



Dissertation zur Erlangung des akademischen Grades

eines Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

an der Universität Passau

Three Essays on Price Discovery, Stock Liquidity, and Crash

Risk

eingereicht von

Marco Seruset, M. Sc.

September 2022

Disputation am:

21.12.2022

Erstgutachter: Prof. Dr. Oliver Entrop
Lehrstuhl für BWL mit Schwerpunkt Finance und Banking
Universität Passau

Zweitgutachter: Prof. Dr. Niklas Wagner
Lehrstuhl für BWL mit Schwerpunkt Finanzcontrolling
Universität Passau

Vorsitz: Prof. Dr. Dirk Totzek
Lehrstuhl für BWL mit Schwerpunkt Marketing und Services
Universität Passau

Danksagung

An erster Stelle möchte ich Herrn Prof. Dr. Oliver Entrop, dem Betreuer dieser Doktorarbeit und Koautor von zwei Beiträgen, für seine kontinuierliche Unterstützung und Ermutigungen in zahlreichen Diskussionen danken. Nicht nur von seiner höchsten fachlichen Expertise konnte ich während meiner gesamten Lehrstuhlzeit enorm profitieren, auch seine ruhige und besonnene Art trugen zu einer äußerst angenehmen Arbeitsatmosphäre bei.

Auch danke ich meinen beiden Koautoren, Herrn Prof. Dr. Bart Frijns sowie Herrn PD Dr. Martin Rohleder, für die inhaltliche Unterstützung. Insbesondere der konstruktive Austausch und die Gespräche auf fachlicher und persönlicher Ebene waren eine große Hilfe für mich.

Ein ganz besonderer Dank geht außerdem an alle aktuellen und ehemaligen Mitglieder des Lehrstuhls, insbesondere an Christa Listl, Carl-Friederich Grösbrink sowie Dr. Fabian Fuchs. Gesondert möchte ich hierbei nochmal Christa Listl hervorheben, die mir jederzeit mit ihrem unermüdlichen Einsatz in jeglichen Belangen und ihrer Hilfsbereitschaft den Rücken freigehalten hat. Ihre lebensfrohe Art und uneingeschränkte Unterstützung haben mich stets positiv beeinflusst und ermutigt. Auch danke ich meinem Weggefährten, Carl-Friederich Grösbrink, für die mühevollen Arbeit des Korrekturlesens und die Zusammenarbeit während unserer gemeinsamen Zeit am Lehrstuhl. Ein ganz besonderer Dank gilt auch meiner „Steuerkollegin“, Mia Kamper. Ihre intensiven Lektoratstätigkeiten und hilfreichen Kommentare konnten die Qualität der vorliegenden Dissertation deutlich heben.

Mein innigster Dank gilt schließlich meinen Eltern Monika & Wolfgang, die mir meinen bisherigen Lebensweg ermöglicht und mich in jeder Situation bedingungslos und mit größter Hingabe unterstützt haben. Ebenso danke ich meiner Schwester, Ramona, von ganzem Herzen für die aufmerksame, liebevolle und vielseitige Unterstützung während meines gesamten

Lebens. Auch meinem Schwager und „großem Bruder“, Philip, möchte ich an dieser Stelle für die zahlreichen gemeinsamen Erlebnisse danken.

Nicht zuletzt bedanke ich bei meinen allerbesten Freunden, „den Udos“, für all die schönen Ausflüge, Feiern und sonstigen Ablenkungen, die mir stets einen Lichtschimmer gegeben haben. Auch die Passauer Clique, Mia, Sina und Carl, war stets ein großer Rückhalt für mich, der als wichtiger Teil zum Erfolg meiner Dissertation beigetragen hat.

Table of Contents

Einleitung	1
1 The Determinants of Price Discovery on Bitcoin Markets	22
1 Introduction	23
2 Data	26
3 Price discovery	31
3.1 Vector error-correction model and price discovery measures	31
3.2 Empirical analysis	33
4 Determinants of price discovery	36
4.1 Potential determinants and summary statistics	36
4.2 Empirical strategy	40
4.3 Empirical results	41
4.3.1 Market quality and uncertainty	41
4.3.2 Adding investor sentiment, investor attention and macroeconomic news	44
5 Conclusion	45
Appendix	48
Appendix A: Component Share (CS) measure	48
Appendix B: Information Share (IS) measure	49
References	53
Figures and Tables	59
2 Local Religiosity and Stock Liquidity	69
1 Introduction	70
2 Hypothesis development	74

3	Data and summary statistics	78
3.1	Firm level sample.....	78
3.2	County level variables.....	79
3.3	Dependent variable - bid-ask spread.....	80
3.4	Firm controls	81
3.5	Descriptive statistics	84
4	The relation between religiosity and liquidity.....	86
4.1	Baseline regression results	86
4.2	Additional control variables.....	89
4.3	Robustness tests	91
4.3.1	Different variable and model specifications	91
4.3.2	Alternative measures of religiosity	94
4.4	Religious subgroups.....	94
4.5	Endogeneity	95
4.5.1	Reverse causality and instrumental variable analysis	96
4.5.2	Placebo test.....	97
4.5.3	Entropy balancing and propensity score matching	98
4.5.4	Company relocation	100
5	Information environment and religiosity.....	103
6	Other aspects of liquidity and the trading environment of a firm	105
7	Local religiosity and firm value	106
8	Conclusion.....	107
	Appendix.....	109
	References.....	116
	Figures and Tables	127
	References for Internet Appendix.....	150
3	Geographic Proximity and Stock Price Crash Risk: Evidence from Institutional Investors	152
1	Introduction	153

2	Related literature and hypothesis development.....	156
3	Data and univariate analyses	160
3.1	Firm level sample.....	160
3.2	Measures for shareholder proximity	162
3.3	Measures of stock price crash risk	163
3.4	Empirical model.....	165
3.5	Descriptive statistics	167
3.6	Correlation analysis	169
3.7	Sorting results	170
4	The relation between shareholder proximity and stock price crash risk	171
4.1	Baseline regression results.....	171
4.2	Endogeneity	174
4.2.1	Additional control variables.....	174
4.2.2	Entropy balancing	176
4.2.3	Firm and high dimensional fixed effects regression	177
4.2.4	Change analysis.....	178
4.2.5	Instrumental variable regression	179
4.2.6	Placebo test.....	181
4.3	Alternative proxies and additional analysis	182
4.3.1	Alternative definition of local monitors.....	182
4.3.2	Alternative measures of future stock price crash risk	182
4.3.3	Estimation of the expanded market model using daily returns	184
4.3.4	Other robustness tests.....	185
4.4	Monitoring and crash risk: the role of corporate governance	185
4.5	Effect of SOX issued in 2003 and the information advantage of local institutions	187
4.6	Bad news hoarding tests.....	188
4.7	Do all proximate investors monitor?.....	189
5	Conclusion.....	191
	Appendix.....	193
	References.....	200

Tables..... 209

Einleitung

Die heutige Bedeutung der empirischen Finanzmarktforschung und die bemerkenswerten Fortschritte der zu diesem Zweck angewendeten Methoden sind unbestritten. Beispielsweise verweisen Brooks et al. (2019) in ihrer Studie auf einen Anstieg des durchschnittlichen Anteils quantitativ ausgerichteter empirischer Analysen in den Top 5 Finance-Journals¹ von ursprünglich 53% in den 1970er Jahren auf 84% in den 2010er Jahren.² Der weit verbreitete Zugang zu großen Datenbanken, die damit einhergehende Informationsfülle sowie die Möglichkeit, anspruchsvolle Modelle und rechenintensive Techniken über geeignete (Statistik-) Programme in rascher Geschwindigkeit zu verarbeiten, ermöglichen es den Wissenschaftlern, eine enorme Bandbreite an ökonomischen Fragestellungen empirisch zu untersuchen. Nicht zuletzt durch die stetig besser werdende Ausbildung der Forschenden in fortgeschrittenen quantitativen Techniken hat sich die empirische Seite des Finanzwesens zu einem dominierenden Bereich in der finanzwirtschaftlichen Forschung entwickelt (vgl. Eckbo, 2011; Bell et al., 2013; Mitton, 2022).

Derartige Entwicklungen ermöglichen eine umfassende Analyse der sogenannten Marktstruktur (Garman, 1976), die sich vor allem auf die Interaktion zwischen den Mechanismen des Handelsprozesses und dessen Ergebnissen konzentriert. Konkret fokussiert sich dieser Forschungsstrang auf die Kosten des Wertpapierhandels sowie auf die Auswirkungen der Handelskosten auf das kurzfristige Verhalten der Wertpapierkurse (siehe

¹ Hierbei führt die referenzierte Studie folgende Finance-Journals an: *Journal of Finance*, *Journal of Financial Economics*, *Review of Financial Studies*, *Journal of Financial and Quantitative Analysis* sowie *Review of Finance*. Ausgehend von den aktuellsten Werten haben dabei im *Journal of Financial Economics* knapp neun von zehn Arbeiten (89%) einen empirischen Fokus, wohingegen dieser Anteil im *Journal of Financial and Quantitative Analysis* 73% beträgt.

² Für eine kritische Einordnung dieser Entwicklung siehe auch Brooks et al. (2019).

auch Easley und O'Hara, 1995; Stoll, 2003). Durch die Betrachtung des reinen Handelsprozesses, also der Abbildung jedes einzelnen Handels und jedes gestellten Kurses, basiert die empirische Anwendung in der Regel auf großen Datensätzen im hochfrequenten Intraday-Datenformat (vgl. z.B. O'Hara, 2015).

Die Frage nach dem kurzfristigen Verhalten von Wertpapierkursen im Sinne der Preisbildung ist hierbei ein weit begutachtetes Feld im Finance-Bereich (vgl. z.B. Pagnottoni und Dimpfl, 2019) und auch Bestandteil des ersten Papiers dieser Dissertation (Entrop et al., 2020). Lehmann (2002, S. 259) beschreibt in diesem Zuge die Preisfindung als „die effiziente und rechtzeitige Einbeziehung der Informationen, die im Handel der Anleger enthalten sind, in die Marktpreise“ (Übersetzung des Verfassers). Wenn zwei oder mehrere Preisreihen über ein gemeinsames Underlying gekoppelt sind, wird der Beitrag einer Preisreihe zur Preisbildung typischerweise als das Ausmaß angesehen, in dem sie als erste neue Informationen über den „wahren“ Wert des Basiswerts widerspiegelt (vgl. auch Putniņš, 2013). In der empirischen Literatur wird der Preisbildungsprozess unter anderem für eine Aktie, die an mehreren Handelsplätzen, auch länderübergreifend, gehandelt wird, untersucht (vgl. z.B. Hasbrouck, 1995; Eun und Sabherwal, 2003; Frijns et al., 2015a, 2015b). Auch Rohstoff- und Währungsmärkte (siehe z.B. Hauptfleisch et al., 2016; Dimpfl et al., 2017; Chen und Gau, 2010) sind Gegenstand dieser empirischen Analysen. Basierend auf Arbitragebeziehungen können ebenso Preise derivativer Finanzinstrumente (z.B. Optionen oder Futures) implizit in Spot-Preise umgewandelt werden, die dann mit den tatsächlichen Spotpreisen verglichen werden (vgl. Hasbrouck, 1995 für die Definition, und, unter anderem, Garbade und Silber, 1983; Tse, 1999; Booth et al., 1999; Chakravarty et al., 2004 für Studien in diesem Kontext).

Zwei einflussreiche Arbeiten, die den Informationsbeitrag im Preisbildungsprozess verschiedener, miteinander agierender Märkte oder Wertpapiere quantifizierbar machen, gehen auf Hasbrouck (1995; Information Share) sowie Gonzalo und Granger (1995; Component

Share) zurück, wobei deren Dekompositionsidee von Harris et al. (2002) weiterentwickelt wird (vgl. Putniņš, 2013). Beide vorgeschlagenen Informationsmaße basieren auf der Anwendung eines Vektor-Fehlerkorrekturmodells, wobei die Ansätze unterschiedliche Definitionen der Preisbildung liefern. Während sich Gonzalo und Granger (1995) nur mit dem Fehlerkorrekturprozess befassen, misst Hasbrouck (1995) die Preisfindung anhand der Varianz der Innovationen des gemeinsamen Faktors. Ökonomisch ausgedrückt erfassen beide Informationsmaße eine Kombination aus Präzision und Geschwindigkeit der Anpassungsreaktion der betrachteten Underlyings. Simulationsstudien zeigen dabei, dass der Component Share (Gonzalo und Granger, 1995; Harris et al., 2002) eher die Präzision der Anpassungsreaktion abdeckt, wohingegen der Information Share (Hasbrouck, 1995) vorrangig die Geschwindigkeit misst (vgl. hierzu Yan und Zivot, 2010; Putniņš, 2013). Hinsichtlich der Informationsverarbeitungseffizienz gilt, dass eine relativ gesehen exaktere und schnellere Berücksichtigung einer Information im gemeinsamen impliziten Preis eines Marktes (Wertpapiers) zu einem höheren Informationsanteil im Preisbildungsprozess führt, d.h. für diesen Markt (dieses Wertpapier) werden höhere Informationsmaße zu beobachten sein. Neuere Studien adressieren auch die Frage, wie sich die beschriebene Preisfindung im Laufe der Zeit verändert und was diese Veränderung determiniert. In diesem Kontext zeigen Frijns et al. (2010, 2015a, 2015b) für Cross-Listings, dass vor allem die Geld-Brief-Spanne sowie das Handelsvolumen wesentliche Treiber des Informationsverarbeitungsprozesses sind. Fernandez-Perez et al. (2018) finden ähnliche Resultate in Bezug auf die Werttreiber für Volatilitätsprodukte.

Neben dem Preisbildungsprozess stellt die Sicherstellung der Liquidität eine weitere wichtige Marktfunktion dar und ist ebenfalls dem Forschungsstrang der Marktmikrostruktur zuzuordnen (vgl. hierzu O'Hara, 2003; Brogaard et al., 2014). In der empirischen Literatur ist die Aktienliquidität ein häufig untersuchtes Instrument, wobei sich die Wissenschaftler

zunächst auf die Geld-Brief-Spanne und deren Dekomposition in ihre Bestandteile fokussieren. Bagehot (1971) weist darauf hin, dass die Handelskosten, die von sogenannten Market Makern gesetzt werden, grundsätzlich aus drei Komponenten bestehen: Kosten für Auftragsabwicklung, Bestandkosten und Kosten der adversen Selektion. Auf dieser Grundlage und auf Basis von Überlegungen zur Informationsasymmetrie werden in der empirischen Anwendung Ansätze vorgeschlagen, diese einzelnen Komponenten im Rahmen von Intraday-Daten aus dem Bid-Ask Spread zu schätzen. In diesem Kontext liefern unter anderem Kyle (1985), Glosten und Harris (1988) sowie Huang und Stoll (1996) wichtige Beiträge. Von großem Interesse ist es außerdem, potentielle Determinanten der Aktienliquidität zu begutachten. Neben den klassischen Marktgrößen wie Aktienpreisen, Handelsvolumen oder Aktienvolatilität (vgl. z.B., Easley und O'Hara, 1987; Kyle 1985; Glosten 1987) werden in der neueren Literatur auch zunehmend nicht-finanzielle Einflussgrößen, wie zum Beispiel die Qualität der internen Governance-Strukturen (vgl. Chung et al., 2010) oder lokale Konjunkturzyklen (vgl. Bernile et al., 2015), der Aktienliquidität untersucht. Außerdem zeigt die vorhandene Literatur, dass eine höhere Aktienliquidität³ eines Unternehmens mit niedrigeren Kapitalkosten (Amihud und Mendelson, 1986; Diamond und Verrecchia, 1991), höheren Aktienkursen und einem höheren Unternehmenswert (Fang et al., 2009) sowie einer höheren institutionellen und ausländischen Beteiligung (Ferreira und Matos, 2008; Gompers und Metrick, 2001) verknüpft ist.

Die Finanzierungsstruktur von Unternehmen und deren Einfluss auf den Unternehmenswert sind weitere relevante Fragestellungen, die im Zusammenhang mit der empirischen Kapitalmarktforschung stehen. In diesem Rahmen konzentriert sich ein Anfang der 2000er aufkommender Forschungsstrang auf das extreme Abwärtsrisiko des Eigenkapitals. Dieses ist für Investoren von höchster Relevanz, da ein Aktienkurs-Crash mit einem enormen Einbruch des Aktienwerts einhergeht und somit zu einem erheblichen Verlust der Aktienanleger führt.

³ Im Kontext dieser Studien wird die Liquidität als erklärende Variable verwendet.

Anekdotische Evidenz liefern beispielsweise Fälle wie Enron, Tyco und WorldCom. Zwei einflussreiche Forschungsarbeiten (Chen et al., 2001; Jin und Myers, 2006) schlagen dabei vor, die Schiefe der firmenspezifischen Aktienrenditeverteilung als Maß für die Crashanfälligkeit einer Aktie zu verwenden. Wenn es Managern gelingt, den Fluss negativer Informationen in den Aktienmarkt zu blockieren, sollte die Verteilung der Aktienrenditen asymmetrisch sein. Diese negativen Nachrichten stauen sich solange auf, bis ein gewisser Schwellenwert überschritten wird und die Nachrichten auf einmal öffentlich werden. Dies führt dann zu einem starken negativen Rückgang der Aktienkurse (vgl. z.B. Hutton et al., 2009). Auch im Rahmen dieses Forschungsfelds werden einige Anstrengungen unternommen, um zu verstehen, welche Faktoren das Crash-Risiko determinieren. An dieser Stelle reicht die empirische Forschung von klassischen Determinanten, wie die Besitzverhältnisse oder die Finanzierungsstruktur (vgl. z.B. Callen und Fang, 2013; An und Zhang, 2013; Andreou et al., 2016), bis hin zu weichen Faktoren wie Religiosität (Callen und Fang, 2015) und Internet-Recherche (Xu et al., 2021).

Neuere Fragestellungen der empirischen Finanzmarktforschung berücksichtigen auch vermehrt weniger naheliegende, jedoch relevante Einflussgrößen ökonomischer Outputvariablen, wie zum Beispiel demografische und geografische Faktoren des Unternehmensumfelds. So finden beispielsweise Hilary und Hui (2009) in ihrer einflussreichen Arbeit im Kontext sozialer Normen auf Finanzmärkten, dass der Grad der Religiosität des Counties, in dem das Unternehmen seinen Hauptsitz hat, die Risikoneigung sowie die Performance des Unternehmens beeinflusst. Nachfolgende Studien dokumentieren für solche Unternehmen auch eine bessere Qualität der Informationsoffenlegung (vgl. z.B. McGuire et al., 2012; Dyreng et al., 2012; Omer et al., 2018), ein geringeres Crash-Risiko (Callen und Fang, 2015) sowie geringere Kreditzinsen und bessere Ratings (Jiang et al., 2018). Zudem zeigen Coval und Moskowitz (1999), Grinblatt und Keloharju (2001) sowie Ivkovic und Weisbenner (2005), dass Anleger überproportional Aktien ihnen geographisch lokaler Unternehmen

besitzen und handeln. Anknüpfend daran dokumentieren Chhaochharia et al. (2012), dass auch die geografische Nähe zu institutionellen Investoren eine bedeutende Rolle spielt. Die Autoren liefern Evidenz für eine Überwachungsfunktion lokaler Investoren, was zu besseren internen Governance-Strukturen führt.

Die bisherige Übersicht arbeitet deutlich heraus, dass es aktuelle Entwicklungen der empirischen Finanzmarktforschung⁴ möglich machen, verschiedene Forschungsdisziplinen und Bestandteile zu vereinen und daraus geeignete Forschungsfragen abzuleiten. An dieser Stelle setzt die vorliegende Arbeit an, mit dem Ziel, offene Fragestellungen in den vorgestellten Forschungssträngen zu identifizieren und damit relevante Lücken zu schließen. Das erste Papier untersucht den Mechanismus des Preisbildungsprozesses zwischen dem Spot- und reguliertem Terminmarkt für den Bitcoin und stellt sich die Frage, welche Faktoren diesen Prozess determinieren. Ein weiteres Ziel dieser Dissertation ist es, nicht-finanzielle Informationen auf Kapitalmärkten zu analysieren. Auf diese Weise schließt das zweite Papier die Lücke in der Liquiditätsliteratur, indem gezeigt werden kann, dass vorherrschende soziale Normen im Umfeld eines Unternehmens die Aktienliquidität beeinflussen. An die Relevanz weicher Faktoren auf Kapitalmärkten anknüpfend, untersucht das dritte Papier das distanzabhängige Monitoring-Verhalten institutioneller Investoren und dessen Einfluss auf das Crash-Risiko von Aktien.

