Three essays on corporate finance and research methodology

Li, Wei

2020

Li, W. (2020). Three essays on corporate finance and research methodology. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/138252>

<https://doi.org/10.32657/10356/138252>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 15:47:46 SGT

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**THREE ESSAYS ON CORPORATE FINANCE AND RESEARCH
METHODOLOGY**

WEI LI

NANYANG BUSINESS SCHOOL

2020

**THREE ESSAYS ON CORPORATE FINANCE AND RESEARCH
METHODOLOGY**

WEI LI

NANYANG BUSINESS SCHOOL

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Doctor of Philosophy

2020

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

26/03/2020

.....

Date

Handwritten signature of Wei LI in blue ink, consisting of stylized Chinese characters followed by the name 'Wei LI' in cursive script.

.....

Wei LI

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate with amendments, changes and improvements as suggested by me as the Supervisor. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

26/03/2020

.....

Date

.....

Xin Chang (Simba)

Authorship Attribution Statement

This thesis contains material from 3 papers accepted at conferences in which I am listed as an author.

Chapter 1, Investor-paid Credit Rating Agency and Stock Price Informativeness (sole paper), is presented at the Singapore Scholars Symposium.

Chapter 2, *P*-hacking in Experimental Accounting Studies (with Xin CHANG (Simba) and Huasheng GAO), is presented at the 2018 FMA Asia/Pacific Conference and the 2019 MIT Asia Conference in Accounting. The contributions of the co-authors are as follows:

1. Prof. Gao suggested the study of *p*-hacking in an experimental setting.
2. I collected the data and conducted the empirical analyses.
3. Prof. Chang proposed the simulation analyses and the solutions for *p*-hacking.
4. I wrote the first draft of the manuscript.
5. The manuscript was revised under the supervision of Prof. Chang.

Chapter 3, Local Gambling Preference and Corporate Cash Holding (sole paper), is presented at the 2017 China International Conference in Finance and the 2018 FMA Asia/Pacific Conference.

26/03/2020

.....

Date

.....

Wei LI

Acknowledgments

First of all, I would like to thank my two supervisors – Professor Xin Chang (Simba) and Professor Huasheng Gao – for their kindness and exceptional guidance. They have taught me essential knowledge in financial research, instilled to me the practice of excellent scholars, and showed me the joy of pursuing an academic career. They have set lifetime examples for me to follow.

I thank my dissertation committee – Professor Jiang Luo, Hong Ru, and Rui Shen – for their collective insights and guidance on my thesis. I would also like to express my gratitude to several other faculties – Professor Tao Chen, Zhanhui Chen, Steve Dimmock, Chuan Yang Hwang, Jun-Koo Kang, Huai Zhang – for creating a nurturing environment for learning and research.

I am indebted to Professor Qifei Zhu – my coauthor and mentor. As a brilliant researcher, he has inspired me with his creativity and empirical skills. As a trusted mentor, he helped me at every stage of my job search. I feel fortunate to have the opportunity to work with and learn from him.

Lastly, I am grateful for the continuous support from my parents and my wife, Lu Xie. Their love and trust have always been the driving force for me to overcome the challenges and difficulties in life. I also want to thank my PhD friends in NBS, their friendship and companion created many happy memories for me during the last five years.

Table of Contents

Summary.....	1
Chapter 1: Investor-paid Credit Rating Agency and Stock Price Informativeness	3
1. Introduction.....	4
2. Related literature	11
2.1 CRAs' incentives and their informational role	12
2.2 Credit rating agencies and corporate information disclosure	14
3. Data, variable, sample description	15
3.1 Data and sample selection	15
3.2 Variables	16
3.3 Determinants of coverage.....	19
4. Empirical results	20
4.1 The effect of EJR coverage on price non-synchronicity	20
4.2 Tests on endogeneity	22
5. Additional tests.....	28
5.1 Evidence on the information channel.....	28
5.2 Evidence on the disciplining channel	34
5.3 Propensity-score-matched samples.....	36
6. Conclusion	37
Appendix A: Variable Definitions	44
Appendix B: Alternative measures of stock price informativeness.....	46
Chapter 2: <i>P</i> -hacking in Experimental Accounting Studies	58
1. Introduction.....	59
2. Background information and related literature.....	65
2.1. Understanding p-values: what are we hacking?	65
2.2. Literature on p-hacking.....	68
3. Our sample and empirical strategy	71
3.1. Sample construction.....	71
3.2. Methodology	75
4. Empirical results	76
4.1. The cross-sectional distribution of p-values	76
4.2. Fitting the p-curve to an exponential distribution	79
4.3. P-values reported after inequality operator	82
5. What explain the extent of <i>p</i> -hacking?.....	84
5.1. Pressure.....	85

5.2. Opportunity	86
5.3. Rationalization	87
6. Additional tests	89
6.1. P-hacking and future citations	89
6.2. The extent of p-hacking for main results	90
6.3. JAR's updated data and code sharing policy	91
7. Potential remedies to guard against <i>p</i> -hacking	93
7.1. Multiple hypotheses testing	93
7.2. Simulations	97
7.3. The Bayesian approach	100
7.4. Disclosure-based solution	103
8. Conclusion	106
References	108
Appendix A: The distribution of p-values from archival-empirical studies	113
Chapter 3: Local Gambling Preference and Corporate Cash Holding	129
1. Introduction	130
2. Related literature	137
2.1. Religion-induced gambling preference	137
2.2. Corporate risk-taking and cash holding	138
3. Data and research design	139
3.1. County-level religious and demographic characteristics	139
3.2. Firm-level cash holding and other control variables	140
4. Main results	141
4.1. Local gambling preference and cash holding	141
4.2. Tests on endogeneity	142
4.3. Gambling preference, corporate investment, and financial constraint	145
4.4. Gambling preference and the value of excessive cash holding	148
5. Additional analyses	150
5.1. Excessive cash and speed of adjustment	150
5.2. Spurious relation	151
5.3. General robustness checks	152
5.4. Using lottery sale to measure gambling preference	153
6. Conclusion	153
References	158

List of Tables

Table 1.1: Summary statistics	47
Table 1.2: Determinants of EJR and S&P's coverage	48
Table 1.3: Effect of EJR coverage on stock price informativeness	50
Table 1.4: Tests of omitted variable bias and reverse causality	51
Table 1.5: Using quasi-natural experiments that shock EJR's importance	53
Table 1.6: Testing the information channel	54
Table 1.7: EJR coverage and forecasting price efficiency	55
Table 1.8: Testing the disciplining channel	56
Table 1.9: Propensity-score-matched samples	57
Table 2.1: Annual number of published experimental articles	123
Table 2.2: Summary statistics at the article-level	124
Table 2.3: Deviations from the exponential distribution	125
Table 2.4: The magnitude and statistical significance of <i>p</i> -hacking.	126
Table 2.5: Comparing the extent of <i>p</i> -hacking across various subsamples.	127
Table 2.6: Rejection rates after adjusting for multiple hypotheses testing.	128
Table 3.1: Summary statistics	162
Table 3.2: Gambling preference and corporate cash holding	163
Table 3.3: Role of local seniors	164
Table 3.4: Headquarter relocation.....	165
Table 3.5: Instrumental variable approach.....	166
Table 3.6: Gambling preference and corporate risk-taking	167
Table 3.7: Gambling preference and financial constraints.....	168
Table 3.8: Gambling preference and the value of cash holding.....	169
Table 3.9: Speed of Adjustment to target cash level.....	170
Table 3.10: Tests of spurious relations	171
Table 3.11: Robustness checks	172
Table 3.12: Measuring gambling preference using lottery income.....	173

List of Figures

Figure 2.1: Empirical distribution of all p -values from 1990 to 2016.	115
Figure 2.2: Isolating p -values reported with two decimal places.....	116
Figure 2.3: Fitting histogram of p -values to exponential distribution.....	117
Figure 2.4: The distribution of vague p -values	118
Figure 2.5: Including vague p -values.....	119
Figure 2.6: P -hacking and citations	120
Figure 2.7: Do the main results of an article show more p -hacking?.....	121
Figure 2.8: JAR's 2014 policy and the distribution of p -values.	122
Figure 3.1: State-level variation in local firms' cash holding.	130
Figure 3.2: State-level dispersion of local gambling preference.....	161

Summary

My dissertation includes topics in corporate finance and research methodology. It contains three chapters.

Chapter one aims to understand whether investor-paid credit rating agency (CRA) can improve rated firms' information transparency. It has two main findings. First, the coverage of Egan Jones Ratings (EJR) – an investor-paid CRA – leads to an increase in rated firms' stock price informativeness. Second, EJR coverage increases price informativeness by producing incremental information and by disciplining managerial information disclosure. Moreover, the coverage of S&P's – a CRA that is compensated by rated firms – has no impact on rated firms' transparency. These results lead to the conclusion that investor-paid CRAs play a meaningful governance role due to the absence of conflict of interest that plagues issuer-paid CRAs.

Chapter two studies a type of questionable researcher practice: *p*-hacking. It contains three sets of analyses. First, among a broad set of experimental accounting articles, my co-authors and I document that there is an over-abundance of *p*-values right below the significance threshold ($p = 0.05$), and an under-representation of $p = 0.06$. The finding is consistent with researchers relying on *p*-hacking to turn insignificant results into significant ones. Second, cross-sectional analyses show that *p*-hacking is more prevalent among untenured authors, male authors, sole authors, and authors from higher-ranking schools. Lastly, we discussed multiple hypotheses testing methods and disclosure-based solution to guard against *p*-hacking.

Chapter three proposes that gambling preference is a determinant of the significant geographical variation in U.S. firms' cash holding. Local gambling preference, measured as the ratio of local Catholics to Protestants, is positively associated with local firms' cash holding. In addition, firms in gambling-prone regions hold more cash only when they are financially constrained, and they invest cash in risky projects such as research & development, which leads to more volatile cash flows. Overall, the results suggest that more cash holding enables gambling-prone firms to take more risks while staying away from financial distress.

Chapter 1: Investor-paid Credit Rating Agency and Stock Price

Informativeness

Abstract

Unlike issuer-paid credit rating agencies (CRAs), investor-paid CRAs are compensated by investors for their rating services. This paper documents that the coverage of an investor-paid rating agency (the Egan Jones Ratings, EJR) increases rated firms' stock price informativeness. To alleviate endogeneity concerns, I rely on two quasi-natural experiments that shock the importance of EJR for investors' information acquisition. EJR coverage increases price informativeness through two channels, namely, by producing incremental information and by disciplining managerial information disclosure. Moreover, the positive effect of EJR's coverage on price informativeness is more pronounced when the conflict of interest faced by issuer-paid rating agencies is more severe. Lastly, the coverage of an issuer-paid CRA (Standard & Poor's, S&P) is not associated with changes in price informativeness. I conclude that investor-paid CRAs function as a type of effective information intermediary to discipline firm managers and improve corporate transparency.

JEL Classification: G14, G24, M40

Keywords: Investor-paid credit ratings; Stock price informativeness; Firm disclosure

1. Introduction

Financial economists have long recognized that stock prices can aggregate and reveal information among many investors (e.g., Hayek, 1945), and therefore guide the decision making of various market participants.¹ Thus, financial institutions that promote price informativeness by producing and disseminating information play a critical role in the functioning of capital markets. As a type of financial intermediary, credit rating agencies (CRAs) collect information regarding the risks of debt issuers or debt obligations and reveal the information in the form of rating opinions. To the extent that investors rely on credit ratings to evaluate risks and allocate capital, they help incorporate the rating information into security prices (e.g., Hand, Holthausen, and Leftwich, 1992). As a result, CRAs can potentially contribute significantly to improving price informativeness.

CRAs, however, are criticized for their poor performance during multiple financial failures.² Observers believe that the failure of CRAs is driven by the conflict of interest in the issuer-paid rating model (e.g., Bolton, Freixas, and Shapiro, 2012; Jiang, Stanford, and Xie, 2012). Indeed, prior works find that issuer-paid CRAs tend to assign inflated ratings in favor of issuers (see, e.g., Griffin and Tang, 2012; Beck and Milbourn, 2011; Bongaerts, Cremers, and Goetzmann, 2012). Inflated ratings could distort the information on the market and corporate investments (Goldstein and Huang, 2018; Chu and Rysman, 2019).

¹ Prior studies document that stock prices contain information that guide the decisions of various market participants, including managerial investment activities (Luo, 2005; Chen, Goldstein, and Jiang, 2007) and shareholders' governance choices (Ferreira, Ferreira, and Raposo, 2011; Lin, Liu, and Sun, 2019).

² In the Enron scandal and the recent financial crisis, credit ratings failed to predict defaults and warn investors against significant downside risks in a timely fashion (see, for example, Healy and Palepu, 2003; He, Qian, and Strahan, 2011).

Therefore, it is natural to conjecture that, absent from the conflict of interest in the issuer-paid model, CRAs would function as efficient information producers to promote price informativeness.

In this paper, I study the informational role of *investor*-paid CRAs. Unlike issuer-paid CRAs, investor-paid rating agencies are compensated by investors who request their rating services. Thus, they are not under pressure from debt issuers to assign inflated ratings to gain a higher market share and rating revenue. Although previous studies find that investor-paid ratings are timelier and more accurate than issuer-paid ratings (e.g., Cornaggia and Cornaggia, 2013; Bruno, Cornaggia, and Cornaggia, 2016), it remains unclear whether investor-paid CRAs contribute incremental information.³ Brealey and Myers (2003) argue that the influence of rating agencies is exaggerated because they are “as much following investor opinion as leading it.” Adding to this “information piggybacking” concern is the lack of investors’ interest in the investor-paid business model in other markets around the world.⁴

I investigate whether investor-paid ratings promote information efficiency by studying the relationship between investor-paid rating coverage and rated-firms’ stock price informativeness. The focus on the stock market instead of the bond market is due to the illiquidity and sparse trading in the bond market that precludes the construction of price informativeness measure.⁵ More

³ Prior studies show that investor-paid CRAs’ rating lead issuer-paid ratings (Beaver, Shakespeare, and Soliman, 2006; Bruno, Cornaggia, and Cornaggia, 2016), and track concurrent equity price-based default probability more closely (Cornaggia and Cornaggia, 2013; Xia, 2014).

⁴ For instance, two attempts to set up an investor-paid rating model in Europe did not succeed. The failure was partly due to the lack of investors’ interest in the ratings from investor-paid CRAs. The first initiative was from Markus Krall at Roland Berger. Another attempt was from the French credit insurer Coface.

⁵ As detailed in Section 3, constructing the price informativeness measure (i.e., non-synchronicity) requires daily price information. However, bonds trade infrequently. For example, Goldstein and Hotchkiss (2019) document that the median bond trades 3.1 times per month, and trades with a size of more than 100 bonds occur only 0.6 times per month. In contrast, the stock market is much

importantly, I conjecture that credit ratings can have cross-market implications above and beyond the bond market, through the two potential channels explained below.

First, equity returns incorporate default risks (e.g., Vassalou and Xing, 2004). To the extent that investor-paid CRAs provide incremental default-related information to investors, we expect equity investors to incorporate the information into stock prices through trading and thus increase price informativeness. I label this channel the “information channel.” Consistent with the information channel, Bhattacharya, Wei, and Xia (2019) find that a group of small institutional investors follow ratings from an investor-paid rating agency (The Egan Jones Ratings, EJR hereafter), and these investors’ trading profits outperform non-followers. Bruno, Cornaggia, and Cornaggia (2016) find that EJR’s rating updates are accompanied by a greater stock price reaction than Moody’s rating updates.

The second and more nuanced channel focuses on the potential disciplining effect of investor-paid CRAs on corporate information disclosure. Prior studies show that firm managers use earnings management to influence the ratings from issuer-paid CRAs (e.g., Liu et al., 2018). To the extent that issuer-paid CRAs are more likely to condone issuers’ earnings management due to conflicts of interest, investor-paid CRAs should weaken the earnings management incentive by being less likely to tolerate such activity. A lower level of earnings management leads to more informative stock prices (Hutton, Marcus, and Tehranian, 2009).

more liquid, and early studies also document that stock and bond prices move similarly in response to rating changes (e.g., Hand, Holthausen, and Leftwich, 1992; Kliger and Sarig, 2000).

In addition, prior works document that issuer-paid ratings are less responsive to issuers' negative information than to positive information (Bruno, Cornaggia and Cornaggia, 2016). The delayed revelation of negative news likely facilitates managerial bad news hoarding.⁶ In contrast, investor-paid CRAs have been shown to provide ratings that reveal both good and bad news symmetrically, which makes it more difficult for managers to withhold bad news. Furthermore, even if issuer-paid CRAs rate firms truthfully, these firms may choose not to disclose unfavorable ratings when they purchase ratings from multiple CRAs (Sangiorgi and Spatt, 2016).⁷ In contrast, the ratings from investor-paid CRAs go directly to investors who pay for the ratings; thus, firms cannot "censor" unfavorable ratings from investor-paid CRAs. I label the second channel the "disciplining channel."⁸

More informative ratings and a better quality of information disclosure will lead to a greater magnitude of firm-specific information being incorporated into stock prices. A stream of research has established that R^2 from index-model

⁶ For example, the hearing on "Enron and The Credit Rating Agencies" by the Committee on Governmental Affairs suggested that Enron's perpetuation of its accounting scandal relied heavily on above investment-grade credit ratings, because a downgrade to below investment-grade level would require Enron to be explicitly responsible for a large amount of liability hidden in its Special Purpose Entities (The Committee on Governmental Affairs, 2002 p. 25).

⁷ Anecdotal evidence suggests that corporate managers follow investor-paid ratings. For example, EJR downgraded Jefferies Group Inc. on 2 Nov 2011 based on concerns about the investment bank's exposure to sovereign debt. As a rebuttal to the downgrade, Jefferies pointed out that the rating downgrade did not consider its short position in sovereign debt and later disclosed all of its positions in sovereign debt to investors. Another example is PennantPark Investment Corporation, which received a rating of "BBB+" from EJR on 29 July 2019; this rating was an upgrade from the "BBB" rating that it received in August 2017. The CEO of the company responded to the rating upgrade as follows: "Since PennantPark Investment Corporation's initial Egan-Jones rating two years ago, we have substantially reduced risk and optimized the portfolio.....We are pleased that Egan-Jones has recognized our disciplined underwriting approach."

⁸ Prior literature documents that information intermediaries in general can deter managerial misbehaviors that clog information flow from insiders to outside investors. For example, Yu (2008), Irani and Oesch (2013), and Chen, Harford, and Lin (2015) show that security analysts following deters managerial misbehaviors such as earnings management; Miller (2006), Dyck, Morse, and Zingales (2010), Dai, Parwada, and Zhang (2015) document that financial media enhances transparency by detecting and exposing corporate fraud and insider trading.

regressions measures the overall firm-specific information contained in stock prices. Roll (1988) argues that R^2 captures the relative importance of firm-specific information to market-wide information in stock prices. Jin and Myers (2006) and Hutton, Marcus, and Tehranian (2009) find that better accounting information leads to higher firm-specific information in stock prices (low R^2). Moreover, previous studies (e.g., Gul, Srinidhi, and Ng) suggest that R^2 summarizes two parts of information transparency, the part due to firm-specific information disclosure by managers and the other part due to investors' informed trading. I therefore use R^2 as a measure of stock price informativeness to reflect the disciplining channel and the information channel outlined above.⁹

I test the informational role of investor paid CRAs by examining how the rating coverage of a leading investor-paid credit rating agency – the Egan Jones Ratings Company – affects rated firms' stock price informativeness. I find that compared with other firms, EJR-rated firms experience a significant improvement in stock price informativeness; in contrast, S&P's coverage is associated with a decrease in the information content of stock prices, albeit the effect is insignificant. In terms of economic significance, after a firm is rated by EJR, its price non-synchronicity increases by 0.091 (or 9.1%), on average, compared with its non-EJR rated counterparts. The main finding is robust to controlling for potential omitted variables and to propensity-matched samples.

Since EJR's coverage is not randomly assigned, the main result could still be relational instead of causal. To further guard against endogeneity concerns, I exploit two quasi-natural experiments that lead to an exogenous increase in the

⁹ In the Appendix, I also use additional measures of price informativeness developed by Durnev et al. (2003).

importance of EJR ratings for investors' information acquisition. To the extent that EJR's coverage causally increases price informativeness by disciplining corporate information disclosure, we should expect that its disciplining impact becomes stronger when investors rely more on its ratings. The first experiment is based on the mergers and closures of brokerage that result in exogenous decreases in the number of analysts following for some firms, aggravating information asymmetry between corporate insiders and outsiders (Kelly and Ljungqvist, 2012). The losses in analysts following for a firm likely enhance the importance of EJR as an alternative information intermediary and, thus, the effect of EJR coverage on stock price informativeness. The results confirm this prediction.

Next, I adopt an additional quasi-natural experiment that exploits EJR's designation as a "Nationally Recognized Statistical Rating Organization" (NRSRO) in December 2007. The NRSRO designation dictates that EJR's ratings can be used as a benchmark for regulatory and contracting compliance. By certifying EJR as a "nationally recognized" CRA that provides accurate and reliable ratings, NRSRO designation increases investors' reliance on EJR's ratings for credit risk evaluation and investment decisions, and thus amplifies the impact of EJR coverage. I document that the association between EJR coverage and stock price informativeness becomes stronger after EJR received the NRSRO designation. Overall, the endogeneity tests support a causal impact of EJR coverage on stock price informativeness.

In subsequent analyses, I investigate the mechanism through which EJR coverage drives the main results. The findings suggest that both the direct information channel and the indirect disciplining channel contribute to the

positive impact of EJR ratings on rated firms' stock price informativeness. First, the effect of EJR coverage is stronger in scenarios where issuer-paid CRAs are plagued more by conflicts of interest and are thus more likely to compromise their rating quality to cater to issuers. In addition, consistent with the information channel, I document that EJR's ratings are more sensitive to firms' earnings quality in the cross-section than S&P's ratings, suggesting that EJR's ratings incorporate earnings quality information to a greater extent.

Second, consistent with the presence of the disciplining channel, I document that EJR coverage is associated with a decrease in corporate earnings management and an increase in management bad news forecasts. Combined with the finding that EJR's ratings are more responsive to earnings quality, this result suggests that EJR's presence deters earnings manipulation. In addition, although the information channel predicts symmetric responses of price informativeness to EJR's coverage when its ratings are either more or less issuer-friendly than S&P's ratings, I document that the increase in price informativeness is stronger when EJR assigns less favorable ratings than S&P, a situation that signals that potential negative news is kept undisclosed to investors. Moreover, the relationship between EJR coverage and price informativeness strengthens after EJR received the NRSRO designation, which does not alter the information content of ratings but increases the importance of EJR's ratings for the decision-making of corporate managers and investors.

This paper contributes to the literature in two ways. First, it establishes a new mechanism through which CRAs affect corporate transparency. Prior research largely treats CRAs as an initial information provider on corporate creditworthiness. In contrast, this paper provides evidence that investor-paid

CRAAs can be an external monitor of firms' financial reporting by influencing the disclosure incentive of firm managers. In this regard, investor-paid CRAAs can play a corporate governance role similar to other financial intermediaries such as security analysts (see, e.g., Yu, 2008; Irani and Oesch, 2013). Moreover, CRAAs' monitoring role can be significant, considering the prevalent usage of credit ratings in investment and contracting.

Second, this paper extends the literature on the real consequences of credit rating quality. Existing works typically compare the rating properties between issuer-paid and investor-paid rating agencies, for example, rating timeliness (Beaver, Shakespeare, and Soliman, 2013; Cornaggia and Cornaggia, 2013) and accuracy (Xia, 2014; Bonsall IV, Koharki, and Neamtiu, 2017). As pointed out earlier, this comparison does not establish that investor-paid CRAAs' provide incremental information to the market. My paper extends this line of research by studying the real effect of investor-paid rating agencies on the corporate information environment, and thus offers a direct test on the informational role of investor-paid rating agencies.

The rest of this paper proceeds as follows. Section 2 reviews the related literature and develops the hypothesis. Section 3 describes the data and sample. Section 4 reports the main results and endogeneity tests. Section 5 presents additional tests, and Section 6 concludes.

2. Related literature

This paper builds on two strands of literature. The first strand focuses on the incentive of CRAAs and their informational role, and the second strand focuses on CRAAs and corporate information disclosure.

2.1 CRAs' incentives and their informational role

There is a large literature on the information role of credit rating agencies. Earlier studies focus on issuer-paid rating agencies and generally examine stock and bond market reactions to bond rating updates. More recent studies pay closer attention to investor-paid rating agencies and compare the rating properties and methodology between the two types of rating agencies.

Early studies at best offer inconclusive evidence on whether issuer-paid CRAs contribute incremental information to the market. Weinstein (1977) finds no evidence that bond prices react to rating changes. Hand, Holthausen, and Leftwich (1992) distinguish rating changes that are “contaminated” by concurrent corporate news from other sources, and find that “uncontaminated” rating changes do not trigger abnormal bond returns. Kliger and Sarig (2000) rely on Moody’s credit rating refinement that provides additional credit information unrelated to corporate events. They find that bond prices react to the new information conveyed in the refined ratings. Dichev and Piotroski (2001) document substantial negative stock return up to three years after rating downgrades but no reliable abnormal returns following upgrades. Norden and Weber (2004) show that both the stock and credit default swap markets anticipate rating actions such as rating changes and reviews for downgrade. Bongaerts, Cremers, and Goetzmann (2012) find that Fitch’s ratings play a regulatory certification role, without conveying additional information relative to what is already reflected in the ratings from S&P and Moody’s. Kraft (2015a) finds that Moody’s makes upward rating adjustments to help firms avoid rating triggers in their borrowing contracts, suggesting that conflicts of interest may be detrimental to rating informativeness. Badoer and Demiroglu (2018) show that when the

corporate bond market becomes more transparent due to the public dissemination of bond prices, the importance of credit ratings as an information source diminishes dramatically.

The role of the Big Three issuer-paid CRAs in several high-profile financial failures draws more attention to investor-paid rating agencies. Because investor-paid CRAs are not subject to the conflicts of interest in the issuer-paid model, they are expected to generate more accurate ratings as long as investors in general demand for accurate information. A stream of research finds evidence consistent with this expectation. Beaver, Shakespeare, and Soliman show that EJR's rating updates lead Moody's rating updates and are accompanied by a stronger stock price reaction. Stanford and Xie (2012) find that after S&P's switched from the investor-paid model to the issuer-paid model in 1974, it started to assign more favorable ratings to issuers, especially when issuers pay high rating fees. Cornaggia and Cornaggia (2013) find that Rapid Ratings – another investor-paid rating agency – also beat Moody's ratings in terms of timeliness and accuracy to predict default. Xia (2014) further shows that EJR's rating can discipline S&P's ratings: following EJR's coverage, S&P's ratings for the same firms become more accurate and timelier. These studies confirm that CRAs' incentive is an important determinant of their rating characteristics.

Overall, the comparison between investor-paid and issuer-paid ratings shows that investor-paid CRAs possess several desirable rating characteristics such as timeliness and accuracy and are accompanied by a stronger market reaction. However, it is ultimately unclear whether investor-paid CRAs provide incremental information to the market, or if they are simply faster than issuer-paid rating agencies in following other sources of corporate economic news. This

paper informs the literature by studying how the rating coverage of investor-paid CRAs affect stock price informativeness because if investor-paid CRAs matter, their information production should lead to more informative prices.

2.2 Credit rating agencies and corporate information disclosure

The relation between CRAs and corporate information disclosure is multifaceted. First, CRAs rely on firms' public disclosure as an information source for their rating process and claim that they adjust issuer ratings based on issuers' information quality. For example, the corporate rating criteria of S&P states that it pays attention to rated firms' financial reporting quality, especially situations where “[t]he company is particularly aggressive in the application of accounting standards, or demonstrates opaqueness in its financial reporting” (S&P, 2008).¹⁰ Yu (2005) finds that rating agencies assign better ratings to firms with higher information quality. Jorion, Shi, and Zhang (2009) argue that the time-series decrease in accounting quality drives the decline in credit ratings over time. Kraft (2015b) finds that Moody's makes quantitative adjustments to firms' financial statement numbers, and these adjustments help Moody's ratings better capture credit risk. These papers support CRAs' claim that they incorporate accounting quality and transparency into their ratings.

Second, to the extent that CRAs do not fully adjust their ratings for issuing firms' accounting quality, managers may take certain actions to influence their firms' ratings, including earnings manipulation. Alissa et al. (2013) demonstrate that rated firms use earnings management to influence their credit ratings. Jung, Soderstrom, and Yang (2013) provide additional evidence that

¹⁰ For example, EJR states that “Adequate information must be available to reach a view on the creditworthiness of the issuer, entity, or transaction in question. This includes publicly available information on the issuer, such as company financial and operational statistics, reports filed with regulatory agencies.....” (EJR, 2017).

firms smooth long-term earnings to lower rating agencies' perception of credit risk. Liu et al. (2017) find that firms manage earnings upward when they are on a negative credit watch. Moreover, firms that manage earnings to a greater extent are less likely to be downgraded, suggesting that earnings management is effective in influencing the decisions of rating agencies. Beck (2018) documents that municipalities with lower credit ratings use more accounting discretions before bond issuance.

Taken together, although CRAs screen issuers' financial reports before assigning them ratings, prior studies suggest that issuers can successfully influence CRAs' rating decisions by using accounting manipulations. Importantly, the CRAs in these studies are generally issuer-paid, implying that conflicts of interest likely contributes to the success of accounting manipulations; on the other hand, it is also possible that CRAs fail to detect those accounting gimmicks. If conflicts of interest are partly responsible, then the presence of an investor-paid rating agency may discipline corporate disclosure. If rating agencies, in general, tend to be misled by disclosure manipulations, then we expect that the presence of investor-paid CRAs should also not matter. Therefore, whether the disciplining channel works or not is ultimately an empirical question that I attempt to answer in this paper.

3. Data, variable, sample description

3.1 Data and sample selection

Following prior studies (e.g., Cornaggia and Cornaggia, 2013; Liu et al., 2018) that view issuer-paid credit rating agencies as a homogenous group, I use S&P's as the representative issuer-paid CRAs, and EJR as the representative investor-paid rating agencies. EJR has become an active and important player in

the credit rating industry and rates around 1000 listed firms in the U.S. In addition, EJR's ratings are publicly available. It grants researchers access to its historical ratings for all firms upon request.

EJR started to provide credit rating at the request of investors in 1995. However, its ratings were not available before July 1999. I exclude the firms that EJR begins to rate before this date because the exact starting dates for these firms are unavailable; I also exclude the firms that obtain an initial rating but are not followed by EJR since (e.g., Xia, 2014). The EJR credit rating data set is then manually matched by company names with Compustat. I successfully matched 994 firms covered by EJR. I also obtain S&P's long-term corporate credit ratings from Compustat. To compare the effect of EJR's coverage with the effect of S&P's coverage, I combine firms from these two rating databases.

I retrieve the financial statement data for these firms from Compustat and the common stock return data from CRSP. Institutional ownership and trading data are from the Thomson Reuters ownership database compiled from 13f filing. Lastly, I obtain analysts following and management forecasts from the IBES database. Following prior studies (e.g., Ferreira and Laux, 2007), I exclude financial firms (SIC code 6000-6999). The firm-year observations with missing values for the variables employed in the regressions are also excluded. These restrictions resulted in a final sample that contains 14,407 firm-year observations from 1,463 firms from 1999 to 2016, among which 791 are covered by EJR, 1,394 are covered by S&P's, and 669 firms are covered by both.

3.2 Variables

To identify the EJR coverage initiation date for the firms in my sample, I use the first EJR credit rating date for a firm as the coverage initiation date. I then

define a dummy variable, *EJR Cover*, which equals one in a year if a firm has been rated by EJR in a year, and zero otherwise. Firms' stock price informativeness, as measured by the firm-specific component of their stock return, is based on regressing a firm's daily stock return in year t on the concurrent returns of the Fama-French three factors:

$$r_{i,d} = \alpha_i + \beta_1 MKT_d + \beta_2 SMB_d + \beta_3 HML_d + \varepsilon_{i,d} \quad (1),$$

where $r_{i,d}$ is stock i 's return excess of the risk-free rate on day d , MKT_d , SMB_d , and HML_d are the excess daily returns for the market factor, size factor, and the value factor, respectively.¹¹ From this projection, the idiosyncratic variance of firm i 's return is $Var(\varepsilon_{i,d})$. The ratio of idiosyncratic variance to total volatility, $Var(\varepsilon_{i,d})/Var(r_{i,d})$, captures the proportion of variance due to value-relevant firm-specific information (Morck, Yeung, and Yu, 2000) and equals $(1-R_{i,t}^2)$ from regression (1). Given that $R_{i,t}^2$ is bounded between 0 and 1, I follow prior literature (e.g., Gul, Srinidi, and Ng, 2011) to use the logistic transformation of $(\frac{1-R_{i,t}^2}{R_{i,t}^2})$ as the final measure of idiosyncratic volatility. Formally, the non-synchronicity measure $\Psi_{i,t}$ is defined as:

$$\Psi_{i,t} = \text{Ln}\left(\frac{1 - R_{i,t}^2}{R_{i,t}^2}\right) \quad (2).$$

A higher value of $\Psi_{i,t}$ indicates higher firm-specific return variation, and thus a more informative stock price.

