm (*logQueue*), and the WACC for 0 year, 1 year, 2 years, and 3 years after the firm went public. For Panel B, the dependent variables are post-IPO R&D expenditure and capital expenditure in the current year, first year, second year, and third year. For Panel C, the dependent variable is a dummy variable equal to 1 if the stock price of the firm drops below its offer price at 1 day, 5 days, 10 days, 20 days, and 30 days and 0 otherwise. For Panel D, the dependent variable is the post-IPO Discretionary Accrual in the current year, first year, second year, and third year after the firm went public. All regressions control for year and industry-fixed effects. In parentheses are t-stats based on robust standard errors adjusted for heteroscedasticity and firm-level clustering. ***,**, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. All variable definitions are in Table A1.

***Panel A: Willingness of IPO and Weighted Average Cost of Capital (WACC)***

| | (1) inQueue | (2) log(1+inQueue) | (3) WACC in (0,1)Year | (4) WACC in (1,2)Year | (5) WACC in (2,3)Year | (6) WACC in (3,4)Year |
|---|---|---|---|---|---|---|
| Treated×Post | 0.08 | -0.01 | -0.01*** | -0.01*** | -0.01*** | -0.01*** |
| | (0.24) | (-0.89) | (-27.50) | (-20.06) | (-13.89) | (-12.03) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 5070 | 5070 | 6961 | 6907 | 6756 | 6539 |
| adj. $R^2$ | 0.645 | 0.879 | 0.720 | 0.436 | 0.471 | 0.393 |

***Panel B: Investment***

| | (1) R&D in (0,1)Year | (2) R&D in (1,2)Year | (3) R&D in (2,3)Year | (4) R&D in (3,4)Year | (5) CAPX in (0,1)Year | (6) CAPX in (1,2)Year | (7) CAPX in (2,3)Year | (8) CAPX in (3,4)Year |
|---|---|---|---|---|---|---|---|---|
| Treated×Post | 0.01*** | 0.00*** | 0.00** | 0.00*** | 0.04*** | 0.02 | 0.01 | -0.00 |
| | (6.03) | (3.01) | (2.45) | (3.09) | (4.30) | (1.37) | (0.85) | (-0.31) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 6961 | 6907 | 6756 | 6539 | 6961 | 6907 | 6756 | 6539 |
| adj. $R^2$ | 0.366 | 0.394 | 0.396 | 0.396 | 0.099 | 0.090 | 0.111 | 0.091 |

***Panel C:Price Cascade***

| | (1) Cascade in (0,1) day | (2) Cascade in (0,5) days | (3) Cascade in (0,10) days | (4) Cascade in (0,20) days | (5) Cascade in (0,30) days |
|---|---|---|---|---|---|
| Treated×Post | 0.00 | -0.43*** | -0.43*** | -0.46*** | -0.51*** |
| | (.) | (-7.20) | (-6.79) | (-8.20) | (-7.17) |
| Controls | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| $N$ | 3703 | 6070 | 6404 | 6359 | 6295 |

***Panel D: Discretionary Accrual***

| | (1) DAccrual in (-1,0) Year | (2) DAccrual in (0,1) Year | (3) DAccrual in (1,2) Year | (4) DAccrual in (2,3) Year | (5) DAccrual in (3,4) Year |
|---|---|---|---|---|---|
| Treated×Post | -0.04*** | 0.00 | 0.00 | 0.01*** | 0.01*** |
| | (-6.95) | (0.00) | (1.17) | (3.19) | (2.72) |
| Controls | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
| $N$ | 6961 | 6961 | 6907 | 6756 | 6539 |
| adj. $R^2$ | 0.225 | 0.000 | 0.226 | 0.214 | 0.233 |