Den zentralen Untersuchungsgegenstand des ersten Papiers (“The Determinants of Price Discovery on Bitcoin Markets”) bilden die Determinanten des Preisbildungsprozesses auf dem Bitcoin-Markt. Dabei verfolgt das Papier einen zweistufigen Ansatz. In der ersten Stufe wird

⁴ Letztlich ist jedoch darauf hinzuweisen, dass das wohl wichtigste und gegenwärtigste Problem, mit dem derartige Studien zur empirischen Forschung konfrontiert sind, die Endogenität ist. Grundsätzlich wird unter Endogenität die Korrelation zwischen den erklärenden Variablen und dem Fehlerterm in einer Regression verstanden. Endogenität führt zu verzerrten und inkonsistenten Parameterschätzungen, die zuverlässige Schlussfolgerungen praktisch unmöglich machen. Unter der Anwendung verschiedener ökonometrischer Techniken können etwaige Bedenken zumindest reduziert werden (siehe auch Roberts und Whited, 2013). Demnach ist es gerade im Kontext neuartiger empirischer Fragestellungen unabdingbar, potentielle endogene Zusammenhänge zu identifizieren und durch geeignete Methoden entsprechende Bedenken zu reduzieren.

unter Gültigkeit der Cost-of-Carry-Arbitrage⁵ zwischen Spot- und Futures-Markt zunächst ein Vektor-Fehlerkorrekturmodell geschätzt. Um den Preisfindungsprozess⁶ quantifizierbar zu machen, verwendet das Papier in dieser Stufe sowohl den von Gonzalo und Granger (1995) vorgeschlagenen Component Share (CS) als auch den Information Share (IS) von Hasbrouck (1995).⁷ Grundsätzlich gilt, dass sich ein Markt in relativer Informationsführerschaft befindet, wenn das relative Informationsmaß (CS oder IS) über 0,5 liegt. Allerdings ist die Verwendung des IS als relatives Maß für die Preisfindung problematisch, wenn sich die Liquidität auf den Märkten im Laufe der Zeit ändert (siehe hierzu auch Frijns et al., 2015). Dieses Problem betrifft auch andere vorgeschlagene Informationsmaße, wie zum Beispiel den Information Leadership Share (ILS) von Yan und Zivot (2010) und Putniņš (2013), weshalb die Studie in der zweiten Stufe auf den Component Share fokussiert.

Die Untersuchung basiert auf Intraday-Quotes für Bitcoin-Futures, die an der CME gehandelt werden sowie den Spot-Quotes der Bitstamp Börse (BTSP). Die Daten stammen von der Thomson Reuters Tick History (TRTH). Die Quotedaten bestehen aus dem Geld- und Briefkurs und dem genauen Zeitstempel, zu dem ein neuer Kurs gestellt wird. Für die Analyse potentieller Determinanten wird der Datensatz um weitere Variablen wie Unsicherheitsmaße (VIX und EPU) sowie Stimmungsindikatoren von Refinitiv MarketPsych ergänzt.

Für jeden Tag der Stichprobe, die von Dezember 2017 bis März 2019 reicht, wird der am aktivsten gehandelte Futures-Kontrakt verwendet, wobei sich die Aktivität auf das Handelsvolumen bezieht. Die aus dem Vektor-Fehlerkorrekturmodell abgeleiteten relativen Informationsmaße, die auf der Basis von Mid-Quotes und einminütigen Zeitintervallen geschätzt werden, offenbaren, dass der Futures-Markt in neun Kontraktmonaten die

⁵ Es gilt: $F_t^T = S_t e^{r(T-t)}$.

⁶ Bei der Preisfindungsanalyse wird versucht, die Genauigkeit und Geschwindigkeit zu messen, mit der alle Produkte, hier Spot und Futures, neue Informationen in ihren gemeinsamen Preis einfließen lassen. Märkte, die neue Informationen schneller einbeziehen und gleichzeitig besser in der Lage sind, Rauschen zu vermeiden, haben einen höheren Anteil an der Preisfindung (vgl. zum Beispiel Putniņš, 2013).

⁷ Die Berechnung beider Informationsmaße folgt Baillie et al. (2002).

Preisführerschaft innehat, wobei hiervon drei Monate statistisch signifikant sind. Der Kassa-Markt hingegen übernimmt die Preisführerschaft in den verbleibenden sechs Monaten, wobei hier zwei Monate statistisch signifikant sind.⁸

In der zweiten Stufe werden verschiedenartige Determinanten wie Marktqualität (Spread, Handelsvolumen und Anzahl an Quotes), Unsicherheit, Sentiment und Aufmerksamkeit sowie makroökonomische News untersucht (vgl. z.B. Eun und Sabherwal, 2003; Chakravarty et al., 2004; Frijns et al., 2010, 2015a, 2015b). Im Rahmen einer zweistufigen Kleinste-Quadrate-Schätzung (2SLS) zeigt sich in Übereinstimmung mit bisheriger Literatur für andere Märkte, dass die Handelskosten den wesentlichen Treiber des Preisfindungsprozesses darstellen. Zusätzlich determinieren das relative Handelsvolumen sowie die Volatilität die Preisführerschaft, jedoch in geringerem Ausmaß. Weitere Analysen zeigen außerdem, dass die auf Nachrichten basierende Bitcoin-Stimmung als relevantes Maß für den Preisfindungsprozess auftritt, wohingegen die Aufmerksamkeit und makroökonomische Nachrichten keine signifikante Rolle bei der Erklärung der zeitlichen Variation der Informationsmaße spielen. Zusammenfassend kann festgehalten werden, dass trotz der bestehenden Unklarheiten des Bitcoin-Universums zumindest die Untersuchung der Determinanten der Preisfindung zu ökonomisch sinnvollen Ergebnissen führt, die auch in anderen Anlageklassen zu finden sind. Insgesamt können daher Bemühungen zur Verbesserung der Marktqualität und der damit verbundenen Preisbildung die Attraktivität eines Marktes erhöhen.

Das zweite Papier ("Local Religiosity and Stock Liquidity") verlagert den Fokus auf die Aktienliquidität als weitere essentielle Größe der empirischen Kapitalmarktforschung. Die Studie untersucht explizit die Auswirkung von sozialen religiösen Normen als eine Art „weiche“ Information auf die Liquidität von Aktien. Neben klassischen marktbasieren

⁸ Die statistische Signifikanz bezieht sich hierbei auf die Analyse, ob der durchschnittliche Component Share signifikant unterschiedlich von 0,5 ist. Für den Information Share ergeben sich ähnliche Ergebnisse.

Determinanten, wie Handelsvolumen, Preis oder Kursvolatilität, hat die Literatur bereits gezeigt, dass die Aktienliquidität auch mit der Sichtbarkeit und Bekanntheit eines Unternehmens (Grullon et al., 2004; Loughran und Schultz, 2005) und mit der Eigentümerstruktur (Attig et al., 2006) verknüpft ist. Außerdem zeigen neuere Studien, dass die Liquidität von Unternehmen auch von der politischen Stabilität und der Effizienz der Justiz (Eleswarapu und Venkataraman, 2006), der Qualität der internen Corporate Governance (Chung et al., 2010) und Governance Standards (Roy et al., 2022), lokalen Konjunkturzyklen (Bernile et al., 2015) sowie der Bildung des CEOs (Pham et al., 2020) abhängt. Insgesamt sind die allgemeine Transparenz und das Informationsumfeld von Unternehmen wichtige Determinanten für deren Liquidität am Aktienmarkt (siehe z.B. Healy und Palepu, 2001; Roy et al., 2022).

Basierend auf dieser Ausgangsüberlegung sollten Unternehmen mit einer besseren (schlechteren) Qualität der Informationsflüsse und -offenlegung ein höheres (niedrigeres) Niveau der Aktienliquidität aufweisen (vgl. Diamond und Verrecchia, 1991; Kurlat, 2018; Roy et al., 2022). Insbesondere kann das Auftreten von Problemen der adversen Selektion und der damit einhergehenden Informationsasymmetrien zu einer Ausweitung der Geld-Brief-Spanne führen (vgl. Kyle, 1985; Glosten und Milgrom, 1985; Huang und Stoll, 1996; Bhattacharya et al., 2013). Da Religiosität aufgrund ihres antimanipulativen Ethos (vgl. Barro und McCleary, 2003; Callen und Fang, 2015) wahrscheinlich das Vertrauen in das Verhalten und die Handlungen von Unternehmen fördert, und dazu beiträgt, Unternehmen mit glaubwürdigeren Informationsflüssen zu identifizieren (vgl. McGuire et al., 2012), sollte sich dies positiv auf die Aktienliquidität auswirken. Erwartungsgemäß ist dieser vertrauensbasierte Aspekt der Religiosität dann von besonderer Relevanz, wenn Unternehmen in einem Umfeld tätig sind, in dem ansonsten relativ wenig Informationen über das Unternehmen verfügbar sind. Unter diesen

Umständen wirken religiöse Normen wahrscheinlich als ein informeller Mechanismus, der die Bedeutung der Informationsasymmetrie verringert.

Das zweite Papier erweitert die bestehende Literatur in zwei wesentlichen Punkten: Erstens zeigt die Studie, dass die Religiosität mit ihrem antimanipulativen Ethos (vgl. z.B. Barro und McCleary, 2003; Callen und Fang, 2015) die Art und Weise beeinflusst, wie ein Unternehmen von Außenstehenden, d.h. von Liquiditätsanbietern, wahrgenommen wird. Zweitens offenbart die Analyse, dass Liquiditätsanbieter in ihrem Entscheidungsprozess nicht nur „harte“ Marktinformationen berücksichtigen, sondern ebenfalls nicht-finanzielle, „weiche“ Größen eine bedeutende Rolle spielen. Insbesondere dann, wenn relativ wenige Informationen über ein Unternehmen vorliegen, schürt der Grad der Religiosität das Vertrauen in das unternehmerische Verhalten und den Informationsfluss.

Die empirische Analyse verwendet mehrere verschiedene Datenquellen, um der Forschungsfrage nachzugehen. Ausgangspunkt sind alle aktiven und inaktiven Unternehmen mit Hauptsitz in den USA, die im Zeitraum von 1973 bis 2020 an der NYSE, AMEX, und NASDAQ gelistet sind. Zur Berechnung des Grades der Religiosität wird die Datenbank von der Association of Religion Data Archives (ARDA) verwendet, wobei die Hauptvariable durch den relativen Anteil der Gläubigen in einem County repräsentiert wird. Um den Effekt der Religiosität zu isolieren, verwendet das Papier weitere demografische Variablen, die hauptsächlich vom US Census Bureau stammen.⁹ Marktdaten zur Berechnung des Bid-Ask Spreads werden aus Refinitiv, die Bilanzdaten aus Datastream entnommen. Daten zu den Eigentumsverhältnissen kommen von Refinitiv Ownership Profile.

Die empirische Analyse, die einen Pooled OLS-Ansatz mit fixen Effekten auf Industrie- und Jahresebene verwendet, zeigt einen handelskostenreduzierenden Effekt für Unternehmen, die

⁹ Aufgrund der zehnjährigen Frequenz der Daten, wird zwischen den Nicht-Umfragejahren linear interpoliert (vgl. hierzu auch Hilary und Hui, 2009; Jiang et al., 2018).

ihren Hauptsitz in hoch-religiösen Counties haben, der auch ökonomisch signifikant ist. Entsprechend liefert eine Verschiebung der Religiositätsvariable um einen Interquartils-Abstand eine Verringerung des relativen Spreads um 6,75%. Zudem ist der gefundene Effekt robust gegenüber der Berücksichtigung weiterer Kontrollvariablen, wie z.B. Governance-Kennzahlen, sowie zusätzlicher Sensitivitäts- und Endogenitätsanalysen. Die Ergebnisse weisen außerdem darauf hin, dass der Kanal zwischen Religiosität und Liquidität vor allem durch das Informationsumfeld der Unternehmen getrieben ist. Demnach ist die Religiosität als eine Art informelle Verpflichtung des antimanipulativen Handelns offenkundig, wenn Liquiditätsanbieter wenige Informationen zu einem Unternehmen haben und hoher Informationsasymmetrie gegenüberstehen. Auch bestätigen die Resultate die Vermutung, dass die Wahrscheinlichkeit des informierten Handelns geringer ausgeprägt ist für Unternehmen in religiöseren Counties. Schließlich deutet die Studie auf eine potentielle (indirekte) Wertimplikation der lokalen Religiosität hin, die durch ihre Auswirkungen auf die Aktienliquidität zum Tragen kommt.

Insgesamt können sich aus den präsentierten Ergebnissen auch mögliche praktische Konsequenzen für Regulatoren und Liquiditätsanbieter ergeben. Da Religiosität wahrscheinlich einen informellen Mechanismus zur Verringerung von Agency-Kosten (vgl. z.B. McGuire et al., 2012; Leventis et al., 2018) und der Relevanz von Informationsasymmetrie darstellt, bietet der Grad der Religiosität möglicherweise eine neue Perspektive für Investitions-, Überwachungs- und Informationssuchstrategien, insbesondere für Unternehmen, die in einem schwachen Informationsumfeld fungieren. Zusammengefasst unterstreicht die Studie die Bedeutung „weicher“ Informationen für die Verbesserung der Marktqualität.

Das dritte Papier dieser Dissertation („Geographic Proximity and Stock Price Crash Risk: Evidence from Institutional Investors“) konzentriert sich auf den Einfluss der geografischen Nähe von institutionellen Investoren auf das Crash-Risiko von Aktien. In der empirischen

Finanzmarktliteratur ist bekannt, dass physische Nähe vor allem den Zugang zu formellen und informellen unternehmensspezifischen Informationen erleichtert. Auch ist es wahrscheinlich, dass lokale Einrichtungen (z.B. Golf-Clubs) zu den sozialen Netzwerken lokaler Firmenmanagern gehören und sich somit direkt auf das Verhalten auswirken (siehe auch Huang und Kang, 2017). Darüber hinaus baut die Argumentation des vorliegenden Papiers auf der Monitoringtheorie auf (vgl. z.B. Callen und Fang, 2013). Diese besagt, dass institutionelle Investoren mit einem relativ großen Investitionsanteil im Unternehmen eher davon profitieren, Unternehmensaktivitäten zu überwachen, um Kurzsichtigkeit des Managements und andere Agency-Konflikte abzuschwächen. Da Monitoringkosten mit zunehmender geografischer Entfernung steigen (siehe z.B. Coval und Moskowitz, 1999; Lin and Png, 2003; Degryse und Ongena, 2005), werden dahingehende Bestrebungen durch die Distanz abgeschwächt. Insgesamt knüpft das Papier an diese Erkenntnisse an und beschäftigt sich mit zwei neuen Fragestellungen: (a) Welchen Effekt hat die Distanz zwischen institutionellen Investoren und Unternehmen auf das Crash-Risiko? und (b) Inwieweit wirkt sich der Investorentyp auf den Zusammenhang zwischen geographischer Nähe und Absturzrisiko aus?. Die Studie liefert demnach weitere wichtige empirische Erkenntnisse zur Auswirkung „softer“ Informationen auf Kapitalmärkten, in dem Sinne, dass die räumliche Distanz zwischen Stakeholdern und Unternehmen die Aktienkurse und damit den Wohlstand von Investoren beeinflusst.

Die Basis der empirischen Untersuchung bildet ein umfassender Datensatz, der sich aus vier Quellen speist. Aus Datastream werden alle aktiven und inaktiven US-Unternehmen extrahiert, die an der NYSE, AMEX und NASDAQ von 1973 bis 2020 erfasst sind und ihren Hauptsitz in den USA haben. Da Datastream lediglich Informationen zur aktuellen Adresse der Hauptsitze liefert, wird zusätzlich der SEC-Datensatz von Bill McDonald¹⁰ verwendet. Unter anderem beinhaltet dieser alle Informationen zum Kopfteil von 10-K/Q-Berichten (und allen Varianten),

¹⁰ <https://sraf.nd.edu/data/augmented-10-x-header-data/>, zuletzt aufgerufen am 01.09.2022.

die auf EDGAR eingereicht werden. Die Daten aus Datastream und von der EDGAR-Datenbank werden über den CIK-Code zusammengefügt. Um die Distanz zwischen den Hauptsitzen und den jeweiligen Investoren zu messen, bedient sich die Studie außerdem der Investorendaten aus Refinitiv Ownership Profile. Dieser Sub-Datensatz beinhaltet die aktuellen Adressen der Hauptsitze der Investoren sowie Informationen zu deren Besitzverhältnissen. Schließlich komplettieren Informationen zu Bilanz- und Marktdaten aus Datastream und Refinitiv den Datensatz, um unter anderem die Crash-Risiko-Maße zu schätzen.

Unter der Verwendung eines Pooled OLS-Ansatzes mit fixen Effekten auf Industrie- und Jahresebene zeigt die Studie einen positiven Zusammenhang zwischen der wertgewichteten Distanz und dem Crash-Risiko einer Aktie. Die ökonomische Signifikanz dieses Effekts ist ebenfalls gegeben. Demnach führt eine Erhöhung der wertgewichteten Distanz um einen Interquartilsabstand zu einer relativen Erhöhung des künftigen Crash-Risikos um mindestens 13%. Auch in dieser Studie werden Endogenitätsbedenken durch geeignete ökonometrische Methoden reduziert. Des Weiteren zeigt die Analyse, dass der Zusammenhang zwischen geografischer Nähe lokaler institutioneller Investoren und dem Crash-Risiko von Aktien ausgeprägt ist, wenn das entsprechende Unternehmen über eine schwache interne Governance verfügt. Zudem bleibt der Effekt auch mit der Einführung des Sarbanes-Oxley Acts persistent, wobei die Stärke abnimmt. Weiterhin bestätigt das Papier die direkte Beziehung zwischen dem Zurückhalten negativer Informationen von Managern und der Distanz zu den institutionellen Investoren. Abschließend dokumentiert das Papier empirische Evidenz für die Heterogenität in den Monitoring-Aktivitäten zwischen den Investoren-Typen. Es zeigt sich, dass die Distanz zu Investoren mit einem aktiv ausgerichteten Investitionsansatz eine entscheidendere Rolle spielt, verglichen mit der Distanz zu passiven Investoren. Bei Investoren mit einem passiven Investitionsansatz zeigt die wertgewichtete Distanzvariable keinen signifikanten Effekt. Insgesamt ergeben sich aus diesen Erkenntnissen potentielle Implikationen für Aktienanleger

und Regulierungsbehörden. Da das Zurückhalten von negativen Informationen auch von dem Standort institutioneller Investoren und deren Monitoringaktivitäten abhängt, bietet die Studie möglicherweise eine zusätzliche Perspektive bei der Vorhersage der Wahrscheinlichkeit von zukünftigen Aktien-Crashes.

Zusammenfassend beinhaltet die vorliegende Arbeit verschiedene neuartige Fragestellungen in relevanten Disziplinen der empirischen Kapitalmarktforschung und trägt damit in jedem Forschungsstrang zu einem besseren Verständnis der ökonomischen Zusammenhänge bei. Die Analyse der Preisbildung auf dem Bitcoin-Markt (1. Paper) liefert empirische Evidenz, dass die zeitlich variierende Preisführerschaft vor allem von den relativen Handelskosten des Spot- und Futures-Markt abhängig ist. Hinzu kommen das relative Handelsvolumen, die Markt-Volatilität sowie die nachrichtenbasierte Bitcoin-Stimmung, die die Informationsverarbeitung beeinflussen. Diese gefundenen Zusammenhänge für den Kryptowährungsmarkt sind damit überwiegend im Einklang mit den Ergebnissen für andere Asset-Klassen. Hinsichtlich der Aktienliquidität zeigt sich, dass Unternehmen, die ihren Hauptsitz in religiöseren US-Counties haben, niedrigere Spreads aufweisen (2. Papier). Dies ist vor allem darin begründet, dass Liquiditätsanbieter auf Handlungen und Informationsflüsse von Unternehmen vertrauen, die in religiöseren Counties ihren Hauptsitz haben. Dieser vertrauensbasierte Zusammenhang ist dann von großer Bedeutung, wenn Unternehmen in einem wenig informativem Umfeld agieren. Zuletzt offenbart sich, dass eine distanzindizierte Reduzierung der Monitoring-Aktivitäten institutioneller Investoren mit einem erhöhten Crash-Risiko einhergeht, da es Managern ermöglicht wird, einfacher negative Nachrichten zurückzuhalten (3. Paper). Dieser Zusammenhang hält vor allem für institutionelle Investoren mit einem aktiven Managementansatz, wobei für deren passive Peers kein signifikanter Kanal zu finden ist.