To isolate the effect of EJR coverage on stock price informativeness, I include in the regression a standard set of control variables used in Ferreira and

¹¹ In Appendix B, I employ alternative measures of price informativeness to assess the robustness of the baseline results. These measures include price non-synchronicity estimated from the market factor model, the market factor with industry factor, the Fama-French five-factor model, and the future earnings incremental explanatory power measure (*FINC*) of Durnev et al. (2003)

Laux (2007) and Gul, Srinidhi, and Ng (2011). Specifically, the control variables include the natural logarithm of market capitalization (*CAP*), the ratio of total debt to the book value of assets (*Leverage*), the market value of equity plus the book value of debt scaled by the book value of total assets (*MB*), the ratio of earnings before extraordinary items to the book value of equity (*ROE*), the variance of quarterly *ROE* over the past three years (*VROE*), the change in ownership of institutional investors ($\Delta Inst$),¹² a dividend payer dummy (*DD*), and a diversification dummy that equals one if a firm operates in more than one two-digit SIC industry (*DIVER*). I also control for S&P's coverage by including a dummy variable *S&P Cover*, which equals 1 for firm-years with S&P's credit ratings, and 0 otherwise. Bonsall, Koharki, and Neamtiu (2017) find that firms covered by EJR appear to be lower credit risk stable than firms covered by S&P, I thus include Altman's Z-score in the regression (*Z-score*). Finally, I additionally control for firms' debt financing activity (*Debt_Issue*) as it may drive investors' demand for credit rating and firms' disclosure simultaneously. All dollar values are adjusted for inflation based on 2010 constant dollars. The summary statistics based on the regression sample are reported in Table 1.1. The firms in the sample are on average larger than a typical Compustat firm, because larger firms are more likely to have credit ratings. Turning to the untransformed main measure, R^2 , I find that approximately 27.2% of the firm stock return variation is idiosyncratic, and this value is close to 25.0% in Hutton, Marcus, and Tehranian (2009). 33.1% of the firm-years are rated by EJR, and 71.3% are rated by S&P's.

[Insert Table 1.1 here]

¹² The firm-years with total institutional ownership larger than 100% are discarded.

3.3 Determinants of coverage

Both EJR and S&P are profit-driven corporations. Therefore, their decisions to rate a specific firm are by no means exogenous. To better understand the determinants that drive their coverage and inform the subsequent endogeneity tests, I estimate a probit model by using rated-firms and all nonrated firms to forecast the coverage initiation of either EJR or S&P's. Furthermore, because of their different compensation structure and incentive, I estimate the following probit model separately for EJR's coverage and S&P's coverage:

$$Prob(Initiation_{i,t+1} = 1) = \Phi(\alpha_0 + \alpha_1 X_{i,t} + \alpha_2 Industry + \alpha_3 Year) \quad (3),$$

where Φ is the cumulative distribution function of the standard normal distribution. The dependent variable is either *EJR Initiation*_{*i,t+1*} or *SP Initiation*_{*i,t+1*}, which equals zero for the firms that are not rated by EJR or S&P in all years, respectively. To estimate the likelihood of rating initiation, I exclude the post-rating initiation years for the estimations. Previous studies find that the firms rated by EJR tend to differ from these firms rated by S&P's in terms of size, credit risk, and information environment (Xia, 2014; Bonsall IV et al., 2017). I include proxies for these characteristics in the probit regression. Specifically, the regression controls for firms' credit quality (*Leverage*, *Tangibility*, and *Z-score*), five variables (*Market Cap*, *Sale*, *Ln (Analysts)*, *R&D*, and *MB*) to account for the effect of firm size and information environment on the demand for credit ratings, firms' financing activities (*Debt_Fin* and *Equity_Fin*), as well as other control variables introduced in Section 3.2. I also include firms' past stock price informativeness as an explanatory variable to test if it is a significant determinant of CRAs' coverage decisions.

[Insert Table 1.2 Here]

The probit model regression results are reported in Table 1.2. The R-squared from the regressions is high (0.328 for EJR Initiation and 0.225 for S&P's), suggesting that the model explains their coverage decisions reasonably well. The estimation results indicate that both EJR's and S&P's coverage are unrelated to firms' pre-existing stock price informativeness. This finding mitigates the concern of reverse causality. For the other explanatory variables, both EJR and S&P's are more likely to rate high-leverage firms, large firms, low market-to-book firms, profitable firms, and firms with higher institutional ownership. However, there are also some differences in the rating decisions between the two. For example, EJR is more likely to cover older firms and firms with more analysts following at the request of investors, suggesting that EJR's ratings could be value relevant even for the more visible firms. S&P is more likely to rate firms that tap external debt or equity financing, a result consistent with the "certification role" of issuer-paid ratings (Bongaerts, Cremers, and Goetzmann, 2012). In contrast, EJR's rating decision is not significantly related to firms' financing activities.

4. Empirical results

4.1 The effect of EJR coverage on price non-synchronicity

A vast literature defines price informativeness as the extent to which stock prices comove with the market return. If the coverage of EJR leads to better corporate transparency, more of the price movement will be due to firm-specific news, rather than market-wide news. I validate this prediction by showing that firms' stock prices experience a decrease in price synchronicity with index returns after EJR initiates the rating coverage. Specifically, I adopt the following baseline regression specification:

$$\Psi_{i,t} = \beta_0 + \beta_1 EJR_Cover_{i,t} + \gamma Y_{t-1} + Firm_i + Year_t + \varepsilon_{i,t} \quad (4),$$

where $\Psi_{i,t}$ is the logarithm transformed price non-synchronicity as defined in Section 3.2, with a higher value reflecting less non-synchronicity, and thus higher price informativeness. The key independent variable $EJR_Cover_{i,t}$, which equals one if a firm is rated by EJR in a year and 0 otherwise, captures the within-firm difference in price informativeness that can be attributed to EJR coverage. Firm dummies are included to control for unobservable time-invariant firm effects. Year fixed effects are included to account for time-variant trends in the dependent variable. The standard error of the estimated coefficients is clustered at the firm-level. The results, however, are unaffected if I allow clustering by both firm and year.

Table 1.3 presents the baseline results. Column (1) investigates the effect of the EJR rating decision without controlling for S&P's coverage. The coefficient of *EJR Cover* is positive and significant at the 1% level, suggesting that EJR coverage is accompanied by an increase in stock price non-synchronicity. In column (2), I replace *EJR Cover* with *S&P Cover* in the regression. *S&P Cover* is negatively associated with firms' informativeness measure, although its coefficient is statistically insignificant. In column (3), I include both *EJR Cover* and *S&P's Cover* into the regression and find that the coefficient on *EJR Cover* continues to be significant at the 1% level. The economic impact of EJR coverage is also sizeable: in column (3), EJR coverage increases the relative idiosyncratic volatility by 0.091, or 9.1%. To put this economic effect into perspective, I compare the effect of *EJR Cover* with the effect of change in institutional ownership ($\Delta Inst$), which has been shown by prior studies to be a significant determinant of price informativeness (e.g.,

Piotroski and Roulstone, 2004; Ferreira and Laux, 2007). A one standard deviation rise in $\Delta Inst$ increases price informativeness by 8.3%.

[Insert Table 1.3 here]

4.2 Tests on endogeneity

While the analysis above shows a robust positive relation between EJR coverage and stock price informativeness, the causal relationship between the two remains hypothetical. In particular, the documented results could be subject to two types of endogeneity concerns, namely, omitted variable bias and reverse causality. The estimated coefficient of EJR coverage could be spurious if some omitted variable drives both EJR coverage and stock price informativeness. Furthermore, the causal relation between EJR coverage and price informativeness could be bidirectional if EJR conditions its coverage decision on firms' information availability. I address these concerns in the following analyses.

4.2.1 Omitted variable bias

CRAAs are profit-driven corporations, and they make calculated choices in deciding whether to rate an issuer. Therefore, it is likely that the factors that affect their decisions are also related to firms' stock price informativeness, and omitting such variables from the regression analysis would introduce bias in the estimated coefficients. To alleviate this concern, I describe the factors that are most likely to drive both EJR coverage and informativeness and augment the baseline model with variables that capture these factors.

First, when both investors and EJR observe and react to the arrival of the same firm-specific news, the rating coverage of EJR can be positively related to stock price informativeness. To account for the intensity of the arrival of corporate news, I control for the number of days with news coverage from *Dow*

Jones Newswire during each fiscal year for the firms and re-estimate the main results in Table 1.3.¹³ Furthermore, since EJR's ratings may also follow recent unusual volatility or extreme returns, I control for the standard deviation and the mean of the weekly return. The results in columns (1) and (2) show that the association between EJR coverage and stock price non-synchronicity remains after controlling for potential information piggybacking.

Second, prior literature suggests that institutional investors tend to invest in firms with a higher quality of information disclosure (e.g., Bushee and Noe, 2000), and their closer monitoring pressures managers to increase the frequency and accuracy of management forecasts (Ajinkya, Bhojraj, and Sengupta, 2005). Because institutional investors are also the primary customers of EJR, the monitoring of institutional investors can be an omitted variable, which drives both EJR's rating coverage and corporate transparency. Although the regression model (4) controls for the change in institution ownership, which accounts for the trading channel through which institutional investors incorporate their private information into stock prices, I further augment it with the level and the concentration of institutional ownership to account for the effect of institutional monitoring. The results in column (3) of Panel A in Table 1.4 show that the impact of EJR coverage is robust after controlling for institutional monitoring.

Third, prior literature also suggests that high stock liquidity can encourage investors to acquire more private information by increasing the profitability of informed trading (e.g., Subrahmanyam and Titman, 2001). More informed trading will lead to an increase in stock price informativeness

¹³ The data on daily news coverage comes from Boudoukh et al. (2019) who construct a data set that classifies trading days according to whether there are news articles that covering a stock on a given day. The news articles that they use are from *Dow Jones Newswire*.

(Brogaard, Li, and Xia, 2017). If higher stock liquidity incentivizes institutional investors to acquire private information through various sources (including EJR's credit ratings), the coverage initiation of EJR could be positively correlated with informed trading or stock price informativeness through stock liquidity. To alleviate this concern, I explicitly control for firms' stock liquidity by including in the regression in column (4) the illiquidity measure in Amihud (2002). The coefficient on *EJR Cover* remains significant and positive in this test, which suggests that the effect of EJR coverage is not driven by an increase in stock liquidity.

Lastly, Since EJR's ratings are regarded as trading signals by institutional equity investors (Bhattacharya, Wei, and Xia, 2017), investors may request rating services from EJR when firms issue new equity, which is typically accompanied by more corporate information disclosure. To account for firms' equity financing as a potential omitted variable, I augment Eq. (4) with firms' net equity financing, as measured by the ratio of the difference between stock issuance and stock repurchases over total assets (*Equity_Issue*). The results reported in column (5) show that the coefficient of *Equity_Issue* is significantly positive. However, the coefficient of *EJR Cover* remains unchanged.

[Insert Table 1.4 here]

4.2.2 Reverse causality

Bonsall, Koharki, and Neamtiu (2017) point out that unlike the Big Three issuer-paid CRAs who have access to firms' private information during their rating process, EJR relies on firms' public disclosure to evaluate firms' credit quality. It is possible that EJR decides to rate a firm at investors' requests when more firm-specific information becomes publicly available, which results in a causal relationship running from price informativeness to EJR's rating coverage.

As such, the causal relation between EJR coverage and stock price informativeness can be bidirectional.

I conduct two tests to alleviate this reverse causality concern. First, I check the dynamics of the effect of EJR coverage over the periods that surround EJR's coverage initiation. If reverse causality drives the results, we should expect an increase in stock price informativeness prior to EJR's coverage. To test this possibility, I replace the *EJR Cover* dummy in Eq. (4) with 5 year dummies, which are EJR^{-2} , EJR^{-1} , EJR^0 , EJR^1 , and EJR^{+2} . The first four dummies indicate the t^{th} year relative to the year of EJR coverage initiation, and the last dummy captures all the years that start from the 2nd year of EJR coverage and onwards. The results in Panel B of Table 1.4 show evidence against reverse causality. The coefficients for the year dummies before EJR coverage are insignificant and close to zero, suggesting that EJR firms do not differ significantly from non-EJR firms in terms of price informativeness before coverage. Furthermore, the coefficients on EJR^1 and EJR^{+2} are both significantly positive. Therefore, the effect of EJR coverage on stock informativeness manifests after the year of coverage initiation. In the second test, I include firms' past relative idiosyncratic volatility as a direct control for firms' recent information environment and re-estimate Eq. (4). The results in column (2) of Panel (B) show that the effect of EJR coverage is still positive and significant.

4.2.3 The quasi-natural experiments

Ideally, to rule out endogeneity concerns, one would conduct an experiment in which EJR is paid to rate a randomly selected group of firms, and then, the researchers can observe how firms' stock price informativeness evolves

compared with non-rated firms. This ideal setting is, unfortunately, unavailable to empirical researchers. Instead, I exploit two quasi-natural experiments that shock the importance of EJR ratings for investors and corporate decision-making. The logic is that a greater reliance on EJR's ratings for decision-making should amplify the causal impact of EJR's coverage, if any, on price informativeness.

First, I use exogenous shocks to analyst coverage due to brokerage mergers to test EJR's effect on price informativeness. Hong and Kacperzyck (2010) document that when two brokerages houses merge, redundant analysts that cover the overlapping stocks are often let go, which decreases the analyst coverage for those stocks. Kelly and Ljungqvist (2012) further show that the decreases in analyst coverage increase information asymmetry for the affected firms. If investor-paid rating agency causally improves price informativeness, we should expect this causal impact strengthens after other information sources become less available.

I construct an indicator variable (*Analyst Loss*) that equals 1 if a firm's analyst coverage is affected by brokerage mergers in a given year, and 0 otherwise. I then augment Eq. (4) with this indicator and its interaction with *EJR Cover* and report the results in column (1) of Table 1.5. As expected, *Analyst Loss* decreases the informativeness of stock prices. However, the coefficient of *Analyst Loss* $\times$ *EJR Cover* is positive and significant at 1% level, with a magnitude (0.173) close to the magnitude of *Analyst Loss* (-0.176), which suggests that EJR coverage can roughly “undo” the negative impact of the decrease in analyst coverage on informativeness.

[Insert Table 1.5 here]

Next, I adopt an additional identification strategy that exploits EJR's designation as a "Nationally Recognized Statistical Rating Organization" (NRSRO) in December 2007. The NRSRO designation from SEC makes EJR's ratings eligible for certain regulations that guide firm and investor decisions. For example, the SEC allows bond issuers to provide a shorter prospectus when issuing bonds if an issuer meets several standards, and one of these standards is obtaining an NRSRO credit rating above a certain level. The SEC also requires some investment companies such as money market funds and issuance companies to hold bonds with very high ratings from NRSRO.¹⁴ This regulatory certification thus enhances the importance of EJR ratings for credit risk evaluation and corporate financing decisions, without affecting the information content of EJR ratings (Bruno, Cornaggia, and Cornaggia, 2016). If EJR coverage causally improves stock price informativeness through its disciplining effect on corporate information disclosure, we should expect its impact becomes larger when its ratings command greater market attention.¹⁵

The NRSRO designation thus presents a convenient way to substantiate the hypothesized causal link between EJR coverage and price informativeness. I construct an indicator variable, *NRSRO*, which equals one for the years after 2007, and zero otherwise. I then augment Eq. (4) with the interaction term of *NRSRO* and *EJR Cover*, and report the estimation results in column (2) of Table 1.5. The coefficient estimated for *EJR Cover* is positive and significant at the 10% level

¹⁴ Bhattacharya, Wei, and Xia (2019) show that the proportion of institutional investors that respond to EJR's rating changes increases after the NRSRO designation.

¹⁵ Bruno, Cornaggia, and Cornaggia (2016) find that EJR's rating characteristics (timeliness, accuracy, and average rating) does not change after its NRSRO designation. Therefore, the positive effect of the NRSRO designation on the relationship between EJR coverage and informativeness is more likely due to the disciplining channel than to the information channel. For example, after it receives NRSRO designation, EJR's rating changes are more influential for the attractiveness of corporate bond to investors because of regulatory requirement, and therefore draw more market pressure to the firms.

with a smaller magnitude (0.052) than in the baseline model (0.089). Moreover, the interaction term $EJR\ Cover \times NRSRO$ carries a coefficient of 0.080 that is statistically significant at the 5% level, which confirms the prediction that the impact of EJR coverage on price informativeness is enhanced after the NRSRO designation.

To summarize, while endogeneity is a perennial issue that no empirical test can entirely rule out, I conduct a battery of tests to alleviate the endogeneity concerns and find that my main conclusion holds. Each test can be subject to criticism; however, the balance of the evidence points to a causal relation that runs from EJR rating coverage to stock price informativeness.

5. Additional tests

In this section, I probe the specific mechanisms through which EJR coverage exerts an impact on informativeness. I first design tests to assess the relevance of the information channel, and then link EJR coverage to managerial information disclosure. Finally, I use propensity score-matched samples as robustness checks for the main results.

5.1 Evidence on the information channel

First, I examine whether the impact of EJR coverage depends on the difference in ratings between EJR and S&P. To the extent that EJR's ratings contain incremental information, I expect the amount of information to be larger when its ratings differ from S&P's ratings. I construct a dummy variable (*Diff*) that equals one if EJR's rating is different from S&P's rating for a given firm-year and re-estimate Eq. (4) by including $EJR\ Cover \times Diff$ as an explanatory variable. The results are reported in Panel A of Table 1.6. Column (1) shows that the positive impact of EJR coverage on price informativeness is more

pronounced when EJR's ratings disagree with S&P's. Among the firm-years with the same EJR and S&P's ratings, the *EJR Cover* coefficients are not significantly different from zero and are much smaller in magnitude.

Furthermore, the information channel suggests that EJR's contribution to price informativeness will be higher when S&P's ratings are more plagued by conflicts of interest and are thus less informative. I identify two such scenarios. The first scenario is S&P's ratings for frequent bond issuers. Firms that issue bonds to a broad investor base are required to purchase ratings from major issuer-paid CRAs. If a firm issues bonds more frequently, it will contribute more revenues to issuer-paid CRAs than do other issuers. Hence, to compete for the revenues from frequent issuers, CRAs may assign more favorable ratings to them. The second scenario is where S&P's rating differs from EJR's around the investment grade-high yield boundary (IG-HY). Due to the discontinuous increase in financing costs when the firms' ratings slide cross the IG-HY, firms rated near the IG-HY have a strong incentive for rating-shopping or pressing for better ratings from CRAs. I follow Badoer, Demiroglu, and James (2019) to construct two indicators, *Frequent Issuer* and *IG-HY*, for frequent issuers and firms with ratings near the investment-high yield boundary, respectively.¹⁶ Consistent with the information channel, the regression results in Column (2) and (3) from the Panel A of Table 1.6 suggest that the impact of EJR coverage is indeed stronger for *Frequent Issuers* and *IG-HY* firms.¹⁷

¹⁶ Specifically, for frequent issuers, in each year, I count the number of bonds issued for all firms in the past three years and classify a firm as a frequent issuer if its number of bonds issued is above the median of number of issuance across all firms. To identify IG-HY firms, I construct a dummy variable (IG-HY) that equals one if either S&P's or EJR rates a firm at BBB- or higher while the other rates the same firm BB+ or lower, and the variable is zero otherwise.

¹⁷ In column (1), the variable *Diff* itself is subsumed by its interaction term with *EJR Cover*, because *Diff* is constructed conditional on a firm being rated by EJR. Similarly, the variable IG-HY is also subsumed by its interaction term with *EJR Cover* in column (3).

[Insert Table 1.6 here]

Second, I examine the cross-sectional sensitivity of EJR's and S&P's ratings to the firms' accounting quality. As discussed in Section 2, CRAs claim that they assess the quality of accounting information and financial transparency, and they raise concerns when firms demonstrate opaqueness in financial reporting (e.g., S&P, 2008). If EJR's ratings incorporate accounting quality to a greater extent than do S&P's ratings, I expect that EJR's ratings will be more responsive to firms' accounting quality than S&P's ratings. Moreover, because EJR's entry disciplines S&P's to provide ratings with higher informational quality (Xia, 2014), I expect S&P's ratings to become more sensitive to firms' accounting quality *after* EJR's entry. To test the two predictions, I follow Yu (2005) to run a cross-sectional regression where the dependent variable is EJR's or S&P's average ratings in numerical grades (1-Aaa, 22-C) assigned to a firm in a year. I follow Hutton, Marcus, and Tehranian (2009) to measure the key independent variable, accounting quality (*ACCQ*), as the three-year moving average of the absolute value of discretionary accrual estimated from the modified Jones model.¹⁸ Also included in the regression are the same set of control variables as in Eq. (4).

Panel B of Table 1.6 presents the results. Column (1) shows that when EJR's rating is the dependent variable, the coefficient of *ACCQ* carries a positive sign and is significant at the 10% level, suggesting that EJR assigns lower ratings to firms with more opaque financial information. In column (2), when S&P's ratings replace EJR's ratings as the dependent variable, the coefficient of *ACCQ* is positive but insignificant; furthermore, its magnitude is around half of the

¹⁸ See the Appendix for the definition of the modified Jones model.

magnitude in column (1). This outcome is consistent with EJR's ratings incorporating rated firms' accounting quality to a greater extent than S&P's ratings. In column (3), to test whether EJR's entry disciplines S&P's into improving its ratings' sensitivity to *ACCQ*, I include in the regression the *EJR Cover* dummy and its interaction term with *ACCQ*. The regression results show that *EJR Cover*ACCQ* loads significantly with a positive sign, which is consistent with the argument that EJR's entry makes S&P assigns less favorable ratings to firms with lower accounting quality. Accordingly, the results in Table 1.6 lend support to the information channel of EJR's coverage. Moreover, EJR's entry may indirectly improve corporate transparency by making S&P's ratings be more responsive to firms' accounting quality. The evidence also complements the finding in Xia (2014) that S&P's ratings become more responsive to credit risk following EJR's coverage.

Lastly, I check if EJR's coverage improves stock prices' ability to track contemporaneous cash flows. Specifically, I validate that EJR's ratings contain incremental information about corporate performance that is later released at the earnings announcements. When more information about future corporate performance is reflected in stock prices, subsequent earnings announcements should command weaker market reactions (see, e.g., McNichols and Trueman, 1994). I test this prediction by setting up a regression model of the following form:

$$\begin{aligned}
 ABS_AR[0,2] = & \beta_0 + \beta_1 EJR_Cover_{i,q} + \beta_2 EJR_Cover_{i,q} \\
 & \times ABS_SUE_{i,q} \\
 & + \beta_3 ABS_SUE_{i,q} + \delta Controls_{i,q} + \theta Controls_{i,q} \\
 & \times ABS_SUE_{i,q} \\
 & + Firm_i + Year_t + \varepsilon_{i,q}
 \end{aligned} \tag{5}$$

where the dependent variable is the absolute value of the abnormal size characteristic portfolio-adjusted return over the trading days $[0, +2]$ of the quarterly earnings announcement. $EJR\ Cover_{i,q}$ is an indicator that equals 1 if a firm is rated by EJR during the 60 days prior to quarterly earnings announcements, and 0 otherwise. $ABS_SUE_{i,q}$ is the absolute value of the unexpected earnings relative to the median value of analyst earnings forecasts.

Given that EJR's ratings are first made available to investors who pay for the ratings, sell-side analysts likely do not have timely access to EJR's ratings. Moreover, because the forecasts by analysts presumably reflect available public information (including corporate public disclosure), the coefficient of $EJR\ Cover_{i,q} \times ABS_SUE_{i,q}$ thus captures the relative pre-announcement information sets of investors who purchase ratings from EJR and the investors who do not. A negative coefficient suggests that EJR's coverage decreases the stock price reaction to unexpected earnings, consistent with incremental information being conveyed through EJR's ratings in anticipation of public announcements.

The control variables in the regression include S&P's coverage indicator ($S\&P\ Cover$), a loss indicator for below zero quarterly earnings per share ($Loss$), the fourth fiscal quarter indicator ($QTR4$), firm size (CAP), the book-to-market ratio (MB), the number of analysts ($Ln\ (analysts)$), institutional ownership (IO), stock price volatility (VOL), leverage (Lev), the stock price beta ($Beta$), earnings persistence ($Persistence$), and earnings volatility ($EVOL$). I also include the

interactions of these control variables with $ABS_SUE_{i,q}$ to account for the impact of these variables on the slope coefficient of $ABS_SUE_{i,q}$.¹⁹

Table 1.7 reports the regression results of model (5). Column 1 finds that the coefficient of $EJR\ Cover_{i,q} \times ABS_SUE_{i,q}$ is significantly negative, suggesting a reduction in stock price reaction per unit of unexpected earnings when a firm is rated by EJR prior to the earnings announcement. In contrast, the coefficient on $S\&P\ Cover_{i,q} \times ABS_SUE_{i,q}$ is insignificant. In column 2, I replace $EJR\ Cover_{i,q}$ ($S\&P\ Cover_{i,q}$) with $EJR\ Update_{i,q}$ ($S\&P\ Update_{i,q}$), which indicates that EJR (S&P) updates its rating during the 60 days before earnings announcements, and re-estimate model (5). The coefficient of $EJR\ Update_{i,q} \times ABS_SUE_{i,q}$ is almost twice the size of the coefficient on $EJR\ Cover_{i,q} \times ABS_SUE_{i,q}$, which suggests that when EJR updates its ratings, it helps to incorporate into stock price more information related to upcoming earnings announcements. Overall, the results confirm Bhattacharya, Wei, and Xia (2014) that investors seem to trade on EJR's ratings before corporate earnings announcements and thus incorporate earnings-related information into stock prices. In addition, consistent with the results from Table 1.3, the coverage of S&P's is not associated with a significant change in stock prices' ability to predict future earnings.

[Insert Table 1.7 here]

¹⁹ *Vol* is calculated as the standard deviation of stock returns over the previous quarter. *Beta* is the stock market beta calculated over trading days [-252, -5] relative to the earnings announcement, *Persistence* is the AR(1) coefficient of regressing current earnings on prior-year earnings in the same quarter calculated using the past four years. *EVOL* is the standard deviation of seasonal difference in EPS calculated using the past four years.

5.2 Evidence on the disciplining channel

In this section, I examine in greater detail the disciplining channel through which EJR coverage makes stock prices more informative. Specifically, the disciplining mechanism predicts that when making disclosure decisions, firm managers respond strategically to the incentive of CRAs. Issuer-paid CRAs' incentives to cater to issuers by delaying revealing (unfavorable) information could enable managers to obfuscate disclosure. In contrast, investor-paid CRAs' incentive to cater to *investors* by providing timely ratings could deter managerial obfuscation. I test the two types of managerial obfuscation here.

First, I check the potential impact of EJR coverage on earnings quality, as measured by the absolute level of discretionary accruals. The focus on earnings quality is motivated by prior works (e.g., Liu et al. 2018) that document that firms exploit earnings management (EM) to swing issuer-paid ratings. If EM can effectively shape ratings because of issuer-paid CRA's conflicts of interest, we would expect EM to have a weaker influence on EJR's ratings.²⁰ Therefore, EJR coverage may weaken managers' ex-ante incentive to conduct EM. I estimate three measures of discretionary accruals as the absolute value of the residuals from three accrual models. They are the modified Jones model, the modified Jones model with ROA, and performance-matched discretionary accrual based on the modified Jones model (see, e.g., Kothari, Leone, and Wasley, 2005).²¹ The analyses in Panel A of Table 1.8 provide evidence consistent with this expectation. EJR coverage results in decreases in all of the three most commonly used measures of EM. The reductions in EM are also sizable: taking

²⁰ Moreover, EJR's coverage may also deter other issuer-paid CRA to condone EM by increasing their reputation costs. This is consistent with the evidence in the Panel B of Table 7.

²¹ The detailed estimations of discretionary accruals based the three methods are provided in the Appendix.

the coefficient on *EJR Cover* in column (3) as an example, the coverage of EJR results in a 12.42% reduction in the value of the absolute performance-matched accrual from its mean.

[Insert Table 1.8 here]

Second, I investigate how EJR coverage influences firm managers' incentive to provide negative news forecasts. The focus on negative news forecasts is based on prior findings that compared with investor-paid CRAs, issuer-paid CRAs tend to be slower in incorporating negative news (e.g., Cornaggia and Cornaggia, 2013; Bruno, Cornaggia, and Cornaggia, 2016). Slower revelation of negative news could enable managers to withhold bad news for a more extended period (e.g., for better credit ratings or private benefits). I reason that investor-paid CRAs' timelier downgrades relative to issuer-paid CRAs may thus deter managerial bad news hoarding by informing investors earlier. To test this prediction, I obtain management forecast data from the IBES database and construct two variables – *Bad Forecasts* and *Good Forecasts* – to measure the annual frequencies of negative and positive management forecasts, respectively.²² The two measures are thus used as the dependent variables in Eq. (4). The regression results in Panel B of Table 1.8 show that EJR coverage increases the frequency of managers making bad news forecast by 7.5% while exerting a weaker impact on the likelihood of good news forecast. I interpret the results as suggesting that EJR's timelier revelation of negative news weakens *ex-ante* the bad news hoarding incentive from managers.²³ Overall, the results in this

²² *Bad Forecasts* (*Good Forecasts*) is the natural logarithm of one plus the frequency of negative (positive) management forecast in a given fiscal year, where a negative (positive) forecast is defined as a forecast that fall short (beat) analysts' forecasting consensus (Li and Zhang, 2015).

²³ The positive and significant relationship between *Bad Forecasts* and *EJR Cover* cannot simply be due to reverse causality, i.e., investors are more likely to request EJR's rating services when there is a greater likelihood of bad news. First, in my sample, EJR's ratings are more issuer-

section suggest that investor-paid rating agencies can improve price informativeness through the disciplining impact on corporate information disclosure.

5.3 Propensity-score-matched samples

In a concluding analysis, I assess whether the main results are robust to propensity score-matched samples. The main sample of this paper includes firms rated by either EJR or S&P's to contrast their differential association with price informativeness. However, the two CRAs rate firms with different characteristics, e.g., EJR tends to rate larger firms, and S&P tends to rate the firms that are tapping external financing. To ensure covariates balance between the firms rated by EJR and these firms not rated by EJR, I rely on propensity score matching based on the likelihood of EJR coverage, which is estimated from the probit model in Section 3.3. For each EJR-rated firm in the year prior to EJR's first rating initiation, I use the following four different matching criteria to choose matching firms: (1) the matching firm with the closest propensity score to an EJR-rated firm; (2) the matching firm with the closest propensity score to an EJR-rated firm and within a difference of 1%; (3) two matching firms with the closest propensity score to an EJR-rated firm; and (4) two matching firms with the closest propensity score to a EJR-rated firm and within a difference of 1%.²⁴

Table 1.9 presents the regression results based on the four propensity score-matched samples. In all of these specifications, the coefficient estimates for *EJR*

friendly than S&P's ratings for 2,098 firm-years; in contrast, EJR's ratings are less issuer-friendly for 1,587 firm-years. Given that S&P's ratings respond to news more slowly than EJR's ratings, these results are inconsistent with the argument that EJR is more likely to enter when bad news occurs. Second, Bhattacharya, Wei, and Xia (2018) document that institutional investors are net buyers after EJR's rating changes, suggesting that EJR ratings do not simply respond to the arrival of bad news.

²⁴ A firm that has never been rated by EJR can be matched to more than one EJR rated firms. For example, the first matching criteria identifies 318 unique matched firms for the 712 EJR rated firms.

Cover are significant and positive. Moreover, the economic magnitude is also quite similar to the result based on the main sample in Table 1.3. These findings confirm that the firms rated by EJR still experience a significant increase in stock price informativeness when compared with otherwise similar firms.

[Insert Table 1.9 here]

6. Conclusion

Credit rating agencies play a critical role in the financial market, and the information content of credit ratings has long been the subject of extensive academic research. More recently, the poor performance of issuer-paid CRAs has also spurred research on an alternative type of CRA, that is, investor-paid CRAs. The general conclusion from the literature is that investor-paid CRAs outperform their issuer-paid counterparts, in terms of rating accuracy, timeliness, etc. Notably, it is still unclear whether investor-paid CRAs contribute incremental information to the market, or simply respond to other sources of information faster than issuer-paid CRAs. This paper aims to answer this question by studying the relationship between CRAs coverage and price informativeness.

I find that the coverage of an investor-paid CRAs (EJR) leads to greater stock price informativeness. Additional analyses suggest that EJR increases price informativeness through an information channel and a disciplining channel. In contrast, the coverage of a representative issuer-paid CRA (S&P's) has no meaningful impact on price informativeness. I employ two quasi-natural experiments to alleviate endogeneity concerns. The findings in this paper thus suggest that investor-paid CRAs provide incremental information to the financial market. More importantly, similar to other financial intermediaries, such as security analysts, investor-paid CRAs can play a meaningful monitoring role on

rated firms' disclosure. This monitoring role may be particularly useful as a pre-emptive force to mitigate the accumulation of information asymmetry and asset overvaluation in the financial market. Given that the credit rating industry is still dominated by three issuer-paid CRAs, this paper also generates policy implications that investor-paid ratings can play more than just a marginal role in the rating industry.