Quellen

- Amihud, Y., Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics* 17(2), 223-249. [https://doi.org/10.1016/0304-405X\(86\)90065-6](https://doi.org/10.1016/0304-405X(86)90065-6)
- An, H., Zhang, T. (2013). Stock price synchronicity, crash risk, and institutional investors. *Journal of Corporate Finance* 21, 1-15. <https://doi.org/10.1016/j.jcorpfin.2013.01.001>
- Andreou, P. C., Antoniou, C., Horton, J., Louca, C. (2016). Corporate Governance and Firm-specific Stock Price Crashes. *European Financial Management* 22(5), 916-956. <https://doi.org/10.1111/eufm.12084>
- Attig, N., Fong, W.-M., Gadhoun, Y., Lang, L. H. P. (2006). Effects of large shareholding on information asymmetry and liquidity. *Journal of Banking & Finance* 30, 2875-2892. <https://doi.org/10.1016/j.jbankfin.2005.12.002>
- Bagehot, W. (1971). The Only Game in Town. *Financial Analysts Journal* 27(2), 12-14. <https://doi.org/10.2469/faj.v27.n2.12>
- Baillie, R. T., Booth, G. G., Tse, Y., Zobotina, T. (2002). Price Discovery and common factor models. *Journal of Financial Markets* 5(3), 309-321. [https://doi.org/10.1016/S1386-4181\(02\)00027-7](https://doi.org/10.1016/S1386-4181(02)00027-7)
- Barro, R. J., McCleary, R. (2003). Religion and Economic Growth. NBER Working Paper Series 9682. <https://doi.org/10.3386/w9682>
- Bell, A., Brooks, C., Prokopczuk, M. (2013). Handbook of Research Methods and Applications in Empirical Finance. Volume 1, Eds. Edward Elgar Publishing Limited, Cheltenham (UK) and Northampton (US), Preface.
- Bernile, G., Korniotis, G., Kumar, A., Wang, Q. (2015). Local Business Cycles and Local Liquidity. *Journal of Financial and Quantitative Analysis* 50(5), 987-1010. <https://doi.org/10.1017/S0022109015000447>
- Bhattacharya, N., Desai, H., Venkataraman, K. (2013). Does Earnings Quality Affect Information Asymmetry? Evidence from Trading Costs. *Contemporary Accounting Research* 30(2), 482-516. <https://doi.org/10.1111/j.1911-3846.2012.01161.x>
- Booth, G. G., So, R. W., Tse, Y. (1999). Price discovery in the German equity index derivatives markets. *Journal of Futures Markets* 19, 619-643. [https://doi.org/10.1002/\(SICI\)1096-9934\(199909\)19:6<619::AID-FUT1>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1096-9934(199909)19:6<619::AID-FUT1>3.0.CO;2-M)

- Brogaard, J., Hendershott, T., Riordan, R. (2014). High-Frequency Trading and Price Discovery. *Review of Financial Studies* 27(8), 2267-2306. <https://doi.org/10.1093/rfs/hhu032>
- Brooks, C., Fenton, E., Schopohl, L., Walker, J. (2019). Why does research in finance have so little impact. *Critical Perspectives on Accounting* 58, 24-52. <https://doi.org/10.1016/j.cpa.2018.04.005>
- Callen, J. L., Fang, X. (2013). Institutional investor stability and crash risk: Monitoring versus short-termism? *Journal of Banking & Finance* 37(8), 3047-3063. <https://doi.org/10.1016/j.jbankfin.2013.02.018>
- Callen, J. L., Fang, X. (2015). Religion and Stock Price Crash Risk. *Journal of Financial and Quantitative Analysis* 50(1/2), 169-195. <https://doi.org/10.1017/S0022109015000046>
- Chakravarty, S., Gulen, H., Mayhew, S. (2004). Informed trading in stock and option markets. *Journal of Finance* 59, 1235-1257. <https://doi.org/10.1111/j.1540-6261.2004.00661.x>
- Chen, J., Hong, H., Stein, J. C. (2001). Forecasting crashes: trading volume, past returns, and conditional skewness in stock prices. *Journal of Financial Economics* 61(3), 345-381. [https://doi.org/10.1016/S0304-405X\(01\)00066-6](https://doi.org/10.1016/S0304-405X(01)00066-6)
- Chen, Y.-L., Gau, Y.-F. (2010). News announcements and price discovery in foreign exchange spot and futures markets. *Journal of Banking & Finance* 34, 1628-1636. <https://doi.org/10.1016/j.jbankfin.2010.03.009>
- Chhaochharia, V., Kumar, A., Niessen-Ruenzi, A. (2012). Local investors and corporate governance. *Journal of Accounting and Economics* 54, 42-67. <http://dx.doi.org/10.1016/j.jacceco.2012.03.002>
- Chung, K. H., Elder, J., Kim, J.-C. (2010). Corporate Governance and Liquidity. *Journal of Financial and Quantitative Analysis* 45(2), 265-291. <https://doi.org/10.1017/S0022109010000104>
- Coval, J., Moskowitz, T. (1999). Home bias at home: local equity preference in domestic portfolios. *Journal of Finance* 54(6), 2045-2073. <https://doi.org/10.1111/0022-1082.00181>
- Degryse, H., Ongena, S. (2005). Distance, lending relationships, and competition. *Journal of Finance* 60(1), 231-266. <https://doi.org/10.1111/j.1540-6261.2005.00729.x>
- Diamond, D., W., Verrecchia, R. E. (1991). Disclosure, Liquidity, and the Cost of Capital. *The Journal of Finance* 46(4), 1325-1359. <https://doi.org/10.1111/j.1540-6261.1991.tb04620.x>

- Dimpfl, T., Flad, M., Jung, R. C. (2017). Price discovery in agricultural commodity markets in the presence of futures speculation. *Journal of Commodity Markets* 5, 50-62. <https://doi.org/10.1016/j.jcomm.2017.01.002>
- Dyreng, S. D., Mayew, W. J., Williams, C. D. (2012). Religious Social Norms and Corporate Financial Reporting. *Journal of Business Finance & Accounting* 39(7/8), 845-875. <https://doi.org/10.1111/j.1468-5957.2012.02295.x>
- Easley, D., O'Hara, M. (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics* 19(1), 69-90. [https://doi.org/10.1016/0304-405X\(87\)90029-8](https://doi.org/10.1016/0304-405X(87)90029-8)
- Easley, D., O'Hara, M. (1995). Market Microstructure. In *Handbooks in Operations Research and Management Science: Finance*, R. A. Jarrow, V. Maksimovic, and W. T. Ziemba, Eds. Elsevier B.V., Volume 9, 357-383. [https://doi.org/10.1016/S0927-0507\(05\)80056-8](https://doi.org/10.1016/S0927-0507(05)80056-8)
- Eckbo, B. E. (2007). Handbook of Corporate Finance – Empirical Corporate Finance. Volume 1, Eds. Elsevier, North-Holland, Preface.
- Eleswarapu, V. R., Venkataraman, K. (2006). The Impact of Legal and Political Institutions on Equity Trading Costs: A Cross-Country Analysis. *Review of Financial Studies* 19(3), 1081-1111. <https://doi.org/10.1093/rfs/hhj026>
- Entrop, O., Frijns, B., Seruset, M. (2020). The determinants of price discovery on bitcoin markets. *Journal of Futures Markets* 40(5), 816-837. <https://doi.org/10.1002/fut.22101>
- Eun, C. S., Sabherwal, S. (2003). Cross-Border Listings and Price Discovery: Evidence from U.S.-Listed Canadian Stocks. *Journal of Finance* 58(2), 549-575. <https://doi.org/10.1111/1540-6261.00537>
- Fang, V. W., Noe, T., H., Tice, S. (2009). Stock market liquidity and firm value. *Journal of Financial Economics* 94(1), 150-169. <https://doi.org/10.1016/j.jfineco.2008.08.007>
- Fernandez-Perez, A., Frijns, B., Gafiatullina, I., Tourani-Rad, A. (2018). Determinants of intraday price discovery in VIX exchange traded notes. *Journal of Futures Markets* 38, 535-548. <https://doi.org/10.1002/fut.21907>
- Ferreira, M. A., Matos, P. (2008). The colors of investors' money: The role of institutional investors around the world. *Journal of Financial Economics* 88, 499-533. <https://doi.org/10.1016/j.jfineco.2007.07.003>
- Frijns, B., Gilbert, A., Tourani-Rad, A. (2010). The dynamics of price discovery for cross-listed shares: Evidence from Australia and New Zealand. *Journal of Banking & Finance* 34(3), 498-508. <https://doi.org/10.1016/j.jbankfin.2009.08.014>

- Frijns, B., Gilbert, A., Tourani-Rad, A. (2015a). The determinants of price discovery: Evidence from US-Canadian cross-listed shares. *Journal of Banking & Finance* 59, 457-468. <https://doi.org/10.1016/j.jbankfin.2015.07.011>
- Frijns, B., Indriawan, I., Tourani-Rad, A. (2015b). Macroeconomic news announcements and price discovery: Evidence from Canadian-U.S. cross-listed firms. *Journal of Empirical Finance* 32, 35-48. <https://doi.org/10.1016/j.jempfin.2014.05.001>
- Garbade, K. D., Silber, W. L. (1983). Price Movements and Price Discovery in Futures and Cash Markets. *Review of Economics and Statistics* 65(2), 289-297. <https://doi.org/10.2307/1924495>
- Garman, M. B. (1976). Market microstructure. *Journal of Financial Economics* 3(3), 257-275. [https://doi.org/10.1016/0304-405X\(76\)90006-4](https://doi.org/10.1016/0304-405X(76)90006-4)
- Glosten, L. R. (1987). Components of the Bid-Ask Spread and the Statistical Properties of Transaction Prices. *Journal of Finance* 42(5), 1293-1307. <https://doi.org/10.1111/j.1540-6261.1987.tb04367.x>
- Glosten, L. R., Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14(1), 71-100. [https://doi.org/10.1016/0304-405X\(85\)90044-3](https://doi.org/10.1016/0304-405X(85)90044-3)
- Glosten, L. R., Harris, L. E. (1988). Estimating the components of the bid/ask spread. *Journal of Financial Economics* 21(1), 123-142. [https://doi.org/10.1016/0304-405X\(88\)90034-7](https://doi.org/10.1016/0304-405X(88)90034-7)
- Gompers, P. A., Metrick, A. (2001). Institutional Investors and Equity Prices. *Quarterly Journal of Economics* 116(1), 229-259. <https://doi.org/10.1162/003355301556392>
- Gonzalo, J., Granger, C. (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Studies* 13, 27-35. <https://doi.org/10.1080/07350015.1995.10524576>
- Grinblatt, M., Keloharju, M. (2001). How Distance, Language, and Culture Influence Stockholdings and Trades. *Journal of Finance* 56(3), 1053-1073. <https://doi.org/10.1111/j.1540-6261.2005.00730.x>
- Grullon, G., Kanatas, G., Weston, J. P. (2004). Advertising, Breadth of Ownership, and Liquidity. *Review of Financial Studies* 17(2), 439-461. <https://doi.org/10.1093/rfs/hhg039>
- Ivkovic, Z., Weisbenner, S. (2005). Local Does as Local Is: Information Content of the Geography of Individual Investors' Common Stock Investments. *Journal of Finance* 60(1), 267-306. <https://doi.org/10.1111/j.1540-6261.2005.00730.x>

- Harris, F. H., McInish, T. H., Wood, R. A. (2002). Common factor components versus information shares: a reply. *Journal of Financial Markets* 5, 341-348. [https://doi.org/10.1016/S1386-4181\(02\)00030-7](https://doi.org/10.1016/S1386-4181(02)00030-7)
- Hasbrouck, J. (1995). One Security, Many Markets: Determining the Contributions to Price Discovery. *Journal of Finance* 50(4), 1175-1199. <https://doi.org/10.1111/j.1540-6261.1995.tb04054.x>.
- Hauptfleisch, M., Putniņš T. J., Lucey, B. (2016). Who sets the price of gold? London or New York. *Journal of Futures Markets* 36, 564-586. <https://doi.org/10.1002/fut.21775>
- Healy, P. M., Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of empirical disclosure literature. *Journal of Accounting and Economics* 31(1-3), 405-440. [https://doi.org/10.1016/S0165-4101\(01\)00018-0](https://doi.org/10.1016/S0165-4101(01)00018-0)
- Hilary, G., Hui, K. W. (2009). Does religion matter in corporate decision making in America? *Journal of Financial Economics* 93(3), 455-473. <https://doi.org/10.1016/j.jfineco.2008.10.001.S>
- Huang, R. D., Stoll, H. R. (1996). Dealer versus auction markets: A paired comparison of execution costs on NASDAQ and the NYSE. *Journal of Financial Economics* 41, 313-357. [https://doi.org/10.1016/0304-405X\(95\)00867-E](https://doi.org/10.1016/0304-405X(95)00867-E)
- Hutton, A. P., Marcus, A. J., Tehranian, H. (2009). Opaque financial reports, R², and crash risk. *Journal of Financial Economics* 94(1), 67-86. <https://doi.org/10.1016/j.jfineco.2008.10.003>
- Jiang, F. W., John, K., Wei Li, C., Qian, Y. (2018). Earthly Reward to the Religious: Religiosity and the Costs of Public and Private Debt. *Journal of Financial and Quantitative Analysis* 53(5), 2131-2160. <https://doi.org/10.1017/S002210901800039X>
- Jin, L., Myers, S. C. (2006). R² around the world: New theory and new tests. *Journal of Financial Economics* 76(2), 257-292. <https://doi.org/10.1016/j.jfineco.2004.11.003>
- Kurlat, P. (2018). Liquidity as Social Expertise. *Journal of Finance* 73(2), 619-656. <https://doi.org/10.1111/jofi.12606>
- Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica* 53(6), 1315-1335. <https://doi.org/10.2307/1913210>
- Lehmann, B. N. (2002). Some desiderata for the measurement of price discovery across markets. *Journal of Financial Markets* 5(3), 259-276. [https://doi.org/10.1016/S1386-4181\(02\)00025-3](https://doi.org/10.1016/S1386-4181(02)00025-3)

- Leventis, S., Dedoulis, E., Abdelsalam, O. (2018). The Impact of Religiosity on Audit Pricing. *Journal of Business Ethics* 148, 53-78. <https://doi.org/10.1007/s10551-015-3001-x>
- Lin, C.-C., Png, I. (2003). Monitoring costs and the mode of international investment. *Journal of Economic Geography* 3(3), 261-274. <https://doi.org/10.1093/jeg/3.3.261>
- Loughran, T., Schultz, P. (2005). Liquidity: Urban versus rural firms. *Journal of Financial Economics* 78, 341-374. <https://doi.org/10.1016/j.jfineco.2004.10.008>
- McGuire, S., T., Omer, T. C., Sharp, N. Y. (2012). The Impact of Religion on Financial Reporting Irregularities. *Accounting Review* 87(2), 645-673. <https://doi.org/10.2308/accr-10206>
- Mitton, T. (2022). Methodological Variation in Empirical Corporate Finance? *Review of Financial Studies* 35(2), 527-575. <https://doi.org/10.1093/rfs/hhab030>
- O'Hara, M. (2003). Presidential Address: Liquidity and Price Discovery. *Journal of Finance* 58(4), 1335-1354. <https://doi.org/10.1111/1540-6261.00569>
- O'Hara, M. (2015). High frequency market microstructure. *Journal of Financial Economics* 116(2), 257-270. <https://doi.org/10.1016/j.jfineco.2015.01.003>
- Omer, T. C., Sharp, N. Y., Wang, D. (2018). The Impact of Religion on the Going Concern Reporting Decisions of Local Audit Offices. *Journal of Business Ethics* 149, 811-831. <https://doi.org/10.1007/s10551-016-3045-6>
- Pagnotoni, P., Dimpfl, T. (2019). Price discovery on Bitcoin markets. *Digital Finance* 1, 139-161. <https://doi.org/10.1007/s42521-019-00006-x>
- Pham, M. H. (2020). In law we trust: Lawyer CEOs and stock liquidity. *Journal of Financial Markets* 50, forthcoming. <https://doi.org/10.1016/j.finmar.2020.100548>
- Putniņš, T. J. (2013). What do price discovery metrics really measure? *Journal of Empirical Finance* 23, 68-83. <https://doi.org/10.1016/j.jempfin.2013.05.004>
- Roberts, M. R., Whited, T. M. (2013). Endogeneity in Empirical Corporate Finance. In *Handbook of the Economics of Finance*, G. M. Constantinides, M. Harris, and R. M. Stulz, Eds. Elsevier B.V., Netherlands, Volume 2, Part A, 493-572. <https://doi.org/10.1016/B978-0-44-453594-8.00007-0>
- Roy, P. P., Rao, S., Zhu, M. (2022). Mandatory CSR expenditures and stock market liquidity. *Journal of Corporate Finance* 72, 102158. <https://doi.org/10.1016/j.jcorpfin.2022.102158>

- Stoll, H. R. (2003). Market Microstructure. In *Handbook of the Economics of Finance*, G. M. Constantinides, M. Harris, and R. M. Stulz, Eds. Elsevier B.V., Netherlands, Volume 1, Part A, 553-604. [https://doi.org/10.1016/S1574-0102\(03\)01013-6](https://doi.org/10.1016/S1574-0102(03)01013-6)
- Tse, Y. (1999). Price Discovery and Volatility Spillovers in the DIJA Index and Futures Markets. *Journal of Futures Markets* 19(8), 911-930. [https://doi.org/10.1002/\(SICI\)1096-9934\(199912\)19:8<911::AID-FUT4>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1096-9934(199912)19:8<911::AID-FUT4>3.0.CO;2-Q)
- Yan, B., Zivot, E. (2010). A structural analysis of price discovery measures. *Journal of Financial Markets* 13, 1-19. <https://doi.org/10.1016/j.finmar.2009.09.003>
- Xu, Y., Xuan, Y., Zheng, G. (2021). Internet searching and stock price crash risk: Evidence from a quasi-natural experiment. *Journal of Financial Economics* 141(1), 255-275. <https://doi.org/10.1016/j.jfineco.2021.03.003>

1 | The Determinants of Price Discovery on Bitcoin Markets*

Oliver Entrop[†], Bart Frijns[‡], Marco Seruset[#]

Journal of Futures Markets 40(5), 819-837, 2020.

Abstract

This paper investigates whether market quality, uncertainty, investor sentiment and attention, and macroeconomic news affect bitcoin price discovery in spot and futures markets. Over the period December 2017 – March 2019, we find significant time variation in the contribution to price discovery of the two markets. Increases in price discovery are mainly driven by relative trading costs and volume, and by uncertainty to a lesser extent. Additionally, medium-sized trades contain most information in terms of price discovery. Finally, higher news-based bitcoin sentiment increases the informational role of the futures market, while attention and macroeconomic news have no impact on price discovery.

Keywords: Bitcoin; Price Discovery; Futures

JEL classification: G12; G13; G14

* We thank the participants of the 2019 Derivative Markets Conference in Queenstown, of the 2019 HypoVereinsbank Doctoral Seminar in Dießen am Ammersee, of the research seminar at the University of Passau and especially Stefan Greppmair (discussant) for very helpful comments and suggestions on an earlier draft of this paper. Parts of this research was done while Oliver Entrop was visiting the Auckland University of Technology. He thanks Aaron Gilbert and the academic and administrative staff for their hospitality and support.

[†] Oliver Entrop, University of Passau, Chair of Finance and Banking, Innstraße 27, 94032 Passau, Germany, phone: +49 851 509 2460, E-mail: oliver.entrop@uni-passau.de

[‡] Bart Frijns, Professor of Finance, Auckland University of Technology, Department of Finance, Private Bag 92006, 1142 Auckland, New Zealand, phone: +64 9 921 9999 5706, email: bfrijns@aut.ac.nz

[#] Marco Seruset, University of Passau, Chair of Finance and Banking, Innstraße 27, 94032 Passau, Germany, phone: +49 851 509 2463, email: marco.seruset@uni-passau.de

1 Introduction

Cryptocurrencies¹ and especially bitcoin² have received increasing attention in the academic finance literature in recent years. Much of this research focuses on issues such as long- and short-term determinants of the exchange value of bitcoin (e.g., Kristoufek, 2015; Li and Wang, 2017; Mai et al., 2018), the market efficiency of bitcoin (e.g., Urquhart, 2016; Köchling et al., 2019), the diversification effects and connectedness of bitcoin with other financial assets (e.g., Brière et al., 2015; Dyhrberg, 2016; Bouri et al., 2017; Corbet et al., 2018), illegal activities (e.g., Foley et al., 2019), or the price discovery process among bitcoin trading venues (e.g., Brandvold et al., 2015; Pagnottoni and Dimpfl, 2019).

In December 2017, the CME and CBOE introduced bitcoin futures, enabling investors to trade and hedge bitcoin on regulated markets. The introduction of this new market raises two important questions related to price discovery. First, which market, i.e., spot or futures, lead the bitcoin price discovery process?³ Second, what are the determinants of price discovery? The first question has been the focus of three recent studies. Corbet et al. (2018), and Baur and Dimpfl (2019) explore price discovery leadership using high-frequency transaction data and find that the spot market incorporates information into prices first and thus dominates in terms of price discovery. In contrast, using daily data, Kapar and Olmo (2019) find that the futures market is the price discovery leader. To the best of our knowledge, the second question on determinants of price discovery for regulated bitcoin futures has not been addressed yet.⁴

¹ According to coinmarketcap.com, over 2,000 cryptocurrencies exist with a total market capitalization surpassing 172 billion US Dollar as of 11 April, 2019.

² A detailed description of the Bitcoin technology is provided in Nakamoto (2008), Kroll et al. (2013) and Boehme et al. (2015).

³ This standard microstructure analysis between spot and futures markets has already been subject for various asset classes, such as stocks (e.g., Hasbrouck, 1995; Booth et al., 1999), exchange rates (e.g., Chen and Gau, 2010) and commodities (e.g., Dimpfl et al., 2017).

⁴ Alexander et al. (2020) analyze unregulated BitMEX bitcoin futures and find that the relative trading volume and the relative bid-ask spread are important determinants of price discovery.

Our study extends the extant literature in two important directions. First, while the studies mentioned above examine price discovery using the full contract term of each separate futures contracts, we consider the liquidity of each contract on each day. Specifically, we determine the daily contribution to price discovery based on the most actively traded futures contract, which allows us to capture the potential dynamics in the relation between spot and futures markets on a day-to-day basis. Using a sample of high-frequency midquotes over the period December 2017 to March 2019, this first-stage analysis demonstrates that price discovery in bitcoin markets is subject to time variation. Using the Gonzalo and Granger (1995) Component Share and Hasbrouck (1995) Information Share, we find that, on average, the futures market leads the price formation process in nine (contract) months, while the spot market is the leader in the remaining (six) months. We further observe that the price discovery measures get closer to 0.5 when increasing time intervals. One of the critical points we raise in this stage is that the spot market does not lead the price discovery process exclusively.

Second, we analyze potential determinants of daily price discovery. First, we examine the effect of market quality and uncertainty (and of some controls). Frijns et al. (2015a) argue that the relation between price discovery and measures of market quality, such as trading costs and trading activity, is potentially endogenous, where an enhancement in price discovery may attract investors to a market, while an increase in liquidity, trading activity, and lower trading costs may improve price discovery. We, therefore, implement 2SLS time-series regressions to control for potential endogeneity. Our results show that trading costs, captured by the relative bid-ask spread, are negatively associated with price discovery, while relative trading volume is positively related to price discovery. Thus, an increase in relative spread (relative trading volume) in one market relative to the other market, decreases (increases) the contribution to price discovery of that market. Quoting activity does not affect price discovery. Furthermore, measures of uncertainty such as volatility of the spot market and VIX partially reveal a

significant shift of price discovery to the futures market. In additional analyses we find that the relative number of medium-sized trades carries most information for the price discovery process.

Besides considering variables of market quality and uncertainty, we also examine the effect of investor sentiment (see Lin et al., 2018), of investor attention, and of macroeconomic news announcements (see, e.g., Chen and Gau, 2010; Fricke and Menkhoff, 2011; Frijns et al., 2015b; Chen and Tsai, 2017) on bitcoin price discovery. The results show that price discovery on bitcoin markets is affected by news-based investor sentiment rather than by investor attention and macroeconomic news.

Baur and Dimpfl (2019) point out that the analysis of bitcoin price discovery may be somewhat different from other asset classes. Given the absence of a bitcoin pricing model, the ambiguity to which asset class the bitcoin even belongs,⁵ as well as the different design of spot (unregulated) and futures markets (regulated), one cannot expect that the results of other asset classes also hold for the bitcoin market ex-ante. The time variation in price discovery we observe in our first stage is in line with the findings in the DAX ETF and DAX futures market (see Schlusche, 2009), in the VIX short-term futures ETN and inverse VIX short-term ETN (see Fernandez-Perez et al., 2018), and in the Indian stock and futures markets (see Karmakar and Inani, 2019). However, studies on price discovery between spot and futures markets usually find the futures market to lead (see, e.g., Hauptfleisch et al., 2016, for gold spot and futures markets; Theissen, 2012, for the DAX spot and DAX futures; Dimpfl et al., 2017, for spot and futures of eight agricultural commodities).

In our second-stage analysis, we observe a significant effect of trading costs and trading volume on price discovery. This is consistent with other studies that have also focused on the

⁵ In fact, there is an ongoing discussion as to whether bitcoin is a speculative asset or a currency (see, e.g., Glaser et al., 2014; Yermack, 2015; Baur et al., 2018). Contributing to this discussion is beyond the scope of this paper.

relation between market quality and price discovery on spot and derivatives markets (see, e.g., Chakravarty et al., 2004, for stocks and stock option markets; Fernandez-Perez et al., 2018, for VIX short-term futures ETN and inverse VIX short-term ETN). Our results concerning uncertainty suggest that the relative contribution of the futures market to price discovery is higher when volatility on the bitcoin spot market and stock markets is higher. For spot market volatility, our findings contrast the stock and stock options markets (see Chakravarty et al., 2004), but are in line with the foreign exchange spot and futures markets (see Chen and Gau, 2010). The mechanism of the VIX relating to bitcoin price discovery is difficult to assess and has not been addressed in such a setting. Finally, our results suggest that news-based bitcoin sentiment is important for the incorporation of information in futures markets. Overall, our findings imply that the price discovery on bitcoin markets is not too different from other asset classes.

The remainder of this paper is organized as follows. Section 2 describes the data and presents summary statistics. In Section 3, we present the model used to evaluate price discovery, present our empirical results, and discuss several robustness tests. Section 4 focuses on the determinants of price discovery and reports results of our second-stage analysis. We conclude in Section 5.

2 Data

This study concentrates on the dynamic relation between bitcoin spot and futures prices from December 17, 2017 to March 31, 2019. We consider intraday trade and quote data for bitcoin futures traded on Chicago Mercantile Exchange (CME) as well as the corresponding spot of the

Bitstamp (BTSP) exchange. We obtain these data from the Thomson Reuters Tick History (TRTH) database.⁶

The transaction data include the timestamp to the nearest millisecond, the traded price, and associated volume. The quote data consist of the bid and ask quotes, and the exact timestamp a new quote is issued. From this, we calculate the midpoint (average of bid and ask quotes) for spot and futures.

CME bitcoin futures (RIC: BTC) are US dollar-denominated cash-settled contracts, based on the CME CF Bitcoin Reference Rate (BRR), having a contract size of five bitcoins. The BRR aggregates the weighted median USD price for four major exchanges (Bitstamp, Coinbase, itBit, and Kraken) once a day. Trading in expiring futures contracts terminates at 4 pm London Time on the expiration day. The trading hours for CME futures contracts are between 5 pm and 4 pm Chicago Time (CT) from Sunday to Friday with a 60-minute break each day beginning at 4 pm CT.⁷

We follow Baur and Dimpfl (2019) and select the Bitstamp spot as the spot price (we do not use the daily available Bitcoin Reference Rate (BRR) nor its continuous version (Bitcoin Realtime Index – BRTI) because investors cannot trade these indices). Bitstamp is one of the largest cryptocurrency spot trading platforms, where bitcoin can be traded against USD (RIC: BTC=BTSP).⁸

The analysis of the daily behavior of price discovery requires a continuous futures time series. We follow Fricke and Menkhoff (2011) and Hauptfleisch et al. (2016) and use the most actively traded futures contract on each day in our sample. An alternative procedure in empirical studies is to use the nearest-to-maturity futures contract (e.g., Booth et al., 1999; Cabrera et al.,

⁶ Note that we do not consider futures contracts traded on Chicago Board Option Exchange (CBOE). First, CBOE stopped futures listing as of March 2019. Second, notional trading volume on CME is superior to CBOE from March 2018 onwards. Therefore, we assume the CME to be the relevant futures market.

⁷ See https://www.cmegroup.com/trading/equity-index/us-index/bitcoin_contract_specifications.html for more details.

⁸ See <https://www.bitstamp.net/> for more information.

2009). In our case, however, there are only minor differences when comparing the time series resulting from both methods. In particular, the most actively traded futures contract equals the nearest-to-maturity contract until one business day before maturity. At that point, volume shifts to the second-nearby contract, implying that the closest-to-maturity contract is no longer the most actively traded.

Another important issue of data preparation relates to the trading hours of the futures contracts. Similar to Grammig et al. (2005), we consider overlapping trading hours between spot and futures only. We further follow the procedure of Hauptfleisch et al. (2016) and delete all entries before 0 am and after 8 pm GMT. This avoids the need to deal with market closures on CME and time zone adjustments, thus simplifying our two-stage analysis. Finally, we remove all observations on holidays according to CME holiday calendar.

[Table 1 about here]

Column 2 of Table 1 shows the time interval in which the respective futures contract (RIC) is the most actively traded. Column 4 presents the total daily volume of the most-traded futures (MTF) in the respective time period. Interestingly, volume increases nearly monotonically until August 2018, while we observe a more volatile behavior of volume after August 2018 until the end of the sample. Column 5 emphasizes the importance of using the most actively traded futures contracts for analyzing the dynamic price discovery process. For example, BTCQ8 exhibits an average proportion of 96.65%, indicating that there is almost no trading in other contracts at that time. This strong shift in liquidity between futures contracts may favor previous empirical results of spot-driven price discovery (see, e.g., Corbet et al., 2018; Baur and Dimpfl, 2019) when futures contracts are considered over their whole life span.

Finally, the analysis of price discovery between spot and futures can be conducted on either quotes or transaction prices. Several studies have already discussed the advantages of using midquotes over transactions data (see, e.g., Shyy et al., 1996; Eun and Sabherwal, 2003; Grammig et al., 2005; Theissen, 2012). The use of quote midpoints implies three main advantages. First, quotes can be updated in the absence of transactions. Second, midquotes mitigate the problem of infrequent trading, which is normally observed in transaction prices. Third, midquotes are not affected by the bid-ask bounce. Hence, we base our analysis on midquotes.

We estimate the contribution to price discovery of the spot and futures separately for each day in our sample period to capture the dynamic behavior of the price formation process. Since midquotes of bitcoin spot and futures are not uniformly spaced in time, we construct synchronized time intervals to align the spot and futures data. Within each time interval, we keep the last observed midquote. If no midquote is observed, we fill missing intervals with the most recent non-missing value (see, e.g., Chan, 1992; Chen and Gau, 2010).⁹ The choice of sampling interval is an important issue when studying price discovery. Brandvold et al. (2015) and Jin et al. (2018) point out that it is important to keep time intervals short enough to ensure information is not lost between sampling intervals, but also long enough to avoid noise due to stale prices. Following Jin et al. (2018), we consider various sampling frequencies. In particular, we compute the non-synchronous quoting probability, as well as the frequency of zero-returns as zero-returns are an important indicator of liquidity differences between spot and futures markets (see Theissen, 2012). It should be noted, however, that different trading activity and different liquidity does not necessarily have to be an indication of the leading market (see, e.g., Theissen, 2012; Jin et al., 2018).

⁹ For an alternative procedure of constructing a matched sample of midquotes see Harris et al. (1995).

Table 2 reports the trading frequency and the proportion of zero-returns. We observe a lower proportion of missing quotes on the spot market. On average, the non-synchronous quoting for one-minute intervals is 0.35% and 4.40% for the spot and futures market, respectively. Non-synchronous quoting decreases as we increase the time interval. When we consider the proportion of zero returns, however, figures increase substantially. Zero returns for spot and futures prices occur in 22.20% and 43.37% of the one-minute return intervals, respectively. Thus, midquotes change more frequently in the spot market than in the futures market. We proceed with our price discovery analysis using one-minute intervals, but also consider five-, ten- and fifteen-minute intervals for robustness purposes in our first stage.

[Table 2 about here]

Table 3 presents summary statistics for one-minute intervals based on midquotes. The average quote midpoint is 7,035 for spot and 7,031 for futures. Bitcoin spot and futures midquotes show a declining trend, which results in a negative return of almost 80% from the start to the end of our sample period.

The non-synchronicity between spot and futures is remarkably low for all contracts in our sample, which again supports our decision to analyze price discovery on a one-minute frequency. However, figures increase when we consider the evolution of zero returns, where futures always exhibit a higher percentage of zero returns than the spot. In terms of percentage changes, however, the pattern is not uniform over the sample period. The percentage of zero returns increases fivefold between the January (BTCH8) and June contract (BTCM8) for spot and futures. In the subsequent contract months, the percentage of zero returns increase for the spot market, while the futures market's zero returns decrease. After the September contract

(BTCU8), the spot and futures market reveal nearly a doubling in the percentage of zero returns until March 2019 (BTCH9). The growth in the zero returns is more volatile than before.

[Table 3 about here]

3 Price discovery

To study the dynamics of the price discovery process between bitcoin spot and futures prices, we apply the standard approach of estimating a vector error correction model (VECM) and deriving our price discovery measures directly from the outcome of the VECM. We use two of the most important price discovery measures for non-stationary price series, the Gonzalo and Granger (1995) Component Share (CS), and the Hasbrouck (1995) Information Share (IS). Subsequently, we present the results of the VECM as well as the price discovery measures.

3.1 Vector error-correction model and price discovery measures

We are interested in questions related to the intra-day relation between bitcoin spot and futures prices. Suppose Bitstamp spot has a log US dollar price s_t , and f_t denotes the log US dollar price of the CME futures. Let $y_t = (s_t \quad f_t)'$ be the vector of these price series. Given the cost-of-carry relation between spot and futures prices, the respective log price series should be integrated of order one, $I(1)$, with cointegrating vector $\beta' = (1 \quad -1)$ (see Baur and Dimpfl, 2019). Therefore, price changes can be expressed as an error correction equation of the form

$$\Delta y_t = \alpha(\beta' y_{t-1} + \mu) + \sum_{i=1}^p \Gamma_i \Delta y_{t-i} + \varepsilon_t, \quad (1)$$

where Δy_t is the (2 x 1) vector of changes in the log series of the spot and futures price at time t . α is a (2 x 1) vector for the bitcoin spot and futures prices measuring the speed of adjustment of short-term deviations from the long-term equilibrium. Our specification of β' implies that we expect $\alpha^{Spot} \leq 0$ and $\alpha^{Futures} \geq 0$. μ is a constant term¹⁰ in the cointegrating equation, and Γ_i are (2 x 2) matrices of autoregressive prices, representing the short-term transitory effects due to market imperfections. ε_t is a zero-mean vector of serially uncorrelated innovations with the following covariance matrix:

$$\Omega = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad (2)$$

where σ_1^2 (σ_2^2) is the variance of spot market innovations (futures market innovations) and ρ is the correlation between these innovations.

Appendices A and B outline the calculation of the Component Share (CS) and Information Share (IS) from the outcome of Equation (1). Values above (below) 0.5 suggest that the spot (futures) market leads the price formation process.

Frijns et al. (2015a) point out that the IS may be biased when liquidity increases over time. In such a case, a rise in liquidity increases the contemporaneous correlation and widens the lower and upper bound.¹¹ This bias causes the IS to move towards 0.5 for both markets (see Yan and Zivot, 2010). Indeed, we observe that liquidity of the spot and futures market has changed over time (see Section 2). For this reason, we calculate the CS and IS for each day in

¹⁰ Note that this constant term refers to the restricted constant specification as defined by Johansen (1995). According to Hansen and Juselius (1995) this is the minimum deterministic component recommended by Johansen (1995). This allows the cointegrating equations to be stationary around a constant mean, which seems appropriate for daily estimation procedure. We conduct all analysis based on this specification pointing out, however, that our results are robust to the choice of the deterministic component, i.e. results of price discovery remain qualitatively and quantitatively the same for constant or restricted trend specifications. For a more detailed discussion, see Ahking (2002).

¹¹ The daily contemporaneous correlation between the residuals is 79%, on average.

our sample, but focus only on the CS in our second-stage analysis. As this problem also affects the Information Leadership Share (ILS) (see Yan and Zivot, 2010; Putniņš, 2013), we do not consider this measure.

3.2 Empirical analysis

Analyzing the price discovery process of two time series requires data to be cointegrated. For this purpose, we determine the number of cointegrating equations by Johansen's (1995) trace statistic method.¹² We determine the lag length included in the model by the multivariate version of Schwartz's Bayesian Criterion (SBIC).¹³ Our first step is to test whether there are at most zero cointegrating vectors for each day in our sample. The null hypothesis of $r = 0$ cointegrating vectors is rejected for around 91% of the days at the 1% level. In the next sequence, the null hypotheses of $r = 1$ cointegrating vectors cannot be rejected for about 79% of those days. We thus discard 21% of days from our data set as the inclusion of days where we observe no cointegration can produce misleading and distorted results (see Fricke and Menkhoff, 2011). The mean cointegrating equation is $\beta' = (1 \quad -0.89307)$. However, we cannot reject the null hypothesis that the cointegrating relation is $\beta' = (1 \quad -1)$ at the 5% level.¹⁴

¹² We additionally perform unit root tests for both series for each day in our sample. Results of Augmented Dickey-Fuller tests for the log-levels of spot and futures reveal that roughly 82% of the days are non-stationary (at the 1% level), while first differences are always stationary.

¹³ The average lag length for each day is $p = 3$.

¹⁴ For detailed results of the VECM estimation see Table A1 in the Appendix. By definition of the VEC model stated in (1), β_{Spot} is 1 and, by theory, $\beta_{Futures}$ is -1 . Due to outliers in beta estimations, we observe that the mean beta significantly deviate from the theoretical value in contract months M8 and F9. Additionally, t-values of beta estimates are significant at the 1% level in six out of fifteen contract months, indicating that the cointegrating vector does not hold. These indistinct results, however, are in line with the findings of Baur and Dimpfl (2019). The median value turns out to be the better indicator in this case, where we observe a reasonably tight range of median figures. Therefore, we assume that the theoretical cointegrating equation $\beta' = (1 \quad -1)$ holds for all days in our sample.

We confirm the presence of one cointegrating relation on almost 80% of the days in our sample. Our next aim is to investigate the price discovery dynamics between bitcoin spot and futures using two measures of price discovery, the Component Share (CS) and the Information Share (IS). Once again, it is worth noting that the results of price discovery refer to the spot market and that values above (below) 0.5 indicate that the spot (futures) market is the leading market.

Table 4 reports the CS and IS for each most-traded futures (see Table 1) in our sample, based on one-minute intervals (Panel A). We document that the futures market leads the spot market in nine contract months (price discovery measures < 0.5), while three months are significant at the 1% level (Column 2). The spot market is the leading market in the remaining months with two significant months (price discovery measures > 0.5). Over the full sample period (Panel B), we, on average, observe that price discovery measures are close to 0.5. Overall, the IS produces similar results with respect to the price discovery leader, however, with one more significant contract month at the 5% level (BTCQ8). In summary, the importance of spot and futures market in incorporating new information changes over time. The variability is also visualized by the 5-day moving average in Figure 1, which is calculated from the daily CS and IS. We, again, point out that the time variation in price discovery can also be observed in other asset classes, such as DAX ETF and DAX futures market (see Schlusche, 2009) as well as in the VIX short-term futures ETN and inverse VIX short-term ETN (see Fernandez-Perez et al., 2018) and in the Indian stock market (see Karmakar and Inani, 2019).

[Table 4 about here]

[Figure 1 about here]

Considering the distributional properties of the price discovery measures over the contract months (Table 4, Panel A), CS compared to IS, is more volatile and reveals a wider difference between the 95th and the 5th percentiles. Moreover, on average, we observe a lower standard deviation for significant contract months, ranging from 15.3% to 22.7% for CS, and from 6.4% to 12.2% for IS, respectively.

For robustness purposes, we replicate our analysis for five-, ten- and fifteen-minute intervals. Table 5 documents the CS and IS for the different sampling intervals. In line with Jin et al. (2018), price discovery shares get closer to 0.5 when lower-frequency intervals are used, on average. Stated differently, the differences in price discovery shares between the spot and futures market are less when increasing time-intervals (see Tse et al., 2006, for similar results). This fact confirms that information transmission between the spot and futures market takes less than fifteen minutes.

[Table 5 about here]

Finally, while the bitcoin price declines over our sample period on average, there are still sub-periods of rising prices. To assess whether there are differences between periods showing (short-term) rising and falling prices, we define a dummy variable which is one (= short-term bear market), if the mean return over the last five days is negative and zero otherwise (see, e.g., Chen, 2009). Comparing price discovery for both states of the dummy variable, we find no evidence that price discovery significantly differs between short-term falling and rising markets in our sample.¹⁵

¹⁵ Note that this conclusion does not change when using the mean returns over the last 2 or 10 days.

4 Determinants of price discovery

4.1 Potential determinants and summary statistics

In our second-stage analysis, we examine different variables that may explain our previous price discovery findings. For this purpose, we consider six sets of variables.