References

- Ajinkya, B., Bhojraj, S., Sengupta, P., 2005. The association between outside directors, institutional investors and the properties of management earnings forecasts. *Journal of Accounting Research* 43, 343-376
- Alissa, W., Bonsall IV, S.B., Koharki, K., Penn Jr, M.W., 2013. Firms' use of accounting discretion to influence their credit ratings. *Journal of Accounting and Economics* 55, 129-147
- Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets* 5, 31-56
- Badoer, D.C., Demiroglu, C., 2018. The Relevance of Credit Ratings in Transparent Bond Markets. *The Review of Financial Studies*, 32(1), 42-74
- Badoer, D.C., Demiroglu, C., James, C.M., 2019. Ratings quality and borrowing choice. *The Journal of Finance*.
- Beaver, W.H., Shakespeare, C., Soliman, M.T., 2006. Differential properties in the ratings of certified versus non-certified bond-rating agencies. *Journal of Accounting and Economics* 42, 303-334
- Beck, A.W., 2018. Opportunistic financial reporting around municipal bond issues. *Review of Accounting Studies* 23, 785-826

- Bhattacharya, U., Wei, K., Xia, H., 2019. Follow the money: Investor trading around investor-paid credit rating changes. *The Journal of Corporate Finance*, 58, 68-91
- Bolton, P., Freixas, X., Shapiro, J., 2012. The credit ratings game. *The Journal of Finance* 67, 85-111
- Brealey, R., and S. Myers, 2003, *Principles of Corporate Finance* (7th ed.), Irwin McGraw Hill, New York.
- Brogaard, J., Li, D., Xia, Y., 2017. Stock liquidity and default risk. *Journal of Financial Economics* 124, 486-502
- Bruno, V., Cornaggia, J., Cornaggia, K.J., 2015. Does regulatory certification affect the information content of credit ratings? *Management Science* 62, 1578-1597
- Bushee, B.J., Noe, C.F., 2000. Corporate disclosure practices, institutional investors, and stock return volatility. *Journal of Accounting Research*, 171-202
- Chu, C. S., Rysman, M., (2019). Competition and Strategic Incentives in the Market for Credit Ratings: Empirics of the Financial Crisis of 2007. *American Economic Review*, 109(10), 3514-55
- Committee on Governmental Affairs. 2002. Rating the raters: Enron and the credit rating agencies. Washington: United States Senate.
- Cornaggia, J., Cornaggia, K.J., 2013. Estimating the costs of issuer-paid credit ratings. *The Review of Financial Studies* 26, 2229-2269
- Dai, L., Parwada, J.T., Zhang, B., 2015. The governance effect of the media's news dissemination role: Evidence from insider trading. *Journal of Accounting Research*, 53(2), 331-366
- Dichev, I. D., Piotroski, J. D., 2001. The long-run stock returns following bond ratings changes. *The Journal of Finance*, 56(1), 173-203.
- Durnev, A., Morck, R., Yeung, B., Zarowin, P., 2003. Does greater firm-specific return variation mean more or less informed stock pricing? *Journal of Accounting Research* 41, 797-836

- Dyck, A., Morse, A., Zingales, L., 2010. Who blows the whistle on corporate fraud?.
The Journal of Finance, 65(6), 2213-2253
- Egan Jones Ratings, 2015. Procedures and Methodologies for Determining Credit Ratings
https://www.sec.gov/Archives/edgar/data/1651331/000165133115000003/NR_SRO_Exhibit_2_093015.pdf
- Ferreira, D., Ferreira, M. A., Raposo, C. C., 2011. Board structure and price informativeness. Journal of Financial Economics, 99(3), 523-545
- Ferreira, M.A., Laux, P.A., 2007. Corporate governance, idiosyncratic risk, and information flow. The Journal of Finance 62, 951-989
- Fracassi, C., Petry, S., Tate, G., 2016. Does rating analyst subjectivity affect corporate debt pricing? Journal of Financial Economics 120, 514-538
- Francis, J., Nanda, D., Olsson, P., 2008. Voluntary disclosure, earnings quality, and cost of capital. Journal of Accounting Research 46, 53-99
- Goldstein, I., Huang, C., 2018. Credit Rating Inflation and Firms' Investments. Available at SSRN 3082428.
- Goldstein, M. A., Hotchkiss, E. S., 2019. Providing liquidity in an illiquid market: Dealer behavior in US corporate bonds. Journal of Financial Economics
- Griffin, J.M., Tang, D.Y., 2012. Did subjectivity play a role in CDO credit ratings? The Journal of Finance 67, 1293-1328
- Gul, F.A., Srinidhi, B., Ng, A.C., 2011. Does board gender diversity improve the informativeness of stock prices? Journal of Accounting and Economics 51, 314-338
- Hand, J. R., Holthausen, R. W., Leftwich, R. W., 1992. The effect of bond rating agency announcements on bond and stock prices. The Journal of Finance, 47(2), 733-752
- He, J., Qian, J., Strahan, P.E., 2011. Credit ratings and the evolution of the mortgage-backed securities market. American Economic Review 101, 131-35

- Healy, P. M., Palepu, K. G., 2003. The fall of Enron. *Journal of economic perspectives*, 17(2), 3-26
- Hong, H., Kacperczyk, M., 2010. Competition and bias. *The Quarterly Journal of Economics* 125, 1683-1725
- Jiang, J.X., Stanford, M.H., Xie, Y., 2012. Does it matter who pays for bond ratings? Historical evidence. *Journal of Financial Economics* 105, 607-621
- Jin, L., Myers, S.C., 2006. R2 around the world: New theory and new tests. *Journal of Financial Economics* 79, 257-292
- Jorion, P., Shi, C., Zhang, S., 2009. Tightening credit standards: the role of accounting quality. *Review of Accounting Studies* 14, 123-160
- Jung, B., Soderstrom, N., Yang, Y.S., 2013. Earnings smoothing activities of firms to manage credit ratings. *Contemporary Accounting Research* 30, 645-676
- Hayek, F. A., 1945. The use of knowledge in society. *The American economic review*, 35(4), 519-530
- Kelly, B., Ljungqvist, A., 2012. Testing asymmetric-information asset pricing models. *The Review of Financial Studies* 25, 1366-1413
- Kisgen, D.J., 2006. Credit ratings and capital structure. *The Journal of Finance* 61, 1035-1072
- Kliger, D., Sarig, O., 2000. The information value of bond ratings. *The Journal of Finance*, 55(6), 2879-2902
- Kothari, S.P., Leone, A.J., Wasley, C.E., 2005. Performance matched discretionary accrual measures. *Journal of Accounting and Economics* 39, 163-197
- Kraft, P., 2015a. Rating agency adjustments to GAAP financial statements and their effect on ratings and credit spreads. *The Accounting Review*, 90(2), 641-674
- Kraft, P., 2015b. Do rating agencies cater? Evidence from rating-based contracts. *Journal of Accounting and Economics* 59, 264-283

- Li, Y., Zhang, L., 2015. Short selling pressure, stock price behavior, and management forecast precision: Evidence from a natural experiment. *Journal of Accounting Research*, 53(1), 79-117
- Lin, T. C., Liu, Q., Sun, B., 2019. Contractual managerial incentives with stock price feedback. *American Economic Review*, 109(7), 2446-68
- Liu, A.Z., Subramanyam, K., Zhang, J., Shi, C., 2017. Do firms manage earnings to influence credit ratings? Evidence from negative credit watch resolutions. *The Accounting Review* 93, 267-298
- Luo, Y., 2005. Do insiders learn from outsiders? Evidence from mergers and acquisitions. *The Journal of Finance*, 60(4), 1951-1982
- Manso, G., 2013. Feedback effects of credit ratings. *Journal of Financial Economics* 109, 535-548
- Miller, G. S., 2006. The press as a watchdog for accounting fraud. *Journal of Accounting Research*, 44(5), 1001-1033
- Norden, L., Weber, M., 2004. Informational efficiency of credit default swap and stock markets: The impact of credit rating announcements. *Journal of Banking & Finance*, 28(11), 2813-2843
- Opp, C.C., Opp, M.M., Harris, M., 2013. Rating agencies in the face of regulation. *Journal of Financial Economics* 108, 46-61
- Piotroski, J. D., Roulstone, D. T., 2004. The influence of analysts, institutional investors, and insiders on the incorporation of market, industry, and firm-specific information into stock prices. *The Accounting Review*, 79(4), 1119-1151
- Sangiorgi, F., Spatt, C., 2016. Opacity, credit rating shopping, and bias. *Management Science* 63, 4016-4036
- Securities and Exchange Commission (SEC). 2018. Annual report on nationally recognized statistical rating organizations. Washington, DC: SEC.

- Standard & Poor's (S&P), 2008. Corporate ratings criteria. Available at:
http://regulationbodyofknowledge.org/wp-content/uploads/2013/03/StandardAndPoors_Corporate_Ratings_Criteria.pdf
- Subrahmanyam, A., Titman, S., 2001. Feedback from stock prices to cash flows. *The Journal of Finance* 56, 2389-2413
- Tang, T.T., 2009. Information asymmetry and firms' credit market access: Evidence from Moody's credit rating format refinement. *Journal of Financial Economics* 93, 325-351
- Weinstein, M. I., 1977. The effect of a rating change announcement on bond price. *Journal of Financial Economics*, 5(3), 329-350
- Xia, H., 2014. Can investor-paid credit rating agencies improve the information quality of issuer-paid rating agencies?. *Journal of Financial Economics*, 111(2), 450-468.
- Yu, F., 2005. Accounting transparency and the term structure of credit spreads. *Journal of Financial Economics* 75, 53-84
- Yu, F.F., 2008. Analyst coverage and earnings management. *Journal of financial economics* 88, 245-271

Appendix A: Variable Definitions

Variables	Definition
Panel A: Informativeness variables	
R^2	The R^2 from the annual regression of firm's daily stock return on the Fama-French three-factor model
Ψ	Annual Logistic relative Idiosyncratic volatility estimated from the Fama-French three-factor model
$\Psi_{mkt\ factor}$	Annual Logistic relative Idiosyncratic volatility estimated from the one-factor market model
$\Psi_{FF\ 5\ factor}$	Annual Logistic relative Idiosyncratic volatility estimated from the Fama-French five-factor model
$FINC$	The increase in R^2 of the annual regression on each two-digit industry in the form of $r_{j,t} = \beta_0^b + \beta_1^b \Delta EPS_{j,t} + \sum_{\tau=1}^3 \beta_{2,\tau}^b \Delta EPS_{j,t+\tau} + \sum_{\tau=1}^3 \beta_{3,\tau}^b r_{j,t+\tau}$, relative to base regression: $r_{j,t} = \beta_0^c + \beta_1^c \Delta EPS_{j,t} + \varepsilon_j^c$

Panel B: Disclosure quality variables

$MJones$	Absolute discretionary accrual based on Modified Jones model: $TA_{it} = \beta_0^{da} + \beta_1^{da}(1/Assets_{i,t-1}) + \beta_2^{da}((\Delta SALE_{i,t} - \Delta AR_{i,t})/Assets_{i,t-1}) + \beta_3^{da}PPE_{it}/Assets_{i,t-1} + \varepsilon_{i,t}$
$MJones_ROA$	Absolute discretionary accrual based on the Modified Jones model with ROA: $TA_{it} = \beta_0^{da} + \beta_1^{da}(1/Assets_{i,t-1}) + \beta_2^{da}((\Delta SALE_{i,t} - \Delta AR_{i,t})/Assets_{i,t-1}) + \frac{\beta_3^{da}PPE_{it}}{Assets_{i,t-1}} + ROA_{i,t} + \varepsilon_{i,t}$
$Perf_Matched$	Performance-matched absolute discretionary accrual. Specifically, I match each firm-year observation with another from the same two-digit SIC code with the closest return on assets (net income divided by total assets) in the current year, and define $Perf_Matched$ as the Modified Jones model discretionary accrual minus the matched firm's Modified Jones model accrual.
Bad_News	The frequency of negative management forecasts during a year
$Good_News$	The frequency of positive management forecasts during a year

Panel C: Credit rating coverage variables	
<i>EJR Cover</i>	A dummy that switches on if a firm is rated by Egan Jones Ratings in a year
<i>S&P's Cover</i>	A dummy that switches on if a firm is rated by S&P in a year
Panel D: Control variables	
<i>Leverage</i>	The book value of long-term debt and debt in current liabilities scaled by the book value of assets
<i>MB</i>	The Market-to-Book ratio calculated as the market value of equity plus book value of debt scaled by the book value of total assets.
<i>CAP</i>	The market capitalization of common shares outstanding
<i>ROE</i>	Return on equity measured as the ratio of earnings before extraordinary items to book value of equity
<i>VROE</i>	The variance of quarterly ROE over the past three years
<i>ΔInst</i>	The average change in institutional investor's holding scaled by yearly average trading volume
<i>DD</i>	An annual dividend dummy, which equals one if the firms pay dividend, and zero otherwise (Compustat annual item 32)
<i>AGE</i>	The number of years since the stock inclusion into the CRSP database
<i>DIVER</i>	A dummy variable that equals one for firms operating in more than one segments, and zero otherwise
<i>Z-score</i>	Modified Altman's Z-score, calculated using the following equation: $(1.2*wcap + 1.4*re + 3.3*ebit + 0.999*sale)/at$. I exclude the ratio of market value of equity to book value of debt, because I already have a similar term, MB, in the regression.
<i>Debt_Issue</i>	Net debt financing calculated as Compustat annual item 111 minus – annual item 114, with both items scaled by total assets.

Appendix B: Alternative measures of stock price informativeness

This table presents the results on alternative measures of stock price informativeness. In column (1), the dependent variable is the price non-synchronicity estimated from a single factor market model. In column (2), non-synchronicity is estimated using a two-factor index model that includes the market return and SIC two-digit industry return. In column (3), Fama-French five-factor model is used to estimate non-synchronicity. The t -statistics in parentheses from column (1) to (3) are based on robust standard errors clustered at the firm level. In column (4), the dependent variable is the *FINC* measure from Durnev et al. (2001); the regression is at the two-digit SIC industry level, t -statistics are based on robust standard errors at the industry level. The definitions of other control variables are provided in the Appendix. Superscript *, **, and *** denote significance level at 10%, 5%, and 1%, respectively.

	(1)	(2)	(3)	(4)
	$\Psi_{mkt\ factor}$	$\Psi_{mkt\ \&\ ind\ factors}$	$\Psi_{FF\ 5\ factor}$	<i>FINC</i>
<i>EJR Cover</i>	0.104*** (2.982)	0.080*** (2.770)	0.090*** (3.463)	0.064** (2.170)
<i>S&P's Cover</i>	-0.049 (-1.250)	-0.086** (-2.533)	-0.053* (-1.764)	-0.029 (-0.730)
<i>LEV</i>	-0.055 (-0.519)	-0.016 (-0.178)	-0.067 (-0.846)	-0.037 (-0.545)
<i>MB</i>	-0.050*** (-3.130)	-0.050*** (-3.966)	-0.043*** (-3.570)	-0.007 (-0.558)
<i>CAP</i>	-0.416*** (-15.955)	-0.450*** (-21.167)	-0.305*** (-17.020)	-0.020* (-1.743)
<i>ROE</i>	-0.071 (-1.408)	-0.056 (-1.476)	-0.047 (-1.495)	-0.040 (-0.562)
<i>VROE</i>	0.002 (0.152)	-0.009 (-0.841)	-0.005 (-0.454)	0.030** (2.053)
$\Delta Inst$	1.153*** (5.835)	1.223*** (7.770)	0.866*** (7.119)	0.205 (1.510)
<i>DD</i>	0.073 (1.567)	0.049 (1.258)	0.037 (1.107)	0.023 (0.786)
<i>AGE</i>	0.030 (0.861)	-0.018 (-0.560)	0.008 (0.323)	-0.061*** (-4.385)
<i>DIVER</i>	0.001 (0.033)	-0.036 (-1.124)	-0.010 (-0.362)	0.089*** (2.914)
<i>Z-score</i>	-0.089*** (-3.905)	-0.058*** (-3.151)	-0.060*** (-3.542)	0.056*** (6.269)
<i>Debt_Issue</i>	-0.048 (-0.466)	0.054 (0.676)	-0.001 (-0.010)	0.324** (1.979)
Firm FE	Y	Y	Y	N
Year FE	Y	Y	Y	Y
Industry FE	N	N	N	Y
N/ Adj. R-squared	14,408/0.718	14,407/0.790	14,406/0.729	743/0.170

Table 1.1: Summary statistics

This table provides summary statistics for the variables used in regressions. The sample consists of Compustat firms that are rated by either EJR or S&P's. Financial firms, firms that are covered by EJR before the third quarter of 1999, and firms with missing variables used in regression Eq. (4), are all excluded. All dollar values are in 2010 dollars. The definitions of variables are provided in the Appendix. All continuous variables are winsorized at the 1st and 99th percentiles.

Variable	Mean	P25	P50	P75	SD	N
<i>Panel A: Informativeness variables</i>						
R^2	0.272	0.129	0.256	0.391	0.174	14,407
Ψ	1.279	0.444	1.067	1.905	1.208	14,407
$\Psi_{mkt\ factor}$	1.660	0.629	1.304	2.27	1.584	14,407
$\Psi_{FF\ 5\ factor}$	1.140	0.347	0.971	1.769	1.119	14,406
$FINC$	0.139	0.024	0.085	0.200	0.15	743
<i>Panel B: Disclosure quality variables</i>						
$MJones$	0.234	0.027	0.076	0.222	0.429	12,997
$MJones_ROA$	0.220	0.026	0.072	0.213	0.374	13,037
$Perf_Matched$	0.330	0.039	0.101	0.286	0.634	12,938
Bad_News	0.733	0.000	0.693	1.099	0.621	5,268
$Good_News$	0.478	0.000	0.000	0.693	0.556	5,268
<i>Panel C: Credit rating coverage variables</i>						
$EJR\ Cover$	0.331	0	0	1	0.470	14,407
$S\&P's\ Cover$	0.713	0	1	1	0.452	14,407
<i>Panel D: Control variables</i>						
Lev	0.318	0.163	0.296	0.436	0.221	14,407
MB	1.851	1.139	1.451	2.068	1.224	14,407
CAP	7.328	6.167	7.353	8.466	1.705	14,407
ROE	0.003	-0.004	0.026	0.050	0.216	14,407
$VROE$	0.225	0.014	0.030	0.078	0.847	14,407
$\Delta Inst$	0.040	0.004	0.012	0.035	0.090	14,407
DD	0.427	0.000	0.000	1.000	0.495	14,407
Age	2.515	1.946	2.639	3.332	0.954	14,407
$DIVER$	0.478	0.000	0.000	1.000	0.500	14,407
$Z\text{-}score$	1.460	0.755	1.523	2.321	1.456	14,407
$Debt_Issue$	0.015	-0.016	0.000	0.033	0.090	14,407

Table 1.2: Determinants of EJR and S&P's coverage

This table reports the probit regression results on the determinants of EJR and SP's coverage initiation. The sample consists of Compustat firms from 1999 to 2016 with non-missing values for variables used in the model. All explanatory variables are lagged for one year. In the first (second) column, the dependent variable is a dummy variable indicating the year when EJR (S&P's) starts to rate a firm; firms are dropped from the sample once they are covered by EJR (first column) or S&P (second column). *Sale* is the natural logarithm of firms' sales. *Equity_Issue* is firm's net equity financing calculated as Compustat annual item 108 – item 115, with both items scaled by total assets. *DTURN* is the detrended average monthly stock turnover. *IO* is the total institutional ownership measured in the last quarter of each year. *Ln (Analysts)* is the natural logarithm of the number of analysts covering a firm in a year. *Tangibility* is net property, plant, and equipment scaled by total assets. *R&D* is research and development expenditure scaled by total assets. The definitions of other explanatory variables are provided in the legend of Table 1.1. Both regressions control for year and industry fixed effects. The *z*-statistics are based on robust standard errors clustered at the firm-level and are reported in parentheses. Superscript *, **, and *** denote significance level at 10%, 5%, and 1%, respectively.

	(1)	(2)
	<i>EJR Coverage Initiation_{t+1}</i>	<i>S&P's Coverage Initiation_{t+1}</i>
Ψ	0.019 (0.644)	0.005 (0.178)
<i>Sale</i>	0.145*** (4.086)	0.126*** (4.461)
<i>LEV</i>	1.571*** (12.939)	1.074*** (7.714)
<i>MB</i>	-0.066*** (-3.136)	-0.039** (-2.298)
<i>CAP</i>	0.260*** (6.716)	0.261*** (7.357)
<i>ROE</i>	-0.103 (-1.299)	-0.056 (-0.563)
<i>VROE</i>	-0.006 (-0.191)	-0.092** (-2.431)
$\Delta Inst$	-0.281 (-1.299)	-0.149 (-0.835)
<i>DD</i>	-0.087* (-1.655)	-0.178*** (-2.952)
<i>AGE</i>	0.059** (2.233)	-0.079*** (-3.254)
<i>DIVER</i>	0.012 (0.260)	-0.095* (-1.853)
<i>Z-score</i>	-0.025** (-2.552)	0.013 (1.109)
<i>Debt_Issue</i>	-0.148 (-0.632)	0.851*** (3.179)
<i>Equity_Issue</i>	-0.076 (-0.282)	0.690*** (3.631)
<i>DTURN</i>	-0.188 (-0.842)	0.280 (1.385)
<i>IO</i>	0.312***	0.231**

	(2.822)	(2.043)
<i>Ln (Analysts)</i>	0.086**	0.025
	(2.544)	(0.688)
<i>Tangibility</i>	0.063	0.124
	(0.436)	(0.833)
<i>R&D</i>	-0.738**	-0.693
	(-1.977)	(-1.578)
Industry & Year FE	Yes	Yes
N/Pseudo R-squared	37,946 /0.328	33,023/0.225

Table 1.3: Effect of EJR coverage on stock price informativeness

This table reports the baseline regression results on how EJR coverage affects rated firms' stock price informativeness. The sample consists of firms that are either covered by EJR or S&P's during 1999 and 2016, and financial firms are excluded. The dependent variable is the logistic transformed relative idiosyncratic volatility estimated from the Fama-French three-factor model. The key independent variable *EJR Cover* is a dummy variable that equals one if a firm is rated by EJR during the year and zero otherwise. Regression in column (1) excludes *S&P's Cover*, which is a dummy indicating S&P's rates a firm in a year. In column (2), *S&P's Cover* replaces *EJR Cover* as the key independent variable. In the last column, both dummy variables are included. The definitions of other control variables are provided in the Appendix. The *t*-statistics in parentheses allow for clustering at the firm-level, and superscript *, **, and *** denote significance level at 10%, 5%, and 1%, respectively.

	$\Psi_{i,t+1}$		
	(1)	(2)	(3)
<i>EJR Cover</i>	0.089*** (3.211)		0.091*** (3.275)
<i>S&P's Cover</i>		-0.049 (-1.493)	-0.052 (-1.600)
<i>LEV</i>	-0.136 (-1.637)	-0.090 (-1.070)	-0.101 (-1.201)
<i>MB</i>	-0.028** (-2.253)	-0.032** (-2.520)	-0.032** (-2.474)
<i>CAP</i>	-0.329*** (-16.458)	-0.320*** (-15.879)	-0.323*** (-16.049)
<i>ROE</i>	-0.059 (-1.646)	-0.062* (-1.722)	-0.061* (-1.680)
<i>VROE</i>	-0.005 (-0.453)	-0.005 (-0.461)	-0.005 (-0.510)
$\Delta Inst$	0.931*** (6.338)	0.923*** (6.261)	0.923*** (6.285)
<i>DD</i>	0.032 (0.891)	0.031 (0.851)	0.030 (0.825)
<i>AGE</i>	-0.009 (-0.350)	-0.006 (-0.224)	-0.007 (-0.271)
<i>DIVER</i>	-0.005 (-0.157)	-0.004 (-0.116)	-0.004 (-0.145)
<i>Z-score</i>	-0.067*** (-3.653)	-0.066*** (-3.600)	-0.066*** (-3.630)
<i>Debt_Issue</i>	0.005 (0.067)	-0.002 (-0.021)	0.007 (0.084)
Firm & Year FE	Yes	Yes	Yes
N/Adjusted R-squared	14,407 /0.691	14,407 /0.690	14,407 /0.691

Table 1.4: Tests of omitted variable bias and reverse causality

Panel A addresses issues associated with potential omitted variable bias. In column (1), the regression controls for the natural log of one plus the number of firm-level news articles from *Dow Jones Newswire*. In column (2), the control variables include the standard deviation and average of weekly return during the past fiscal year to account for past extreme returns or events that drive both EJR coverage and stock comovement. In column (3), the model further controls for the level (*IO*) and concentration of institutional ownership (*IO_HHI*). In column (4), I augment the model with *ILLQ*, the Amihud illiquidity measure. Lastly, column (5) includes *Equity_Issue*, which is firm's net equity financing calculated as Compustat annual item 108 – item 115, with both items scaled by total assets. Panel B tests the reverse causality by verifying the parallel trend assumption in column (1), and by controlling for one year lagged informativeness measure in column (2). All regressions control for firm and year fixed effects. S&P's coverage and other control variables are included in all regressions but not reported. The *t*-statistics in parentheses are calculated based on robust standard errors clustered at the firm level. ***, **, * denotes significance level at 1%, 5%, and 10%, respectively.

Panel A. Tests for omitted variable bias

	$\Psi_{i,t+1}$				
	(1)	(2)	(3)	(4)	(5)
<i>EJR Cover</i>	0.080*** (2.597)	0.091*** (3.344)	0.087*** (3.172)	0.089*** (3.271)	0.089*** (3.217)
<i>News</i>	0.045*** (4.446)				
<i>Vol</i>		-1.080*** (-3.015)			
<i>Return</i>		-0.048*** (-6.836)			
<i>IO</i>			-0.774*** (-8.485)		
<i>IO_HHI</i>			2.071*** (3.433)		
<i>ILLQ</i>				2.942*** (9.405)	
<i>Equity_Issue</i>					-0.671*** (-5.439)
Firm & Year FE	Yes	Yes	Yes	Yes	Yes
N/Adjusted R-squared	10,414/0.676	14,407/0.693	14,407/0.695	14,313/0.693	14,407/0.692

Panel B. Tests for reverse causality

	$\Psi_{i,t+1}$	
	(1)	(2)
EJR_Cover^{-2}	0.003 (0.071)	
EJR_Cover^{-1}	0.022 (0.499)	
EJR_Cover^0	0.049 (1.159)	
EJR_Cover^1	0.124*** (2.849)	

EJR_Cover^{2+}	0.108*** (2.951)	
$EJR\ Cover$		0.078*** (3.409)
$\Psi_{i,t-1}$		0.276*** (20.889)
Firm & Year FE	Yes	Yes
N/ Adjusted R-squared	14,407/0.691	14,407/0.715

Table 1.5: Using quasi-natural experiments that shock EJR's importance

This table presents the regression results of using brokerage mergers as an exogenous shock to analyst coverage. The sample consists of firms covered either by EJR or S&P's between 1999 and 2016. Brokerage mergers between 1999 and 2008 are obtained from Hong and Kacperzyck (2011), and Kelly and Ljungqvist (2012). The dependent variable is the relative idiosyncratic volatility of stock return. *Analyst Loss* is a dummy variable equals one for a firm-year if that firm experiences analyst coverage reduction due to brokerage mergers less than one year ago. *NRSRO* is an indicator variable that equals one for years after Dec 2007, when EJR received the NRSRO designation from SEC. The definitions of other control variables are provided in the Appendix. The *t*-statistics in parentheses are calculated based on robust standard errors clustered at the firm level. ***, **, * denotes significance level at 1%, 5%, and 10%, respectively.

	$\Psi_{i,t+1}$	
<i>EJR Cover</i>	0.078*** (2.729)	0.052* (1.690)
<i>Analyst Loss</i>	-0.176*** (-3.954)	
<i>EJR Cover</i> $\times$ <i>Analyst Loss</i>	0.173*** (3.043)	
<i>EJR Cover</i> $\times$ <i>NRSRO</i>		0.080** (2.142)
<i>S&P's Cover</i>	-0.051 (-1.577)	-0.048 (-1.493)
<i>LEV</i>	-0.105 (-1.251)	-0.102 (-1.203)
<i>MB</i>	-0.030** (-2.404)	-0.032** (-2.507)
<i>CAP</i>	-0.321*** (-15.868)	-0.322*** (-16.044)
<i>ROE</i>	-0.061* (-1.699)	-0.060* (-1.662)
<i>VROE</i>	-0.005 (-0.479)	-0.006 (-0.531)
$\Delta Inst$	0.911*** (6.207)	0.913*** (6.227)
<i>DD</i>	0.029 (0.798)	0.029 (0.807)
<i>AGE</i>	-0.006 (-0.244)	-0.003 (-0.108)
<i>DIVER</i>	-0.003 (-0.095)	-0.004 (-0.132)
<i>Z-score</i>	-0.068*** (-3.734)	-0.066*** (-3.627)
<i>Debt_Issue</i>	0.007 (0.083)	0.007 (0.095)
Firm & Year FE	Yes	Yes
N/Adjusted R-squared	14,407/0.691	14,407/0.691

Table 1.6: Testing the information channel

Panel A presents results on how the effect of EJR coverage on informativeness depends on the differences between EJR's rating and S&P's rating for the same firm. The dependent variable is the relative idiosyncratic volatility at the firm-year level. *Diff* is a dummy variable that equals one for a firm-year if EJR's rating is different from S&P's for that firm-year. *Frequent_Issuer* is a binary indicator that equals one if a firm issues an above-median number of bonds during the past three years in the Mergent database. *IG_HY* is a binary indicator for firm-years with an investment-grade rating (BBB+ and above) from one of the two rating agencies, while high yield rating (BB+ and lower) from the other. *EJR Cover* is a dummy variable that equals one if a firm-year is rated by EJR, and zero otherwise. S&P's coverage and other control variables in Eq. (4) are included in the regressions but not reported. Panel B reports the second-stage results from the Fama-MacBeth style regression on the sensitivity of S&P's ratings to rated firms accounting quality. *EJR Rating* (*S&P's Rating*) is the average of EJR's (S&P's) rating for a firm in a fiscal year. *ACCQ* is the three-year moving average of a firm's absolute value of the discretionary accrual from the modified Jones model. All independent variables are lagged by one year. The definition of other variables can be found in the Appendix. All regressions control for firm and year fixed effects. The *t*-statistics based on robust standard errors clustered at the firm-level are reported in parentheses. Superscript *, **, and *** denote significance level at 10%, 5%, and 1%, respectively.

<i>Panel A: Cross-sectional variation in the impact of EJR coverage</i>			
	(1)	(2)	(3)
		$\Psi_{i,t+1}$	
<i>EJR Cover</i>	0.0396 (1.203)	0.070** (2.444)	0.082*** (2.978)
<i>EJR Cover</i> $\times$ <i>Diff</i>	0.0766*** (2.695)		
<i>EJR Cover</i> $\times$ <i>Frequent_Issuer</i>		0.091** (2.263)	
<i>EJR Cover</i> $\times$ <i>IG_HY</i>			0.081** (2.368)
Controls	Yes	Yes	Yes
Firm & Year FE	Yes	Yes	Yes
N/Adjusted R-squared	14,407/0.723	14,407/0.723	14,407/0.723
<i>Panel B: Cross-sectional sensitivity of credit ratings to accounting quality</i>			
	<i>EJR Rating</i>	<i>S&P's Rating</i>	<i>S&P's Rating</i>
<i>ACCQ</i>	0.234* (1.790)	0.092 (1.290)	0.064 (0.957)
<i>EJR Cover</i> $\times$ <i>ACCQ</i>			0.152*** (3.072)
<i>EJR Cover</i>			0.162*** (4.840)
N/Average R-squared	7927/0.703	15,271 /0.737	15,271/0.739

Table 1.7: EJR coverage and forecasting price efficiency

This table reports the results of estimating Model 5. Observations are firm-quarters. The dependent variable is the abnormal size characteristic portfolio-adjusted return over the trading days [0, +2] of the quarterly earnings announcement. *EJR Cover* (*S&P's Cover*) is a dummy variable that equals 1 for a firm-quarter if the firm is rated by EJR (S&P's) over the days [-60, -1] of the earnings announcement date, and zero otherwise. Similarly, *EJR Update* (*S&P's Update*) indicates that EJR (S&P's) updates its rating for the firm over the days [-60, -1] of the earnings announcement date. Control variables include loss indicator, fourth fiscal quarter indicator, firm size, book-to-market ratio, number of analysts, institutional ownership, stock price volatility, leverage, beta, earnings persistence, and earnings volatility. Also included are the control variables' interaction term with *ABS_SUE*. The regressions control for firm and year fixed effects. Standard errors are clustered by firm. Superscript *, **, and *** denote significance level at 10%, 5%, and 1%, respectively.

	<i>ABS_AR[0,2]</i>	
	(1)	(2)
<i>EJR Cover</i>	0.004*** (4.313)	
<i>EJR Cover</i> × <i>ABS_SUE</i>	-0.244*** (-3.022)	
<i>S&P's Cover</i>	0.001 (0.579)	
<i>S&P's Cover</i> × <i>ABS_SUE</i>	0.137 (1.521)	
<i>EJR Update</i>		-0.000 (-0.269)
<i>EJR Update</i> × <i>ABS_SUE</i>		-0.418*** (-3.505)
<i>S&P's Update</i>		0.002 (1.254)
<i>S&P's Update</i> × <i>ABS_SUE</i>		-0.044 (-0.411)
<i>Controls</i> × <i>ABS_SUE</i>	Yes	Yes
<i>Firm & Year FE</i>	Yes	Yes
N/Adjusted R-squared	42,611/0.213	42,611/0.213

Table 1.8: Testing the disciplining channel

Panel A presents the OLS regression results on the effects of EJR coverage on three measures of discretionary accruals. The sample consists of firms covered by either EJR and S&P's during 1999 and 2016 with non-missing discretionary accruals, which is measured by the modified Jones model, the modified Jones model with ROA, and performance-matched discretionary accrual based on the modified Jones model. Panel B reports the results on how EJR coverage affects the frequency of good and bad news forecasts. The sample period begins in 2003 when voluntary management forecast data starts to be available. $\text{Log}(\text{Good News}+1)$ and $\text{Log}(\text{Bad News}+1)$ is the natural logarithm of one plus the number of bad news forecasts during a fiscal year. *EJR Cover* (*S&P's Cover*) is a dummy variable that equals one if a firm is rated by EJR (S&P) during the year and zero otherwise. The definitions of other control variables are provided in the Appendix. The regressions in both Panel control for firm and year fixed effects. *T*-statistics in parentheses are calculated from standard errors clustered by firm. Superscript *, **, and *** denote significance level at 10%, 5%, and 1%, respectively.