Market Quality

The first set of variables capture various aspects of market quality, such as trading activity or trading costs of the bitcoin spot and futures market. Following earlier studies (e.g., Frijns et al., 2015a; Fernandez-Perez et al., 2018), we consider the relative number of quotes ($rel_num_Quotes_t$), which is the number of quotes on the spot market divided by the number of quotes on the futures market on day t . We also take into account the relative trading volume ($rel_vol_trades_t$), which is the volume of contracts traded on the Bitstamp spot market divided by the volume of traded contracts on the CME futures market on day t . The variable rel_BAS_t is defined as the daily average percentage bid-ask spread on the Bitstamp spot market divided by the daily average percentage bid-ask spread on the CME futures market.

We also consider the relative size of each trade in a sub-analysis. In particular, we decompose the relative traded volume into small, medium, and large trades. Large trades ($rel_num_large_trades_t$) are those of five futures contracts¹⁶ or five bitcoins, respectively, or more; small trades ($rel_num_small_trades_t$) are defined with a respective number of less or equal one, while medium-sized trades ($rel_num_medium_trades_t$) are those with a respective number of more than one and less than five.¹⁷

¹⁶ This boundary refers to the block trading limit of CME, where trades are negotiated manually between the exchange and investors. See <https://www.cmegroup.com/education/bitcoin/cme-bitcoin-futures-frequently-asked-questions.html> for more details.

¹⁷ Note that the definition of different trading sizes is not homogenous in literature. Some researchers define the trading sizes according to the contract volume (e.g., Barclay and Warner, 1993; Eun and Sabherwal, 2003;

Uncertainty

Our second set of variables contains several measures of uncertainty. We include the Bitstamp spot market volatility ($vola_{t,spot}$), which is defined as the square root of the sum of the squared 1-min returns for each day in our sample, similarly done by Chakravarty et al. (2004), Chen and Gau (2010), and Lin et. al (2018). This variable serves as a proxy of the uncertainty on the bitcoin market. We also include the daily log-return of VIX (ret_VIX_t), which is often used as a proxy of fear on stock markets, or even as a general fear measure for capital markets. In addition, we consider the daily log-return of the economic policy index (ret_EPU_t), which was developed by Baker et al. (2013) for the US. It serves as a proxy of real economic policy uncertainty.

Data on market quality and uncertainty are either calculated from the data as described in Section 2, collected from Thomson Reuters Eikon or downloaded from the economic policy uncertainty website (<https://www.policyuncertainty.com>, EPU, 2019).

Sentiment

Our third set of variables contains daily bitcoin sentiment data from Thomson Reuters MarketPsych (TRMI). Sentiment is defined as the overall positive references net of negative references (see Nooijen and Broda, 2016). As already done in literature (see, e.g., Nooijen and Broda, 2016), we consider contents derived both from news and social media, respectively, to capture the sentiment of both professional and retail investors. The first sentiment category covers more sophisticated sources, such as *The New York Times* or *The Wall Street Journal*, while the latter sources news from less formal sites, such as *Yahoo! Finance* or *Blogger* (for

Frijns et al., 2015a), while others consider also the transaction volume of each trade (e.g., Lee and Radhakirshna, 2000).

more information on TRMI see Nooijen and Broda, 2016; Griffith et al., 2019; Drobetz et al., 2019). For our analysis, we follow Lin et al. (2018) and investigate the impact of high sentiment periods on price discovery. For this purpose, we define two sentiment dummy variables ($Dummy_HighSent_t$), which take the value of one, if the respective TRMI on day t is above the median TRMI during the previous 20 days, zero otherwise (see Ben-Rephael et al., 2017, for a similar approach), where we consider news-based and social-based sentiment separately.

Attention

We consider (retail) investor attention by using Google Trends Search Volume Index (SVI, 2019) (website: trends.google.com; key word: “bitcoin”; region: US; period: December 01, 2017 to March 31, 2019).¹⁸ In the spirit of Da et al. (2011), we define a high attention dummy ($Dummy_HighAtt_t$), which takes the value of one, if the SVI in the current week is above the median SVI during the previous 8 weeks, and zero otherwise. As Google Trends provides only weekly data and reports the respective value on each Saturday, we use this value for each day in the subsequent week (see Mai et al., 2018, for a similar approach).¹⁹

Macroeconomic news announcements

There is some literature that also investigates the effect of announced macroeconomic news releases on price discovery such as Chen and Gau (2010) in foreign exchange markets, Fricke and Menkhoff (2011) for the Euro bund market, Frijns et al. (2015b) for Canadian-U.S. cross-listed firms and Chen and Tsai (2017) for the VIX spot and futures markets. We follow this approach and consider some key macroeconomic news listed in Table A2 (MA, 2019). We

¹⁸ A more detailed discussion on different issues regarding the identification of the key word is given in Da et al. (2011).

¹⁹ Our results remain unchanged if we move from daily to weekly data by taking weekly averages of all variables.

define a macroeconomic announcement dummy variable ($Dummy_MEA_t$), which takes the value of one if a news is released on day t , and zero otherwise.

Controls

The final set of variables represents two control variables. In particular, we use the daily log-returns on Bitstamp exchange (ret_BTSP_t) to assess whether the direction of the spot returns affects price discovery. Finally, we include the daily log-returns of the front-end contract of the Gold futures (COMEX), denoted as ret_Gold_t , serving as a proxy for the demand for financial safety in times of economic turmoil. These data are collected from Thomson Reuters Eikon.

Table 6 reports descriptive statistics for all variables that we consider in our second-stage analysis. We focus on characteristics of the market quality variables here (see the upper part of Table 6). The spot market (Panel A) has a lower quoting and trading activity than the futures market (Panel B) over the full sample period. In particular, the daily average number of quotes is 33,524 and 56,604 for the spot and futures market, respectively. Moreover, the average traded volume is higher on the futures market (14,242) than on the spot market (9,762). For trading costs, we find that the spot market is the cheaper market. Finally, we report summary statistics for the different trading size groups. These figures reveal that the number of trades is much higher on the spot market than on the futures market. The explanation underlying this result refers to the fact that bitcoin is divisible into smaller units, while this is not possible on the futures market. Especially the number of small trades is exceptionally high on the spot market. The possibility of trading bitcoin contracts in smaller fractions potentially attracts retail investors allowing them to participate with a small investment.²⁰

[Table 6 about here]

²⁰ The minimum unit of bitcoin is the “Satoshi”, which is 0.00000001 bitcoin.

4.2 Empirical strategy

To assess the influence of the six sets of variables on price discovery measured by the Component Share as discussed in Section 3.1, we estimate the following equation:

$$\begin{aligned} \text{logit_CS}_t = & \beta_0 + \delta' \text{MarketQuality}_t + \gamma' \text{Uncertainty}_t + \lambda' \text{Sentiment}_t \\ & + \eta' \text{Attention}_t + \tau' \text{Macro_News}_t + \theta' \text{Controls}_t + \varepsilon_t, \end{aligned} \quad (3)$$

where logit_CS_t is the logit transformation of the spot market Component Share, which allows the mapping of the original variable, which was bounded between zero and one, to the other variables. MarketQuality, Uncertainty, Sentiment, Attention, Macro_News, and Controls are the respective vectors of variables presented in Section 4.1, where we apply the natural logarithm on the market state variables (i.e., on quotes, trading volume, spread, and spot market volatility).²¹ We further use the Variance Inflation Factor (VIF) to test for multicollinearity in Equation (3). The VIF is always below 2.88 for all presented subsequent analyses, suggesting that multicollinearity is not an issue in our setting.

We analyze the relation between Component Shares and the above mentioned explanatory variables for two periods. First, we consider the whole sample period, which ranges from December 2017 to March 2019. Second, we look at the period from end of March 2018 through March 2019, which leaves out the establishment stage of the CME bitcoin futures market (futures transactions volume started very low (see Hale et al., 2018) and we avoid any liquidity issues by leaving out the first three months of trading). As this reduced sample seems to be more reliable with regard to the explanatory power, we primarily focus on these results.

²¹ When adding a dummy for the halt of futures trading on CBOE (15 March 2019) to Equation (3) we do not find any change in our results.

Following Frijns et al. (2015a), we consider potential endogeneity issues when investigating the determinants of price discovery. In particular, we expect reverse causality between variables of market quality and CS. An improvement in price discovery may enhance several aspects of market quality. Concurrently, lower trading costs, increased liquidity, or trading volume may improve price discovery as well. Since the presence of simultaneity would produce biased estimates in an OLS framework, we employ a 2SLS estimator to capture the influence of market quality on CS.

Unreported tests reveal that the relative number of quotes, as well as the relative trading volume, are potentially endogenous in Equation (3).²² We use lag one of relative number of quotes, relative trading volume, and CS, as internal instruments (see Wintoki et al., 2012; Frijns et al., 2015a, for a similar procedure). Various specification statistics show that our instruments are valid and that we cannot reject the null hypothesis of exogeneity of our instruments (see Frijns et al., 2015a, for similar results on diagnostic statistics).

4.3 Empirical results

Firstly, we present in Section 4.3.1 the results when only market quality and uncertainty variables and controls are considered. Subsequently, in Section 4.3.2, we incorporate all variables as presented in Equation (3).

4.3.1 Market quality and uncertainty

Market quality

²² The detailed results of our pre-analysis are available upon request. We particularly find potential endogeneity in our reduced sample period. A comprehensive description of the underlying intuition of the conducted tests is provided in Wintoki et al. (2012) and Frijns et al. (2015a).

The results for the impact of market quality on price discovery (CS) in Table 7 show that the number of quotes is insignificant in all model specifications and considered time periods, indicating that there is no relation between price discovery and quoting activity. In contrast, the relative trading volume exhibits positive and significant coefficients at the 10% levels in models (1) and (2), respectively. This result indicates that an increase in trading volume on Bitstamp spot market relative to the CME bitcoin futures market is associated with an increase in price discovery on the spot market. For the whole sample period, however, the significance of relative trading volume disappears. Recall that, as discussed before, the relations between price discovery and explanatory variables may be distorted in the full sample period due to the maturing stage of CME futures.

[Table 7 about here]

For the relative spreads all four specifications show negative and significant coefficients, suggesting that a decrease of trading costs in the spot relative to the futures market leads to an increase in price discovery of the spot market and vice versa. These results confirm that the cost of trading is an important determinant of where (informed) traders execute their trades and where information enters the market.

Overall, these findings are in line with the results on other asset classes, such as stocks and stock options markets (see, e.g., Chakravarty et al., 2004), foreign exchange rates (see, e.g., Chen and Gau, 2010), or volatility products (see, e.g., Fernandez-Perez et al., 2018).

Uncertainty

The uncertainty variables exhibit negative coefficients that are significant for the volatility of the spot market in model (1) and (2) and for the VIX in model (1), while the return on EPU

has no (significant) effect. This implies that higher market volatility in the bitcoin spot market and higher fear in the stock market tend to increase price discovery on the bitcoin futures market.

The significant negative impact of spot market volatility on price discovery indicates that during times of high spot volatility, (informed) traders prefer to trade in the futures market. This finding could be a result of the hedging role of the bitcoin futures market when risk increases on the spot market. Chen and Gau (2010) find similar results on foreign exchange spot and futures markets, while Chakravarty et al. (2004) discover the opposite channel on stock and option markets. For the significantly negative relation between the return of the VIX and the CS of the spot market, there is no straightforward ex-ante intuition as to why information enters bitcoin futures markets during times of high stock market volatility. Given the negative relation between the levels of VIX and bitcoin price (around a correlation of -41% in our reduced sample period) and the interpretation of the VIX as a general fear measure, the underlying reason may also be related to hedging demand.

Once we target to the other models, however, these coefficients are no longer significantly related to price discovery, as before.

Controls

The control variables are insignificant in our reduced sample period, while the Bitstamp return turns significant in the full sample period (see model (3) and (4)). However, this result between return on Bitstamp and price discovery is unstable, i.e., if we consider the results in Section 3.2, we do not find a significant effect of short-term rising and falling prices on price discovery.

Additionally, we consider the influence of the trade size (small/medium/large trades) on price discovery. This analysis refers to the question of which trades have the highest price

impact. Previous studies (e.g., Barclay and Warner, 1993; Chakravarty, 2001; Eun and Sabherwal, 2003) document that most information is conveyed by institutional investors, who use medium-sized orders. The so-called stealth trading hypothesis (Barclay and Warner, 1993) indicates that investors avoid to give away their information too easily by splitting large trades into smaller orders. Medium-sized orders emerge as an optimal point between trading costs and the price impact of transactions (e.g., Chakravarty, 2001).

We report the results of the different trading volume groups in Table 8 for the period March 2018 through March 2019. In line with the previous studies, the relative number of medium-sized trades is significant, while the relative number of small and large trades are insignificant. In addition, Bitstamp returns turn significant in specifications (1), and (3). The results of the different trading volumes suggest that medium-sized orders are more informative than small and large trades. Hence, the more medium-sized trades occur in one market relative to the other market, the higher, on average, is the price discovery in the respective market. This finding is consistent with the stealth trading hypothesis mentioned above. Due to our data structure, however, we cannot evaluate which (informed) investors (e.g., bitcoin miners, banks, or exchanges) conduct these medium-sized trades.

[Table 8 about here]

4.3.2 Adding investor sentiment, investor attention and macroeconomic news

This section considers the variables from Equation (3) together. Table 9 presents the results. We first note that the results for the market quality and uncertainty variables remain unchanged.

[Table 9 about here]

Sentiment

High news-based bitcoin sentiment has a significantly negative impact on the CS of the spot market in the reduced sample period (model (1) and (4)), which implies that the futures price contribute more to the bitcoin price discovery process when news sentiment is high.²³ Given the interpretation of news-based indices as a measure for sentiment of professional investors (see Nooijen and Broda, 2016), this may suggest that informed traders are more willing to trade on the futures market during high news-sentiment periods, thus makes futures prices relatively more informative during such periods. This is in line with Ben-Rephael et al. (2017), who find that institutional attention triggers more trading and responds more quickly to news. Turning our focus to social-based sentiment, the coefficients are not statistically significant.

Attention

The high attention dummy has a positive impact on price discovery, however, with no coefficient being statistically significant (model (2) and (4)). Thus, we conclude that high attention periods cannot explain the time variation in price discovery on bitcoin markets.

Macroeconomic News

Among US news, there is no significant effect of announcement days on bitcoin price discovery (model (3) and (4)).

5 Conclusion

This paper examines the evolution of bitcoin price discovery as well as the determinants of the calculated price discovery measure. Using Component Share and Information Share in our first

²³ If the whole sample period is considered, only the coefficient of the relative BAS remains statistically significant, as in the previous analyses.

stage of the analysis, we find that the price discovery measures are subject to time variation, suggesting that the leading market has changed over time. These findings reveal that price discovery is not limited to the spot market when considering the most liquid futures contract on each day. In particular, our results show a clear price leadership of the futures market from the start of the sample until the middle of 2018. On the contrary, we find evidence that the spot market is the leading market at the end of our sample. Our robustness analysis with increased time intervals shows that the information transmission between spot and futures market takes less than fifteen minutes.

In our second stage, we find strong evidence that the relative bid-ask spread negatively affects price discovery. Furthermore, we show that the relative trading volume has a positive effect on price discovery that is, however, not always statistically significant. For the relative number of quotes, we find no effect on price discovery. We further document a negative relation between spot market volatility and price discovery, which we attribute to the hedging demand of informed investors in times of high spot market volatility. Among the control variables, we do not find a stable effect on price discovery. In a further sub-analysis, we report that medium-sized trades affect the price discovery process most, suggesting that institutional investors potentially split large trades into medium-sized trades. Our results thus imply that an enhancement in market quality, such as lower trading costs and higher trading activity, has a positive causal effect on price discovery.

In extending the determinants of price discovery we complement market quality and uncertainty variables by sentiment, attention and macroeconomic news variables. News-based bitcoin sentiment has been found to be a relevant measure in our reduced sample period, while the attention and macroeconomic news play no significant role in explaining the time variation in CS.

The bitcoin, as an emerging innovation in recent years, has received much attention due to its unique features. Despite the still existing ambiguity of the bitcoin universe, our research shows that, at least, the analysis of determinants on price discovery leads to economically reasonable results, which can also be found in other asset classes. However, the causal channel between VIX and price discovery is still unclear at this point.

Of course, comprehensive data on participating traders, and their classification into informed and uninformed traders, would allow us to even better explain the observed time variation in price discovery. For example, there is anecdotal evidence that bitcoin miners participate in the bitcoin futures market when prices move towards the mining costs. This may cause the futures market to lead the price discovery in this phase as miners potentially hedge downside risk. Unfortunately, we cannot address the underlying structure in price discovery shifts in more-depth as our data on the does not allow us to uncover the involved players.

Appendix

Appendix A: Component Share (CS) measure

Following Baillie et al. (2002) we compute the daily Component Share as

$$\gamma_{Spot,t} = \frac{\alpha_t^{Futures}}{\alpha_t^{Futures} - \alpha_t^{Spot}}, \quad (4)$$

where γ_{1t} is the Component Share of the spot market on day t . Likewise,

$$\gamma_{Futures,t} = 1 - \gamma_{Spot,t}. \quad (5)$$

The CS equation does not prevent the error-correction coefficients from being negative. Since the size, and not the sign, plays an important role in the price discovery process, we follow Cabrera et al. (2009) and restrict the factor weights to be positive. In our case of a two-market system, we define the CS as

$$CS_{1,t}^{Spot} = \gamma_1 = \frac{|\alpha_t^{Futures}|}{|\alpha_t^{Futures}| + |\alpha_t^{Spot}|} \text{ and } CS_{2,t}^{Futures} = \gamma_2 = \frac{|\alpha_t^{Spot}|}{|\alpha_t^{Futures}| + |\alpha_t^{Spot}|}, \quad (6)$$

where $CS_{1,t}^{Spot}$ is the daily Component Share for the bitcoin spot market, and $CS_{2,t}^{Futures}$ is the daily Component Share for the bitcoin futures market. The sum of the Component Shares equals one.

Appendix B: Information Share (IS) measure

Skipping the VMA representation, Hasbrouck (1995) defines $\psi\Omega\psi'$ as the variance of the common factor shocks. If we assume that two markets of interest are uncorrelated, then Ω is diagonal, and the information share IS_j of the distinct market j to the total variance is given by

$$IS_j = \frac{\psi_j^2 \Omega_{jj}}{\psi\Omega\psi'}, \quad (7)$$

where ψ_j is the contribution of the corresponding market to the total variance. Following Baillie et al. (2002), we compute the Information Share directly from the results of the VECM. The authors show that $\psi = (\psi_1 \ \psi_2)$ is directly related to the common factor component, which means that

$$\frac{\psi_1}{\psi_2} = \frac{\gamma_1}{\gamma_2}. \quad (8)$$

Thus, we can substitute (8) into (7) and receive the contribution of the market shocks on one market to the total variance, i.e., the information share, as

$$IS_j = \frac{\gamma_j^2 \sigma_j^2}{\gamma_1^2 \sigma_1^2 + \gamma_2^2 \sigma_2^2}, \quad (9)$$

where j represents the market of interest, and σ_1^2 and σ_2^2 is the variance of the bitcoin spot and futures, respectively. If the innovations of the two markets are contemporaneously correlated, i.e., $\rho \neq 0$, Hasbrouck (1995) uses the Cholesky factorization of $\Omega = MM'$ to adjust for the correlation. The Information Shares can be expressed in our bivariate market system as

$$IS_{1,t}^{Spot} = \frac{(\gamma_1 m_{11} + \gamma_2 m_{12})^2}{(\gamma_1 m_{11} + \gamma_2 m_{12})^2 + (\gamma_2 m_{22})^2} \text{ and } IS_{2,t}^{Futures} = \frac{(\gamma_2 m_{22})^2}{(\gamma_1 m_{11} + \gamma_2 m_{12})^2 + (\gamma_2 m_{22})^2} \quad (10)$$

where $M = \begin{pmatrix} m_{11} & 0 \\ m_{12} & m_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{\sigma_2(1-\rho^2)} \end{pmatrix}$, and γ_j is the contribution of each market to the total innovations. Since the calculation of the Information Shares is impacted by the order of the market price series in the Cholesky factorization, we follow Baillie et al. (2002) transposing the order of the bitcoin spot and futures markets, and take the simple average of the lower and upper bound.

Table A1: VEC model results

	(1) Obs.	(2) α_{spot}	(3) α_{fut}	(4) β_{mean}	(5) β_{median}
Panel A: Contract by contract					
BTCF8	23	-0.0892 (-4.767)	0.0542 (3.692)	-0.9982 (0.121)	-0.9908
BTCG8	12	-0.0705 (-2.025)	0.0722 (2.668)	-1.0177 (-0.857)	-1.0093
BTCH8	21	-0.0963 (-5.269)	0.0395 (2.017)	-1.0970 (-0.768)	-0.9867
BTCJ8	19	-0.0965 (-5.092)	0.0569 (1.834)	-0.8704 (1.699)	-0.9647
BTCM8	14	-0.0911 (-7.710)	0.0836 (3.320)	-0.9667 (2.612)	-0.9694
BTCN8	19	-0.1376 (-7.152)	0.0142 (0.990)	-0.4741 (1.077)	-0.9550
BTCO8	11	-0.1264 (-9.341)	0.0171 (1.027)	-0.9606 (2.555)	-0.9889
BTCQ8	22	-0.1067 (-5.267)	0.0492 (2.162)	-0.9686 (3.999)	-0.9723
BTCU8	15	-0.1414 (-7.820)	0.0084 (0.322)	-0.9125 (3.918)	-0.9425
BTCV8	14	-0.0461 (-5.812)	0.0284 (2.749)	-0.9566 (1.085)	-0.9506
BTCX8	16	-0.0549 (-3.985)	0.0548 (2.330)	-0.9291 (2.721)	-0.9460
BTCZ8	17	-0.0501 (-3.066)	0.1170 (4.341)	-0.9538 (2.072)	-0.9653
BTCF9	15	-0.0868 (-3.165)	0.0776 (2.843)	-0.3182 (1.052)	-0.9552
BTCG9	13	-0.0527 (-5.063)	0.1182 (5.414)	-0.9752 (0.660)	-0.9441
BTCH9	19	-0.0362 (-5.785)	0.1299 (11.752)	-0.9700 (2.540)	-0.9673
Panel B: All Data (December 18, 2017 – March 31, 2019)					
	250	-0.0858 (-16.944)	0.0615 (10.253)	-0.8931 (1.943)	-0.9653

This table reports the results of the VECM as presented in (1), based on one-minute sampled midquotes on CME. The VEC model is estimated every day, and the average coefficients, as well as the respective t-statistics (in parentheses), are presented for each considered contract. Additionally, we present the median of the Beta estimation. Rank of co-integration is estimated by Likelihood-Ratio test. SBIC is used to identify the daily lag length.

Table A2: Macroeconomic news releases for the US

Number	Announcement	Frequency	Announcement time	Source
1	Monetary (Federal Funds Rate - FOMC)	6-weeks	02:00 p.m.	FED (www.federalreserve.gov/newsevents/pressreleases.htm)
2	CPI	Monthly	08:30 a.m.	Bureau of Labour Statistics (www.bls.gov/schedule/news_release/cpi.htm)
3	UMP	Monthly	08:30 a.m.	Bureau of Labour Statistics (www.bls.gov/schedule/news_release/empisit.htm)
4	PPI	Monthly	08:30 a.m.	Bureau of Labour Statistics (www.bls.gov/schedule/news_release/ppi.htm)
5	GDP Final	Quarterly	08:30 a.m.	Bureau of Economic Analysis (www.bea.gov/news/current-releases)

This table provides a summary of the macroeconomic news announcements used in our study. Abbreviation: FED, Federal Reserve System; FOMC, Federal Open Market Committee; CPI, Consumer Price Index; UMP, Unemployment Rate; PPI, Producer Price Index, GDP, Gross Domestic Product.

References

- Ahking, F. W. (2002). Model mis-specification and Johansen's co-integration analysis: an application to the US money demand. *Journal of Macroeconomics* 24(1), 51-66. [https://doi.org/10.1016/S0164-0704\(02\)00017-4](https://doi.org/10.1016/S0164-0704(02)00017-4)
- Alexander, C., Choi, J., Park, H., Sohn, S. (2020). BitMEX bitcoin derivatives: Price Discovery, informational efficiency, and hedging effectiveness. *Journal of Futures Markets* 40(1), 23-43. <https://doi.org/10.1002/fut.22050>
- Baillie, R. T., Booth, G. G., Tse, Y., Zobotina, T. (2002). Price Discovery and common factor models. *Journal of Financial Markets* 5(3), 309-321. [https://doi.org/10.1016/S1386-4181\(02\)00027-7](https://doi.org/10.1016/S1386-4181(02)00027-7)
- Baker, S. R., Bloom, N., Davis, S. J. (2013). Measuring economic policy uncertainty. Chicago Booth Paper 13-02. Retrieved from http://www.policyuncertainty.com/media/EPU_BBD_2013.pdf
- Barclay, M. J., Warner, J. B. (1993). Stealth trading and volatility: Which trades move prices? *Journal of Financial Economics* 34(3), 281-305. [https://doi.org/10.1016/0304-405X\(93\)90029-B](https://doi.org/10.1016/0304-405X(93)90029-B)
- Baur, D. G., Hong, KH., Lee, A. D. (2018). Bitcoin: Medium of exchange or speculative assets? *Journal of International Financial Markets, Institutions and Money* 54, 177-189. <https://doi.org/10.1016/j.intfin.2017.12.004>
- Baur, D. G., Dimpfl, T. (2019). Price discovery in bitcoin spot and futures? *Journal of Futures Markets* 39(7), 803-817. <https://doi.org/10.1002/fut.22004>
- Ben-Rephael, A., Da, Z., Israelsen, R. D. (2017). It depends on where you search: Institutional investor attention and underreaction to news. *Review of Financial Studies* 30(9), 3009-3047. <https://doi.org/10.1093/rfs/hhx031>
- Boehme, R., Christin, N., Edelman, B., Moore, T. (2015). Bitcoin: Economics, technology, and governance. *Journal of Economic Perspectives* 29(2), 213-238. <https://doi.org/10.1257/jep.29.2.213>
- Booth, G. G., So, R. W., Tse, Y. (1999). Price discovery in the German equity index derivatives markets. *Journal of Futures Markets* 19(6), 619-643. [https://doi.org/10.1002/\(SICI\)1096-9934\(199909\)19:6<619::AID-FUT1>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1096-9934(199909)19:6<619::AID-FUT1>3.0.CO;2-M)

- Bouri, E., Jalkh, N., Molnár, P., Rouband, D. (2017). Bitcoin for energy commodities before and after the December 2013 crash: diversifier, hedge or safe haven? *Applied Economics* 49(50), 5063-5073. <https://doi.org/10.1080/00036846.2017.1299102>
- Brandvold, M., Molnár, P., Vagstad, K., Valstad, O. C. A. (2015). Price discovery on Bitcoin exchanges. *Journal of International Financial Markets, Institutions & Money* 36, 18-35. <https://doi.org/10.1016/j.intfin.2015.02.010>
- Brière, M., Oosterlinck, K., Szafarz, A. (2015). Virtual currency, tangible return: Portfolio diversification with bitcoin. *Journal of Asset Management* 16, 365-373. <https://doi.org/10.1057/jam.2015.5>
- Cabrera, J., Wang, T., Yang, J. (2009). Do futures lead price discovery in electronic foreign exchange markets? *Journal of Futures Markets* 29(2), 137-156. <https://doi.org/10.1002/fut.20352>
- Chakravarty, S. (2001). Stealth-trading: Which traders' trades move stock prices? *Journal of Financial Economics* 61(2), 289-307. [https://doi.org/10.1016/S0304-405X\(01\)00063-0](https://doi.org/10.1016/S0304-405X(01)00063-0)
- Chakravarty, S., Gulen, H., Mayhew, S. (2004). Informed trading in stock and option markets. *Journal of Finance* 59(3), 1235-1257. <https://doi.org/10.1111/j.1540-6261.2004.00661.x>
- Chan, K. (1992). A further analysis of the lead-lag relationship between the cash market and stock index futures market. *Review of Financial Studies* 5(1), 123-152. <https://doi.org/10.1093/rfs/5.1.123>
- Chen, S.-S. (2009). Predicting the bear stock market: Macroeconomic variables as leading indicators. *Journal of Banking & Finance* 33(2), 211-223. <https://doi.org/10.1016/j.jbankfin.2008.07.013>
- Chen, Y.-L., Gau, Y.-F. (2010). News announcements and price discovery in foreign exchange spot and futures markets. *Journal of Banking & Finance* 34(7), 1628-1636. <https://doi.org/10.1016/j.jbankfin.2010.03.009>
- Chen, Y.-L., Tsai, W.-C. (2017). Determinants of price discovery in the VIX futures market. *Journal of Empirical Finance* 43, 59-73. <https://doi.org/10.1016/j.jempfin.2017.05.002>
- Corbet, S., Lucey, B., Peat, M., Vigne, S. (2018). Bitcoin Futures – What use are they? *Economics Letters* 172, 23-27. <https://doi.org/10.1016/j.econlet.2018.07.031>
- Da, Z., Engelberg, J., Gao, P. (2011). In search of attention. *Journal of Finance* 66(5), 1461-1499. <https://doi.org/10.1111/j.1540-6261.2011.01679.x>

- Dimpfl, T., Flad, M., Jung, R. C. (2017). Price discovery in agricultural commodity markets in the presence of futures speculation. *Journal of Commodity Markets* 5, 50-62. <https://doi.org/10.1016/j.jcomm.2017.01.002>
- Drobetz, W., Momtaz, P. P., Schröder, H. (2019). Investor sentiment and initial coin offerings. *Journal of Alternative Investments* 21(4), 41-55. <https://doi.org/10.3905/jai.2019.1.069>
- Dyhrberg, A. H. (2016). Hedging capabilities of bitcoin. Is it the virtual gold? *Finance Research Letters* 16, 139-144. <https://doi.org/10.1016/j.frl.2015.10.025>
- EPU (2019). *Economic policy uncertainty measures*. Retrieved from <https://www.policyuncertainty.com>. Accessed as of December 2019.
- Eun, C. S., Sabherwal, S. (2003). Cross-border listings and price discovery: Evidence from U.S.-listed Canadian stocks. *Journal of Finance* 58(2), 549-575. <https://doi.org/10.1111/1540-6261.00537>
- Fernandez-Perez, A., Frijns, B., Gafiatullina, I., Tourani-Rad, A. (2018). Determinants of intraday price discovery in VIX exchange traded notes. *Journal of Futures Markets* 38(5), 535-548. <https://doi.org/10.1002/fut.21907>
- Foley, S., Karlsen, J. R., Putniņš T. J. (2019). Sex, drugs, and Bitcoin: How much illegal activity is financed through cryptocurrencies? *Review of Financial Studies* 32(5), 1798-1853. <https://doi.org/10.1093/rfs/hhz015>
- Fricke, C., Menkhoff, L. (2011). Does the “Bund” dominate price discovery in Euro bund futures? Examining information shares. *Journal of Banking & Finance* 35(5), 1057-1072. <https://doi.org/10.1016/j.jbankfin.2010.09.022>
- Frijns, B., Gilbert, A., Tourani-Rad, A. (2015a). The determinants of price discovery: Evidence from US-Canadian cross-listed shares. *Journal of Banking & Finance* 59, 457-468. <https://doi.org/10.1016/j.jbankfin.2015.07.011>
- Frijns, B., Indiawan, I., Tourani-Rad, A. (2015b). Macroeconomic news announcements and price discovery: Evidence from Canadian-U.S. cross-listed firms. *Journal of Empirical Finance* 32, 35-48. <https://doi.org/10.1016/j.jempfin.2014.05.001>
- Glaser, F., Zimmermann, K., Haferkorn, M., Weber, M., Siering, M. (2014). Bitcoin – Asset or currency? Revealing users’ hidden intentions. Proceedings of the 22nd European Conference on Information Systems.

- Gonzalo, J., Granger, C. (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Studies* 13(1), 27-35. <https://doi.org/10.1080/07350015.1995.10524576>
- Grammig, J., Melvin, M., Schlag, C. (2005). Internationally cross-listed stock prices during overlapping trading hours: price discovery and exchange rate effects. *Journal of Empirical Finance* 12(1), 139-164. <https://doi.org/10.1016/j.jempfin.2003.10.004>
- Griffith, J., Najand, M., Shen, J. (2019). Emotions in the stock market. *Journal of Behavioral Finance* 21(1), 42-56. <https://doi.org/10.1080/15427560.2019.1588275>
- Hale, G., Krishnamurthy, A., Kudlyak, M., Shultz, P. (2018). How futures trading changed Bitcoin prices. FRBSF Economic Letter. Retrieved from <https://www.frbsf.org/economic-research/files/el2018-12.pdf>
- Hansen, H., Juselius, K. (1995). Cats in rats: Co-integration analysis of time series. Estima, Evanston, Illinois.
- Harris, F. H. deB., McInish, T. H., Shoesmith G. L., Wood, R. A. (1995). Cointegration, error correction, and price discovery on informationally linked security markets. *Journal of Financial and Quantitative Analysis* 30(4), 563-579. <https://doi.org/10.2307/2331277>
- Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *Journal of Finance* 50(4), 1175-1199. <https://doi.org/10.1111/j.1540-6261.1995.tb04054.x>
- Hauptfleisch, M., Putniņš T. J., Lucey, B. (2016). Who sets the price of gold? London or New York. *Journal of Futures Markets* 36(6), 564-586. <https://doi.org/10.1002/fut.21775>
- Jin, M., Li, Y., Wang, J., Yang, Y. C. (2018). Price discovery in the Chinese gold market. *Journal of Futures Markets* 38(10), 1262-1281. <https://doi.org/10.1002/fut.21938>
- Johansen, S. (1995). Likelihood-based inference in cointegrated vector autoregressive models. Oxford: Oxford University Press.
- Kapar, B., Olmo, J. (2019). An analysis of price discovery between Bitcoin futures and spot markets. *Economics Letters* 174, 62-64. <https://doi.org/10.1016/j.econlet.2018.10.031>
- Karmakar, M., Inani, S. (2019). Information share and its predictability in the Indian stock market. *Journal of Futures Markets* 39(10), 1322-1343. <https://doi.org/10.1002/fut.22041>

- Köchling, G., Müller, J., Posch, P. N. (2019). Does the introduction of futures improve the efficiency of Bitcoin? *Finance Research Letters* 30, 367-370. <https://doi.org/10.1016/j.frl.2018.11.006>
- Kristoufek, L. (2015). What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. *PLoS ONE* 10(4), 1-15. <https://doi.org/10.1371/journal.pone.0123923>
- Kroll, J. A., Davey, I. C., Felten, E. W. (2013). The economics of Bitcoin mining or, Bitcoin in the presence of adversaries. Working Paper, Princeton University.
- Lee, C. M. C., Radhakrishna, B. (2000). Inferring investor behavior: Evidence from TORQ data. *Journal of Financial Markets* 3(2), 83-111. [https://doi.org/10.1016/S1386-4181\(00\)00002-1](https://doi.org/10.1016/S1386-4181(00)00002-1)
- Li, X., Wang, C. A. (2017). The technology and economic determinants of cryptocurrency exchange rates: The case of Bitcoin. *Decision Support Systems* 95, 49-60. <https://doi.org/10.1016/j.dss.2016.12.001>
- Lin, C.-B., Chou, R. K., Wang, G. H. K. (2018). Investor sentiment and price discovery: Evidence from the pricing dynamics between the futures and spot markets. *Journal of Banking & Finance* 90, 17-31. <https://doi.org/10.1016/j.jbankfin.2018.02.014>
- MA (2019). *Macroeconomic announcement dates*. Retrieved from multiple website listed in appendix A2. Accessed from December 2019.
- Mai, F., Shan, Z., Bai, Q., Wang, X. S., Chiang, R.H.L. (2018). How does social media impact Bitcoin value? A test of the silent majority hypothesis. *Journal of Management Information Systems* 35(1), 19-52. <https://doi.org/10.1080/07421222.2018.1440774>
- Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. Retrieved from <https://bitcoin.org/bitcoin.pdf>.
- Nooijen, S. J., Broda, S. A. (2016). Predicting equity markets with digital online media sentiment: Evidence from Markov-switching model. *Journal of Behavioral Finance* 17(4), 321-335. <https://doi.org/10.1080/15427560.2016.1238370>
- Pagnottoni, P., Dimpfl, T. (2019). Price discovery on Bitcoin markets. *Digital Finance* 1, 139-161. <https://doi.org/10.1007/s42521-019-00006-x>
- Putniņš T. J. (2013). What do price discovery metrics really measure? *Journal of Empirical Finance* 23, 68-83. <https://doi.org/10.1016/j.jempfin.2013.05.004>

- Schlusche, B. (2009). Price formation in spot and futures markets: Exchange traded funds vs. index futures. *Journal of Derivatives* 17(2), 26-40. <https://doi.org/10.3905/JOD.2009.17.2.026>
- Shyy, G., Vijayraghavan, V., Scott-Quinn, B. (1996). A further investigation of the lead-lag relationship between the cash market and stock index futures market with the use of bid/ask-quotes: The case of France. *Journal of Futures Markets* 16(4), 405-420. [https://doi.org/10.1002/\(SICI\)1096-9934\(199606\)16:4<405::AID-FUT3>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1096-9934(199606)16:4<405::AID-FUT3>3.0.CO;2-M)
- SVI (2019). *Google trends bitcoin search volume*. Retrieved from trends.google.com (keywords used: Bitcoin; US). Accessed on December 2019.
- Theissen, E. (2012). Price discovery in spot and futures markets: A reconsideration. *European Journal of Finance* 18(10), 969-987. <https://doi.org/10.1080/1351847X.2011.601643>
- Tse, Y., Xiang, J., Fung, J. K. W. (2006). Price discovery in the foreign exchange futures market. *Journal of Futures Markets* 26(11), 1131-1143. <https://doi.org/10.1002/fut.20229>
- Urquhart, A. (2016). The inefficiency of Bitcoin. *Economics Letters* 148, 80-82. <https://doi.org/10.1016/j.econlet.2016.09.019>
- Wintoki, M. B., Linck, J. S., Netter, J. M. (2012). Endogeneity and the dynamics of internal corporate governance. *Journal of Financial Economics* 105(3), 581-606. <https://doi.org/10.1016/j.jfineco.2012.03.005>
- Yan, B., Zivot, E. (2010). A structural analysis of price discovery measures. *Journal of Financial Markets* 13(1), 1-19. <https://doi.org/10.1016/j.finmar.2009.09.003>
- Yermack, D. (2015). Is Bitcoin a real currency? An economic appraisal. In: *Handbook of Digital Currency*, 31-43, Elsevier.

Figures and Tables

Figure 1: Five-day moving average of price discovery measures for Bitstamp spot market

This figure plots the five-day moving average of the Component Share and Information Share on the spot market. Component Share, as well as Information Share, are calculated from one-minute sampled midquotes.