<i>Panel A: Effects of EJR coverage on earnings management</i>			
	(1)	(2)	(3)
	<i>MJones</i>	<i>MJones_ROA</i>	<i>Perf_Matched</i>
<i>EJR Cover</i>	-0.023*	-0.027**	-0.041**
	(-1.654)	(-2.348)	(-2.019)
<i>S&P's Cover</i>	-0.011	0.001	-0.019
	(-0.703)	(0.057)	(-0.884)
Controls	Yes	Yes	Yes
Firm & Year FE	Yes	Yes	Yes
N/R-Square	12,997/0.216	13,037/0.240	12,937/0.161

<i>Panel B: Effects of EJR coverage on the sign of management forecast</i>		
	(1)	(2)
	$\text{Log}(\text{Bad News}+1)$	$\text{Log}(\text{Good News}+1)$
<i>EJR Cover</i>	0.075**	-0.054
	(2.236)	(-1.620)
<i>S&P's Cover</i>	0.027	-0.031
	(0.750)	(-0.874)
Controls	Yes	Yes
Firm & Year FE	Yes	Yes
N/Pseudo R-squared	5,268/0.241	5,268/167

Table 1.9: Propensity-score-matched samples

This table presents the estimation results of the impact of EJR coverage on price informativeness using propensity-score-matched samples from 1999 to 2016. The dependent variable is relative idiosyncratic volatility at the firm-year. The overall sample of firms is taken from Compustat. Propensity-score-matched firms are selected based on the propensity scores estimated from Model (3) in Section 3.3. Four different matching procedures are used to search for the nearest firms. In column (1), I find a control firm that has the nearest propensity score with EJR-rated firms. In column (2), the nearest control firm is further required to have a propensity score difference within 1% from the EJR-rated firms. In column (3), each EJR-rated firm are matched to two nearest control firms in terms of propensity score. In column (4), the two nearest control firms are further required to have a propensity score difference within 1% from the EJR-rated firm. Variable definitions are provided in the Appendix. The numbers in parentheses are t -statistics based on standard errors clustered at the firm-level. Superscript *, **, and *** denote significance level at 10%, 5%, and 1%, respectively.

	$\Psi_{i,t}$			
	(1) Nearest one matching	(2) Nearest one PS diff < 1%	(3) Nearest two matching	(4) Nearest two PS diff < 1%
<i>EJR Cover</i>	0.063** (2.041)	0.067** (2.152)	0.071** (2.157)	0.077** (2.318)
<i>S&P's Cover</i>	-0.025 (-0.489)	-0.019 (-0.372)	-0.041 (-0.678)	-0.031 (-0.513)
<i>LEV</i>	-0.095 (-0.689)	-0.117 (-0.848)	-0.075 (-0.533)	-0.080 (-0.586)
<i>MB</i>	-0.057*** (-3.708)	-0.053*** (-3.516)	-0.066*** (-4.126)	-0.059*** (-3.888)
<i>CAP</i>	-0.287*** (-9.250)	-0.289*** (-9.152)	-0.278*** (-9.032)	-0.285*** (-9.372)
<i>ROE</i>	0.025 (0.472)	0.043 (0.823)	-0.002 (-0.039)	0.011 (0.218)
<i>VROE</i>	-0.008 (-0.461)	-0.008 (-0.459)	-0.014 (-0.742)	-0.013 (-0.701)
<i>$\Delta Inst$</i>	0.647*** (4.444)	0.641*** (4.408)	0.640*** (5.471)	0.623*** (5.380)
<i>DD</i>	0.019 (0.392)	-0.007 (-0.156)	0.024 (0.515)	0.007 (0.148)
<i>AGE</i>	-0.038 (-0.904)	-0.029 (-0.696)	-0.059 (-1.230)	-0.041 (-0.907)
<i>DIVER</i>	-0.046 (-1.038)	-0.046 (-1.015)	-0.070 (-1.429)	-0.062 (-1.273)
<i>Z-score</i>	-0.050*** (-3.002)	-0.051*** (-3.066)	-0.056*** (-3.507)	-0.057*** (-3.615)
<i>Debt_Issue</i>	0.053 (0.415)	0.042 (0.329)	0.005 (0.044)	-0.008 (-0.068)
Firm & Year FE	Yes	Yes	Yes	Yes
N/Adjusted R-squared	13,879/0.705	13,447/0.707	20,469/0.709	19,535/0.713

Chapter 2: *P*-hacking in Experimental Accounting Studies

Abstract

We provide evidence of *p*-hacking in experimental accounting studies. Extracting *p*-values from experimental studies published in the three top accounting journals from 1990 to 2016, we find that the frequency of *p*-values equal to 0.05 far exceeds what would be expected based on the frequency of neighboring *p*-values. The discontinuity at 0.05 is more pronounced among papers written by male authors, authors who face greater career concerns, and who have more opportunities for *p*-hacking. Papers that show stronger *p*-hacking receive more citations after publication. We find that the updated data and code sharing policy adopted by the Journal of Accounting Research since 2014 appears to deter *p*-hacking. Lastly, we suggest several potential solutions to guard against *p*-hacking in experimental accounting studies.

Keywords: *P*-hacking; Experimental accounting; Statistical significance

“Given the competition for top journal space, there is an incentive to produce ‘significant’ results”.

Campbell Harvey (2017)

1. Introduction

There is growing consensus among various disciplines that due to intensifying competition for top journal spaces, researchers may shift their focus away from producing scientific findings to hit publication standards (e.g., Simmons, Nelson, and Simonsohn, 2011; Masicampo and Lalande, 2012; Head et al., 2015; Krawczyk, 2015; Harvey, 2017). Chief among the standards for publication is statistical significance, which helps journals attract more readers and citations (Fanelli, 2012). When it comes to search for statistical significance, researchers can be quite “resourceful.” For instance, researchers can: try many tests and statistical methods, collect less or more data, combine or transform specific measures, delete or keep certain observations, employ more or fewer dependent variables or control variables, and so on. The behavior of exploiting these discretions to search and report the most significant results is, however, often considered controversial or even unethical, and is generally referred to as *p*-hacking.

P-hacking may benefit researchers and their affiliated institutions by inflating the chance of top journal publications. But it impedes progress in a research area in the long run by increasing the false-positive rates. Furthermore, higher false positives rates, combined with the undisclosed decisions made by researchers during the data analysis process, will lead to a higher likelihood that these findings are irreproducible in the future. Such concern is not groundless. Open Science Collaboration (2015) initiated large-scale replication of 100 experimental studies in psychology. The main finding is somewhat surprising:

only 36% of the replications have significant results, a decrease from 97% in the original articles.¹ Yet reproducibility is regarded as “a core principle of scientific progress” (Open Science Collaboration, 2015). Consequently, discussions of *p*-hacking and its potential remedies have attracted considerable attention, and *p*-hacking has been found to be prevalent in many disciplines (e.g., psychology, communication, biomedical research, finance, and economics). However, there are important cross-disciplinary variations in the extent of *p*-hacking.

P-hacking is most notable in fields that rely on laboratory experiments to conduct analyses. Such fields include psychology, pharmacology, and toxicology (e.g., Head et al., 2015; Harvey, 2017; Fanelli, 2012). Several reasons have been proposed to explain this phenomenon. First, unlike areas that rely on public (i.e., archival) data sources, experimental studies often create their own data sets. This gives experimentalists additional discretions regarding sample selection (Wicherts et al., 2016). Second, to the extent that these data sets are often kept proprietary, follow-up replications are costly as it requires constructing a new sample. The difficulties in replication may *ex ante* enhance the incentive to engage in *p*-hacking because results are less likely to be cross-validated later. Lastly, studies with smaller sample sizes are more subject to *p*-hacking because low statistical power makes it harder to obtain significant results (Head et al., 2015). Furthermore, results based on a small sample is particularly sensitive to data manipulations, rendering the effect of *p*-hacking on significance levels more powerful.

¹ Moreover, the mean effects size from the replication is merely half the mean effects size of the original results.

In this paper, we investigate the prevalence of *p*-hacking among experimental accounting studies. Aside from the above reasons that experimental accounting studies may be prone to *p*-hacking, our focus on this specific area is motivated by two additional reasons. First, experimental studies in accounting have been growing and taken a stronghold in the U.S. accounting academy.² This is perhaps due to an important comparative advantage experimental studies have over archival-empirical research. That is, by randomizing the treatments and controlling for confounding factors, experimentalists can easily draw causal conclusions and even identify the underlying mechanism behind a causal relationship. The growing importance of experimental accounting research strengthens the need to guard against *p*-hacking in this area. Second, a methodological advantage in focusing on experimental studies is that papers in this area commonly report *p*-values, instead of *t*- or *F*- statistics, in text. This greatly facilitates our data collection and analyses (introduced later).

We examine the extent of *p*-hacking by revealing a discontinuity in the pooled distribution of reported *p*-values. The reliance on the *p*-value distribution to detect *p*-hacking exploits an essential feature of the Null Hypotheses Significance Testing (NHST): for an effect to be considered significant in experimental accounting studies, its corresponding *p*-value must be ≤ 0.05 . This arbitrary standard for significance is made to facilitate the discrimination between rejected and accepted hypotheses, a selection also dubbed as *publication bias* (see, e.g., Sterling, 1959; Greenwald, 1975). However, a potential consequence of this standard is that researchers may consider it as a hurdle for

² Libby, Bloomfield, and Nelson (2002) discuss several factors that promote progress in experimental accounting research since 1990s.

their papers to be published, and rely on p -hacking to turn insignificant results into significant ones, just like managers engage in earnings management to turn small losses into small gains.

How does p -hacking introduce discontinuity into the distribution of p -values? Consider a researcher who works on a project that may yield results with uncertain p -values. Under the condition that her null hypothesis is no effect, the researcher will clearly prefer and keep p -values below or equal to 0.05. From the publication point of view, the remaining p -values can be divided into two groups, those that are “promising,” i.e., close to 0.05 but not there yet, and those “beyond hope,” i.e., much larger than 0.05. Suppose the researcher has some discretion in data collection and analyses, for example, she can conduct another experiment to gather more data, or delete some seemingly-outlier participants from the sample. Now, if she got a p -value of 0.06, the marginal cost of adjusting her sample or methods to achieve a p -value below 0.05 should be much lower than the marginal benefit of doing so, which is a higher chance of publication. However, this may not be the case if the initial p -value is very large. If such a practice is widespread, we should observe an unusually high frequency of p -values right below or equal to 0.05, followed by a significant decline in the frequencies of p -values slightly larger than 0.05.

Using a sample of 2,404 p -values reported in 240 experimental accounting articles published in three top accounting journals (i.e., Journal of Accounting and Economics, Journal of Accounting Research, and The Accounting Review), we document that the distribution of these p -values exhibits a sharp discontinuity at 0.05. Moreover, the frequency of “just” significant p -value (0.05) is higher than the frequencies of 0.04 and 0.03, as if researchers

prefer “just” significant results to more significant ones. Interestingly, the frequency of 0.06 is unusually low: it is lower than the frequency of 0.07, suggesting that at least part of the marginally significant results from the articles are inflated into significant results, resulting in an over-representation of just-significant p -values.

It is worthwhile to point out that publication bias cannot explain the discontinuity in the distribution of p -values. If journals selectively publish papers with more significant results, lower p -values should be more preferred by editors (e.g., $p = 0.03$ is more significant than $p = 0.05$). Therefore, unless $p = 0.05$ is more likely to occur in statistical analyses without selective reporting from authors, publication bias *alone* will lead to a monotonic decreasing frequency of p -values as we move from 0 to 1. This monotonic selection turns out to be a useful benchmark that portrays the “natural” distribution of p -values in the absence of p -hacking, and thus us to decompose the observed frequency of p -values into two parts: the part due to publication bias, and the part due to p -hacking.³

Theoretically, any distribution that is monotonically decreasing in $[0, +\infty)$ can be potential candidates for the benchmark. We follow Masicampo and Lalande (2012) and Harvey, Liu, and Zhu (2016) to choose an exponential distribution to fit our distribution. The exponential curve matches well with the empirical distribution of p -values in our sample: the R-squared from the fitting exercise is, on average, above 98%. The discontinuity around 0.05 generates a

³ There may be also other potential selection procedures that shape the distribution of p -values. For examples, authors might simply censor themselves; there is also a potential feedback effect as outlined by Henry (2009): editors and referees prefer significant results, pushing authors to generate more stars, this may further increase the requirement from editors and referees. However, the monotonic decreasing benchmark distribution are able to capture all those non-increasing selections (i.e., lower p -values are more likely to be selected).

large residual that cannot be explained by publication bias, and we thus attribute this residual to inflation from *p*-hacking. After removing the inflation component, our conservative estimate from the fitting shows that the expected frequency of 0.05 should be 23% less than observed. The residual of fitting for 0.05 is also much larger than the mean residuals of other *p*-values.

Having shown the existence and extent of *p*-hacking, we turn to examine what predicts the extent of *p*-hacking. By exploiting the cross-sectional variations in authors' incentives and costs to conduct *p*-hacking, we identify five author and paper characteristics. For example, we find that *p*-hacking is more prevalent when researchers face greater career concern, have more opportunities for *p*-hacking, and are more likely to rationalize *p*-hacking. These large cross-sectional variations in the extent of *p*-hacking again confirm that the discontinuity is indeed caused by *p*-hacking; more importantly, the absence of *p*-hacking in certain subsamples indicates the prevalence of good research practice among many researchers. We also document stronger results among articles written by authors working in (or graduated as PhD from) higher ranking schools. Additional test also shows that more cited articles exhibit a greater extent of *p*-hacking.

We have argued that the extent of *p*-hacking partly depends on the degree of researchers' discretion over data collection and analyses, with higher discretion indicating more opportunities for *p*-hacking. We empirically test this conjecture by exploiting the adoption of the updated data and code policy by the Journal of Accounting Research (JAR) in December 2014. The policy requires stricter data and code disclosure from authors to enhance research transparency and replicability. We argue that the more stringent data and code sharing policy reduces the discretion available to researchers during the data collection and

analyses and thus the extent of p -hacking, leading to a smaller discontinuity in the distribution of p -values around 0.05. We find results consistent with this prediction.⁴

Lastly, we provide some solutions to help guard against false positives resulted from p -hacking. We first introduce three broad solutions developed from the statistics literature to adjust for multiple testing and then generate simulated “experiments” to demonstrate the application and effectiveness of these methods. The simulation results show that most of the significant findings from p -hacking are unlikely to survive after adjusting for multiple testing or applying the Bayesian perspective. We also introduce the Bayesian approach and discuss its potential application, and provide some guidelines for journals that try to improve the disclosure transparency from authors. The rest of this article proceeds as follows. In Section 2, we clarify a basic confusion on the definition of p -value and review the related literature on p -hacking. We introduce our sample construction and empirical strategy in Section 3, and present evidence on the existence and extent of p -hacking in Section 4; Section 5 further explores the factors that shape researchers’ incentive to conduct p -hacking, Section 6 presents additional tests. Section 7 includes the proposed remedies for p -hacking, and we conclude in Section 8.

2. Background information and related literature

2.1. Understanding p -values: what are we hacking?

The use of p -value null hypothesis testing is widespread in science. Initially introduced by Fisher (1925), it is designed to separate findings from

⁴ In Appendix A, we also display the distribution of p -values from archival-empirical studies, which arguable use more public data to conduct research.

noises by measuring the extent to which the data contradicts a hypothesis (i.e., the null hypothesis, and no alternative hypothesis are defined). Formally, the Fisherian p -value is the probability of observing an outcome or more extreme outcomes, given the *null hypothesis being true*. Users of the Fisherian approach may evaluate a hypothesis based on the corresponding p -values, with smaller p -values indicating stronger evidence against the null. However, such evaluation can only be applied to the *particular* data set under examination, implying that inferences based on this approach may not generalize. To draw from Neyman and Pearson (1933): “as far as a particular hypothesis is concerned, no test based upon the theory of probability can by itself provide any valuable evidence of the truth or falsehood of that hypothesis.”

Aware of the above limitation of the p -value, Neyman and Pearson (1933) came up with an alternative framework. The key argument of Neyman and Pearson is as follows. While we may not be able to learn from a single p -value (or any other test statistics) on the truth or falsehood of a given hypothesis, there is one way to make sure that *in the long run*, we make the right decisions most of the time. That is, if we always follow the rule of rejecting the null when the p -value is lower than a predetermined Type I error rate, say 0.05, then *over repetitive experiments*, we will reject the null when it's true around 5% of the time. However, over the years after the Fisherian p -value is combined with the Neyman and Pearson (1933) framework to form the current Null Hypothesis Significance Testing framework, p -value is often misinterpreted even among scientists in many fields (see, Pashler and Harris, 2012). Specifically, in NHST, the p -value of a specific test is calculated, and then compared with a predetermined Type I error rate (e.g., $\alpha=0.05$). If $p\text{-value} < \alpha$, it is usually

interpreted as evidence rejecting the null hypothesis and accepting the alternative; otherwise, accept the null and reject the alternative. This comparison creates one confusion: people often treat p -value as equivalent to the Type I error rate (Harvey, 2017). The confusion would temper many to think that a p -value gives the probability that the null hypothesis is true or that a result is driven by chance alone.⁵ This is the most crucial mistake in the use of p -value.

A better understanding of p -value enables us to say more about the nature and consequence of p -hacking. Consider the following practices of p -hacking from a researcher, for example, running many experiment sessions and cherry-pick the most significant session to report, collecting multiple (related) outcome variables, and choose the one that gives the more significant results. Finally, the researcher ends up with a p -value of 0.01. This low p -value certainly does *not* mean that there is only 1% percent probability that the null hypothesis is true, or that if other researchers repeat the same experiment for a large number of times, they will obtain significant results 99% of the time. Quite differently, the p -value of 0.01 merely means that, given the null being true, there is only a 1% probability that we observe the result or more extreme results. However, in this example, the extreme result obtained by the researcher is very likely due to p -hacking: because the researcher has sampled the data long enough and hard enough, he or she is almost bound to find some extreme observations.⁶ Therefore, in the end, p -hacking will lead to higher false-positive rates.

⁵ In mathematical terms, the definition of p -value can be written as $P(D|H_0)$, which is the probability of observing an effect (or more extreme) D given the null is true. However, p -value is often misinterpreted as $P(H_0|D)$, which is the probability of the null being true given an observed (or more extreme) effect D .

⁶ There are also other reasons why p -value can be low. For example, it could be due to the poor ability of the hypothesis to explain the data, or a mis-specified test.

2.2. Literature on *p*-hacking

The literature generally defines *p*-hacking as the process of exploiting undisclosed flexibility in data collection and analyses to cherry-pick the most significant results, while filing away the rest (for examples of *p*-hacking, see, e.g., Simmons, Nelson, and Simonsohn, 2011; Head et al., 2015; Harvey, 2017).⁷ The concerns over the negative effect of *p*-hacking can be traced back to Sterling (1959), who points out an unconscious type of *p*-hacking. When journals only publish papers with significant results (i.e., the publication bias), those experiments that do not work will not be observable to other researchers, who may independently repeat such experiments until a significant finding occurs by chance. The consequence is a higher false-positive error than expected.

While *p*-hacking has been recognized for many years, the concern that it can be a widespread problem has only recently aggravated. Two main reasons are discussed in the literature. First, there is increasing publication bias against negative findings from academic journals over the recent decades. Fanelli (2010) documents that in social science (such as Economics and business, psychology), it is more difficult to see published results that do not support the main hypothesis developed in the paper than in other “hard” science (such as Space Science and Physics). This publication bias incentivizes researchers to only report papers with significant findings or research specifications that give the most significant results in a paper. Fanelli (2012) and Head et al. (2015) point out that negative results are disappearing, and significant ones are more likely to be published in many fields. Second, the competition for top journal space may have intensified.

⁷ In addition to *p*-hacking, the practice of trying many tests and analyses until none significant results become significant has received other names, for example, data dredging or data fishing (Selvin and Stuart, 1966), data snooping (White, 2000).

Researchers and institutions are increasingly evaluated based on their top journal publications, and the citations these papers generate (Young, Loannidis, and Al-Ubaydli, 2008; Statzner and Resh 2010), which pressures researchers to engage in unethical behavior to increase the chance of publication, including *p*-hacking. Journals' bias against negative findings, combined with the increasing competition for top journal space, likely contribute to researchers' stronger incentive to conduct *p*-hacking.

The growing concerns over the widespread of *p*-hacking have accelerated research on it in many areas. As we have outlined in the introduction, *p*-hacking induces kink around the significance threshold (0.05) in the cross-sectional distribution of *p*-values. Masicampo and Lalande (2012) first apply the distribution of *p*-values to examine *p*-hacking in experimental psychology. Simonsohn, Nelson, and Simmons (2014) term this method as the use of “*p*-curve” and further promote its usefulness to detect *p*-hacking, which is similar to the use of distribution of earnings numbers to detect earnings management in accounting research (see, e.g., Burgstahler and Dichev, 1997). This method subsequently stimulates *p*-hacking research in many fields. Head et al. (2015) employ this method to study the extent of *p*-hacking in 14 disciplines and find that it is widespread. Krawczyk (2015) documents too many “just significant findings” in experimental psychology from more than 5,000 papers. Vermeulen et al. (2015) take this method to communication science and document similar evidence. Moreover, they also find that close to 9% of *p*-values in their sample are misreported, among which 75% are too low. Brodeur et al. (2016) apply a similar methodology to articles published on top economics journals and document strong evidence of *p*-hacking. They also explore how the extent of *p*-hacking

varies across several author and article characteristics. Basu and Park (2014) also examine the distribution of p -value in accounting studies, but they focus on quantifying Type I and II error rates.⁸

The paper closest to ours is Khan and Tronnes (2018). They specifically focus on experimental audit research and find a kink in the distribution of p -values. Our paper extends theirs in the following ways. First, we include all topics in experimental accounting research and focus only on articles that appeared in top accounting journals. It may be more unsettling that p -hacking is widespread in even the most prestigious journals. Second, Khan and Tronnes (2018) exclude vague p -values (e.g., $p < 0.05$) in their sample, which may inflate the relative frequency of 0.05 and extent of p -hacking if many such vague p -values have actual values close to 0.04 or 0.03. Our paper addresses this issue by empirically estimating the distribution of vague p -values using their underlying test statistics and degree of freedom. Third, as mentioned previously, we conduct rich cross-sectional analyses that help to understand the determinants of p -hacking, and especially how p -hacking is related to future citations. Lastly, we also briefly examine evidence of p -hacking in archival papers. While archival research does not display a discontinuity around 0.05 in the distribution of p -values, a caveat to interpreting this result is that archival studies generally rely on large sample of data, and thus have higher statistical power (See Appendix A for more discussions).

⁸ In their paper, the distribution of p -value does not exhibit an over-abundance of p -value right below significance threshold such as 0.05, relative to other smaller p -value such as 0.04 and 0.03. They show that the frequency of p -values right below two-decimal p -values is higher than that right above, e.g., $[0.015, 0.02)$ has higher frequency than $[0.02, 0.025)$. Moreover, this pattern is stronger for testing variables than for control variables. We argue that this finding can also be driving by publication bias, which makes papers with smaller p -values for test variables more likely to be published. We find that 0.05 appears more frequently than other smaller p -values such as 0.03 despite publication bias favoring the latter, thus our finding offers stronger evidence of p -hacking, or p -value management in their paper.

Furthermore, our article extends the above works by making a serious attempt to provide statistical solutions that may help guard against p -hacking. We demonstrate the usefulness of procedures to adjust for multiple comparisons when the chance of false positives is high, and also introduce the Bayesian approach as a potential alternative method. However, we also emphasize that transparent disclosure from researchers is necessary to facilitate the implementation of the strategies we have proposed.

In light of the widely-documented evidence of p -hacking from many disciplines, readers of our paper may question what we can gain by testing a null hypothesis (i.e., no p -hacking) that is almost certain to be rejected. Yet, just as researchers strive to produce scientific and economic evidence to inform accounting standard setters and financial regulators (Leuz, 2018), the modest goal of this paper is to contribute to an *evidence*-based discussion about the existence, incentive, and potential solutions, of p -hacking in accounting research.

3. Our sample and empirical strategy

3.1. Sample construction

Our aim is to include in our sample as many experimental accounting studies from influential accounting journals as possible. We thus use a computer program to search among all articles on three top publishers of experimental studies – The Accounting Review (TAR), Journal of Accounting Research (JAR), and Journal of Accounting and Economics (JAE) – from 1990 to 2016.⁹ We record an article if its abstract contains the keyword “experiment” but not

⁹ Our sample period starts from 1990 because that’s when Dallas Ranking for accounting divisions across the world starts to be available, which we will need to conduct subsample analysis. While the focus on only three top journals may seem narrow, it actually biases against us finding evidence of p -hacking. Arguably, top journals should maintain the highest standards for publication, including rigorousness in data analyses and methodology, disclosure requirement from authors. These standards should discourage p -hacking.

“(quasi-) natural experiment”.¹⁰ A total of 317 papers meet this criteria and are download for further data collection. We exclude articles that do not mention any p -values in their text (30 articles), and articles that are retracted due to Dr. Hunton’s fraud (6 articles). In the end, we have 281 articles in our final sample, with 193 from TAR, 84 from JAR, and 4 from JAE, accounting for 14.9%, 10.5%, and 0.3% percent of the total number of articles published in the three journals during our sample period, respectively.¹¹ In our main sample, we also exclude papers that only report vague p -values, such as $p < 0.05$ (41 articles). The main sample thus contains 240 articles.

To analyze what affects the extent of p -hacking, we further collect a series of author and article characteristics that may be related to researchers’ marginal benefit and cost of p -hacking. Researchers with greater career concerns are likely to have a stronger incentive to conduct p -hacking (Fanelli, 2012). We thus measure researchers’ career concerns by calculating the proportion of tenured authors for an article (*Tenured*), and by counting the average number of years since a researcher earns PhD at the time of the article’s publication (*Experience*). Head et al. (2012) state that prestigious schools impose higher tenure hurdles, and thus greater publication pressure on researchers. To measure academic ranking, we use the mean ranking of authors’ affiliated institutions (*School Ranking*), as well as the mean ranking of the institutions where each author earns PhD degree (*PhD School Ranking*). We distinguish articles with no female

¹⁰ Alternatively, we can also rely on JEL Classification code to distinguish experimental studies from archival studies, but this information is not provided on the website of the three journals.

¹¹ When counting the total number of articles, we exclude any articles that are generated by the comments and/or discussions of existing papers. In addition, TAR frequently publishes book reviews in our earlier sample period, these are also excluded from the count. The number of experimental articles scraped by our scripted matches the number in Bloomfield, Nelson, and Soltes (2016), who identified 139 such articles on “Top 3” accounting journals during 2003 and 2013, or around 13 articles per year. Our script returns roughly 12 (317/27) articles per year.

author(s) from these with at least one (*Female*), to consider the gender difference of misbehavior (see, e.g., Friesen and Gangadharan, 2012). To account for the coordination costs among researchers, we also count the number of authors for each article (*N_Authors*). We retrieve the total number of participants in an article's experiment(s) as various *p*-hacking "techniques" involve sample manipulation (*N_Participants*). Because researchers may revise their hypotheses *ex post* as directional in order to report lower one-tailed *p*-value, we construct a dummy variable equals to one if all *p*-values in an article are two-tailed, and zero otherwise (*Two-tailed*). Lastly, we collect the number of citations for all articles from both Google Scholar (*Google_Citation*) and Web of Science (*Web_Citation*) to examine the relation between *p*-hacking and citations.

The variables described above come from various sources. To obtain tenure, graduation, and school ranking information for each researcher, we search for "author name" + "affiliated school" on google. Most of these searches deliver an author's latest CV, from which we can determine the year of tenure and graduation. We then collect information on school ranking from The UTD Top 100 Business School Research Rankings. Since our focus is experimental accounting studies, we thus choose to rank the schools by the number of accounting top journal publications in a given year.¹² Note that for each article-author-school combination, the ranking for that school is two years before the publication of that article. We introduce this two-year lag to capture the ranking status while a paper is written rather than when it is published. If a school is not on the ranking list, we replace its ranking with the lowest ranking number on the

¹²We choose three top journals in the accounting field: TAR, JAR, and Journal of Accounting Economics (JAE).

ranking list in a specific year. Finally, if we cannot find the CV for an author, we rely on other credible sources to obtain as much information as possible.¹³ We manually check each article to obtain the number of participants, and in articles where multiple experiments are conducted, we sum the number of subjects from the experiments. Lastly, we read all articles to determine whether all p -values in a paper are based on two-tailed tests.

[Insert Table 2.1 about here]

Table 2.1 reports the number of experimental articles in our sample for each year. The figures suggest considerable uncertainty in the number of published experimental studies, especially before the early 2000s. After 2010, the annual publications of experimental studies have always been above 10. This finding is consistent with Birnberg (2011), who points out that experimental accounting research has been growing from the 1990s, in terms of the number of papers and diversity of topics. Table 2.2 offers a glimpse of the characteristics of experimental studies in our sample. An average article in our sample has 2.2 authors, 117 subjects in the experiments, and mentions 18 p -values in its text. 50.0% of the articles contain at least one female author, 34.3% of the articles report only two-tailed p -values, and 28.7% are sole works. In addition, an average paper is published 9 years after its authors obtain their PhD degree. When we break the total sample by journals, we notice that about two-thirds of the articles in our sample come from TAR, which publishes more papers than JAR does.¹⁴ We also find that the variables do not exhibit significant differences between the two journals.

¹³ Those sources include school website, LinkedIn, and BYU Accounting author search, etc.

¹⁴ Differences in number of observations among variables are due to missing values.

[Insert Table 2.2 about here]

3.2. Methodology

After transforming all downloaded pdf files into txt format, we use a Python script to search for numbers between 0 and 1 immediately preceded by “p =,” “p-value =,” “p_{two-tailed} =,” or “p_{one-tailed} =” regardless of spacing or capitalization. Unlike prior studies such as Head et al. (2015) and Khan and Tronnes (2018), we also collect *p*-values after inequality operators, such as “p <” or “p ≤.” Importantly, we omit any numbers that indicate significance thresholds instead of real *p*-values in an article; for example, numbers appeared in sentences such as “*, **, *** Indicates p < 0.10, p < 0.05, and p < 0.01, respectively” are discarded. Our script cannot retrieve *p*-values within tables, because the formatting of tables in pdf is messy and varies from article to article, rendering the search impractical. However, because researchers are more likely to mention the significance level of their key variables than of other covariates. By focusing only on *p*-values mentioned in the text, we ensure that we collect the right data to test our hypothesis: that researchers have the strongest incentive to overstate the significance level of their *key* results.

Our initial search yields 6,144 *p*-values. However, not all of these *p*-values are informative about the extent of *p*-hacking. We are particularly interested in any irregularities in the cross-sectional distribution of the *p*-values at points such 0.05, and 0.10, and thus exclude all *p*-values larger than or equal to 0.155, which leaves 4,766 *p*-values in our sample. Because we cannot uncover the exact *p*-values for all *p*-values reported after inequality operators (e.g., < or ≤), we focus on exact *p*-values in our main sample, which includes 2,404 (50.4%)

p -values from the 4,766 p -values.¹⁵ We then pool these p -values across articles and time to generate a distribution of p -value.

It is worthwhile to emphasize that our text-scraping methodology has two advantages. First, it does not require us to have further knowledge about the articles in our sample, such as the topic of interest, hypotheses tested, and designs of experiments, allowing us to examine all experimental articles that study any accounting topics. Second, since it is a substitute for hand collection, it minimizes biases from data collectors such as research assistants, either consciously or unconsciously, to collect data that conforms to this paper's prediction.¹⁶

4. Empirical results

We present two sets of results to determine whether p -hacking to turn insignificant results into significant exist or not. First, we provide graphical evidence in the form of the histogram on the cross-sectional distributions of the p -values. Second, we fit the distribution of p -values to an exponential distribution to quantify the extent of p -hacking.

4.1. The cross-sectional distribution of p -values

We first plot a histogram using all p -values in our sample by choosing a bin width of 0.001. The result is shown in Panel A of Figure 2.1. One bin stands

¹⁵ Many p -values reported after inequality operators are placed in footnotes for robustness check or additional tests, where no information about the test statistics or degree of freedom are provided. However, if we simply pool these p -values with our main sample, the main results will only become stronger, because “<” and “≤” are commonly followed by threshold value 0.05. We present analyses to deal with these p -values in Section 4.

¹⁶ Nonetheless, one may argue that our script can still produce some noises. For example, it may capture some irrelevant numbers, but ignore p -values for an article's key results. To alleviate such concerns, we randomly select 30 articles from our sample, and for each article, we identify all the hypotheses tested. We then check if the p -values extracted by the script from an article are the significance level for hypotheses testing in that article. Of the 260 p -values from the 30 articles, 251 (96.12%) are the significance level of the main hypotheses tested, the rest 9 p -values belong to 6 articles. Among the 6 articles that do not have p -values for the main hypotheses, 2 do not discuss the p -values in their text, the other 4 do not report exact p -values (i.e., $p < 0.05$). These results suggest that our script captures relevant p -values generated from hypotheses testing in experimental studies, and is unlikely to generate systematic biases in our sample.

out from the distribution: there is an obvious bump in the frequency of interval right below the significance threshold, i.e., $(0.049, 0.05]$, which we highlight using a bold bar. The frequency of this interval (122) is even higher than the frequency for the interval $(0.039, 0.04]$ and $(0.029, 0.03]$ (88 and 111, respectively), as if less significant p -values are preferred for publication. Interestingly, we observe that the frequency of p -values right below 0.06 is lower than that below 0.07. The two discontinuities around the conventional significance threshold cannot be explained by the publication bias but are in line with the presence of p -hacking to make marginally significant results significant.