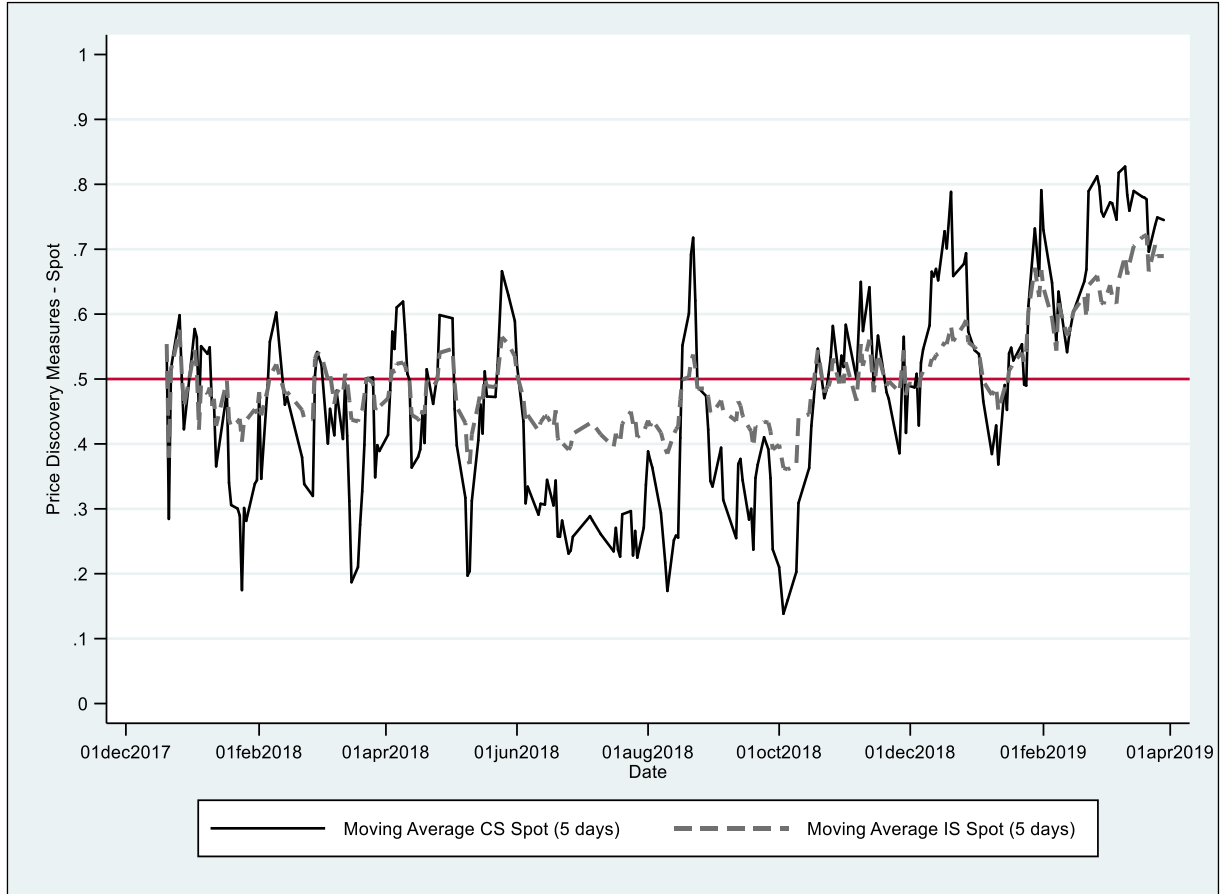


Table 1: Trading volume and average proportion of futures contracts by maturing month

(1)	(2)	(3)	(4)	(5)
RIC	Time interval most traded	Expiration	Volume MTF	Avg. Proportion
BTCF8	18Dec2017 – 25Jan2018	26Jan2018	23,457	84.80
BTCG8	26Jan2018 – 22Feb2018	23Feb2018	19,999	84.50
BTCH8	23Feb2018 – 28Mar2018*	30Mar2018*	46,090	95.27
BTCJ8	29Mar2018 – 26Apr2018	27Apr2018	62,265	94.67
BTCK8	27Apr2018 – 24May2018	25May2018	66,470	94.34
BTCM8	25May2018 – 28Jun2018	29Jun2018	57,637	94.06
BTCN8	29Jun2018 – 26Jul2018	27Jul2018	80,652	95.27
BTCQ8	27Jul2018 – 30Aug2018	31Aug2018	121,796	96.65
BTCU8	31Aug2018 – 27Sep2018	28Sep2018	46,475	94.69
BTCV8	28Sep2018 – 25Oct2018	26Oct2018	38,005	92.72
BTCX8	26Oct2018 – 29Nov2018	30Nov2018	97,089	93.95
BTCZ8	30Nov2018 – 27Dec2018	28Dec2018	56,642	94.28
BTCF9	28Dec2018 – 24Jan2019	25Jan2019	55,432	94.68
BTCG9	25Jan2019 – 21Feb2019	22Feb2019	66,737	94.31
BTCH9	22Feb2019 – 28Mar2019	29Mar2019	74,209	93.78
BTCJ9	29Mar2019 – 31Mar2019	26Apr2019#	4,237	92.43
TOTAL	327 trading days	---	815,888	93.15

This table contains several statistics on our CME futures time series. Time interval covers the days, on which the respective contract (RIC) is the most traded futures (MTF) per day. Expiration refers to the settlement date of the respective futures contract. Volume MTF is the sum of the daily volume during the provided time interval. Finally, the average proportion is defined as the trading volume of the most actively traded futures contract relative to the total trading volume in the respective time interval. * indicates that volume shifts one day earlier as the day before expiration is a holiday, while # marks that the respective contract is outside of our sample period. The sample period is from December 17, 2017 – March 31, 2019.

Table 2: Non-synchronicity and percentage of zero returns

Time Interval	Non-synchronous quoting (%)		Zero Returns (%)	
	Spot	Fut.	Spot	Fut.
1 minute	0.35	4.40	22.20	43.37
5 minutes	0.02	0.23	8.21	24.04
10 minutes	0.02	0.11	5.47	18.22
15 minutes	0.02	0.09	4.63	15.04

This table reports the proportion of non-synchronous quoting and percentage of zero returns of merged spot and futures time series during our sample interval from December 18, 2017 – March 31, 2019. Non-synchronous quoting is defined as the proportion of time intervals in which no quote is observed. Zero Returns (%) is the proportion of no price change. We calculate both measures for one-, five-, ten-, and fifteen-minute intervals.

Table 3: Summary statistics of Bitcoin spot and futures midquotes

Panel A: Contract by contract									
(1)	(2)	(3)		(4)		(5)		(6)	
		N	Mean		Std. Dev.		NQ-Prob. (%)		ZR (%)
		Spot	Fut.	Spot	Fut.	Spot	Fut.	Spot	Fut.
BTCF8	31,200	14,035	14,076	2,275	2,391	0.07	1.78	4.30	8.66
BTCG8	22,800	9,393	9,391	1,368	1,349	0.00	1.23	5.03	12.32
BTCH8	28,799	9,398	9,393	1,101	1,103	0.08	3.03	7.49	17.52
BTCJ8	23,999	7,765	7,769	852	855	0.02	5.75	13.59	30.82
BTCCK8	24,000	8,770	8,777	579	592	0.06	3.40	13.29	28.73
BTCM8	29,998	6,925	6,924	548	560	0.03	6.30	21.98	42.61
BTCN8	22,800	6,935	6,928	709	714	0.04	3.16	15.06	41.50
BTCQ8	29,995	6,865	6,855	585	586	0.17	1.48	18.73	34.56
BTCU8	22,767	6,536	6,521	298	307	0.22	6.85	28.04	46.36
BTCV8	23,909	6,456	6,450	121	125	1.25	10.70	45.86	63.70
BTCX8	28,669	5,591	5,577	988	994	1.07	9.39	33.73	56.88
BTCZ8	21,599	3,649	3,626	274	274	0.02	1.06	15.97	39.53
BTCF9	21,590	3,727	3,703	167	165	0.30	2.44	31.70	67.49
BTCG9	22,788	3,540	3,525	178	188	1.28	2.52	39.12	80.82
BTCH9	29,962	3,883	3,878	85	88	0.61	6.20	41.18	84.76
Panel B: All data (December 19, 2017 – March 31, 2019)									
	386,071	7,035	7,031	2,957	2,978	0.35	4.40	22.20	43.37

This table reports summary statistics of average midquote (mean), standard deviation (Std. Dev.), non-synchronous quoting in percent (NQ-Prob. (%)), and percentage of zero returns (ZR (%)) for each most-traded contract (Panel A) and for the whole sample (Panel B). The statistics are calculated from one-minute sampled midquotes.

Table 4: Price discovery measures for one-minute intervals

	Component Share					Information Share				
	Mean	5 th Per.	Med.	95 th Per.	Std. Dev.	Mean	5 th Per.	Med.	95 th Per.	Std. Dev.
Panel A: Contract by contract										
BTCF8	0.418	0.084	0.247	0.917	0.321	0.466	0.224	0.424	0.760	0.167
BTCG8	0.438	0.030	0.418	0.942	0.300	0.471	0.335	0.473	0.625	0.091
BTCH8	0.406	0.012	0.385	0.781	0.288	0.481	0.343	0.474	0.568	0.100
BTCJ8	0.507	0.164	0.473	0.885	0.250	0.499	0.328	0.488	0.678	0.088
BTCK8	0.442	0.062	0.401	0.784	0.272	0.472	0.316	0.474	0.643	0.106
BTCM8	0.297***	0.010	0.226	0.667	0.175	0.425***	0.269	0.426	0.561	0.071
BTCN8	0.245***	0.036	0.178	0.438	0.153	0.413***	0.314	0.421	0.483	0.064
BTCQ8	0.410	0.068	0.333	0.857	0.279	0.460**	0.360	0.449	0.572	0.072
BTCU8	0.321***	0.050	0.349	0.901	0.227	0.426***	0.331	0.409	0.569	0.066
BTCV8	0.450	0.004	0.461	0.933	0.261	0.485	0.267	0.453	0.790	0.153
BTCX8	0.511	0.043	0.456	0.940	0.299	0.497	0.220	0.482	0.487	0.145
BTCZ8	0.611	0.051	0.667	0.941	0.283	0.539	0.365	0.535	0.719	0.082
BTCF9	0.530	0.005	0.507	0.967	0.314	0.537	0.404	0.500	0.753	0.121
BTCG9	0.670***	0.197	0.716	0.995	0.213	0.608***	0.370	0.620	0.802	0.122
BTCH9	0.774***	0.367	0.868	0.944	0.165	0.682***	0.437	0.719	0.848	0.115
Panel B: All Data (December 18, 2017 – March 31, 2019)										
	0.472	0.050	0.451	0.933	0.288	0.499	0.328	0.479	0.764	0.128

Panel A reports descriptive statistics for daily price discovery measures, referring to the spot market, and estimated for each day in our sample. Panel B presents the results for the whole data set. We estimate the Component Shares (CS) and the Information Shares (IS) for one-minute time intervals. The ***/**/* are used to indicate that an estimate is significantly different from 0.50 at the 1% /5% /10% level.

Table 5: Price Discovery measures for different time intervals

	Component Share			Information Share		
	Five-MI	Ten-MI	Fifteen-MI	Five-MI	Ten-MI	Fifteen-MI
Panel A: Contract by contract						
BTCF8	0.464 (0.367)	0.499 (0.391)	0.467 (0.439)	0.502 (0.482)	0.503 (0.493)	0.496 (0.497)
BTCG8	0.398 (0.416)	0.345** (0.282)	0.471 (0.506)	0.488 (0.491)	0.490* (0.490)	0.499 (0.500)
BTCH8	0.475 (0.496)	0.476 (0.470)	0.589 (0.613)	0.510 (0.500)	0.509 (0.499)	0.502 (0.502)
BTCJ8	0.586* (0.629)	0.574 (0.606)	0.474 (0.486)	0.513 (0.507)	0.505 (0.505)	0.497 (0.500)
BTCK8	0.490 (0.522)	0.514 (0.518)	0.550 (0.565)	0.496 (0.502)	0.498 (0.501)	0.504 (0.505)
BTCM8	0.359*** (0.361)	0.420 (0.383)	0.445 (0.450)	0.482** (0.485)	0.490 (0.494)	0.493 (0.498)
BTCN8	0.321*** (0.240)	0.389 (0.356)	0.471 (0.510)	0.474** (0.478)	0.481 (0.494)	0.488 (0.500)
BTCQ8	0.475 (0.434)	0.488 (0.455)	0.496 (0.471)	0.492 (0.497)	0.497 (0.497)	0.500 (0.500)
BTCU8	0.388 (0.365)	0.427 (0.452)	0.546 (0.502)	0.482* (0.487)	0.491 (0.497)	0.499 (0.499)
BTCV8	0.429 (0.470)	0.417 (0.420)	0.498 (0.526)	0.488 (0.485)	0.478 (0.488)	0.506 (0.500)
BTCX8	0.558 (0.473)	0.489 (0.442)	0.394 (0.374)	0.519 (0.496)	0.496 (0.496)	0.492 (0.496)
BTCZ8	0.567 (0.616)	0.536 (0.573)	0.557 (0.568)	0.506 (0.507)	0.505 (0.501)	0.503 (0.501)
BTCF9	0.468 (0.510)	0.500 (0.508)	0.501 (0.518)	0.507 (0.500)	0.505 (0.500)	0.502 (0.500)
BTCG9	0.688*** (0.649)	0.630 (0.692)	0.566 (0.571)	0.562*** (0.544)	0.522 (0.525)	0.510 (0.513)
BTCH9	0.797*** (0.823)	0.681*** (0.729)	0.629** (0.676)	0.614*** (0.604)	0.552*** (0.539)	0.537** (0.514)
Panel B: All data (December 18, 2017 – March 31, 2019)						
	0.501 (0.495)	0.497 (0.471)	0.516 (0.511)	0.510** (0.500)	0.503 (0.498)	0.503 (0.500)

Panel A of Table 5 reports average results for daily price discovery measures, referring to the spot market, and calculated for each contract in our sample from mid-quotes on CME. We also present the results for the whole data set (Panel B). We estimate the Component Shares (CS) and the Information Shares (IS) for five-, ten-, and fifteen-minute time intervals. The ***/**/* are used to indicate that an estimate is significantly different from 0.50 at the 1% /5% /10% level. Median figures are reported in parentheses.

Table 6: Summary statistics of determinants

	Mean	Median	5% quantile	95% quantile	Std. dev.
<i>Market Quality</i>					
Panel A: Spot market					
Number of Quotes _t	33,524.21	35,738.00	14,814.00	48,923.00	10,962.56
Traded Volume _t	9,762.40	8,047.92	3,109.69	19,895.60	6,825.32
%BAS _t	0.0611	0.0508	0.0243	0.1471	0.0377
Number small trades _t	23,658.78	19756.50	5810.00	58392.00	18896.98
Number medium trades _t	1,741.30	1,511.00	636.00	3,735.00	1,092.36
Number large trades _t	266.16	196.50	50.00	705.00	247.67
Panel B: Futures market					
Number of Quotes _t	56,603.98	45,207.50	23,394.00	100,515.00	69,086.92
Traded Volume _t	14,242.11	12,270.00	3,460.00	32,500.00	9,168.37
%BAS _t	0.1869	0.1412	0.0936	0.4654	0.1176
Number small trades _t	1,685.53	1,452.50	519.00	3,650.00	1,056.43
Number medium trades _t	454.69	376.50	35.00	1,226.00	376.23
Number large trades _t	19.41	12.00	1.00	58.00	21.49
<i>Uncertainty</i>					
Vola _{t,spot}	0.0409	0.0328	0.0129	0.0888	0.0272
ret_VIX _t (in %)	0.1169	-0.8511	-12.22	16.26	9.33
ret_EPU _t (in %)	-0.0102	0.4214	-80.66	92.01	64.92
<i>Sentiment</i>					
Dummy_HighSent _t (news)	0.5288	-	-	-	-
Dummy_HighSent _t (social)	0.5373	-	-	-	-
<i>Attention</i>					
Dummy_HighAtt _t	0.1836	-	-	-	-
<i>Macroeconomic News</i>					
Dummy_MEA _t	0.1215	-	-	-	-
<i>Controls</i>					
ret_Bitstamp _t (in %)	-0.4545	-0.2215	-8.58	7.18	4.79
ret_Gold _t (in %)	0.0094	0	-1.01	1.02	0.6486

This table reports summary statistics of all explanatory variables on price discovery on a daily (except for Google Trends) basis for the full sample period (December 2017 – March 2019). The considered variables are defined in Section 4.1.

Table 7: Determinants of price discovery: market quality and uncertainty

Variable name	<i>March 29, 2018 – March 31, 2019</i>		<i>December 18, 2017 – March 31, 2019</i>	
	(1) logit_CS	(2) logit_CS	(3) logit_CS	(4) logit_CS
<i>Market Quality</i>				
ln_rel_num Quotes _t	-0.981 (-1.402)	-0.870 (-1.278)	-0.197 (-0.456)	-0.198 (-0.459)
ln_rel_Traded Volume _t	1.444* (1.950)	1.340* (1.895)	0.229 (0.736)	0.210 (0.674)
ln_rel_%BAS _t	-0.915** (-2.142)	-0.908** (-2.083)	-0.933*** (-3.096)	-1.098*** (-3.073)
<i>Uncertainty</i>				
ln_volat _{t,spot}	-0.891* (-1.858)	-0.824* (-1.782)	-0.552 (-1.461)	-0.524 (-1.395)
ret_VIX _t	-0.029** (-2.008)		0.002 (0.214)	
ret_EPU _t	0.003 (1.180)		0.002 (1.145)	
<i>Controls</i>				
ret_Bitstamp _t	-0.035 (-1.149)	-0.035 (-1.189)	-0.039* (-1.687)	-0.040* (-1.904)
ret_Gold _t	0.039 (0.210)	0.050 (0.262)	-0.012 (-0.078)	-0.021 (-0.127)
Constant	-3.759*** (-3.038)	-3.534*** (-2.929)	-3.327*** (-3.100)	-3.245*** (-3.029)
Observations	195	195	250	250
Adj. R-squared	0.093	0.080	0.078	0.081
Hansen's J test	0.737	0.674	0.918	0.855
Wooldridge's score test	0.382	0.401	0.547	0.517

This table reports results for variables of market quality and uncertainty where we assess the relationship to the logit transformation of Component Share that refers to the spot market. The model is estimated by 2SLS using robust standard errors, where the relative number of quotes and the relative traded volume are treated as endogenous and the remaining variables as exogenous. We use lag one as instruments. Robust t-figures are reported in parentheses. The ***/**/* indicate that an estimate is statistically significant at the 1% /5% /10% level.

Table 8: Determinants of price discovery with decomposed relative trading volume

Variable name	March 29, 2018 – March 31, 2019			
	logit_CS	logit_CS	logit_CS	logit_CS
	(1)	(2)	(3)	(4)
<i>Market Quality</i>				
$\ln_rel_num\ Quotes_t$	0.982 (1.101)	-0.725 (-1.501)	0.224 (0.347)	0.342 (0.314)
$\ln_rel_num_small_trades_t$	-1.102 (-1.558)			-0.789 (-1.112)
$\ln_rel_num_medium_trades_t$		1.043*** (2.924)		1.128*** (2.961)
$\ln_rel_num_large_trades_t$			-0.248 (-0.741)	-0.412 (-1.199)
$\ln_rel_ \%BAS_t$	-1.131*** (-3.666)	-0.933** (-2.560)	-1.205*** (-3.755)	-0.615* (-1.731)
<i>Uncertainty</i>				
$\ln_vola_{t,spot}$	-0.019 (-0.064)	-0.745** (-2.232)	0.226 (0.361)	-0.066 (-0.101)
ret_VIX_t	-0.019 (-1.085)	-0.026* (-1.916)	-0.028** (-2.039)	-0.024 (-1.492)
ret_EPU_t	0.001 (0.430)	0.003 (1.249)	0.002 (0.779)	0.002 (0.742)
<i>Controls</i>				
$ret_Bitstamp_t$	-0.063* (-1.859)	-0.040 (-1.345)	-0.065* (-1.898)	-0.060 (-1.543)
ret_Gold_t	0.151 (0.843)	0.021 (0.119)	0.094 (0.516)	0.034 (0.192)
Constant	1.375 (0.538)	-5.286*** (-3.854)	-0.018 (-0.006)	0.591 (0.133)
Observations	195	195	193	193
Adj. R-squared	0.0654	0.128	0.102	0.134
Hansen's J test	0.591	0.891	0.426	0.682
Wooldridge's score test	0.433	0.422	0.800	0.549

This table reports results from 2SLS regressions using the decomposed trading sizes as explanatory variables. The dependent variable refers to the logit transformation of the Component Share of the spot market. Relative number of small, medium, and large trades, as well as relative number of quotes, are treated as endogenous variables. We use lag one as instruments. Results are reported for the sample period from March 2018 through March 2019. Robust t-statistics are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Table 9: Determinants of price discovery: all variables

Variable name	March 29, 2018 – March 31, 2019			
	(1) logit_CS	(2) logit_CS	(3) logit_CS	(4) logit_CS
<i>Market Quality</i>				
ln_rel_num Quotes _t	-1.050 (-1.471)	-0.881 (-1.114)	-0.936 (-1.289)	-0.835 (-0.993)
ln_rel_Traded Volume _t	1.512* (1.913)	1.441* (1.935)	1.397* (1.810)	1.437* (1.739)
ln_rel_%BAS _t	-0.877** (-2.058)	-0.944** (-2.126)	-0.943** (-2.165)	-0.957** (-2.117)
<i>Uncertainty</i>				
ln_volat _{t,spot}	-0.899* (-1.843)	-0.946** (-2.019)	-0.872* (-1.786)	-0.950** (-1.976)
ret_VIX _t	-0.029* (-1.939)	-0.030** (-2.086)	-0.029* (-1.960)	-0.029* (-1.965)
ret_EPU _t	0.003 (1.215)	0.003 (1.215)	0.003 (1.156)	0.003 (1.233)
<i>Sentiment</i>				
Dummy_HighSent _t (news-based)	-0.387* (-1.727)			-0.438* (-1.913)
Dummy_HighSent _t (social-based)	0.026 (0.100)			0.045 (0.170)
<i>Attention</i>				
Dummy_HighAtt _t		0.182 (0.510)		0.269 (0.729)
<i>Macroeconomic news</i>				
Dummy_MEA _t			-0.134 (-0.456)	-0.207 (-0.691)
<i>Controls</i>				
Constant	-3.534*** (-2.830)	-3.983*** (-3.231)	-3.713*** (-2.957)	-3.768*** (-3.021)
Observations	195	195	195	195
Adj. R-squared	0.091	0.089	0.092	0.092
Hansen's J test	0.762	0.704	0.761	0.754
Wooldridge's score test	0.324	0.362	0.423	0.330

This table reports results for Equation (3) where we assess the relationship between various explanatory variables and the logit transformation of Component Share that refers to the spot market. The model is estimated by 2SLS using robust standard errors, where the relative number of quotes and the relative traded volume are treated as endogenous and the remaining variables as exogenous. We use lag one as instruments. Robust t-figures are reported in parentheses. The ***/**/* indicate that an estimate is statistically significant at the 1% /5% /10% level.