As noted in Table 2.1, the average article reports 18 p -values in our sample. However, there is substantial variation among articles in terms of the number of p -values: the highest number is 64, while the lowest is 1. Furthermore, the number of p -values reported (or the number of tests conducted) may be related to the overall significance level or robustness of the results. For example, a researcher may conduct and report more tests when these tests have p -values not too far below the significance threshold. Therefore, it is possible that certain articles that report many p -values are driving the main results. To alleviate this concern, in Panel B, we weight each p -value by the inverse of the number of p -values reported in an article, thus giving lower weight to those articles with many p -values. The discontinuity around threshold becomes more striking after weighting: the weighted frequency of p -values right below 0.05 is even higher than that of p -values right below 0.02. Meanwhile, there is also a non-monotonic pattern in the distribution around 0.10, as if it is more desirable than 0.09 and 0.08. Overall, the weighted distribution of p -values shows a greater extent of p -

hacking. This result seems to suggest that articles report *fewer* p -values tend to report a higher fraction of “just” significant results.

[Insert Figure 2.1 about here]

The Figure also shows that articles in our sample prefer reporting p -values with two decimal places, which accounts for 40% of the total number of p -values, despite there are only 16 numbers with two decimal places below or equal to 0.15.¹⁷ This creates the indented histograms we observe in Figure 2.1. In Panel A of Figure 2.1, we use only p -values reported with two decimal places to plot another histogram. The resulting histogram shows that the unusually high frequency of p -values in the interval (0.0475, 0.05] in Figure 2.1 is driven entirely by p -value = 0.05, which was found 122 times in our sample. The same also holds for p -value = 0.04 and 0.03.

In Panel B, we round all p -values to two decimal places, count the frequency of the 16 p -values (0.00 to 0.15), and plot another p -curve. A similar pattern emerges, though the difference in frequency between 0.05 and 0.04 is less salient. This is unsurprising because rounding effectively sets the bin width to 0.01, thus muting the acuteness of the histogram to detect *marginal incentive* of p -hacking at 0.05.

[Insert Figure 2.2 about here]

In sum, the results presented in this subsection provide evidence on the existence of p -hacking among experimental accounting studies to inflate the significant results. Since 0.05 is the maximum p -value for an effect to be considered as statistically significant in experimental accounting research, the

¹⁷ Among the 2,404 p -values, the decimal places of a p -value range from one to five, the frequencies of the five different lengths are 2, 1013, 1257, 130, and 2, respectively.

relative abundance of 0.05, together with the relative scarcity of 0.06, provides strong evidence that there are too many results with “just significant” p -values. Such irregularity in the distribution of p -values presents a challenge for any other potential explanation.

4.2. *Fitting the p -curve to an exponential distribution*

In order to gauge the extent of p -hacking among our sample articles, in this subsection, we follow Masicampo and Lalande (2012), and Harvey, Liu, and Zhu (2016) to fit the histogram of p -values to an exponential distribution. The choice of a rapidly decreasing exponential curve aims to approximate the effects of strong publication bias, which is substantiated by the highly right-skewed distribution of p -value in the previous section. In doing so, we are able to roughly estimate the *expected* frequency of p -values without p -hacking, and quantify its impact on the magnitude of inflation.

Because p -values with two decimal places are more frequently reported, in order to have a smooth histogram for fitting, we use either p -values reported with two decimal places or all p -values rounded to two-decimals. The results are presented in Figure 2.3. In Panel A, the sample only includes p -values with two decimal places. An exponential curve fits the histogram quite well except for two p -values: 0.05 and 0.06, the former lies well above, and the latter well below the curve. Another less noticeable discontinuity appears in p -value 0.10. The frequency of observing 0.11 is merely half the frequency of observing 0.10. We posit that the reason for this discontinuity is that 0.10 is also a target to beat: p -values $\in (0.05, 0.10]$ are deemed marginally significant, while p -values > 0.10 are insignificant. The importance of this target, however, is subordinate to 0.05.

We observe a similar pattern at Panel B, where all p -values are rounded to two decimal places.

[Insert Figure 2.3 about here]

We further tabulate the fitting results in Table 2.3. For each, p -value from 0.01 to 0.15 (i.e., 0.01, 0.02, 0.03... 0.15), we report its observed frequency and expected frequency implied by the fitting from Figure 2.3, and calculate how much the former deviates from the latter. In Panel A of Table 2.3, where only p -values reported with two decimal places are used, we observe that the residual at 0.05 is the highest. However, residuals are also sizable at p -values such as 0.03, 0.04, and 0.06, all of which are negative. This suggests that compared with p -value = 0.05, p -value = 0.03 and p -value = 0.04 should have higher frequencies than what we observe here. We observe similar results using all p -values rounded to two decimal places in Panel B.

[Insert Table 2.3 about here]

We conduct two post hoc pairwise comparisons to determine if the residuals from fitting for 0.05 and 0.06 are significantly different from the mean residuals for the rest p -values in the two histograms. The comparison is based on Dunnett's Test (pairwise versus reference group).¹⁸ Specifically, we treat

¹⁸ Dunnett's test is a many-to-one comparison procedure that compares the mean of every treatment group procedure against the mean of a single control group. Each comparison results in a test statistics that is calculated as below:

$$\frac{\bar{X}_i - \bar{X}_c}{\sqrt{2 \times MS_{error} \left(\frac{1}{n_i} + \frac{1}{n_c} \right)}}$$

where $\bar{X}$ refers to the mean of each group, i indexes each treatment group, and c the control group. MS_{error} is the mean squared error term for the whole sample. Lastly, n_i and n_c refer to the number of observations for treatment group i and control group. The degrees of freedom for this test, $N - k$, equal to the total number of observations in all groups minus the number of groups (including the control). The p -values from Dunnett's test are determined by looking up the test statistics in a table of critical values that already adjusted for family-wise error rate. Compared with other post hoc pairwise comparison methods (e.g., Bonferroni, Tukey's and Scheffé's methods), Dunnett's test is more powerful as it conducts fewer comparisons (if there are k groups, Dunnett's test makes $k-1$ comparison while others make $2 \times k \times (k-1)$ comparisons), and thus does not result in unnecessarily low critical p -values. See Dunnett (1955) for more details..

residuals for p -values except 0.05 and 0.06 from exponential fitting as the reference group, and then test if the residuals for 0.05 and 0.06 is significantly different from the mean residual of this reference group. The significance level obtained is adjusted by times of comparison. The results are reported in Table 2.4. Panel A and B of Table 2.4 are based on the fitting in Panel A and B in Figure 2.3, respectively. Despite the small sample size (16 residuals), we find that in both Panels, the residuals at 0.05 are always much larger than the reference group. Residual at 0.06 is much smaller than the reference group in A, though it is statistically insignificant; in B, residual at 0.06 is indistinguishable from the reference group.

For economic magnitude, we calculate the ratio of residual frequency to the observed frequency at $p = 0.05$, which we label residual ratio. The fitting shows that the residual ratio is between 24.88% and 26.72%, meaning that the expected frequency of 0.05 should be 24.88% to 26.72% lower than the observed frequency; for 0.06, the expected frequency should be 13.16% to 38.46% higher than the observed. To put this magnitude in perspective, we compare the residual ratio in our study to that shown by Basu and Park (2016), in which the sample mainly consists of p -values from archival-empirical research. Their analyses show that the expected frequency of $p = 0.05$ should be 12.73% lower than the observed frequency (Table 2.4).

As we stressed in section 4.1, histogram based on rounded p -values is not sharp enough to detect the strongest incentive to engage in p -hacking at 0.05. Thus the results in Panel B of Table 2.4 can be viewed as a lower bound for our results. Taken together, the results in this subsection show that there is a large proportion of just-significant p -values that cannot be explained by a decreasing

selection process alone, suggesting that p -hacking has a material impact on the distribution of p -values.

[Insert Table 2.4 about here]

4.3. P -values reported after inequality operator

In our initial sample, 2,362 p -values scraped from research articles are reported after the inequality operator (e.g., $p < 0.05$). We term these numbers “vague” p -values. The distribution of these p -values, reported in Panel A of Figure 2.4, shows that the majority of numbers reported after “ $<$ ” are significance thresholds such as 0.001, 0.01, 0.05, and 0.10. The frequency of these four numbers accounts for 69.98% percent of the 2,362 p -values. We are particularly interested in p -values reported as $p < 0.05$ and $p < 0.10$, since redistributing the actual values behind the two numbers may invalidate our main conclusion. For example, p -values reported as $p < 0.05$ can be any number between 0.00 and 0.05, excluding these p -values may cause an under-representation of p -values such as 0.03 and 0.04 in our main sample, thus causing the discontinuity around 0.05 we observed in the distribution of p -values.

[Insert Figure 2.4 about here]

To alleviate this concern, we attempt to calculate the actual values for p -values reported as $p < 0.05$ and $p < 0.10$ (396 such cases), by using the test statistics and degree of freedom information provided in research articles. However, we notice that in many cases, such information is not provided.¹⁹ In the end, we are able to reveal the exact p -values for 104 out of the 396 cases. While it is infeasible to reveal all of the exact values for such p -values, the distribution

¹⁹There are several reasons for not reporting such information. For example, many p -values are reported in footnotes for additional tests or robustness checks; multiple p -values are reported together as “all p -value < 0.05 ”. Papers published in the beginning of our sample period are more likely to omit such information.

of the 104 p -values may provide some insights on when researchers prefer to report “ $p <$ ” instead of “ $p =$.”

As shown in Panel B of Figure 2.4, the distribution of the 104 exact p -values indicates that, for a p -value smaller than 0.05, researchers are more likely to report it as “ $p < 0.05$ ” when it is closer to 0.05 than to 0.00. This distribution is somewhat expected since an investigator has a strong incentive to report a low p -value (e.g., 0.02) *exactly* to demonstrate statistical significance. In contrast, she may find it in her interest to hide the exact value of a p -value and its test statistics when it is very close to 0.05 (e.g., 0.0486). This selective reporting incentive suggests that for the rest p -values that we have no information to calculate exact values, the distribution of their true values is likely to be even more concentrated near 0.05 than the distribution Figure 2.4. We then impose the same distribution in Figure 2.4 to the rest 292 p -values, and stack the estimated distribution of the 396 vague p -values (i.e., $p < 0.05$ and $p < 0.10$) with the distribution of exact p -values used in our main sample. The resulting distribution is reported in Panel A of Figure 2.5, which shows stronger evidence of p -hacking than using the sample of exact p -values alone.

While our results are quite robust to including vague p -values reported as $p < 0.05$ and 0.10, there are also some p -values reported as $p < 0.04$ ($p < 0.03$), which may have true values close to 0.04 (0.03). Thus, omitting them may again inflate the extent of p -hacking. We repeat the procedure to manually compute the exact values for $p < 0.04$ and 0.03. The data shows that exact p -values behind $p < 0.04$ (0.03) are approximately uniformly distributed between 0.03 and 0.04 (0.02 and 0.03). The uniform distribution is then applied to all other vague p -values except $p < 0.001$, $p < 0.01$, $p < 0.05$, and $p < 0.10$. For instance, the exact

value of a $p < 0.09$ can be any number between 0.08 and 0.09. We further include these redistributed vague p -values (i.e., $p < x$, where $x = 0.02, 0.03 \dots 0.15$, excluding 0.05 and 0.10) to the sample used in Panel A. By doing so, we utilize all of the p -values (4,766) scraped from experimental accounting articles to plot a new distribution. The result is reported in Panel B of Figure 2.5. After including all vague p -values, we still observe strong evidence of p -hacking. For example, if we fit the histogram on the right side of Panel B to an exponential distribution, the expected frequency at 0.05 will be 24% less than what we observe. Overall, the main message from Figure 2.5 is that, after taking into account the impact of vague p -values on the distribution of p -values, our main results stand valid.

[Insert Figure 2.5 about here]

5. What explain the extent of p -hacking?

In this section, we study the following question: what determines researchers' decisions to conduct p -hacking? While p -hacking is far from outright academic fraud, such as data fabrication, Bloomfield, Rennekamp, and Steenhoven (2018) suggest that the “fraud triangle,” proposed by Cressey (1953), is useful to understand any self-interested misbehavior, including p -hacking. We thus adopt the framework implied by the triangle in trying to understand the determinants of p -hacking. Specifically, the triangle proposes three factors that are present in situations of misbehavior: pressure (or motive), opportunity, and rationalization. That is, misbehavior is more likely when it can help address the pressure people face, when there are opportunities of misbehaving, and when it can be rationalized as acceptable or appropriate. We discuss how each of the factors can fit into the context of p -hacking, and identify proxies for these causal and enabling factors of p -hacking.

5.1. Pressure

To survive in academia, researchers need to publish papers, especially in top journals. To the extent that papers with more significant findings have a greater chance of publication, researchers will have the motive to overstate statistical significance through p -hacking. Therefore, career concerns could pressure researchers to engage in p -hacking. We rely on two measures to quantify publication pressure: researchers' tenure status and years since obtaining PhD degree. The prediction is that p -hacking should be more prevalent among researchers that have not tenured yet, and those have fewer years of research experience.²⁰

Fanelli (2012) noted that research institutions and other organizations are increasingly evaluated by the number of top journal publications produced by their researchers. We examine whether institutions' rankings — as measured by their numbers of top journal publications — are related to the extent of p -hacking. It is possible that high ranking institutions have higher tenure requirements, which imposes additional pressure on researchers to publish more papers in top journals. Consequently, researchers in high ranking institutions engage in a greater extent of p -hacking.²¹ Institution ranking is measured by the number of publications in three accounting top journals obtained from “The UTD Top 100 Business School Research Rankings.” For each article published in year t , we use the year $t-2$ *School Ranking* as defined in Section 3.1, and separate articles into

²⁰ We determine an author to be tenured only if he or she is tenured two years before the publication of the article of interest.

²¹ Alternatively, researchers affiliated with high ranking schools might be able to come up with research questions that have larger true effects, or they may simply be more rigorously trained, thus reducing the need and incentive to engage in p -hacking.

high and low institution ranking based on its sample median. We also repeat the above procedure to split our sample based on *PhD School Ranking*.

5.2. Opportunity

Researchers' discretions over data collection and analyses can be viewed as opportunities for *p*-hacking, the greater the discretions, the more tests researchers can try to get significant results. Researchers that work together face fewer opportunities, because *p*-hacking may require teamwork and coordination among co-authors. For example, co-authors may need to concur with each other on how to identify and treat outliers, whether to collect more data, and whether to combine or transfer specific measures. If one (undisclosed) decision is controversial but yields the most significant results, the author who is in charge of data analysis might need to convince other authors why they should make that choice. As such, researchers who work alone have more opportunities.

In addition, researchers may have more opportunities when they have access to a greater pool of potential participants for their experiments, which gives researchers flexibility in conducting pilot studies, collecting additional data. This flexibility certainly has its merits, but it may also be a double-edged sword. For example, researchers can easily collect more data until they document significant findings. When referees ask for more tests or experimental sessions, researchers may cater to their requests by collecting more data, and mining additional significant results to increase the chance of publication. Easier access to a large number of participants also facilitates researchers to draw random experimental samples repeatedly until significant findings show up. Hence, a greater pool of participants may present more *p*-hacking opportunities.

5.3. Rationalization

Necker (2014) surveys scientific misbehavior among economists, and documents that men are less likely to regard misbehavior as unjustifiable. The practices considered as misbehavior in Necker (2014) include fabricating data, excluding part of the data without reporting, stopping statistical analysis until desired results show up, etc. Others find that under laboratory experiments, men are also more likely to cheat than women (e.g., Friesen and Gangadharan, 2012). We thus posit that male researchers are more likely to justify the need for p -hacking than female researchers do, and therefore exhibit more aggressive p -hacking. To test this conjecture, we separate articles into two subsamples by whether the author(s) of an article contains at least one female researcher.

After describing the proxies for the three factors of the “fraud triangle,” we test whether these proxies can predict the prevalence of p -hacking in the cross-section. Specifically, for each proxy, we classify the sample articles into two subgroups by their values of this proxy (*high* versus *low* group for continuous variables, and *yes* or *no* for dummy variables), and determine the p -value distribution of which subgroup show a sharper discontinuity at $p = 0.05$. To quantify the discontinuity, we again fit the distribution of p -values against an exponential distribution, and report the residual ratio at 0.05 (the difference between observed frequency and expected frequency scaled by the former). Table 2.5 reports the results. The findings are generally consistent with the predictions derived from the fraud triangle framework. For the pressure factor, we find the residual at 0.05 is much larger for articles without tenured authors, for articles written by authors who obtained PhD degree for a short time before their papers are published. For instance, the residual ratio at $p = 0.05$ – a measure

that quantifies how many of the $p = 0.05$ is due to p -hacking – is 38% among articles written by untenured authors, while the percentage is 9% for articles with tenured authors. Somewhat surprisingly, we observe stronger evidence of p -hacking in the subsample of articles written by authors affiliated with (or graduated as PhD from) higher ranking institutions. While it is hard to pin down the exact reason, a plausible explanation is that more pressure to publish in higher-ranked institutions enhances the incentive to conduct p -hacking. Alternatively, one could also argue that researchers from more prestigious institutions are less likely to be challenged by reviewers and other readers.

[Insert Table 2.5 about Here]

Turning to the opportunity factor, we document stronger evidence of p -hacking for articles that are sole-authored and articles with a higher number of participants in their experiments. Specifically, the residual ratio for the subsample of sole-authored articles is much higher than co-authored articles (36% versus 14%), consistent with the conjecture that working alone presents more p -hacking opportunities. Comparing articles with above-median number of experimental participants to the rest, we find that the former group exhibits stronger evidence of p -hacking (residual ratio 25% versus 14%). Lastly, we find that the residual ratio for articles with only male author(s) is almost twice the magnitude of the residual among articles with at least one female author, consistent with females being less likely to justify p -hacking than males.

Taken together, we identify five factors that seem to be able to predict the extent of p -hacking, they are authors' tenure, experience, authors' affiliated school ranking, number of authors, number of participants in the experiments of articles, and author gender. These cross-sectional variations in the extent of p -

hacking are consistent with the predictions from the fraud triangle framework. However, our results in this section are purely based on univariate analysis, and the results obtained can be relational instead of causal. It is possible that those factors are correlated with each other or other omitted variables. Our results thus need to be interpreted with these limitations in mind. Nonetheless, while some alternative stories may explain part of our results, the collection of the cross-sectional evidence is most consistent with the existence of p -hacking. To sum up, while there are many subsamples of articles that exhibit strong evidence of p -hacking, the flip side of this finding is that there are also many subsamples of articles show weak or even no evidence of p -hacking. This interpretation suggests that good practice from many researchers could dominate questionable research behavior, including p -hacking.

6. Additional tests

6.1. P-hacking and future citations

If p -hacking increases the chance of publication, an interesting question is whether it is related to the citations an article receives in the future. *A priori*, it is unclear how the two will be correlated. On the one hand, to the extent that editors prefer articles with results that are more statistically significant because such articles attract more citations in the future, we may expect p -hacking to increase the number of citations by inflating the level of statistical significance. On the other hand, if p -hacking increases the chance of false-positive findings, these findings are less likely to hold in the future (Open Science Collaboration, 2015). Subsequent studies that build on these findings thus have a lower chance of bearing fruit and publication. Hence, findings coming from p -hacking or data-mining may generate fewer citations than other robust findings.

We use the citation counts from Google Scholar to test how the extent of p -hacking is related to the number of future citations. To control for the truncation bias that articles published earlier have a longer time to accumulate citations, in each year, we split all articles published during that year into two groups based on that year's median number of citations. The final *high (low) citations* group includes all the *high (low) citations* groups from every year. We then check whether more cited articles exhibit stronger or weaker evidence of p -hacking by plotting the distribution of p -values.²² The histograms in Figure 2.6 deliver two findings. First, not surprisingly, more cited articles have more significant results, as indicated by the higher frequency of p -values near zero in the *high citations* group than that in *low citations* group. Second, more cited articles also have a relatively higher proportion of “just” significant results: the ratio of the residual frequency to observed frequency in Panel B (24%) is almost twice the ratio in Panel A (13%). Similar results are obtained if we use the citation counts from Web of Science instead of Google Scholar (Panel I in Table 2.5). Taken together, these results are consistent with the argument from prior works that more significant results attract more readers and citations (e.g., Fanelli, 2012), though some of the results may be due to p -hacking.

[Insert Figure 2.6 about here]

6.2. The extent of p -hacking for main results

If researchers use p -hacking to obtain the most significant results, then they should have the strongest incentive to do so for their *key* results in the article.

²² Ideally, to test how p -hacking may affect future citations, one needs to construct an article-level measure of the extent of p -hacking, and use this measure to forecast future citations. However, due to the limited number of p -values a typical article reports (10 in our sample), article-level measure is hard to construct. That said, the method we adopt is still legitimate, because future citations are usually generated years after the writing of an article, it is implausible that the causal relation between p -hacking and future citations runs from the latter to the former.

To test this conjecture, we slice the p -values in our sample as follows. For an article, we rank its p -values by the sequence they appeared in the text, and then group p -values appeared in the first half of the whole sequence as p -values for the key results. The reason is that the key results of an article are likely to be mentioned earlier rather than later.

Results in Figure 2.7 demonstrate that the kink in the p -value histogram at 0.05 in our full sample is almost entirely driven by the subsample of p -values mentioned in the earlier part of each article. Moreover, in Panel A, the frequency of p -value = 0.01 is almost twice the frequency of p -value = 0.02, which is even lower than the p -value = 0.05. The potential explanation is that researchers have a higher standard of significance level for their main results: many p -values equal to 0.02 are inflated to 0.01. In contrast, in Panel B, where the p -values are mentioned in the latter part of each article, we observe weaker evidence of kink at 0.05. The results from this exercise thus provide strong evidence of p -hacking.

[Insert Figure 2.7 about here]

6.3. JAR's updated data and code sharing policy

In December 2014, JAR implemented a data and code sharing policy that requires more disclosure from researchers. For example, the policy stipulates that at the time of initial submission to JAR, the authors' team needs to describe data collection and analyses in detail. For example, the authors need to describe which author(s) collect(s) the data and conduct(s) analyses, how and when the data used in the submitted paper are obtained, and contact information for editors to verify the integrity of proprietary data. Conditional on acceptance, the authors(s) need to provide detailed computer programs to facilitate replication. Before 2014, such

requirements are absent. The greater disclosure required from researchers likely reduces the opportunities of p -hacking and thus the discontinuity in the distribution of p -values at $p = 0.05$. To test this prediction, we extend our sample to include articles published in JAR from 2017 to July 2019. Because similar policy update did not occur in TAR during the same period, we also include articles in TAR from the same period as a contrasting group.

Importantly, to isolate the effect of the policy in 2014, we exclude five experimental studies published through the Registration-based Editorial Procedure (REP) in 2017, which is likely to influence the distribution of p -values in a way similar to the stringent data and code policy does.²³ As in previous analyses, we assume a typical paper takes two years from initial submission to online publication in JAR. Thus, the policy should take effect for papers published in 2016 and onwards. We choose articles published in JAR and TAR during the 8-year window period around the 2014 policy (i.e., 2012 to 2015 and 2016 to 2019), and plot the distribution of p -values for both journals in the two periods.

[Insert Figure 2.8 about here]

Figure 2.8 reports the results. The upper panel shows that for papers submitted before the policy update in 2014, the distributions of p -values in both journals display a discontinuity at $p = .05$, especially for articles published in TAR. In the lower panel, we observe that the discontinuity at $p = 0.05$ wanes in the distribution of p -values from papers submitted under the new data and code policy. In particular, the frequency of $p = .05$ is no longer higher than $p = .04$

²³ As we will introduce in Section 7, REP also limits researchers' discretion in data collection and analyses.

and .03 for JAR. The overall results thus suggest that the more stringent data and code disclosure policy adopted by JAR curbs p -hacking opportunities. Admittedly, due to the small number of experimental articles published in JAR from 2016 to 2019, we suggest that a longer sample period is needed to draw a more definitive conclusion.

7. Potential remedies to guard against p -hacking

While we have provided evidence of p -hacking among experimental accounting studies, in this section, we further discuss potential remedies that may help to guard against p -hacking. As mentioned in the introduction, p -hacking can hinder academic progress because it increases the chance of false-positive findings, which is largely driven by the problem of Multiple Hypotheses Testing (MHT) accompanied by insufficient disclosure on the research agenda. We first introduce three widely used MHT adjustment methods developed to limit the number of false-positive discoveries, with discussions on the implementation of the methods in experimental accounting research; we then illustrate the application of the methods using simulated experimental data. We also consider the Bayesian approach and provide some guidelines for improving research transparency.

7.1. Multiple hypotheses testing

Multiple Testing refers to situations where researchers test multiple null hypotheses simultaneously (List, Shaikh, and Xu, 2019), and decide which one to reject. It can be due to multiplicity in data, hypothesis, or analysis methods (Bender and Lange, 2001). For instance, data multiplicity can occur when researchers choose among several samples to test the same set of hypotheses; hypothesis multiplicity arises when researchers investigate the effect of a

treatment (or manipulation) on several outcome variables. Multiple Testing could be problematic because it increases Type I errors. This can be illustrated as follows. If a single hypothesis testing is conducted at a significance level α , the probability of not rejecting a true null hypothesis is by definition $(1-\alpha)$. Now consider conducting M independent tests, the probability of not rejecting all true null hypotheses is $(1 - \alpha)^M$, therefore, the probability of rejecting at least one true null hypothesis becomes $1 - (1 - \alpha)^M$, which is also called the *family-wise error rate (FWER)* or *experiment-wise error rate (EER)*. Formally, it is defined as:

$$FWER = Prob \left(\bigcup_{m=1}^M \{ \text{reject } H_{m0} \text{ when it is true} \} \right),$$

where m indexes the hypotheses tested, and H_{m0} is the null for hypothesis m . It is easy to see that $FWER$ is an increasing function of M . When multiple tests are conducted, it becomes necessary to control the FWER at a reasonably low level. The statistical literature has developed ways to control for FWER, and we describe some of the most widely used methods and their implementations.

The first method we consider is the Bonferroni method, which offers an intuitive procedure to control for FWER. Specifically, suppose we need an FWER lower than α , which is also the false positive rate for each test among the M tests. However, according to Bonferroni's inequality, we will have

$$FWER \leq \sum_{m=1}^M Prob(\text{reject } H_{i0} \text{ when it is true}) = \alpha \times M,$$

To guarantee a FWER lower than α , the Bonferroni method sets the false positive rate for each individual test as $\frac{\alpha}{M}$, then

$$FWER \leq \frac{\alpha}{M} \times M = \alpha.$$

Consequently, under the Bonferroni method, an investigator compares the p -values of all tests to a single significance threshold α/M , and rejects a null

hypothesis H_m if $p_m \leq \alpha/M$, where m indexes each test. By doing so, she controls for FWER at the level α .

The Bonferroni method has been used in experimental accounting studies since the 1980s, but its application is quite limited.²⁴ However, in situations where the chances of false positives are high, the Bonferroni method offers a simple adjustment to guards against results that are driven by chance. Specifically, if the key hypothesis in an experimental study is examined by means of multiple comparisons, then Bonferroni adjustment is necessary. For example, it is not uncommon for experimental studies to compare the mean of a control group with each mean of four treatment groups, without specifying clearly the *ex ante* priorities among the comparison. In this specific case, the Bonferroni-adjusted p -values for significance at the 5% level should be 1.25% ($5\% \div 4$).

Widely known for its simplicity, the Bonferroni method also lacks power when the number of tests is high because, for a large number of tests, the critical p -value will be extremely small. In addition, it treats each test as independent from each other. In practice, however, the test statistics from a family of tests are almost always correlated, and Bonferroni methods will over-correct the FWER. We introduce our second method to account for the interdependence of test statistics.

The second method we consider is the resampling-based multiple testing proposed by Westfall and Young (1993), which essentially derives the adjusted p -value estimated by permutation. This method is implemented as follows. Suppose there are M tests, each with an actual p -value of p_m , where $1 \leq m \leq M$. The

²⁴ See Margheim (1986), Peecher (1996), and Libby and Brown (2013) for examples of the application of Bonferroni method.

first step is to obtain a permutation sample by randomly re-allocating observations between control and treatment groups.; second, for the M tests, calculate their respective p -values $\{p_1^p, p_2^p, \dots, p_M^p\}$ using the permutation sample, and get the minimum value p_{min}^b , where superscript b denotes that these p -values come from the permutation data; third, for any test m , create a dummy variable, C_m^p , which equals to one if $p_{min}^p \leq p_m$ and zero otherwise (so we will have M counters for each permutation). Lastly, repeat the above steps for P times, and the adjusted p -value for an actual p -value is calculated as $\frac{\sum_{p=1}^P C_m^p}{P}$, which is the ratio of times when the minimum p -value is lower than or equal to the actual p -value over the total times of permutation. The adjusted p -values are compared against the 5% significance threshold. In our setting, we set P equals to 500.

As Westfall, Young, and Wright (1993) point out, the adjusted p -value measures how extreme a given p -value is relative to the distribution of the most extreme p -values from the permutation distribution of the statistic. Compared with the Bonferroni method, the biggest advantage of the bootstrapping adjusted p -values is that it takes into account the dependence structure among individual test statistics, thereby improving the power of the multiple testing procedure. This method is thus particularly useful when the number of tests in an experiment is high. In such cases, the Bonferroni method will give extremely small critical p -values. For example, an investigator might examine the effect of two treatments on five outcome variables, generating a total of ten hypotheses.²⁵

²⁵ Readers may notice that the permutation test resembles the placebo test used in quasi-natural experiment studies. In this exercise, we permute the simulated data by randomizing observations over all cells in the experiment, which is also referred as the unrestricted permutation of observations, and is recommended by Manly (2007). Westfall and Young (1993) show that this resampling-based multiple testing procedure provides a strong control of FWER.

The last method we consider to adjust for multiple comparisons is controlling for False Discovery Ratio (FDR), which controls for the “proportion,” rather than the “number,” of false rejections. The distinction between the “proportion” and the “number” is particularly important when we conduct a large number of tests are thus more willing to tolerate a higher number of false rejections. We adopt the method in Benjamini and Hochberg (1995) to control for the False Discovery Ratio δ at the 5% level. Specifically, we first rank a family of M p -values from minimum to maximum, and define: $j^* = \max\{j: p_j \leq \frac{j \times \delta}{M}\}$, j^* indexes the largest p -value that satisfies this inequality; Next, we reject all H_i , $i = 1, 2, \dots, j^*$. It is helpful to consider an example. If $M = 5$, and the five p -values are 0.01, 0.02, 0.03, 0.05, and 0.06. In this example, $j^* = 3$ ($\frac{3 \times 0.05}{5} = 0.03$). Therefore, although the fourth hypothesis has a p -value of 0.05, it is still higher than 0.04 ($\frac{4 \times 0.05}{5}$), we thus only reject the first three hypotheses. Benjamini and Hochberg (1995) theoretically show that this procedure controls for FDR at δ for independent test statistics.

7.2. Simulations

To demonstrate the application of the multiple test procedures outlined above, we follow Simmons, Nelson, and Simonsohn (2011) to generate simulated experimental data, and show two sets of results: 1) how the rejection rates change as the number of tests increases; 2) the rejection rates after adjusting for multiplicity. Each “experiment” is a 2×2 design with 20 “participants” per cell, and an outcome variable drawn from a standard normal distribution. Since we are interested in the magnitude of false-positive rates, “participants” are

randomly assigned to each of the 4 cells to construct a zero treatment effect. We label this design as the baseline case.

We then add three discretionary choices on top of the baseline case to mimic the process of *p*-hacking. They are choosing among different dependent variables, selecting the sample size, and using covariates.²⁶ A researcher can choose any combination of the three “degree of freedom” to search for the one that gives the highest level of statistical significance. For instance, one can choose to use the first dependent variable, the second, or the average of the two, and whether to add another experiment session or not. There are 12 unique combinations of all the choices available; in each experiment, we thus run 12 Analyses of Variance (ANOVAs) to exhaust all possible combinations of the options. To simplify the analysis, suppose we are only interested in finding a statistically significant *interaction effect* of any pair of independent variables. We then perform 1,000 simulated experiments and count — for each combination — how often the minimum *p*-value among all the interaction effects is below 0.05. The purpose of the simulation is to show that, if one looks into a given data set long enough and hard enough, it will be much easier to find statistically significant results by chance (i.e., spurious findings).

Column (1) of Table 2.5 reports each situation and its corresponding false-positive rates. In the baseline case, the false-positive rate is 3.8%, which is a bit lower than 5%. In situation 2, a researcher has the flexibility to analyze one of two dependent variables (correlation coefficient 0.5), or their mean, and choose whichever gives the highest statistical significance. This simple

²⁶ We choose these three researchers’ “degree of freedom” because these are the commonly made decisions while a researcher is collecting and analyzing data, and also because they are easily implementable. We can add more choices, but the fundamental takeaway from the exercise will be the same.

flexibility doubles the chance of obtaining a false-positive finding. If the researcher combines all choices, the rejection rate increases to 34.3% (situation 7), which implies the researcher can find at least one statistical significance in every three simulations. Because we only consider three choices, these estimates may still be conservative.

Columns (2) to (4) of Table 2.5 provide the rejection rates under the multiple testing procedures introduced. The Bonferroni method is the simplest but also the most conservative one, as evidenced by the low rejection rates compared with other methods. For example, in situation 1, there are 76 times among the 1000 simulations with at least one significant interaction effect; under the Bonferroni method, the number drops to 27. Turning to column (3), the rejection rates across all situations under the Bootstrapping adjusted p -values are generally higher than the rejection rates in column (2), especially for situations with a higher number of tests. Therefore, taking into account the dependence among test statistics indeed gives less conservative critical p -values. The rejection rates under Bootstrap adjusted p -values are also close to the 5% significance level. Column (4) reports the result of the BH method. By tolerating a higher number of false rejections when the total number of tests is high, BH also method yields a higher rejection rate than the Bonferroni method.

To summarize, we find that flexibility in data collection and analyses allow researchers to easily report statistically significant findings. After accounting for multiple testing, significant findings are much less frequent, and the rejection rates move towards 5%. Given that experimental accounting studies generally rely on small samples with low statistical power (higher p -values), we recommend the use of Bootstrap adjusted p -value and the BH method, which are

more powerful than Bonferroni method. However, a critical limitation in implementing multiple testing procedures is that, in reality, we can only observe those tests that are disclosed by researchers, and the reported tests may only be a subgroup of all performed tests. Hence, the use of multiple tests procedure cannot fully protect against the bias caused by p -hacking.

[Insert Table 2.5 about here]

7.3. The Bayesian approach

In our simulation, the treatment effect is designed to be zero and has no theoretical underpins. Consequently, the prior probability of running into a real discovery is also low, i.e., the prior odds ratio between the null hypothesis and the alternative hypotheses will be high. However, the computation of p -value does not incorporate this prior odds ratio. Recall that the p -value indicates $p(D|H_0)$, which is the probability of observing the data given the null being true, not $p(H_0|D)$, which is the probability of null being true given the data. If, however, we can incorporate this prior belief into our hypotheses testing, then we can potentially have a method to help guard against extreme findings that result from p -hacking. This is the underlying logic of the Bayesian approach.

Essentially, the Bayesian approach assesses the viability of a hypothesis based on a posterior odds ratio between the null hypothesis and the alternative hypotheses, given data likelihood ratio and prior odds ratio. According to the Bayesian theorem, the posterior odds ratio is defined as:

$$\underbrace{\frac{f(\theta_0|\text{data})}{f(\text{alternative}|\text{data})}}_{\text{Posterior odds ratio}} = \underbrace{\frac{f(\text{data}|\theta_0)}{\int (data|\theta)\pi_A(\theta)d\theta}}_{\text{Bayesian factor}} \times \underbrace{\frac{\pi_0}{\pi_1}}_{\text{prior odds ratio}},$$

where $f(x|y)$ is the probability density function of x conditional on y , $\pi_A(\theta)$ is the probability density under the alternative hypotheses, and $\int (data|\theta)\pi_A(\theta)d\theta$ is

thus the data likelihood under the alternative hypotheses. The odds ratio between the data likelihood under the null and the alternative, $\frac{f(\text{data}|\theta_0)}{\int (\text{data}|\theta)\pi_A(\theta)d\theta}$, is also called the Bayesian factor. After multiplying the prior odds ratio by the Bayesian factor, we have the posterior odds ratio on the left-hand side. Economically, the Bayesian factor measures the extent to which our prior belief changes after observing the data, with a smaller Bayesian factor providing stronger evidence *against* the null.

As can be seen from the above equation, calculating the Bayesian factor requires knowledge on the probability density of the alternative hypotheses, i.e., $\pi_A(\theta)$. To calculate $\pi_A(\theta)$, we need to specify priors on all possible alternative hypotheses, and for each of these hypotheses, we calculate the data likelihood under the null. However, in financial and accounting studies, we usually cannot specify all possible alternative hypotheses, making it is infeasible to calculate the Bayesian factor. Fortunately, there is one type of Bayesian factor that does not rely on the prior specification of alternatives: the minimum Bayesian factor (MBF). Harvey (2017) provides an excellent description of MBF, which we draw upon here.

Specifically, MBF is calculated for the case when a researcher focuses on a specific alternative hypothesis that turns out to be the maximum likelihood estimate of the data, i.e., the hypothesis that is most supported by the data. In other words, the MBF is calculated under the situation that what we believe (the alternative hypothesis) coincides with what we see (the data). Therefore, MBF provides the strongest evidence against the null hypothesis and gives an alternative hypothesis the best shot by concentrating the prior density at the

alternative rather than the null. Moreover, MBF is easy to calculate under the standard normal distribution assumption:

$$\text{MBF} = e^{\frac{z^2}{2}},$$

where e is the natural exponent, and z is the corresponding z -value from the null effect. For example, if $p\text{-value} = 0.05$ (two-tailed), then $z = \pm 1.96$, and $\text{MBF} = 0.15$, or $\frac{1}{6.8}$. As a result, the observed value supports the alternative hypothesis 6.8 times as strong as it does the null. Once MBF is obtained, we can calculate the Bayesianized p -value as follows:

$$\text{Bayesianized } p\text{-value} = \frac{\text{posterior odds ratio}}{1 + \text{posterior odds ratio}} = \frac{\text{MBF} * \text{prior odds ratio}}{1 + \text{MBF} * \text{prior odds ratio}}.$$

Continuing the above numerical example, if we believe that *ex ante*, the null hypothesis and the alternative are equally likely to be true (i.e., prior odds ratio = 1), then the Bayesianized p -value is 0.13, which indicates the chance of the null being true is 13%. If instead, the prior odds ratio = 0.50 (i.e., there is a modest chance in favor of the alternative), then the Bayesianized p -value becomes only 0.07. In both cases, the Bayesianized p -values are higher than the conventional p -value 0.05. However, the Bayesianized p -values carry a fundamentally different interpretation: it is a statement about the *viability* of the null hypothesis, as opposed to the likelihood of observing the data assuming the null being true.

How can the Bayesian approach be helpful in experimental accounting studies? Bloomfield, Rennekamp, and Steenhoven (2018) mention that the Bayesian approach can be used as a complement to the Null Hypothesis Significance Testing (NHST). We follow their line of reasoning and offer some recommendations here. Specifically, experimental researchers can take their

priors into account, and present Bayesianized p -values alongside the usual test statistics. Given the relatively high cost (e.g., recruiting and rewarding participants) of conducting experimental studies and small sample sizes, experimentalists generally do not commence data collection unless they believe that their underlying theories are highly plausible. This suggests that experimental accounting researchers generally have a low prior odds ratio (i.e., the alternative is more likely to be true than the null). The Bayesianized approach can be particularly valuable because it offers a convenient way for researchers to incorporate their prior beliefs, which is unfeasible in NHST. In addition, the choice of the prior odds ratio is transparent and less subject to manipulations.²⁷ Researchers can also choose to report Bayesianized p -values under multiple prior odds ratios for readers with different priors to make their assessments about the viability of the hypotheses given the test results.

7.4. *Disclosure-based solution*

Since p -hacking relies on exploiting *undisclosed* flexibility in data collection and analyses, another natural solution is to require researchers to make reasonably transparent disclosures on their research agendas and methodologies. Such disclosures are likely to curb researchers' *ex ante* incentives to conduct p -hacking because once a carefully planned research agenda is disclosed, researchers have limited room for cherry-picking afterward. Consequently, disclosure-based solutions have been widely discussed and adopted. For example, top journals in many disciplines usually require researchers to disclose their data

²⁷ One might think the choice of prior odds ratio is subjective. However, writing his presidential address, Harvey (2017) emphasize the point from Fisher (1925): "There is no 'objective' evaluation of hypotheses". Researchers from various disciplines relies too much on p -values to assess the viability of hypotheses, while ignoring the fact that p -value was not designed for this type of interpretation.

or code.²⁸ This disclosure could be more relevant to experimental studies, as making data and code public making them easier to be replicated, similar to archival-studies with public data sources. Of course, greater disclosure and transparency are not without costs. For instance, Bloomfield, Rennekamp, and Steenhoven (2018) point out that while JAR's a Registration-based Editorial Process (REP) can potentially enhance reproducibility by mitigating questionable practices, it may also shut down the channel for authors to learn from their data and refine research agendas.²⁹ In addition to the idea of data and code disclosure, we highlight three types of disclosures that journals can require from authors to mitigate *p*-hacking while imposing minimum costs on authors and reviewers.

1) *Specifying sample collection stopping rule*. Simmons et al. (2011), and Button et al. (2013) cite arbitrary stopping rule as a common practice among researchers to inflate results. As such, if journals aim to improve transparency in author reporting and documentation, a critical dimension of transparency is the sample selection procedure. We thus recommend that journals' disclosure policy can include guidelines that encourage authors to decide their sample selection stopping rule *ex ante*. For example, authors can state that "we decide to recruit 120 MBA participants for our experiment", or "we attempt to conduct a pilot study with 60 participants, and include the pilot sample as part of the experiment

²⁸For a partial list of top journals in social science that adopted similar policies, please see: https://web.stanford.edu/~cy10/public/data/Data_Availability_Policies.pdf.

²⁹ Journal of Accounting Research (JAR) experimented a Registration-based Editorial Process (REP) in 2017. Under REP, authors first submit their proposals containing data collection and analyses to JAR. After going through the process of peer review and revisions, editors and reviewers grant in-principal publication to proposals that meet publication standard, therefore guaranteeing publication regardless of the final results obtained under the agenda. The purpose of REP, according to Professor Leuz, is to enhance reproducibility of findings and encourage academic risk-taking (Wiley, 2018).

proper”. The stopping rules can be study-specific, but once chosen, journals and reviewers should make sure that authors stick to the rules.

2) *Report all the variables collected.* Transparent disclosure of outcome variables and conditions constructed in an experiment allows readers to assess the potential number of tests conducted by authors, and interpret their p -values in context.³⁰ Consequently, journals’ disclosure policy should put great emphasis on requiring authors to list all variables collected in a study. Moreover, the survey conducted by Bloomfield, Rennekamp, and Steenhoven (2018) also shows that researchers believe that “choosing analyses to highlight and report” (including choosing different variables) is most likely to “overstate results without improving informativeness.” This evidence suggests that journals’ efforts to address this issue are expected to meet researchers’ cooperation.

3) *Reporting p -values.* Authors should keep consistency in the choice between reporting one-tailed or two-tailed p -values. The articles in our sample can generally choose to report one-tailed p -values if the corresponding hypotheses are directional. However, it is not uncommon to find articles that report one-tailed p -values in the main tests, but two-tailed p -values in follow-up tests. Such discretionary use of tail choice is indicative of cherry-picking, especially in cases where the p -values would be higher than the significance threshold if two-tailed p -values are used.

In addition, journals should ensure that the underlying statistics of p -values (F - or t -stat, and degree of freedom) be available, either in tables or in the text where a p -value is mentioned. This reporting policy is adopted by *The*

³⁰ For example, if authors use three measures to capture the same outcome in an experiment and only report the one that generates strongest results, readers will have no way to assess the multiplicity of the test because the true number of tests is unknown.

Accounting Review. However, in our sample, we still find articles that do not report the degree of freedom in either tables or text, especially when the results are not reported in the form of ANOVA tables. Without such information, authors may have the incentive to misstate p -values when readers cannot easily verify the accuracy of p -values.³¹

To summarize, we discussed three types of researcher discretions used to overstate the importance of results. We believe that greater disclosure on these discretions can greatly enhance transparency without imposing much costs on both authors and reviewers. Notably, by no means are the three types of discretions a complete list of researchers' "degree of freedom." We hope the discussion here can further stimulate the conversation on improving transparency and reproducibility in experimental accounting research.

8. Conclusion

We uncover a discontinuity at 0.05 in the pooled cross-sectional distribution of p -values from experimental accounting studies. We interpret the discontinuity as a result of researchers' practice of p -hacking to generate significant results. The extent of p -hacking is more pronounced when authors face the pressure of tenure, among male researchers, and when a researcher works alone. Furthermore, researchers who graduated from higher-ranking schools seem to be more aggressive in p -hacking. Papers that exhibit a greater extent of p -hacking also attract more citations after publication.

The results in this paper may underestimate the true extent of p -hacking due to the following reasons. First, our sample could be biased because we use

³¹ When we attempt to calculate the exact values for p -values reported as " $p <$ ", we find several cases that authors overstate their p -values. For example, when authors report " $p < 0.10$ ", the true value is $p = 0.12$ or 0.14 . Had the articles not provide the underlying statistics such as t -statistics or F -statistics, we would not be able to discover these overstatements.

p -values mentioned in the text of articles, and researchers decide whether to mention a p -value in the text or not. Arguably, researchers would like to discuss the most significant results in their papers and are reluctant to emphasize a p -value that is just above the 0.05 threshold. Second, we only use articles published from three of the most prestigious accounting journals. It is reasonable to suspect that articles published on lower-tier journals are also subject to similar or even greater extent of p -hacking than what we documented here. Lastly, we focus only on the discontinuity at 0.05, but some researchers may aim for higher p -value targets such as 0.04, 0.03, or even 0.01 to avoid being too close to the 0.05 threshold. However, such practice is much harder to detect because aiming for a higher significance level will cause a more right-skewed p -value distribution, which can also be explained by publication bias.

We also propose several methods to guard against p -hacking. Specifically, we introduce several multiple testing procedure to control for false-positive findings due to conscious or unconscious p -hacking. Moreover, we propose a Bayesian approach for researchers to account for their prior beliefs on the likelihood of their theories. Lastly, we discuss some guidelines on enhancing disclosure with minimum burden on both authors and reviewers. Given the increasing attention on the prevalence and potential consequence of p -hacking in experimental accounting research, we believe it is the right time to discuss potential remedies. With cooperation among journals, reviewers, and authors, experimental accounting research will continue to produce high-quality works that boost scientific progress in accounting.

References

- Basu, S., Park, H.U., 2016. Publication Bias in Recent Empirical Accounting Research.
- Birnberg, J.G., 2011. A Proposed Framework for Behavioral Accounting Research. *Behavioral Research in Accounting* 23, 1-43
- Bloomfield, R., Nelson, M.W., Soltes, E., 2016. Gathering Data for Archival, Field, Survey, and Experimental Accounting Research. *Journal of Accounting Research* 54, 341-395
- Bloomfield, R., Rennekamp, K., Steenhoven, B., 2018. No System Is Perfect: Understanding How Registration-Based Editorial Processes Affect Reproducibility and Investment in Research Quality. *Journal of Accounting Research* 56, 313-362
- Brodeur, A., Lé, M., Sangnier, M., Zylberberg, Y., 2016. Star Wars: The Empirics Strike Back. *American Economic Journal: Applied Economics* 8, 1-32
- Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience* 14, 365-376
- Byrnes, J.P., Miller, D.C., Schafer, W.D., 1999. Gender differences in risk taking: A meta-analysis. *Psychological Bulletin* 125, 367-383
- Charness, G., Gneezy, U., 2012. Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization* 83, 50-58
- Cressey, D.R., 1953. *Other People's Money; A Study of the Social Psychology of Embezzlement*. Free Press, New York, NY.

- Dunnett, C.W., 1955. A Multiple Comparison Procedure for Comparing Several Treatments with a Control. *Journal of the American Statistical Association* 50, 1096-1121
- Faccio, M., Marchica, M.-T., Mura, R., 2016. CEO gender, corporate risk-taking, and the efficiency of capital allocation. *Journal of Corporate Finance* 39, 193-209
- Fanelli, D., 2010. “Positive” results increase down the hierarchy of the sciences. *PloS ONE* 5, e10068
- Fanelli, D., 2012. Negative results are disappearing from most disciplines and countries. *Scientometrics* 90, 891-904
- Fisher, R.A., 1925. Theory of statistical estimation. In: *Mathematical Proceedings of the Cambridge Philosophical Society*, pp. 700-725. Cambridge University Press
- Friesen, L., Gangadharan, L., 2012. Individual level evidence of dishonesty and the gender effect. *Economics Letters* 117, 624-626
- Greenwald, A.G., 1975. Consequences of prejudice against the null hypothesis. *Psychological bulletin* 82, 1-20
- Harvey, C.R., 2017. Presidential Address: The Scientific Outlook in Financial Economics. *The Journal of Finance* 72, 1399-1440
- Harvey, C.R., Liu, Y., Zhu, H., 2016. ... and the cross-section of expected returns. *The Review of Financial Studies* 29, 5-68
- Head, M.L., Holman, L., Lanfear, R., Kahn, A.T., Jennions, M.D., 2015. The extent and consequences of p-hacking in science. *PLoS Biol* 13, e1002106

- Henry, E., 2009. Strategic disclosure of research results: The cost of proving your honesty. *The Economic Journal* 119, 1036-1064
- Khan, M.J., Tronnes, P., Christen, 2018. P-hacking in Experimental Audit Research. *Behavioral Research in Accounting*
- Krawczyk, M., 2015. The Search for Significance: A Few Peculiarities in the Distribution of P Values in Experimental Psychology Literature. *PLoS One* 10, e0127872
- Leuz, C., 2018. Evidence-based policymaking: promise, challenges and opportunities for accounting and financial markets research. *Accounting and Business Research* 48, 582-608
- Libby, R., Bloomfield, R., Nelson, M.W., 2002. Experimental research in financial accounting. *Accounting, Organizations and Society* 27, 775-810
- Libby, R., Brown, T., 2013. Financial statement disaggregation decisions and auditors' tolerance for misstatement. *The Accounting Review* 88, 641-665
- List, J.A., Shaikh, A.M., Xu, Y., 2019. Multiple hypothesis testing in experimental economics. *Experimental Economics*
- Margheim, L.L., 1986. Further evidence on external auditors' reliance on internal auditors. *Journal of Accounting Research*, 194-205
- Masicampo, E.J., Lalande, D.R., 2012. A peculiar prevalence of p values just below .05. *Q J Exp Psychol (Hove)* 65, 2271-9
- Necker, S., 2014. Scientific misbehavior in economics. *Research Policy* 43, 1747-1759
- Neyman, J., Pearson, E.S., 1933. IX. On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. R. Soc. Lond. A* 231, 289-337

- O'Riordan, M., 2018. The Road to Reproducibility: Registered Reports and the Journal of Accounting Research. URL <https://www.wiley.com/network/researchers/being-a-peer-reviewer/the-road-to-reproducibility-registered-reports-and-the-journal-of-accounting-research>
- Pashler, H., Harris, C.R., 2012. Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science* 7, 531-536
- Peecher, M.E., 1996. The influence of auditors' justification processes on their decisions: A cognitive model and experimental evidence. *Journal of Accounting Research*, 125-140
- Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science* 22, 1359-1366
- Simonsohn, U., Nelson, L.D., Simmons, J.P., 2014. P-curve: a key to the file-drawer. *Journal of experimental psychology: General* 143, 534
- Statzner, B., Resh, V.H., 2010. Negative changes in the scientific publication process in ecology: potential causes and consequences. *Freshwater Biology* 55, 2639-2653
- Sterling, T.D., 1959. Publication Decisions and their Possible Effects on Inferences Drawn from Tests of Significance—or Vice Versa. *Journal of the American Statistical Association* 54, 30-34
- Vermeulen, I., Beukeboom, C.J., Batenburg, A., Avramiea, A., Stoyanov, D., van de Velde, B., Oegema, D., 2015. Blinded by the light: How a focus on statistical “significance” may cause p-value misreporting and an excess

- of p-values just below .05 in communication science. *Communication Methods and Measures* 9, 253-279
- Westfall, P.H., Young, S.S., 1993. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment* (Wiley Series in Probability and Statistics).
- Westfall, P.H., Young, S.S., Wright, S.P., 1993. On adjusting P-values for multiplicity. *Biometrics* 49, 941-945
- Wicherts, J.M., Veldkamp, C.L., Augusteyn, H.E., Bakker, M., Van Aert, R., Van Assen, M.A., 2016. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology* 7
- Young, N.S., Ioannidis, J.P., Al-Ubaydli, O., 2008. Why current publication practices may distort science. *PLoS medicine* 5

Appendix A: The distribution of p -values from archival-empirical studies

Potential readers of our paper will naturally concern that p -hacking also exists among archival accounting research, given that archival accounting researchers also have flexibility over data analyses and strong incentives to generate top journal publications. To assess the degree of p -hacking for archival studies, we randomly select 260 articles from each of the top three accounting journals TAR, JAR, and JAE (Journal of Accounting and Economics). Initially, we attempt to collect p -values from the text of these 780 articles using the same Python script as we used for experimental studies. However, we are able to scrape 232 p -values from only 76 articles, because unlike experimental studies, archival studies generally do not mention p -values within the text. The sample is too small to conduct any meaningful analysis of the existence of p -hacking among those articles. We then identify 163 articles that report p -values in tables of regression analyses among the 780 articles. For a regression table, we record p -values that appeared in the first three rows except those for intercepts. Arguably, such p -values are more likely to be the test statistics for key results. This method yields 4,767 p -values.

We find that the distribution of p -values is extremely uneven and right-skewed: there are 1,680 p -values reported as zero. Such a high frequency of zero would dwarf the heights of bins for other p -values, we thus exclude zero p -values. The histogram for the rest p -values is presented in Figure A1. Compared with Figure 1, the distribution of p -values from archival accounting does not show a discontinuity at 0.05. This finding is largely consistent with the distribution of p -values in Basu and Park (2016), who report a general decreasing frequency from 0.01 to 0.05. We can think of two explanations for the difference. First, archival

studies are less prone to *p*-hacking, probably due to the public availability of databases that makes replication easier and provides a monitoring mechanism for *p*-hacking. Alternatively, as archival researchers usually use much larger samples than experimentalists, they have much higher statistical power. The large sample of archival studies may enable them to aim for $p \leq 0.01$ when doing *p*-hacking. The high frequency of *p*-values near zero is probably the result of this. In sum, we do not observe a sharp discontinuity at 0.05 on the distribution of *p*-values from archival accounting articles. But we still cannot rule out the possibility that those studies aim for the highest statistical significance level (*p*-values near zero) while doing *p*-hacking.

Figure A1. Histogram of *p*-values from archival accounting studies.

The figure presents the histogram of 3,087 *p*-values manually collected from 163 archival accounting studies. The bin width is 0.001.

Panel A: Raw distribution of all p -values

Panel B: Raw distribution weighted by article's number of p -values

Figure 2.1: Empirical distribution of all p -values from 1990 to 2016. Bin width is 0.001. The bold bar indicates frequency of p -values $\in (0.049, 0.05]$. In Panel B, the distribution uses the inverse of the number of p -values collected to weight observations.

Panel A: Only including p -values reported with two decimal places

Panel B: Rounding all p -values to two decimal places

Figure 2.2: Isolating p -values reported with two decimal places.

In Panel A, we use only p -values reported with two decimal places; in Panel B we round all p -values and use the rounded values to generate the histogram.

Panel A: Only including p -values reported with two decimal places

Panel B: Rounding all p -values to two decimal places

Figure 2.3: Fitting histogram of p -values to exponential distribution

Red line represents the fitted exponential curve that minimize the squared fitting errors; grey triangular represents the frequency of each observed p -value on x-axis. The exponential function in Panel A and B are $y = 205.49e^{-17.65x}$ and $y = 458.464e^{-21.96x}$, respectively.

Panel A: The distribution of vague p -values

Panel B: The distribution of 104 exact p -values calculated from $p < 0.05$ or $p < 0.10$

Figure 2.4: The distribution of vague p -values

The histogram in Panel A shows the distribution of 2,362 p -values reported as " $p <$ "; the histogram in Panel B shows the distribution of 104 exact p -values we are able to compute for the 396 p -values reported as $p < 0.05$ or $p < 0.10$, which are indicated with the bold bars.

Panel A: Adding the estimated distribution of $p < 0.05$ or $p < 0.10$ to main sample

Panel B: Adding the estimated distribution of all vague p -value to main sample

Figure 2.5: Including vague p -values

In Panel A, we combine the empirically estimated distribution of 396 vague p -values ($p < 0.05, 0.10$) with the distribution of exact p -values; in Panel B we further include other vague p -values by assuming a uniform distribution for $p < x$ between $(x-0.01)$ and x , where x denotes 0.02, 0.03, ..., and 0.15 (except 0.05 and 0.10).

Panel A: Articles in *low citation* group

Panel B: Articles in *high citation* group

Figure 2.6: *P*-hacking and citations

In each year, we split articles into high and low citation groups based on the median number of citations of all articles published in that year. The final low (high) citation group consists of low (high) citation groups from every year.

Panel A: *P*-values appeared in the earlier part of an article

Panel B: *P*-values appeared in the later part of an article

Figure 2.7: Do the main results of an article show more *p*-hacking?

For each article, we sort all of its *p*-values by the sequences they appear in that article, then we identify the median value of the sequential number. Panel A contains those *p*-values that appear before each articles' respective median sequential number. Panel B contains the rest.

Figure 2.8: JAR’s 2014 policy and the distribution of p -values.

The figure shows the distribution of p -values from articles published in two journals (JAR and TAR) during two periods (2012-2015 and 2016-Jul2019). The 2014 policy of JAR is the updated data and code policy that strengthens disclosure from authors. We assume a two-year lag between a paper’s initial submission to JAR and the final publication. Therefore, papers published in 2016 and after are assumed to be submitted under the updated policy. The numbers of articles for each subsample are shown in parentheses.

Table 2.1: Annual number of published experimental articles

The table reports the annual distribution of the 281 articles in our sample. We are able to find 317 experimental articles published on the top 3 accounting journals. After excluding papers that do not report p -values in our sample (30), and those that were retracted (6), we end up with 281 experimental articles.

Year	# of articles	Percent (%)
1990	15	5.34
1991	10	3.56
1992	10	3.56
1993	9	3.20
1994	6	2.14
1995	12	4.27
1996	8	2.85
1997	14	4.98
1998	2	0.71
1999	12	4.27
2000	6	2.14
2001	12	4.27
2002	8	2.85
2003	5	1.78
2004	9	3.20
2005	13	4.63
2006	7	2.49
2007	9	3.20
2008	12	4.27
2009	8	2.85
2010	13	4.63
2011	17	6.05
2012	13	4.63
2013	13	4.63
2014	12	4.27
2015	14	4.98
2016	12	4.27
Total	281	100%

Table 2.2: Summary statistics at the article-level

This table displays summary statistics for author and paper characteristics based on a sample of 240 experimental articles published in Top 3 accounting journals between 1990 and 2016. *N_Authors* is the number of listed author(s) for a paper. *Female* is a dummy variable that equals one if the listed author(s) contain at least one female, and zero otherwise. *N_Participants* is the number of subjects participated in a paper's experiments. *N_p-values* is the total number of p-values scraped from a paper. *Two-tailed* is a dummy that equals one if an article only report two-tailed *p*-values. *Sole* is a dummy variable that equals one if a paper contain only one author, and zero otherwise. *Tenured* is the fraction of authors that are tenured among all authors of a paper. *School Ranking* is the mean ranking of all authors' affiliated schools for a paper. *PhD School Ranking* is the mean ranking of authors' PhD alma mater for a paper. *Experience* is the average number of years passed after an articles' authors obtained their PhD degree, calculated at the publication year of that article. The subsample for JAE are not tabulated separately because there are only three articles from JAE in our sample.

	N	SD	p25	p50	p75
Full sample					
<i>N_Authors</i>	240	0.937	1.000	2.000	3.000
<i>Female</i>	240	0.501	0.000	0.500	1.000
<i>N_Participants</i>	233	87.926	65.000	97.000	138.000
<i>N_p-values</i>	240	10.589	10.000	16.000	23.000
<i>Two-tailed</i>	239	0.476	0.000	0.000	1.000
<i>Sole</i>	240	0.454	0.000	0.000	1.000
<i>Tenured</i>	235	0.373	0.000	0.500	0.667
<i>School Ranking</i>	240	30.504	19.000	40.000	67.167
<i>PhD School Ranking</i>	229	21.739	15.000	22.500	39.000
<i>Experience</i>	230	6.000	4.000	8.000	12.500
JAR					
<i>N_Authors</i>	77	0.875	2.000	2.000	3.000
<i>Female</i>	77	0.501	0.000	0.000	1.000
<i>N_Participants</i>	74	57.891	78.000	106.000	145.000
<i>N_p-values</i>	77	11.397	10.000	17.000	24.000
<i>Two-tailed</i>	77	0.496	0.000	0.000	1.000
<i>Sole</i>	77	0.434	0.000	0.000	0.000
<i>Tenured</i>	75	0.356	0.000	0.500	0.667
<i>School Ranking</i>	77	27.507	17.500	33.000	49.000
<i>PhD School Ranking</i>	75	20.089	13.000	20.000	34.000
<i>Experience</i>	75	6.459	4.000	9.000	14.000
TAR					
<i>N_Authors</i>	160	0.972	1.000	2.000	3.000
<i>Female</i>	160	0.501	0.000	1.000	1.000
<i>N_Participants</i>	156	99.251	63.000	91.000	130.500
<i>N_p-values</i>	160	10.183	11.000	16.000	23.000
<i>Two-tailed</i>	159	0.463	0.000	0.000	1.000
<i>Sole</i>	160	0.462	0.000	0.000	1.000
<i>Tenured</i>	157	0.384	0.000	0.500	0.667
<i>School Ranking</i>	160	31.599	20.250	48.250	70.250
<i>PhD School Ranking</i>	151	22.552	15.500	25.000	43.000
<i>Experience</i>	152	5.753	4.000	7.583	12.333

Table 2.3: Deviations from the exponential distribution

This Table reports the deviations of the frequency of each p-values with two decimal places (0.01, 0.02... 0.15) from its forecasted frequency determined by the exponential fitting. In Panel A, we only use p-values reported with two decimal places; in Panel B we round p-values to two decimal places and use all p-values in our sample. In Column (1), *Freq* is the observed frequency for each p-value; in Column (2), *Residual Freq* is the observed frequency minus the expected frequency of each p-value based on exponential fitting. The last column measures the ratio of forecast residual in (2) to observed frequency in (1).

<i>p</i> -value	(1) Freq	(2) Residual Freq	(3) (Residual Freq)/ Freq
<i>Panel A: Using p-values reported with two-decimal places</i>			
0.01	187	8.229	4.40%
0.02	143	-7.660	-5.36%
0.03	111	-15.969	-14.39%
0.04	88	-19.004	-21.60%
0.05	122	31.822	26.08%
0.06	60	-15.998	-26.66%
0.07	61	-3.047	-5.00%
0.08	58	4.024	6.94%
0.09	54	8.511	15.76%
0.10	44	5.664	12.87%
0.11	20	-12.308	-61.54%
0.12	20	-7.227	-36.14%
0.13	17	-5.946	-34.98%
0.14	21	1.662	7.91%
0.15	12	-4.297	-35.81%
<i>Panel B: Rounding up all p-values to two decimal places</i>			
0	447	-17.143	-3.84%
0.01	432	54.226	12.55%
0.02	293	-14.476	-4.94%
0.03	192	-58.260	-30.34%
0.04	202	-1.691	-0.84%
0.05	214	48.213	22.53%
0.06	121	-13.937	-11.52%
0.07	112	2.172	1.94%
0.08	88	-1.390	-1.58%
0.09	89	16.244	18.25%
0.10	67	7.782	11.62%
0.11	32	-16.198	-50.62%
0.12	30	-9.229	-30.76%
0.13	27	-4.929	-18.26%
0.14	36	10.012	27.81%
0.15	22	0.848	3.85%

Table 2.4: The magnitude and statistical significance of p -hacking.

This table reports the results of the post hoc pairwise comparison for the residuals from the curve fit in Figure 3. We compare the residuals at 0.05 and 0.06 to the residuals of the rest p -values. In Panel A, the sample only includes p -values reported with two decimal places; in Panel B, we use the full sample of exact p -values. Column (1) is the observed frequencies for 0.05 and 0.06, (2) is the residual frequency from the fitted exponential function, computed as the difference between observed frequency and predicted frequency. Column (3) reports the ratio of the residual frequency in (2) to the observed frequency in (1). P -values in Column (4) are the p -values of mean differences between the residual of the focal p -value (0.05 or 0.06) and the reference group, and p -values in the last column are adjusted for multiple comparisons using the Dunnett's test.

	(1)	(2)	(3)	(4)	(5)
p -value	Freq	Residual Freq (Observed-predicted)	(2)/(1)	p -value	p -value adjusted for multiple comparison
Panel A: Only including p-values reported with two decimal places					
0.05	122	32	26.23%	0.002	0.005
0.06	60	-16	-26.67%	0.281	0.340
Panel B: Rounding all p-values to two decimal places					
0.05	214	48	22.43%	0.053	0.113
0.06	121	-14	-11.57%	0.594	0.856

Table 2.5: Comparing the extent of p -hacking across various subsamples.

This table provides results of post hoc pairwise comparison for each histogram in each pair of subsamples. For each histogram in a pair of subsamples, we derive the residual of exponential fit for every p -value from 0.00, 0.01, to 0.15, and then compare the residuals at 0.05 and 0.06 with the mean residuals of the rest 14 p -values. Results for 0.06 are all insignificant and thus omitted. Column (1) is the observed frequencies for 0.05, (2) is the residual frequency from the fitted exponential function, and (3) is the residual ratio computed as residual frequency over observed frequency. Column (4) reports the p -values from tests of difference between the residual at 0.05 and the mean residuals for p -values in the reference group, and p -values in the last column is adjusted for multiple comparison using the Dunnett's test.

Subsamples	(1) Freq	(2) Residual Freq (Observed-predicted)	(3) (2)/(1)	(4) p -value	(5) p -value adjusted for multiple comparison
<i>Panel A: At least one tenured authors</i>					
Yes	104	9	8.65%	0.311	0.526
No	72	27	37.50%	0.016	0.033
<i>Panel B: Experience</i>					
More	93	11	21.88%	0.417	0.648
Less	107	28	16.98%	0.022	0.051
<i>Panel C: School Ranking</i>					
High	100	32	32.00%	0.083	0.151
Low	114	17	14.91%	0.271	0.474
<i>Panel D: PhD School Ranking</i>					
High	96	21	27.38%	0.050	0.107
Low	106	18	9.64%	0.158	0.298
<i>Panel E: Sole authored article</i>					
Yes	67	24	35.82%	0.027	0.06
No	147	20	13.61%	0.157	0.298
<i>Panel F: N_ Participants</i>					
High	111	28	25.23%	0.051	0.110
Low	99	14	14.14%	0.139	0.226
<i>Panel G: Contain female authors</i>					
Yes	109	15	13.76%	0.122	0.238
No	105	28	26.67%	0.083	0.170
<i>Panel H: Google_Citations</i>					
High	129	31	24.03%	0.026	0.059
Low	85	11	12.94%	0.377	0.616
<i>Panel I: Web_Citations</i>					
High	132	34	25.76%	0.026	0.059
Low	62	1	1.61%	0.808	0.974

Table 2.6: Rejection rates after adjusting for multiple hypotheses testing.

The table reports the percentage of at least one significant interaction effect from analysis of variance (ANOVA) based on 1,000 random samples. The Column “Researcher’s degree of flexibility” displays the flexibility available for researchers. The baseline is a 2×2 “experiment” with 20 observations per cell and only one dependent variable. In situation 1, a researcher collects two dependent variable and conduct 3 ANOVA tests, one on each variable and a third on the mean of the two; we report significant result if any one of the ANOVA delivers statistically significant interaction effect. In situation 2, the researcher can choose to whether collect additional 10 observations per cell or not. In situation 3, the researcher can add another condition or not, we report significant results in any two of the conditions have statistically significant interaction effects. Situation 4 to 7 are different combinations of the first 3 situations. In each situation, we get the lowest p -values among all interaction effects, and check if it is lower than 5%.

	% of p -values below significance threshold			
	(1)	(2)	(3)	(4)
Researcher’s degree of flexibility	$p \leq .05$	Bonferroni method	Bootstrap adjusted p -value	BH method
Situation 0: Baseline	3.80%	3.80%	3.90%	3.80%
Situation 1: 2 dependent variables	7.60%	2.70%	3.90%	3.70%
Situation 2: Adding 10 obs	6.30%	3.10%	4.20%	3.40%
Situation 3: Adding another IV	14.10%	4.30%	5.00%	5.80%
Situation 4: Combining 1 and 2	12.30%	2.30%	3.70%	2.90%
Situation 5: Combining 1 and 3	25.60%	2.60%	6.20%	4.80%
Situation 6: Combining 2 and 3	22.00%	4.50%	5.90%	5.30%
Situation 7: Combining 1, 2, 3	34.30%	2.90%	6.60%	4.70%

Chapter 3: Local Gambling Preference and Corporate Cash Holding

Abstract

This paper studies the relation between local attitudes towards gambling and corporate liquidity. I find that firms exposed to stronger local gambling preference hold significantly more cash, as cash enables them to take more risk while staying away from financial distress. Furthermore, I show that gambling-prone firms hold more cash only when they are financially constrained, and these firms also have higher marginal value of cash holding. Corporate headquarter relocation confirms the causal impact of gambling preference on cash holding. Taken together, these findings suggest that greater cash holding in gambling-prone firms is a value-increasing policy due to their strong risk-taking incentive.

Keywords: Gambling preference; Financial constraint; Cash hoarding; Value of cash

JEL classification: G32; G41; Z12

1. Introduction

Figure 3.1 shows the average corporate cash holding for listed firms by their headquarter states. As can be seen, there is a wide variation in firms' average cash holding at the state level. For example, firms that headquarter in Nebraska hold 15% of its assets in cash or near-cash instruments; in contrast, firms in South Dakota, which borders Nebraska, only have a cash-to-asset ratio of 7%. While prior works have made much progress in understanding the dramatic time-series increase in cash holding starting from the 1980s (e.g., Bates, Kahle, and Stulz, 2009), they do not explain the apparent geographic variation in local firms' cash holding policy. Motivated by a growing literature on how local culture affects financial market outcomes, I examine the relation between religion-induced gambling preference and corporate cash policy.

Figure 3.1: State-level variation in local firms' cash holding.

The sample includes all listed firms with non-missing cash ratio (che/at) from 1987 to 2017. A state's cash ratio is calculated by averaging the cash to asset ratio across all firm-years during the sample period.

As an important dimension of cultural norms, local gambling preference appears to be a natural candidate to explain the large variation in regional firms'

propensity to hold cash. To the extent that access to the financial market or corporate governance does not change dramatically from different states within the U.S., state-level variations in capital market frictions or agency conflicts are unlikely to lead to the state-level dispersion in cash holding. In contrast, a growing body of research has shown that local religion-induced attitudes towards gambling greatly shape people's attitudes towards risk and uncertainty, and this, in turn, manifests in local firms' precautionary saving motive, which is the dominant motives for corporate cash holding.

It is theoretically unclear how gambling preference will affect firms' level of cash holding. On the one hand, stronger gambling preference could foster a higher level of tolerance for uncertainty among local stakeholders, including investors and employees, etc., who may have a weaker motivation to push for conservative corporate policies from managers, including a higher level of cash hoarding. Furthermore, given stakeholders' greater appetite for risk-taking, managers may have more discretion to spend the cash on investment opportunities such as merger and acquisition (Harford, 2002) or on capital-intensive innovative projects, to deploy the cash and grow the firm quickly. Thus, if gambling firms are less likely to save cash but deploy cash more quickly, we may observe that *stronger gambling preference leads to a lower level of cash holding in local firms*.

On the other hand, prior literature also documents that firms with riskier cash flow and financially constrained firms tend to hold more cash (Archarya, Davydenko, and Strebulaev, 2012; Chang et al., 2014). Gambling-prone firms are more likely to engage in investment with lottery-like payoffs, i.e., R&D investments (Chen et al., 2014), which are particularly difficult to be financed

through external funds due to the high idiosyncratic risk. Anticipating the increasing idiosyncratic risk and limited access to external financing, managers in gambling-prone firms may optimally maintain a higher level of cash holding to move away from financial distress (Bolton, Chen, and Wang, 2011). Importantly, cash holding is higher *not* because of managerial or shareholder preference for safe assets, but because of financing frictions that restrict gambling-prone firms to hedge their idiosyncratic risk or timely access to external liquidity. The above discussion thus leads to the prediction that *gambling-prone firms hold more cash than firms with strong opposition to gambling behaviors.*¹

Given the opposing forces that gambling preference may exert on corporate cash holding, which one of them prevail is a question that should best be answered empirically. To empirically determine if gambling preference explains the geographical variation in corporate cash holding, I follow previous literature to measure a firm's exposure to local gambling preference using the religious belief of the residents in the region where the firm headquarters. In particular, Catholics Churches are more tolerable towards a moderate level of gambling. However, Protestant Churches are characterized by their strong moral opposition to gambling. The geographic variation in the ratio of Catholics to Protestants at the U.S. county-level thus provides a plausible proxy for local gambling preference. In terms of the validity of this proxy, prior research has documented robust evidence that religious attitudes towards gambling strongly

¹ This hypothesis does not suggests that firms with stronger preference for gambling are actually more conservative than other firms. Gambling-prone firms hold more cash because their greater appetite for risk-taking increase idiosyncratic cash flow volatility, and external financing constraint makes these firms hoard cash to better cope with cash flow shortfall and reduce the probability of liquidation (e.g., Bolton, Chen, and Wang, 2011).

predicts local residents' actual gambling activities, as well as their risk attitudes. Moreover, local residents' gambling preferences can either directly or indirectly manifest in local firms' economic decisions (Section 2 provides a detailed review of the related evidence).

[Insert Figure 3.2 here]

My empirical analysis reveals a strong positive relation between local gambling preference and corporate cash holding. Combining Figure 3.1 and 3.2, I find that the variation in cash ratio in Figure 3.1 seems to go hand-in-hand with the variation in the Catholic-to-Protestant ratio in Figure 3.2. Specifically, states with a higher Catholic-to-Protestant ratio also tend to be states with a high average corporate cash ratio. After controlling for other known firm-level determinants of corporate cash holding and county-level demographic variables, the positive relation continues to hold. The magnitude of the documented effects are both economically and statistically significant: moving from the county at the 25th percentile to the one at the 75th percentile in terms of Catholics to Protestant ratio increases cash to asset ratio by 3.92 percentage points, which amounts to 20.63% of the mean value of cash holding.

While I have documented a significantly positive relation between gambling preference and cash holding, the result could be relational instead of causal. The biggest obstacle to a causal interpretation is perhaps the measurement error problem, that is, local Catholics to Protestants ratio may capture not only gambling preference but also other factors unrelated to gambling attitudes, such as work ethics, religiosity, etc.² Yet given that previous literature has documented

² Catholics and Protestants may differ in other dimensions, not only in preference for gambling. For example, Weber (1905) wrote that Protestant values is more inductive to capitalism, as Protestants have a more embracing view on work and commerce.

that local Catholics to Protestants ratio is robustly linked to market settings where payoff skewness is a salient feature, it is difficult to think of a better and more reasonable explanation to summarize the prior findings. Nonetheless, I conduct several tests to strive for a causal interpretation.

First, I use corporate headquarter relocations as a shock to firms' local gambling preference. The results show that firms relocate to a region with stronger (weaker) gambling preference measure hold more (less) cash after the relocation, suggesting a change in corporate cash policy in line with the local culture. Second, I use an alternative measure of local gambling preference: state-level per capita lottery revenue. I find this measure to be positively correlated with local firms' cash holding. Importantly, after controlling for the religion-induced gambling preference, the significance of the lottery sale goes away. This implies that religion-induced gambling preference is likely to drive the positive relation between lottery demand and cash holding.

Having established a positive relation between local gambling preference and cash holding, I further investigate the potential mechanism that may plausibly drive the main results. The hypothesis suggests that it is because gambling-prone firms' greater risk-taking makes them operate more uncertain businesses, thus increasing their cash flow risk. This results in higher demand for cash to avoid adverse shocks to future cash flow, especially for more financially constraint firms.³ To test this channel, I study the impact of gambling preference on firms' investment policies. Firms located in counties with a higher ratio of

³ For example, Smith and Stulz (1985) argue that greater volatility in cash flow increase the likelihood of costly bankruptcy. Froot, Scharfstein, and Stein (1993) further point out that higher cash flow volatility could lead to underinvestment when external financing is costly. When the cash flow risk is unhedgeable or it is costly to hedge, firms can hold more cash to alleviate bankruptcy risk and underinvestment (Archarya, Davydenko, and Strebulaev, 2012; Denis and Sibilkov, 2010).

Catholics-to-Protestants (CPRATIO) invest more in risky projects, but less in capital expenditures. As a result, they generate more volatile future cash flows. This is consistent with the prior finding that cash holding is positively related to R&D, but negatively to capital expenditure. Furthermore, I test the interaction effect of gambling preference and financial constraint on cash and find that gambling-prone firms hold more cash only when they are financially constrained. These findings indicate that gambling-prone firms hold more cash not because they prefer safe assets, but because cash allows them to take more risk and alleviate financial constraints.

If greater cash holding allows gambling-prone firms to take more risk, and alleviate cash flow risk, how does gambling preference affect the value of cash? Conceptually, an extra dollar will result in a larger (smaller) increase in firm value if it is well (poorly) invested. It is again *ex ante* unclear whether gambling preference will decrease or increase the value of cash. Empirically, the value of excessive cash is significantly higher in firms located in regions with stronger gambling preference. Combined with the positive relation between gambling preference and R&D investment, this result suggests that gambling preference increases the value of cash by alleviating corporate underinvestment in risky projects. These projects might be foregone due to, say, agency conflicts such as managerial short-termism and career concern. This interpretation is broadly consistent with Chen et al. (2014), who find that gambling preference drives firms to invest more in innovation.

The empirical results are robust to a large number of variations to the baseline specification. In particular, I further follow Chen et al. (2014) to perform a two-stage least squares (2SLS) regressions using past religious beliefs at the

local level. To test whether a common, increasing trend in the key dependent and independent variables causes spurious relation, I perform a Fama-MacBeth style regression. I also use only the census year with available religious data to perform regression. The main results still hold.

This paper contributes to the literature in the following ways. First, this paper contributes to the literature on corporate cash holding. Prior works generally focus on explaining the firm-level variation in cash holding, or the time-series of corporate cash holding in the U.S. To my best knowledge, this paper is the first to show the geographic variation of corporate cash holding in U.S. This paper is also related to Chen et al. (2015) and Dudley and Zhang (2016), who study how local cultural characteristics and social norms such as individualism and trust can affect corporate cash policy.

Second, this paper adds to the strand of literature on how gambling preference affects the aggregate financial market. Earlier works such as Brunnermeier, Gollier, and Parker (2005), and Barberis and Huang (2008) develop behavioral models in which individuals overweight the probability that a large positive payoff will realize, which generates investor preference for lottery-like assets, i.e., positively skewed wealth distribution. Kumar, Page, and Spalt (2011) develop the religion-induced measure of gambling preference and find that this measure explains investors' preference for investment opportunities that feature positively skewed returns. My paper adds to this line of research by showing that gambling preference not only affects investor trading behavior but also plays a significant role in shaping corporate decisions, particularly corporate liquidity.

2. Related literature

Building on existing literature that emphasizes the role of preference for gambling in shaping the financial market outcomes, this paper examines its potential impact on corporate cash holding, in doing so, it brings together two strands of literature.

2.1. Religion-induced gambling preference

Motivated by the empirical evidence on how Catholic and Protestant differ in attitudes towards gambling activities and its effect on state- and county-level lottery sales (Berry and Berry, 1990; Ellison and Nybroten, 1999), Kumar et al. (2011) first introduce Catholics-to-Protestants as a proxy for local gambling preference. They find that investors located in regions with high *CPRATIO* hold disproportionately more stocks with lottery features, i.e., stocks that have a low probability of earning extremely high returns.

Subsequent studies use *CPRATIO* to examine the influence of gambling preference on various investment and corporate decisions. Shu, Sulaeman, and Yeung (2012) find that local *CPRATIO* affects the risk-taking of sophisticated investors such as mutual funds. Kumar et al. (2016) further show that correlated trading of investors with gambling preference causes excessive comovement in the stock market. Chen et al. (2014) use the same measure and link local gambling preference to corporate investment and innovation, and document that firms located in high *CPRATIO* areas invest more in research and development (R&D) and experience higher innovation output. Ucar (2015) shows that this religion-induced gambling preference matters for firms' dividend policy. Kumar et al. (2011) find that broad-based employee stock options are more prevalent among employees with strong gambling preference, possibly due to firms

catering to those preferences. Hilary and Hui (2009) report that chief executive officers (CEOs) are more likely to join a firm with a similar religious environment as in their prior employers. The last two papers provide some evidence on the channels through which local religious beliefs might affect corporate decisions. My paper focuses on the role of local gambling preference on firm cash policy. In doing so, this paper complements the works mentioned above by extending the risk-taking implication of gambling preference to the context of corporate liquidity management.

2.2. Corporate risk-taking and cash holding

There is a growing consensus that cash reserve facilitates corporate risk-taking (e.g., Bates, Kahle, and Stulz, 2009). Fundamentally, this is due to financing frictions that make it costly for firms to fund investment opportunities through either external financing (Myers and Majluf, 1984), or converting assets into cash (Keynes, 1936; Shleifer and Vishny, 1993). Moreover, riskier investments generally face greater external financing constraints, as more uncertain future cash flows complicate project valuation and thus aggravate information asymmetry (Opler and Titman, 1994) and leads to a higher likelihood of financial distress. Consistent with this argument, prior works find that firms hold more cash when they invest more R&D (Opler et al., 1999), are in industries with higher cash flow volatility (Han and Qui, 2007), have high credit risk (Acharya, Davydenko, and Strebulaev, 2014). Bates, Kahle, and Stulz (2009) argue that increasing cash flow volatility and greater R&D intensity contribute to the overall upward trend in cash holdings. Subrahmanyam, Tang, and Wang (2017) document that firms hold more cash after the inception of credit default

swap on their debt, due to heightened firm risk stemming from CDS trading transforming lenders into tougher negotiators.

Importantly, the literature does not suggest that firms with higher risk are conservative. For high-risk firms, the marginal benefits of liquidity are higher: a larger cash pile can help avoid costly liquidation and fund investment opportunities (Bolton, Chen, and Wang, 2011). My paper builds on this literature by showing that cultural factors that encourage corporate risk-taking are likely to increase corporate cash holding.

3. Data and research design

3.1. County-level religious and demographic characteristics

The first main data set contains U.S. county-level religious composition, which I retrieve from the “Churches and Church Membership” files from the American Religion Data Archive (ARDA). The data set is compiled by Glenmary Research Center and contains county-level statistics for more than 100 Judeo-Christian church bodies, including the number of churches and each church’s adherents in the county. From 1980 to now, the county-level religion data are available only for 1980, 1990, 2000, and 2010, I follow prior literature (e.g., Hilary and Hui, 2009; Kumar et al., 2011) to linearly interpolate the religion data to obtain the values in the intermediate years.

Three main gambling measures are constructed: (i) the proportion of Catholics among total population in the county (*CATH*); (ii) the proportion of Protestants among total population in the county (*PROT*); (iii) Catholics-to-Protestant ratio (*CPRATIO*). Additional county-level demographic characteristics are retrieved from the U.S. Census Bureau. Five county

characteristics are added as control variables in our analysis. They are total population in the county, education level of the county (the proportion of people above age 25 with a bachelor's degree or higher), the median level of household income in the county, the proportion of household with married couples, and minority population (the proportion of non-white residents).

3.2. Firm-level cash holding and other control variables

My second main data set contains firm-level cash holding, headquarter location, and other control variables from Compustat database. Compustat only reports firms' current state and county headquarter locations, which introduces noise if firms have relocated through time. I obtain historical information on firms' headquarters county from several sources. Before 1994 (when the SEC's EDGAR is not available), I obtain information from Compact Disclosure; from 1994 to 2007, headquarter information is from SEC's EDGAR website⁴; after 2007, such information is from the Compustat database, which provides information on a firm's current headquarter location.

Firm cash holding is measured as the ratio of cash plus marketable securities (che) to total book value of assets (at). Following the literature, control variables included are firm size, leverage, market-to-book ratio, tangibility, R&D intensity, cash flow, cash flow volatility in the past 12 quarters, profitability, dividend payment, net working capital, the share of foreign income, and the number of segments. My main sample contains 57,112 firm-year observations with non-missing values for all the variables from 1987 to 2010.

⁴ The data is from Bill MacDonald's website <http://www3.nd.edu/~mcdonald/10-K-Headers/10-K-Headers.html>.

Table 3.1 report the descriptive statistics. For each variable, I report its mean, median, 25th and 75th percentiles, and standard deviation. Measured at the firm-level, in an average county, the number of Catholics (Protestants) accounts for about 27.20% (19.00%) percent of the total population and the average Catholics-to-Protestants ratio is 1.43. Those numbers are close to Kumar et al. (2011) (Panel B of Table 1). The cash-to-assets ratio is 19.00%, close to 18.80% reported in Duchin (2010). Other variables are largely similar to prior literature.

[Insert Table 3.1 here]

4. Main results

4.1. Local gambling preference and cash holding

I examine the relation between gambling preference and corporate cash holding using the following econometric model based on prior literature (e.g., Fresard, 2010; Gao et al., 2013):

$$\text{Cash}_{i,t+1} = \alpha + \beta \text{Gambling}_{j,t} + \gamma X_{i,t} + \theta \text{Industry}_k + \varphi \text{Year}_t + \epsilon_{i,t}, \quad (6)$$

where *gambling* refers to the three gambling preference measure *CATH*, *PROT*, and *CPRATIO* for county *j* in year *t*. $X_{i,t}$ is firm- and county- level control variables introduced in Section 3. I also include 2-digit SIC industry and year fixed effects in the model.⁵ Standard errors are estimated by clustering by both county and year.

Table 3.2 reports the results of the baseline regression in model (6). I find that corporate cash holding is positively and significantly associated with *CATH*. In contrast, cash holding decreases with *PROT*, the Protestant concentration ratio.

⁵ Median county household income are not included as control variable since it is highly correlated with county education level. Robustness check shows that including income instead of income yields quantitatively similar results.

Economically, increasing *CPRATIO* from its 25th percentile to 75th percentile leads to 3.92 percentage points increase in corporate cash holding level, which amounts to 20.63% of the mean value of *Cash*.⁶ The positive relation between local gambling preference and corporate cash holding is consistent with the notion that gambling preference strengthens firms' precautionary demand for cash. Coefficients of other control variables are largely consistent with prior literature. For example, small firms, R&D intensive firms, and focused firms tend to hold more cash; in contrast, firms with more tangible assets and net working capital hold less cash.

[Insert Table 3.2 here]

4.2. Tests on endogeneity

Although I documented a strong positive relation between local gambling preference and corporate cash holding, the results are only cross-sectional and thus suffer from endogeneity concerns. The results obtained so far are unlikely to be driven by reverse causality since it is rather farfetched that cash holding as a firm policy is shaping local religious culture. However, one particular concern is omitted variable bias. If my cash model omits variables that affect both local gambling preference and corporate cash policy, then the results will be biased. In this section, I provide evidence to mitigate endogeneity concerns and further substantiate my results.

4.2.1 Tests on omitted variables

Prior literature finds that seniors tend to prefer stable cash flow for consumption purpose, and thus hold stocks that pay more dividend (Becker,

⁶ The estimated economic magnitude is larger if I use the natural log of cash ratio as dependent variable.

Ivkovic, and Weisbenner, 2011). Due to such preference, seniors are unlikely to be gambling-prone, and as an important class of local investors, such preference may affect corporate cash holding policy. It is possible that demographic variation in population age is correlated with the variation in the ratio of Catholics-to-Protestants across the U.S., and thus drives the results obtained.

To address this concern, I re-estimate model (6) adding the proportion of local seniors as an additional control variable, which is defined as the number of individuals aged 65 or older in a county divided by the total population of that county. Table 3.3 shows that used alone, the proportion of local senior carries a positive but insignificant sign. Once measures of local gambling preferences are controlled for, however, the effects of local population age are wiped out, and the main results remain unchanged. The results assure that the positive relation between gambling preference and corporate cash holding is not driven by the local population age.

[Insert Table 3.3 here]

4.2.3 Corporate headquarter moves and the instrumental variable

Since the county-level religious measure of gambling preference is persistent through time, I use corporate headquarter relocation as an alternative identification strategy to shock the level of gambling preference a firm exposed to. To ensure that firms are exposed to a different level of local gambling preference after relocation, I only include firms that have moves out of its current headquarter states. From 1987 to 2010, I identified 211 such firms. I further classify these 211 relocations into two groups. The first group includes 88 firms that move to other counties in a different state with a stronger religion-induced gambling preference (up group); the second group includes the rest. This

classification allows me to test the treatment effects and reverse treatment effects using a differences-in-differences (DID) approach.

For each firm in the two groups in the relocation year, I match it with another firm in the same SIC 2-digit industry with past five-year average cash ratio closest to that of the moving firm. In addition, the difference in the past average cash ratio between the two firms should be no larger than 10% of the cash ratio of the treatment firm. Table 3.4 presents the estimation results with the matched samples. In column (1), for firms that relocate to other counties with stronger gambling preference, the DID estimated coefficient shows that the cash ratio increases 2.9 percentage points afterwards, this amounts to a 16% increase from the sample mean. Furthermore, in column (2), I show that firms relocated to counties with a stronger gambling preference also invest more in R&D. For firms that relocate to less gambling-prone counties, column (3) shows that the DID coefficient is -2.2 percentage points. In column (4), we did not observe relocated firms significantly decrease their R&D investments, although the DID coefficient is negative.

In sum, when firms move to more gambling-prone counties, they experience an increase in corporate cash holding and R&D investments, while moving to less gambling-prone counties leads to a decrease in cash holding. The findings support the hypothesis that gambling preference increases cash holding by motivating corporate risk-taking. Moreover, the economic magnitude of the two groups is comparable to each other. These results support the argument that the influence of local gambling preference on corporate cash holding is causal.

[Insert Table 3.4 here]

Though I control for a long list of firm- and county-specific characteristics, it's still possible that some omitted variables or county-level heterogeneities are driving the relation between corporate cash holding and local gambling preference. In addition, firms with a higher level of cash holding may choose to locate in counties with a stronger gambling preference. I thus follow Chen et al. (2014) and run a two-stage least regression (2SLS) model by using the 10-year lagged value of $\text{Ln}(\text{CPRATIO})$ as an instrument for current $\text{Ln}(\text{CPRATIO})$. This approach is also adopted by Hilary and Hui (2009) and Kumar et al. (2011). The rationale is that past religious beliefs persist and shape the current attitudes towards gambling in the local communities. The results are reported in Table 3.5. Column (1) presents the first-stage regression. $\text{Ln}(\text{CPRATIO})_{t-10}$ is significantly and positively related to $\text{Ln}(\text{CPRATIO})$, which is unsurprising due to the slow-moving feature of cultural traits. In column (2), results from the second-stage show that the instrumented $\text{Ln}(\text{CPRATIO})$ is still a significant and positive predictor of firms' cash holding level.

To summarize, while it is difficult to exhaust the list of omitted variables and completely overcome endogeneity in my empirical design, I conduct a battery of tests to alleviate such concerns and find that my main conclusion still holds. Though each test may be subject to criticism, the totality of the evidence points to a causal relation between gambling preference and cash holding.

[Insert Table 3.5 here]

4.3. Gambling preference, corporate investment, and financial constraint

So far, the analysis shows that higher local gambling preference is associated with more cash holding. The precautionary saving hypothesis also

predicts that firms with exposure to stronger gambling preference hold more cash because they are likely to choose riskier projects that yield unpredictable cash flow. I empirically test this prediction in this subsection.

I examine how local gambling preference affect corporate investments in risky as well as safe and traditional projects. Understanding how gambling preference shapes resources spending can shed light on the underlying mechanisms that link gambling preference to cash holding and enables us to make statements about the optimality of this outcome.

4.3.1. Gambling preference and corporate risk-taking

Since we are interested in how gambling preference is related to future corporate decisions, $Ln(CPRATIO)$ and other control variables are entered as lagged values. Model (6) of Table 3.6 shows the results for R&D expenditure. Consistent with Chen et al. (2014), gambling-prone firms invest more in R&D-intensive projects,⁷ which have the probability of generating large positive payoffs. Interestingly, in column (2), stronger gambling preference predicts lower capital expenditure. One interpretation of the results is that gambling preference induces firms to invest more in risky and challenging projects. In contrast, it discourages investments in fixed assets, which are usually additions to traditional projects and unlikely to generate extremely high returns. Consistent with gambling preference motivating corporate risk-taking, in column (3), stronger gambling preference positively forecasts future cash flow volatility.

[Insert Table 3.6 here]

⁷ The difference in significance level between their results and mine is perhaps due to choice of the dimension of clustering. They estimate standard errors by clustering by firm, but I cluster by county and year, which yield a more conservative estimation of standard errors.

4.3.2. Gambling preference, financial constraint and cash holding

We have established the fact that gambling preference is associated with higher R&D investment. Prior literature indicates that R&D investment is hard to value and thus suffers from high degree of information asymmetry (Aboody and Lev, 2000). As a result, R&D investment is also hard to be pledged as collateral in the credit market, rendering gambling-prone firms facing higher costs when they seek external financing. The above reasoning implies that cash reserve in gambling-prone firms may help to mitigate financial constraints and underinvestment.

To test this conjecture, I interact the gambling preference measure with three measures of financial constraints. The three measures used to identify constrained firms are WW index (Whited and Wu, 2006),⁸ HP Index (Hadlock and Pierce, 2010), and credit rating⁹. I expect that the effect of gambling preference on cash holding will be stronger in firms that are identified as financially constrained.

Table 3.7 presents the results. In column (1) and (2), gambling preference increases cash holding only among firms are identified as financially constrained using the WW index and HP index. In column (3), we again observe that gambling-prone firms are only associated with higher cash holding when they do not have a credit rating. The evidence provides support for the notion that firms with stronger gambling preference use cash holding to mitigate financial constraints. The results in Table 3.7 also imply that gambling-prone firms are

⁸ I do not use KZ index of Kaplan and Zingales (1997) because prior studies pointed out that it performs relatively poorly to capture the cross-section variation of cash policy for financially constrained and unconstrained firms (Almeida, Campello, and Weisbach, 2004). Faulkender and Wang (2006), and Denis and Sibilkov (2010) also do not use KZ index for the same reason.

⁹ I follow Farre-Mensa and Ljungqvist (2016) in constructing the three measures of financial constraints.

unlikely to hoard cash for non-liquidity reasons (e.g., managerial agency problems).

[Insert Table 3.7 here]

4.4. *Gambling preference and the value of excessive cash holding*

In this subsection, I explicitly investigate whether cash holding is more or less valuable for firms exposed to stronger gambling preference. *A priori*, the interaction effect of gambling preference on the relation between cash holding and firm value is unclear. On the one hand, stronger gambling preference can induce investments in risky and innovative projects. Moreover, if firms in gambling-prone regions are subject to a higher degree of financial constraint, cash holding will add value by alleviating underinvestments in those firms (Denis and Sibilkov, 2009). On the other hand, gambling-prone firms may also waste valuable resources (e.g., spend cash recklessly) and thus lead to a lower marginal value of cash. Both of the effects may be at play, and the net effect is an empirical question.

I follow Dittmar and Mahrt-Smith (2007) and use the value regression akin to Fama and French (1998):

$$\begin{aligned} \frac{MV_{i,t}}{NA_{i,t}} = & \beta_0 + \beta_1 \text{Gambling}_{i,t} + \beta_2 \text{Cash surplus}_{i,t} \\ & + \beta_3 \text{Gambling}_{i,t} * \text{Cash surplus}_{i,t} \\ & + \text{Controls}_{i,t} + \text{Year}_t + \text{Firm}_i + \epsilon_{i,t} \end{aligned} \quad (7),$$

where $MV_{i,t}$ is market value, calculated as Price (Compustat data item 199) times Shares (25) plus total liabilities (181), $NV_{i,t}$ is net assets (item 6 minus item 1). Control variables are taken from Fama and French (1998), which include past

changes, future changes, and current level of earnings, R&D expenses, dividend, interest expenses, as well as past and future change in firm assets and future changes in market value. All of the control variables are scaled by current net assets.¹⁰ The model is estimated using firm fixed effects. $Cash\ surplus_{i,t}$ is cash at time t minus forecasted target level using the method from Dittmar and Mahrt-Smith (2007).¹¹

Consistent with Dittmar and Mahrt-Smith (2007), I estimate value regression (7) using only firms with cash surplus. Table 3.8 presents the results. I use three gambling measures to interact with the measure of cash surplus. All specifications show a positive interaction effect between gambling preference and the value of excessive cash. Focusing on the main gambling preference measure, $Ln(CPRATIO)$, the results indicate that value of excessive cash in gambling-prone firms is significantly larger. Moving from the first quartile to the third quartile of $Ln(CPRATIO)$, the marginal value impact of excessive cash increases almost 47% percent. These results are broadly consistent with Chen et al. (2014) documenting that gambling-prone firms are more R&D intensive, and generate superior innovation outputs in terms of both quantity and quality.

[Insert Table 3.8 here]

¹⁰ More specifically, the control variables are: $\frac{E_{i,t}}{NA_{i,t}}, \frac{dE_{i,t}}{NA_{i,t}}, \frac{dE_{i,t+2}}{NA_{i,t}}, \frac{RD_{i,t}}{NA_{i,t}}, \frac{dRD_{i,t}}{NA_{i,t}}, \frac{dRD_{i,t+2}}{NA_{i,t}}, \frac{D_{i,t}}{NA_{i,t}}, \frac{dD_{i,t}}{NA_{i,t}}, \frac{dD_{i,t+2}}{NA_{i,t}}, \frac{I_{i,t}}{NA_{i,t}}, \frac{dI_{i,t}}{NA_{i,t}}, \frac{dI_{i,t+2}}{NA_{i,t}}, \frac{dNA_{i,t}}{NA_{i,t}}, \frac{dNA_{i,t+2}}{NA_{i,t}}$, and $\frac{dMV_{i,t+2}}{NA_{i,t}}$. dX_t indicates change in X from time $t-2$ to t .

¹¹ Specifically, the estimation equation is from the Equation (4) in the Appendix of Dittmar and Mahrt-Smith (2007).

5. Additional analyses

5.1. Excessive cash and speed of adjustment

I argue that gambling preference affects corporate cash holding because more cash enables risk-taking and alleviate financial constraint. This argument implies that, when gambling-prone firms hold cash lower than their target level, they will adjust their cash holding more rapidly than other firms. In contrast, when their actual cash level is higher than the target level, they may be reluctant to disgorge cash.

I thus follow Gao, Harford, and Li (2013) to test the speed of adjustment (SOA) using the following partial adjustment model:

$$\begin{aligned}\Delta Cash = & \alpha + \beta_1 CPRATIO \times (Cash^* - Cash_{t-1}) \\ & + \beta_2 (Cash^* - Cash_{t-1}) + \beta_3 CPRATIO + \varepsilon\end{aligned}\quad (8)$$

where $\Delta Cash$ is the change in cash ratio, $Cash^*$ is the predicted cash ratio based on model (6) without gambling preference measures. $(Cash^* - Cash_{t-1})$ thus measures a firms deviation of cash holding from its target level. The coefficient β_2 captures SOA, and β_1 capture SOA conditional on a different level of gambling preference.

In Table 3.9, I first use the full sample to estimate model (8) and find that β_1 is insignificantly positive and close to zero, indicating that there is no meaningful change in SOA as we vary the degree of local gambling preference. However, as pointed by Gao et al., this full sample analysis ignores the potential heterogeneities in firms' SOA when their actual cash holdings fall below or above the target levels. I thus identify “excessive cash” and “cash shortfall” subsamples of firms with cash holding higher and lower than their targets, and

then re-estimate model (8) using the subsamples. “Excessive cash” and “cash short fall” firms are defined using the bottom quartile (p25) and the top quartile (p75) of $(Cash^* - Cash_{t-1})$ for the full sample.¹²

The results reported in Table 3.9 are consistent with the prediction. In column (2), conditional on holding excessive cash, the coefficient of $CPRATIO \times (Cash^* - Cash_{t-1})$ is significantly negative, indicating that firms located in gambling-prone regions are slower at adjusting their cash holdings down to target ratios. In contrast, in column (3), conditional on experiencing cash shortfall, those firms are much faster at replenishing their cash holdings.

[Insert Table 3.9 here]

5.2. Spurious relation

Another concern is that there might be a time-series trend in both cash holding and the measure of local gambling preference. This is very likely, since U.S. firms’ cash holding has exhibited a dramatic increasing trend in the past decades, and linear interpolation of the key independent variables may introduce artificial trends in these variables. I run a Fama-MacBeth style regression to address this concern, which is cross-sectional and unaffected by time-series trends. I also use only the census-year cross-sections in my sample (1990, 2000, and 2010) to re-estimate the main results. In column (1) A of Table 3.10, results from Fama-MacBeth style regression confirm the findings from OLS regressions. In column (2), the regression uses only census year cross-sections yields weaker results, perhaps due to the dramatic shrink in the sample size. Overall, the main

¹² Gao et al. (2013) argue that using the extremes of the distribution alleviate the concern of misclassification due to the fact that we do not know the true model of corporate cash holdings.

message remain qualitatively the same: the observed relation between gambling preference and cash holding is unlikely to be spurious.

[Insert Table 3.10 here]

5.3. General robustness checks

In this subsection, I conduct different tests to evaluate the robustness of the main results in Table 3.11. To ensure that my results are not driven by a small group of firms in certain areas, I perform the main regression using a subsample of regions that exclude firms in southwest and northeast states¹³, where CPRATIO are highest. The main results still hold. Second, diversified firms hold significantly less cash than focused firms (Duchin, 2010). To examine whether my gambling preference measures merely capture the geographic variation in corporate diversification, I drop all multi-segment firms in my sample and find that the main results are unaffected. In un-tabulated results, I confirm that there is no significant relation between firms' number of segments and their headquarter locations' gambling preference. Third, I only include firms with zero reported R&D in the regression and find that the effect of gambling preference on cash is not confined to firms with risky investments. In the last specification, I include firm fixed effects instead of industry fixed effects. This further helps to mitigate the concern that unobservable county or firm characteristics may bias the estimation. The results show that the positive relation between gambling preference and cash level holds even within firms.

[Insert Table 3.11 here]

¹³ The dropped states include CA, AZ, NV, UT, TX, ME, CO, NH, VT, NY, MS, NM, ID, and OR.

5.4. Using lottery sale to measure gambling preference

In addition to the religion-induced gambling measure, I also use the state-level per capita lottery sale as an alternative measure of gambling preference. While higher *CPRATIO* value strongly predicts higher state revenue from lottery sales (Kumar, Page, and Spalt, 2011), an advantage of lottery sale is that it is arguably a more direct measure of local gambling preference than the religion-induced measure, which could be related to other social values in a region in addition to gambling preference. Therefore, I retrieve state-level lottery income from the U.S. Census Bureau, and re-estimate model (6) using the per capita state-level lottery income (*Lottery per capita*, in thousand dollars) and the total state-level lottery income (*Ln (lottery)*) as key independent variables. The estimation results in columns (1) and (2) of Table 3.12 show that firms that headquarter in states where residents spend more on lottery hold significantly more cash. In terms of economic magnitude, a one standard deviation increase in *Lottery per capita* from its sample mean translates into a 4.1% increase in the average cash to asset ratio. These results again support the hypothesis that regional gambling preference increases local firms' cash holding. In column (3) and (4), after controlling for the religion-induced gambling preference measure, the coefficients of state-level lottery income become much smaller and are no longer significant. This is consistent with the finding in Kumar et al. (2011) that lottery sales are strongly influenced by local religion.

[Insert Table 3.12 here]

6. Conclusion

Using a religion-induced measure of local gambling preference, this study examines how locally-held attitudes towards gambling affect an important

corporate hedging policy: cash holding. I document that firms headquartered in counties with a more favorable attitude towards gambling hold significantly more cash. Further analysis shows that local gambling preference increases corporate cash holding because cash facilitates risk-taking and mitigate financial stress. Lastly, I find that the value of excessive cash is higher in firms with stronger gambling preference. The results are robust to a wide range of model specifications.

One caveat is that this study focuses on the geographic variation in cash holding induced by gambling preference. Thus, it cannot explain the dramatic increase in U.S. firm cash holding over the past several decades. Nonetheless, this study reveals that gambling preference – by maintaining high uncertainty and tail risk – may lead to a higher demand for cash slack. Thus, my investigation confirms the precautionary saving motive as the dominant way to understand cash holding (Opler et al., 1999; Duchin, 2010) and also adds to the growing literature on how culture shapes financial market outcomes.

The paper that closest to mine is Hu, Lian, and Zhou (2017). They suggest that local Protestant values are characterized by higher ethical standards and mitigate the free cash flow problem, and thus decrease corporate cash holding. They do not comment on why Catholic values, which differ to Protestant values in some regards, do not play a role in their hypothesis. Moreover, they are unable to explain why Catholic values are associated to a higher value of excessive cash holding because according to their results, Catholic values should cause, rather than mitigate, free cash flow problem.

Most importantly, I provide a completely different interpretation of the results from theirs. They conclude that their results are inconsistent with precautionary saving motive, according to which Protestant values should increase cash holding due to strong risk aversion. My analysis suggests that Protestant values decrease cash holding because they also decrease risk-taking, and consequently, the need for cash as an internal funding source. Such a conclusion fits well into the framework of Gatchev, Pulvino, and Tarhan (2010) and Bolton, Chen, and Wang (2011), where firms' investment, financing, and risk management decisions are interdependent.

This paper also leaves several questions unanswered. First, do gambling preferences affect other aspects of hedging and risk management? For example, the use of derivative hedging. Second, since gambling-prone firms may have a stronger preference for skewed payoffs, does it exacerbate the conflict of interest between shareholders and creditors? Such questions remain to be explored by future research.

Appendix: Variable Definitions

Firm-level variables

- Cash:* Cash and marketable securities/Book value of total assets
- Size:* Firms size, which is the natural logarithm of the book value of total assets
- Lev:* Leverage ratio, computed as (debt in current liabilities + long-term debt)/book assets
- MB:* Market-to-book ratio, computed as (book assets +market value of common equity-book value of common equity)/book assets
- TANG:* Asset tangibility, computed as net value of property, plants, and equipment over book assets
- R_D:* Research and development expenditure over book assets
- Dividend:* A dummy variable that equals one if a firm pays cash dividend in one year
- NWC:* Net working capital, computed as (current assets-current liabilities-cash)/book assets
- CF:* Cash flow/assets, which is (income before extraordinary items + depreciation and amortization)/book assets
- CFVOL:* Past three-year quarterly cash flow standard deviation/book assets
- ForeignInc:* A dummy variable that equals one if a firms' pretax foreign income accounts for more than 20% of its pretax net income
- Seg_N:* Natural log Number of business segments reported by the firm

County-level variables

- CATH:* The proportion of Catholics among a county's total population
- PROT:* The proportion of Protestants among a county's total population
- Ln (CPRATIO):* The natural logarithm of (Catholics/Protestants)

- REL:* The total number of religious adherents over total population
- Ln (POP):* The natural logarithm of total population in a county
- EDU:* The proportion of county population above age 25 that has completed a bachelor's degree or higher
- Minority:* The proportion of non-white residents in a county
- Married:* The proportion of households with a married couple in a county

References

- Aboody, D., & Lev, B. (2000). Information asymmetry, R&D, and insider gains. *The Journal of Finance*, 55(6), 2747-2766.
- Almeida, H., Campello, M., Weisbach, M.S., 2004. The cash flow sensitivity of cash. *The Journal of Finance* 59, 1777-1804.
- Bates, T.W., Kahle, K.M., Stulz, R.M., 2009. Why do US firms hold so much more cash than they used to? *The Journal of Finance* 64, 1985-2021.
- Becker, B., Ivković, Z., Weisbenner, S., 2011. Local dividend clienteles. *The Journal of Finance* 66, 655-683.
- Bernile, G., Bhagwat, V., Rau, P.R., 2017. What doesn't kill you will only make you more risk-loving: Early-life disasters and CEO behavior. *The Journal of Finance* 72, 167-206.
- Berry, F.S., Berry, W.D., 1990. State lottery adoptions as policy innovations: An event history analysis. *American Political Science Review* 84, 395-415.
- Bolton, P., Chen, H., Wang, N., 2011. A unified theory of Tobin's q, corporate investment, financing, and risk management. *The Journal of Finance* 66, 1545-1578.
- Byun, S.-J., Kim, D.-H., 2016. Gambling preference and individual equity option returns. *Journal of Financial Economics* 122, 155-174.
- Campello, M., Graham, J.R., Harvey, C.R., 2010. The real effects of financial constraints: Evidence from a financial crisis. *Journal of Financial Economics* 97, 470-487.
- Chen, Y., Podolski, E.J., Rhee, S.G., Veeraraghavan, M., 2014. Local gambling preferences and corporate innovative success. *Journal of Financial and Quantitative Analysis* 49, 77-106.
- Denis, D.J., Sibilkov, V., 2009. Financial constraints, investment, and the value of cash holdings. *Review of Financial Studies* 23, 247-269.
- Dittmar, A., Mahrt-Smith, J., 2007. Corporate governance and the value of cash holdings. *Journal of Financial Economics* 83, 599-634.
- Dittmar, A., Mahrt-Smith, J., Servaes, H., 2003. International corporate governance and corporate cash holdings. *Journal of Financial and Quantitative analysis* 38, 111-133.

- Doran, J.S., Jiang, D., Peterson, D.R., 2011. Gambling preference and the new year effect of assets with lottery features. *Review of Finance* 16, 685-731.
- Duchin, R., 2010. Cash holdings and corporate diversification. *The Journal of Finance* 65, 955-992.
- Dudley, E., Zhang, N., 2016. Trust and corporate cash holdings. *Journal of Corporate Finance*, 41, 363-387.
- Ellison, C.G., Nybrotten, K.A., 1999. Conservative protestantism and opposition to state-sponsored lotteries: Evidence from the 1997 Texas poll. *Social Science Quarterly*, 356-369.
- Fama, E.F., French, K.R., 1998. Taxes, financing decisions, and firm value. *The Journal of Finance* 53, 819-843.
- Faulkender, M., Wang, R., 2006. Corporate financial policy and the value of cash. *The Journal of Finance* 61, 1957-1990.
- Frésard, L., 2010. Financial strength and product market behavior: The real effects of corporate cash holdings. *The Journal of Finance* 65, 1097-1122.
- Gao, H., Harford, J., Li, K., 2013. Determinants of corporate cash policy: Insights from private firms. *Journal of Financial Economics* 109, 623-639.
- Garrett, T.A., Sobel, R.S., 1999. Gamblers favor skewness, not risk: Further evidence from United States' lottery games. *Economics Letters* 63, 85-90.
- Gatchev, V.A., Pulvino, T., Tarhan, V., 2010. The interdependent and intertemporal nature of financial decisions: An application to cash flow sensitivities. *The Journal of Finance* 65, 725-763.
- Guiso, L., Sapienza, P., Zingales, L., 2006. Does culture affect economic outcomes? *The Journal of Economic Perspectives* 20, 23-48.
- Harford, J., 1999. Corporate cash reserves and acquisitions. *The Journal of Finance* 54, 1969-1997.
- Hilary, G., Hui, K.W., 2009. Does religion matter in corporate decision making in America? *Journal of Financial Economics* 93, 455-473.
- Hu, H., Lian, Y., Zhou, W., 2017. Do Local Protestant Values Affect Corporate Cash Holdings? *Journal of Business Ethics*.
- Hsu, P. H., Li, F., Lin, T. C., 2017. Cash Holdings and Innovative Efficiency: The International Evidence. Available at SSRN: <https://ssrn.com/abstract=2643536>

- Keynes, J.M., 1936. The general theory of employment, interest and money. London, 1936.
- Kumar, A., 2009. Who gambles in the stock market? *The Journal of Finance* 64, 1889-1933.
- Kumar, A., Page, J.K., Spalt, O.G., 2011. Religious beliefs, gambling attitudes, and financial market outcomes. *Journal of Financial Economics* 102, 671-708.
- Kumar, A., Page, J.K., Spalt, O.G., 2016. Gambling and comovement. *Journal of Financial and Quantitative Analysis* 51, 85-111.
- Kraus, A., & Litzenberger, R. H. (1976). Skewness preference and the valuation of risk assets. *The Journal of Finance*, 31(4), 1085-1100.
- Opler, T., Pinkowitz, L., Stulz, R., Williamson, R., 1999. The determinants and implications of corporate cash holdings. *Journal of Financial Economics* 52, 3-46.
- Pirinsky, C., Wang, Q., 2006. Does corporate headquarters location matter for stock returns? *The Journal of Finance* 61, 1991-2015.
- Shu, T., Sulaeman, J., Yeung, P.E., 2012. Local religious beliefs and mutual fund risk-taking behaviors. *Management Science* 58, 1779-1796.
- Thaler, R.H., Ziemba, W.T., 1988. Anomalies: Parimutuel betting markets: Racetracks and lotteries. *The Journal of Economic Perspectives* 2, 161-174.
- Tirole, J., 2010. The theory of corporate finance. Princeton University Press, Princeton, NJ.
- Ucar, E., 2015. Local Culture and Dividends. *Financial Management* 40, 105-140.
- Welte, J.W., Barnes, G.M., Wieczorek, W.F., Tidwell, M.-C., 2004. Gambling participation and pathology in the United States—A sociodemographic analysis using classification trees. *Addictive Behaviors* 29, 983-989.

Figure 3.2: State-level dispersion of local gambling preference.

This figure presents a graphic illustration of the state-level variation in the dependent variable and key explanatory variable. I calculate the Catholic-to-Protestant ratio at the state-level in each year. A darker color indicates a higher concentration of Protestants relative to Catholics.

Table 3.1: Summary statistics

The sample includes all Compustat firm-years with non-missing value for all the variables for firms incorporated in the U.S. Financial firms (SIC codes 6000-6999) and utilities (SIC codes 4900-4999) are excluded from the sample. Variable definitions are provided in the Appendix.

<i>Panel A: County demographic characteristics</i>						
Variable	N	Mean	25 th Pctl	Median	75 th Pctl	Std dev
<i>CATH</i>	57,112	0.272	0.166	0.257	0.375	0.142
<i>PROT</i>	57,112	0.190	0.100	0.150	0.264	0.119
<i>Ln</i> (<i>CPRATIO</i>)	57,112	0.357	-0.379	0.601	1.211	1.121
<i>REL</i>	57,112	0.540	0.438	0.546	0.622	0.123
<i>Ln (POP)</i>	57,112	13.692	13.151	13.721	14.270	1.107
<i>EDU</i>	57,112	0.308	0.246	0.291	0.373	0.094
<i>Minority</i>	57,112	0.320	0.168	0.293	0.471	0.181
<i>Married</i>	57,112	0.754	0.716	0.770	0.798	0.068
<i>Panel B: Firm-level characteristics</i>						
<i>Cash</i>	57,112	0.190	0.026	0.096	0.278	0.222
<i>Size</i>	57,112	4.920	3.457	4.826	6.333	2.055
<i>LEV</i>	57,112	0.196	0.018	0.163	0.320	0.183
<i>MB</i>	57,112	1.988	1.058	1.428	2.195	1.716
<i>TANG</i>	57,112	0.270	0.095	0.207	0.385	0.222
<i>R_D</i>	57,112	0.054	0.000	0.002	0.065	0.105
<i>Dividend</i>	57,112	0.328	0.000	0.000	1.000	0.469
<i>NWC</i>	57,112	0.313	0.145	0.303	0.463	0.209
<i>CF</i>	57,112	0.012	0.003	0.068	0.114	0.219
<i>CFVOL</i>	57,112	0.043	0.009	0.018	0.042	0.075
<i>ForeignInc</i>	57,112	0.157	0.000	0.000	0.000	0.364
<i>Seg_N</i>	57,112	0.278	0.000	0.000	0.693	0.483

Table 3.2: Gambling preference and corporate cash holding

This table presents the estimates of corporate cash holding on three measures of gambling preference. The dependent variable is one-year ahead Cash and marketable securities divided by book assets. Variable definitions are provided in the Appendix. All continuous variables, except county-level demographic characteristics, are winsorized at the 1st and 99th percentile. Standard errors are clustered by county and year. The *t*-statistics are reported in parentheses underneath the coefficients. Superscripts ***, **, and * correspond to statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)
<i>CATH</i>	0.065*** (3.807)		
<i>PROT</i>		-0.068*** (-3.914)	
<i>Ln (CPRATIO)</i>			0.008*** (4.372)
<i>Size</i>	-0.007*** (-5.129)	-0.007*** (-5.027)	-0.007*** (-5.046)
<i>LEV</i>	-0.328*** (-24.517)	-0.328*** (-24.446)	-0.328*** (-24.444)
<i>MB</i>	0.009*** (9.427)	0.009*** (9.404)	0.009*** (9.423)
<i>TANG</i>	-0.317*** (-18.276)	-0.317*** (-18.208)	-0.317*** (-18.264)
<i>Profitability</i>	-0.099*** (-6.163)	-0.099*** (-6.140)	-0.099*** (-6.171)
<i>R_D</i>	0.441*** (14.889)	0.443*** (14.846)	0.442*** (14.840)
<i>Dividend</i>	-0.006 (-1.585)	-0.005 (-1.563)	-0.005 (-1.514)
<i>NWC</i>	-0.415*** (-21.898)	-0.414*** (-21.827)	-0.414*** (-21.910)
<i>CF</i>	0.085*** (4.971)	0.085*** (4.994)	0.085*** (4.991)
<i>CFVOL</i>	-0.044** (-2.090)	-0.045** (-2.160)	-0.045** (-2.143)
<i>ForeignInc</i>	-0.015*** (-4.468)	-0.015*** (-4.536)	-0.015*** (-4.566)
<i>Seg_N</i>	-0.030*** (-7.330)	-0.030*** (-7.313)	-0.030*** (-7.324)
<i>Ln (POP)</i>	-0.003 (-1.356)	-0.004 (-1.406)	-0.004 (-1.600)
<i>REL</i>	-0.049** (-2.439)	0.003 (0.158)	-0.024 (-1.295)
<i>EDU</i>	0.064** (2.418)	0.054* (1.946)	0.056** (2.087)
<i>Minority</i>	0.082** (2.394)	0.079** (2.299)	0.078** (2.319)
<i>Married</i>	0.192*** (3.308)	0.205*** (3.429)	0.195*** (3.387)
Industry & Year FE	Y	Y	Y
No. of obs	57,112	57,112	57,112
Adj. R-square	0.656	0.656	0.656

Table 3.3: Role of local seniors

This table presents the analysis of the alternative interpretation that the local population age drives the main results. The dependent variable is the one-year ahead ratio of cash and marketable securities to total assets. *Local senior* is defined as the number of individuals aged 65 or older in a county divided by the total population of that county. The *t*-statistics are reported in parentheses beneath the coefficients, with standard errors clustered by county and year. Firm- and county-level characteristics are controlled but not reported. Superscripts ***, **, and * correspond to statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)
<i>Local Senior</i>	0.136 (1.563)	0.058 (0.734)	0.049 (0.592)	0.034 (0.444)
<i>CATH</i>		0.061*** (3.242)		
<i>PROT</i>			-0.064*** (-3.481)	
<i>Ln (CPRATIO)</i>				0.008*** (3.942)
Firm & County controls	Y	Y	Y	Y
Year & Industry FE	Y	Y	Y	Y
No. of obs	57,112	57,112	57,112	57,112
Adj. R-square	0.579	0.579	0.579	0.580

Table 3.4: Headquarter relocation

This table displays the estimation results of how shocks to firms' local gambling preference affect their cash holding. In Panel A (B), the sample consists of 88 (123) firms that relocate their headquarters to counties in other states with stronger (weaker) religion-induced gambling preference, and the control firms in the same industry with a 5-year average cash ratio closest to that of relocated firms. *Relocate* is a dummy variable that equals one for relocated firms, and zero otherwise. *Post* is also a dummy indicating years after relocation for treated firms and their respective control firms. All regressions control for 2-digit SIC code industry and year fixed effects. Firm- and county-level characteristics are controlled but not reported. The *t*-statistics are reported in parentheses beneath the coefficients, with standard errors clustered by county. Superscripts ***, **, and * correspond to statistical significance at the 1%, 5%, and 10% levels, respectively.

	Panel A		Panel B	
	(1)	(2)	(3)	(4)
	Cash _{<i>t+1</i>}	R&D _{<i>t+1</i>}	Cash _{<i>t+1</i>}	R&D _{<i>t+1</i>}
<i>Relocate</i>	-0.043*** (-4.812)	-0.0121** (-2.76)	-0.006 (-0.899)	0.001 (0.30)
<i>Post</i>	-0.038*** (-4.175)	-0.006 (-1.27)	-0.006 (-0.723)	0.004 (1.00)
<i>Relocate*Post</i>	0.030*** (2.867)	0.019*** (3.19)	-0.022*** (-3.762)	-0.001 (-0.19)
Firm & County Controls	Y	Y	Y	Y
Year & Industry FE	Y	Y	Y	Y
No. of obs	2,369	2,369	3,286	3,286
Adj. R-square	0.649	0.582	0.657	0.633

Table 3.5: Instrumental variable approach

This table presents the estimation results of instrumenting the gambling preference measure using its historical value. The sample consists of all firm-year observations that have non-missing 10-year lagged value of Ln (CPRATIO) and all other variables from 1987 to 2010. Ln (CPRATIO) is the natural log of the ratio of Catholic adherents to Protestant adherents. *Cash* is the ratio of cash and marketable securities over total assets. Column (1) reports the estimates of the 1st stage regression, and column (2) the 2nd stage. The *t*-statistics in parentheses are calculated from the Huber/White/Sandwich heteroskedastic-consistent errors, which are also clustered by county and year. The symbols ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)
	Ln (CPRATIO _{<i>t</i>})	Cash _{<i>t+1</i>}
<i>Instrumented Ln (CPRATIO_{<i>t</i>})</i>		0.008*** (3.174)
<i>Ln (CPRATIO_{<i>t-10</i>})</i>	0.729*** (24.261)	
<i>Size</i>	-0.005 (-0.872)	-0.010*** (-6.523)
<i>LEV</i>	-0.118** (-2.227)	-0.318*** (-20.988)
<i>MB</i>	0.001 (0.379)	0.011*** (7.063)
<i>TANG</i>	0.004 (0.089)	-0.317*** (-17.530)
<i>Profitability</i>	-0.060 (-0.818)	-0.096*** (-4.516)
<i>R_D</i>	0.158 (1.135)	0.413*** (9.948)
<i>Dividend</i>	-0.040*** (-2.604)	-0.002 (-0.534)
<i>NWC</i>	-0.083* (-1.658)	-0.396*** (-20.652)
<i>CF</i>	0.007 (0.127)	0.078*** (3.474)
<i>CFVOL</i>	0.056 (0.629)	-0.068*** (-3.141)
<i>ForeignInc</i>	0.047*** (2.611)	-0.010*** (-2.970)
<i>Seg_N</i>	-0.012 (-0.887)	-0.024*** (-6.237)
<i>Ln (POP)</i>	0.150*** (4.319)	-0.002 (-0.696)
<i>REL</i>	0.959*** (4.636)	-0.021 (-1.035)
<i>EDU</i>	0.238 (0.751)	0.027 (0.972)
<i>Minority</i>	-0.223 (-0.906)	0.070** (2.112)
<i>Married</i>	0.577 (1.106)	0.192*** (3.187)
Year & Industry dummies	Y	Y
No. of obs	28,328	28,328
Adj. R-square	N/A	0.552

Table 3.6: Gambling preference and corporate risk-taking

This table shows the estimation results on how gambling preference affects corporate risk-taking. In column (1), *R&D* is one-year ahead R&D expenses scaled by the book value of total assets, missing R&D expenditure are replaced by zero; in column (2), *CAPX* is one-year ahead capital expenditure scaled by total assets. In column (3), the dependent variable *Cash flow risk* is the volatility of quarterly cash flow volatility over total assets. The Standard errors are clustered by both county and year. Firm- and county-level characteristics are controlled but unreported. The *t*-statistics based on robust standard errors are reported in parentheses. Superscripts ***, **, and * correspond to statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)
	R&D _{t+1}	CAPX _{t+1}	Cash flow risk _{t+1}
<i>Ln (CPRATIO)</i>	-0.001 (-0.992)	-0.001 (-0.921)	0.021** (2.451)
<i>Ln (CPRATIO)*Cash</i>	0.016** (2.262)	-0.004** (-2.565)	N/A
<i>Cash</i>	0.103*** (12.057)	0.023*** (5.827)	N/A
Firm & County controls	Y	Y	Y
Industry & Year FE	Y	Y	Y
No. of obs	54,782	54,782	54,782
Adj. R-square	0.503	0.410	0.432

Table 3.7: Gambling preference and financial constraints

The table presents the estimation results on how financial constraints interact with gambling preference to affect firms' cash holding. The dependent variable is the one-year ahead ratio of Cash and marketable securities to book assets. Measures of financial constraints are the *WW index*, *HP index*, and *Credit rating* dummy in columns (1), (2), and (3), respectively. All continuous variables, except county-level demographic characteristics, are winsorized at the 1st and 99th percentile. The standard errors are clustered by both county and year. Firm- and county-level characteristics are controlled but not reported. The *t*-statistics are reported in parentheses underneath the coefficients. Superscripts ***, **, and * correspond to statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)
<i>Ln (CPRATIO)</i>	0.002 (0.793)	0.001 (0.390)	0.011*** (4.571)
<i>WW index</i>	-0.004 (-0.557)		
<i>Ln (CPRATIO)*WW index</i>	0.016*** (3.242)		
<i>HP index</i>		-0.004 (-0.508)	
<i>Ln (CPRATIO)*HP index</i>		0.014*** (3.135)	
<i>Credit rating</i>			0.001 (0.233)
<i>Ln (CPRATIO)*Credit rating</i>			-0.011** (-2.779)
Firm & County controls	Y	Y	Y
Industry & Year FE	Y	Y	Y
No. of obs	27,287	32,462	52,129
Adj. R-square	0.656	0.659	0.665

Table 3.8: Gambling preference and the value of cash holding

This table shows the results for the value regressions. Dependent variables in all models are the ratio of firms' market value to assets. The independent variables include Cash surplus, which is excessive cash estimated as Dittmar and Mahart-Smith (2007), and its interactions with three measures of local gambling preference. Control variables (unreported) include two-year lagged change, the 2-year future change, and the current realization of the following variables over assets: Earnings, Assets (current realization excluded), R&D, Interest Expenses, Dividends, and Market Value (Only future change). All models are estimated with firm and year fixed effects. All ratios are winsorized at the 1st and 99th percentiles. Only observations with positive Cash surplus are included. The *t*-statistics are reported in parentheses beneath the coefficients, with standard errors clustered by county and year. Superscripts ***, **, and * correspond to statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)
<i>Cash surplus</i>	5.404*** (4.327)	11.406*** (8.046)	6.760*** (9.208)
<i>CATH</i>	1.079 (0.591)		
<i>Cash surplus*CATH</i>	9.044* (1.899)		
<i>PROT</i>		-0.091 (-0.030)	
<i>Cash surplus*PROT</i>		-20.993*** (-4.360)	
<i>Ln CPRATIO</i>			0.314 (0.957)
<i>Cash surplus*Ln (CPRATIO)</i>			2.061*** (3.547)
Controls from Fama and French (1998)	Y	Y	Y
Year & Firm FE	Y	Y	Y
No. of obs	17,154	17,154	17,154
Adj. R-square	0.422	0.423	0.424

Table 3.9: Speed of Adjustment to target cash level

The table shows the estimation results on how gambling preference affects firms adjusting their cash level when they hold more or less cash than predicted. The dependent variable is the change in cash ratio from the previous year to the current year. $Cash^*$ is the predicted cash ratio using model (6). In column (2) the samples consist of the bottom quartile (P25) and top quartile (P75) of the sample firms by the measure of $(Cash^* - Cash_{t-1})$. The standard errors are clustered by county. All ratios are winsorized at the 1st and 99th percentiles. The t -statistics are reported in parentheses underneath the coefficients. Superscripts ***, **, and * correspond to statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)
	Full sample	Firms with excessive cash: $(Cash^* - Cash_{t-1}) \leq P(25)$	Firms with cash shortfall: $(Cash^* - Cash_{t-1}) \geq P(75)$
	$\Delta Cash$	$\Delta Cash$	$\Delta Cash$
$Ln(CPRATIO)$	0.001** (2.025)	-0.007** (-2.374)	-0.006** (-2.373)
$(Cash^* - Cash_{t-1})$	0.381*** (47.752)	0.629*** (24.540)	0.395*** (17.260)
$Ln(CPRATIO) \times (Cash^* - Cash_{t-1})$	-0.001 (-0.135)	-0.076*** (-3.729)	0.055*** (3.090)
No. of obs	56,469	14,118	14,118
Adj. R-square	0.252	0.197	0.092

Table 3.10: Tests of spurious relations

This table shows the results of testing whether corporate cash holding and local gambling preference are spuriously correlated due to trends in the two variables. The dependent variable is the one-year ahead ratio of cash and marketable securities to total assets. Panel A contains the results of Fama-MacBeth style regression: for each year, a cross-sectional regression is performed using model (6), and the final coefficients are obtained as the average of the first step coefficients. The t -statistics are based on heteroscedasticity and autocorrelation consistent Newey-West (1987) standard error estimates with a lag length of 1 periods. In Panel B, the sample consists of the cross-sections of three census years, and standard errors are clustered at both county and year dimensions. All continuous variables, except county-level demographic characteristics, are winsorized at the 1st and 99th percentile. Firm- and county-level characteristics are controlled but not reported. The t -statistics are reported in parentheses beneath the coefficients. Both models include the full set of control variables, as well as industry dummies. Superscripts ***, **, and * correspond to statistical significance at the 1%, 5%, and 10% levels, respectively.

	Fama-MacBeth Style Regression	Census-Year Only (1990, 2000, 2010)
	(1)	(2)
<i>Ln (CPRATIO)</i>	0.009*** (6.348)	0.005** (2.242)
Firm & County controls	Y	Y
Year & Industry FE	Y	Y
No. of obs	57,112	6,804
Avg. R-square	0.416	N/A
Adj. R-square	N/A	0.572

Table 3.11: Robustness checks

This table presents the robustness checks of the main results. The dependent variable is the one-year ahead ratio of cash and marketable securities to total assets. Variable definitions are provided in the Appendix. In column (1), firms headquartered in the southwest and northeast are excluded; Multi-segments firms are excluded in column (2); in column (3), only firms with zero reported research and development expenses are included; column (4) include firm fixed effects. All continuous variables, except county-level demographic characteristics, are winsorized at the 1st and 99th percentile. Firm- and county-level characteristics are controlled but not reported. Standard errors are clustered by county and year. The *t*-statistics are reported in parentheses underneath the coefficients. Superscripts ***, **, and * correspond to statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)
<i>Ln (CPRATIO)</i>	0.006*** (3.228)	0.009*** (3.934)	0.005*** (2.971)	0.006** (2.094)
Firm & County controls	Y	Y	Y	Y
Year & Industry FE	Y	Y	Y	Y
No. of obs	31,460	41,456	28,007	56,115
Adj. R-square	0.645	0.680	0.475	0.774

Table 3.12: Measuring gambling preference using lottery income

This table reports the regression results of using state-level lottery income as an alternative measure of local gambling preference. The sample includes nonfinancial and nonutility firms in Compustat firms from 1996 to 2010. *Lottery per capita* is the per capita state-level annual income from sales of lottery tickets, and *Ln (lottery)* is the natural logarithm of the state-level annual income from lottery sales. Firm- and county-level characteristics are controlled but not reported. The definitions of other variables can be found in the appendix. Robust *t*-statistics based on standard errors clustered at both county and year level are in parentheses. The symbols ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	Cash _{<i>t+1</i>}			
<i>Lottery per capita</i>	0.055** (2.419)		0.033 (1.257)	
<i>Ln (lottery)</i>		0.005* (1.852)		0.003 (1.002)
<i>Ln (CPRATIO)</i>			0.008*** (3.392)	0.0101*** (3.819)
Firm & County controls	Y	Y	Y	Y
Year & Industry dummies	Y	Y	Y	Y
No. of obs	37,421	37,421	37,421	37,421
Adj. R-square	0.571	0.461	0.572	0.463