2 | Local Religiosity and Stock Liquidity*

Oliver Entrop[†], Martin Rohleder[‡], Marco Seruset[#]

Abstract

We investigate whether local religious norms affect stock liquidity for U.S. listed companies. Over the period 1997–2020, we find that firms located in more religious areas have higher liquidity, as reflected by lower bid-ask spreads. This result persists after the inclusion of additional controls, such as governance metrics, and further sensitivity and endogeneity analyses. Subsample tests indicate that the impact of religiosity on stock liquidity is particularly evident for firms operating in a poor information environment. We further show that firms located in more religious areas have lower price impact of trades and smaller probability of information-based trading. Overall, our findings are consistent with the notion that religiosity, with its antimanipulative ethos, probably fosters trust in corporate actions and information flows, especially when little is known about the firm. Finally, we conjecture an indirect firm value implication of religiosity through the channel of stock liquidity.

Keywords: Religiosity; Liquidity; Information Asymmetry; Bid-Ask Spread; Geographic Location; Religious Social Norms

JEL classification: G31; G32; G40; G41; Z12

* We thank the participants of the 2022 European Financial Management Conference in Rome and the research seminar at the University of Passau for very helpful comments and suggestions on an earlier draft of this paper. We also thank Carl-Friederich Grösbrink for helpful comments and suggestions. All errors are our own.

[†] Oliver Entrop, University of Passau, Chair of Finance and Banking, Innstraße 27, 94032 Passau, Germany, phone: +49 851 509 2460, E-mail: oliver.entrop@uni-passau.de

[‡] Martin Rohleder, University of Augsburg, Chair of Banking and Finance, Universitätsstraße 16, 86159 Augsburg, phone: +49 821 598 4120, email: martin.rohleder@uni-a.de

[#] Marco Seruset, University of Passau, Chair of Finance and Banking, Innstraße 27, 94032 Passau, Germany, phone: +49 851 509 2463, email: marco.seruset@uni-passau.de

1 Introduction

“Culture, more than book rules, determines how an organization behaves” (Buffet, 2011).¹ This statement underpins the importance of considering culture as a relevant determinant of corporate decisions and performance. Further, Hirshleifer (2015) concludes in his comprehensive review on behavioral finance that scholars need to pay more attention to social finance, which includes the consideration of social norms, moral attitudes, religions, and ideologies in the context of financial behavior. Recently, Graham et al. (2017) document that 92% of executives believe that improving corporate culture is an important factor for increasing their firm’s value. Indeed, a growing number of studies empirically investigate the link between culture and firm outcomes (e.g., Kanagaretnam et al. 2014; Guiso et al., 2016; Hilary and Huang, 2021).² Of particular interest in this context is the local religiosity of the county where a firm is located (Jiang et al., 2018).

While the investigation of the impact of religiosity on individuals has begun in the early 80s (e.g., Chiswick 1983, 1985; see, Iannaccone 1998 for a comprehensive overview), the consideration of religiosity on the firm level, however, has only gained importance in recent years. Literature reveals that risk aversion (e.g., Hilary and Hui, 2009; Adhikari and Agrawal, 2016) and honesty (e.g., Dyreng et al., 2012) are the key traits that are associated with religiosity on firm level. According to social norm theory, the prevailing set of behaviors and values in an area influences religious and non-religious individuals in a similar vein (e.g., Hilary and Hui, 2009; Cantrell and Yust, 2018). Explicitly, this affects also senior executives and their decision making, as they tend to live in proximity to their firms’ headquarters (Alam et al., 2014; Zolotoy

¹ Also cited in Cantrell and Yust (2018).

² Even in times of machine learning and big data, corporate culture is perceived to be a key factor for many business decisions, and corporate success and failure (Goldstein et al., 2021). Using a semi-supervised machine-learning approach, Li et al. (2021) for example, try to quantify the strength of corporate culture by investigating earnings call transcripts.

et al., 2019; McGuire et al., 2012). Besides the impact of local religiosity on individual and firm behavior, it also affects how an organization is viewed by corporate outsiders (e.g., Jiang et al., 2018).

The idea of linking stock liquidity to factors that go beyond common stock attributes, such as price, trading volume, or return volatility (e.g., Harris, 1994; Huang and Stoll, 1996; Chung and Charoenwong, 1998), is not new in the literature. Several studies provide empirical evidence that stock liquidity relates to stocks' visibility (Grullon et al., 2004), familiarity (Loughran and Schultz, 2005), and ownership structure (Attig et al., 2006). Moreover, further factors, such as political stability and judicial efficiency (Eleswarapu and Venkataraman, 2006), state-level economic conditions (Bernile et al., 2015), internal corporate governance and mandatory CSR expenditures (Chung et al., 2010; Roy et al., 2022), as well as the education of the CEO (Pham, 2020), play an important role for stock liquidity of firms. Our study extends this line of research and investigates the differences in the level of stock liquidity due to the degree of religiosity. In this context, we interpret religiosity as a type of soft information, which is, for example, not reflected in financial statements (Jiang et al., 2018).³

Admittedly, why should local religious norms matter for stock liquidity of the secondary market, in particular? Religiosity is usually seen as a commitment device conveying the firms' willingness to conduct business in a reliable, predictable, and conservative manner (e.g., Callen and Fang, 2015; Jiang et al., 2018), and is also likely a mechanism for reducing agency costs (e.g., McGuire et al., 2012; Leventis et al., 2018). The literature further indicates that firms headquartered in more religious societies are more likely to produce credible information and avoid irregularities in both financial statements and management forecasts (e.g., Grullon et al., 2010; Dyreng et al., 2012; McGuire et al., 2012; Omer et al., 2018).

³ According to Jiang et al. (2018), soft information includes reputation, managerial integrity, corporate culture, and religiosity.

In terms of liquidity, extant theoretical and empirical research shows that stock liquidity is lower for firms with poor disclosure quality and information environment (e.g., Bernile et al., 2015; and the references inhere; Diamond and Verrecchia, 1991; Kurlat, 2018; Roy et al., 2022). Similarly, studies suggest that liquidity providers demand larger compensation, and, for instance, widen the equities' bid-ask spread, when information is partly credible or the adverse selection risk is increased through differential information among market participants (e.g., Kyle, 1985; Glosten and Milgrom, 1985; Huang and Stoll, 1996; Bhattacharya et al., 2013). Given that religiosity through its antimanipulative ethos (e.g., Barro and McCleary, 2003; Callen and Fang, 2015) probably fosters trust in corporate behavior and actions, which could help to identify firms with more credible information flows (e.g., McGuire et al., 2012), we conjecture a positive relation between the degree of religiosity and stock liquidity. If this trust-based aspect of religiosity affects stock liquidity, we would expect that the relevance of religiosity as a commitment device is especially evident when firms operate in settings, in which relatively less information about the firm is otherwise available. In such circumstances, religiosity probably acts as an informal mechanism that reduces the relevance of information asymmetry. Indeed, we show that firms located in high religious areas tend to have higher liquidity, while our findings further indicate that the degree of religiosity has stronger impact on firms operating in a poor information environment.

We measure religiosity as the fraction of religious adherents to the total population of a county in which the firm is headquartered (Hilary and Hui, 2009). By using a broad sample of U.S. listed firms for the period from 1997 through 2020, we then demonstrate in our baseline analysis, which controls for firm controls as well as demographic characteristics, that firms located in more religious areas tend to have lower bid-ask spreads. The results are also economically significant, comparable to existing studies such as Chung et al. (2010) in terms of magnitude. One of the important points we further raise is that our baseline results are robust

to additional control variables, such as governance metrics, and to a battery of different model specifications. Additionally, we conduct several tests, which help us to confirm our baseline results and establish a potential causal link between religiosity and liquidity.

To test our conjecture whether religiosity reduces the relevance of information asymmetry, we create subsamples based on four commonly used proxies for information asymmetry: analyst coverage, S&P 500 membership, firm size, and location. Our results support the notion that the impact of religiosity on stock liquidity is particularly evident for firms that operate in a poor information environment.

Moreover, we show that other aspects of liquidity and the trading environment, such as the Amihud illiquidity (Amihud, 2002) or the probability of information-based trading (*PIN*) (Easley et al., 2002; Brown and Hillegeist, 2007), are also negatively associated with the strength of religiosity. In this context, we use Amihud (2002) measure as a proxy for the price impact of trades (e.g., Goyenko et al., 2009; Edmans et al., 2013), while the *PIN* may be interpreted as a direct proxy for the degree of insider trading and the adverse selection problems faced by liquidity providers (e.g., Chung and Li, 2003).

Lastly, we posit that there is an indirect link between religiosity and firm value through the channel of stock liquidity, suggesting a potential value implication.

We extend existing research in two important ways. First, we provide evidence that religiosity with its antimanipulative ethos (e.g., Barro and McCleary, 2003; Callen and Fang, 2015) may affect the way how an organization acts and is perceived by corporate outsiders. Honesty and conservatism are likely to improve corporate behavior of firms headquartered in more religious societies, thus increasing trust in firms' corporate actions and information flows.⁴ Second, based on this idea, we also add evidence that liquidity providers not only

⁴ We note that there is one related study to ours. Blau (2018) also examines the effect of religiosity on liquidity. However, the author does not investigate local religiosity, since he focuses on a cross-country setting. Moreover, by using a different source of data and a deviating methodological approach, we reason that our study is unique.

consider conventional information (e.g., information from financial statements) in their risk assessment. Thus, in broader terms, we sharpen the understanding of the mechanisms of soft information processing in capital markets and also add evidence to the determinants of stock liquidity.

The remainder of this paper is organized as follows. Section 2 summarizes literature and develops testable hypotheses. In Section 3, we present the data and summary statistics. Section 4 discusses baseline empirical findings, presents results from including additional controls and conducting several robustness tests, and addresses endogeneity issues. Section 5 discusses the channel of information asymmetry, while Section 6 tests whether other aspects of liquidity and the trading environment are connected to religiosity. Section 7 contributes to the ongoing discussion on the impact of religiosity on firm value, while we conclude in Section 8.

2 Hypothesis development

According to social norm theory, norms rule the way of social interaction between members of a group, even if they are not stated explicitly or come with any sanctions when deviating from them (Cialdini and Trost, 1998). Thus, the set of norms inheriting religiosity, particularly risk aversion and honesty, affect the behavior and values of religious and non-religious individuals in the same way (e.g., Hilary and Hui, 2009; Cantrell and Yust, 2018). Consequently, individuals in and around the firm follow the religious norms in the county where the company is headquartered, regardless of their personal religious background. Therefore, we posit that acting according to local religious norms and beliefs by senior executives, employees, and stakeholders should have a positive impact on corporate behavior and outcomes of the firm (e.g., Zolotoy et al., 2019; Alam et al., 2014; Hilary and Hui, 2009; Cantrell and Yust, 2018; McGuire et al., 2012).

Having manifested how local religiosity affects organizational behavior from a theoretical perspective, numerous papers confirm this notion empirically. Starting with Hilary and Hui (2009), who show in their seminal paper that firms located in more religious areas exhibit lower exposure to risk. Subsequent studies investigate the influence of religiosity on the quality and reliance of corporate actions and disclosures, thus in broader terms, on the credibility of information flows of a firm. Grullon et al. (2010) report that religiosity is associated with fewer incidences of inappropriate corporate behavior, such as option backdating, aggressive earnings management, or being target of class action securities lawsuits. Moreover, Dyreng et al. (2012) as well as McGuire et al. (2012) document that companies headquartered in areas with strong social religious norms experience fewer financial reporting irregularities. Additionally, Omer et al. (2018) show that the degree of religiosity affects the audit quality positively. Likewise, Callen and Fang (2015) provide evidence that religiosity helps to curb bad-news-hoarding activities by managers, which is connected with a lower level of future stock price crash risk. Ultimately, Jiang et al. (2018) show that the level of religiosity is also acknowledged by corporate outsiders. They find that firms located in higher-religiosity counties have higher credit rating and lower debt costs.⁵ Taken together, the aforementioned studies provide evidence that religiosity tends to improve corporate actions and outcomes, thus affecting the credibility of corporate information flows.

Relating to liquidity, literature posits that the overall transparency and information environment tend to be a key determinant of stock liquidity (e.g., Healy and Palepu, 2001; Bernile et al., 2015; Roy et al., 2022). Accordingly, firms with better (poorer) quality of information disclosure should have higher (lower) levels of stock liquidity (e.g., Diamond and

⁵ Further studies in the context of religiosity are: Adhikari and Agrawal (2016) and Chircop et al. (2020) find a negative relation between religiosity and risk-taking for a sample of public banks and venture capital investments, respectively, while Cantrell and Yust (2018) document that religiosity is positively associated with asset risk-taking for private banks. Also, stronger social religious norms are associated with lower cost of equity (El Ghouli et al., 2012), lower cost of debt (Cai and Shi, 2019) as well as higher workplace safety (Amin et al., 2021).

Verrecchia, 1991; Kurlat, 2018; Roy et al., 2022). Also, when little is known about the firm and adverse selection problems increase, this should impact stock liquidity negatively (e.g., Easley and O'Hara, 1987; Glosten and Harris, 1988; Roy et al., 2022). Therefore, opaqueness of the information environment and the related information asymmetries are of particular relevance for stock liquidity.

Based on the above considerations and the presented literature, we expect that the level of liquidity should be higher for firms headquartered in more religious areas compared to firms located in areas with low level of religiosity. Given that religiosity potentially acts as a mechanism for reducing agency costs, and examining the strength of local religious could help to identify firms with more credible information flows (e.g., McGuire et al., 2012), we hypothesize that firms located in more religious areas should reveal higher levels of liquidity, thus lower bid-ask spreads.

Hypothesis 1: Firms headquartered in more religious areas have lower bid-ask spreads.

Our second hypothesis makes direct inference on trust-based aspects of religiosity. Since large companies with high visibility and dense analyst coverage are screened more intensively than smaller and less covered firms, social norms probably are less important. In contrast, for firms operating in a poorer information environment, we expect religiosity to act as a commitment indicator of credible behavior and reliable information flows, thus increasing trust and reducing the relevance of information asymmetry (e.g., Jiang et al., 2018).

Hypothesis 2: The relation between religiosity and liquidity is more emphasized when information environment is poor.

Next, we also examine the relation between religiosity, and liquidity and the trading environment in a more direct way by considering two additional measures: Amihud illiquidity (Amihud, 2002) as a proxy for price impact (Goyenko et al., 2009; Edmans et al., 2013), as well as the probability of information-based trading (*PIN*) (Easley et al., 2002; Brown and Hillegeist, 2007). Chung et al. (2010) document that a firm's enhanced information and trading environment through improved internal corporate governance alleviate price impact and information-based trading. Likewise, and according to the adverse selection model of market making, we expect an increasing *PIN* when private information events happen more frequently and when the absolute and relative intensity of informed trading increases (e.g., Brown and Hillegeist, 2007). Again, given more credible and symmetric information flows for firms headquartered in more religious societies, we also expect a negative relation between religiosity and price impact as well as between religiosity and the probability of information-based trading:

Hypothesis 3: Firms headquartered in counties with higher levels of religiosity reveal lower price impact of trades as well as lower probability of information-based trading.

Finally, we conduct a further analysis to sharpen our understanding of the association between local religiosity and stock liquidity that goes beyond our initial story. In particular, we examine the value implications of religiosity induced higher stock liquidity. For the sake of exposition, we refrain from formulating a hypothesis at this stage and briefly discuss the motivation and results in Section 7.

3 Data and summary statistics

3.1 Firm level sample

Our sample construction begins with all active and dead U.S. companies traded on the NYSE, AMEX, or NASDAQ, which are included in Refinitiv Datastream from 1973 until 2020 (22,420 companies). We then apply the data filtration process proposed by Porter and Ince (2006) and Landis and Skouras (2021). Specifically, we first exclude all non-ordinary shares and minor shares, and shares with missing ISIN codes from our list (10,767 stocks). After removing firms with missing SIC codes (1,326 stocks) (e.g., Ma et al., 2021), we further restrict the sample to firms, which are headquartered in the U.S., and for which we have complete information on their headquarter's location (1,049 stocks are deleted). Furthermore, the utility sector (2-digit SIC code 49) is excluded since it appears as being different from other industries, at least partly, due to regulatory issues. We also omit financials (2-digit SIC codes 60-69) because their balance sheets are different from those of other firms (e.g., Hilary, 2006; Jiang et al., 2018). Finally, we delete all firm-years if any variable of our baseline model is missing, if we observe implausible balance sheet data (e.g., negative book values of equity), or if the yearly average price is below \$5 (e.g., Grullon et al., 2004; Chung et al., 2010; Cai and Shi, 2019).⁶ The final sample consists of 5,365 unique firms for a total of 46,201 firm-year observations for the period from 1997 through 2020.⁷

⁶ We note that the results of our baseline model (see equation 1) are robust when retaining firms with average stock prices above \$1 instead of \$5, or when omitting any price filtration. The baseline results also hold when we keep the companies from the financial and utility sector.

⁷ Note that information on institutional ownership from Refinitiv Ownership Profile (ROP) is available from 1997. Since we detect one singleton observation, our sample size reduces to 46,200 firm-year observations and 5,364 firms in our baseline regression analysis.

3.2 County level variables

We measure the strength of local religious norms, our variable of interest, retrieving data from the Association of Religion Data Archive (ARDA). We use the longitudinal version of the religious congregations and membership files, which contains the adherent and congregation counts of 302 religious groups that participated in at least one the of the 1980-2010 data collections.⁸ We construct the religiosity ratio by summing the number of adherents of all religious denomination in a county and dividing it by the total population (*REL*). A higher level of *REL* indicates stronger religious social norms. In further analyses, we additionally consider religious subgroups, i.e., the number of protestant (*PROT*), catholic (*CATH*), mainline protestant (*MPRT*), or evangelical protestant (*EVAN*) adherents to the total population in the respective county.

Surveys are conducted at approximately ten-year intervals: 1980, 1990, 2000, and 2010. As social norms, and in particular religious adherence tend to change slowly over time, we follow previous studies (e.g., Hilary and Hui, 2009; Jiang et al., 2018; Zolotoy et al., 2019) by linearly interpolating missing values to obtain values for non-survey years, i.e., 1991-1999, and 2001-2009. Since our sample period is from 1997 through 2020, we apply the religious ratios in 2010 for the 2011-2020 period (Shu et al., 2012).⁹

Following previous studies (Pirinsky and Wang, 2006; Hilary and Hui, 2009) we define a firm's location as the location of its headquarters. Refinitiv reports for the (current) location of each firm's headquarter zip (zone improvement plan) codes only, while ARDA use fips (federal

⁸ Data are available at <https://www.thearda.com/Archive/Files/Descriptions/RCMSMGCY.asp>.

⁹ An alternative method is to fill in the data for missing years using the survey value in the preceding year in which the data are available; for example, we fill in missing values from 1991 to 1999 using the religious ratios in 1990. Hasan et al. (2017a and 2017b) apply this practice for data on social capital. As discussed in the following, the second method of filling missing values do not have an impact on our results.

information processing standard) codes to identify the location of a county. To overcome this issue, we use U.S. Census and HUD ZCTA crosswalk files to convert zip codes to fips codes.¹⁰

Additionally, we consider a set of county level demographic factors as these characteristics might affect the degree of religiosity in a county (e.g., Hilary and Hui, 2009). The intention of including these variables is to ensure that the effect of *REL* on liquidity is not contaminated by these factors (Hasan et al., 2017a). Consistent with previous research (e.g., Hilary and Hui, 2009; Kumar et al., 2011; Shu et al., 2012), we control for the size of the population in a county (*TOTPOP*); the county population divided by its area size (*DENSITY*); the percentage of residents aged 25 years or older who hold a bachelor's, graduate, or professional degree (*EDUCATION*);¹¹ the median age of people in the county (*AGE*); the fraction of married people in a county (*MARRIAGE*); the fraction of non-white people in a county (*MINORITY*); the male population to the female population (*MF_RATIO*); and the proportion of republican votes during presidential elections (*ELEC*). We obtain these variables from the 1990, 2000, 2010, and 2020 (at time of writing only *TOTPOP* and *MINORITY* were available from 2020 survey) surveys of the U.S. Census Bureau, while we collect the latter from MIT election data lab (<https://electionlab.mit.edu/data>).¹²

3.3 Dependent variable - bid-ask spread

We use bid-ask spread as our main measure of liquidity. Bid-ask spread has been widely used in prior studies, when determinants of liquidity are investigated (e.g., Grullon et al., 2004; Loughran and Schultz, 2005; Attig et al., 2006; Chung et al., 2010). The bid-ask spread posted

¹⁰ We thank Anthony D'Agostino for providing the Stata routine. The code is available at <https://gist.github.com/a8dx/7e9d5af24101fc66aafa739577713b59>.

¹¹ Although data on per-capita income are available, we do not include this variable as it is highly correlated with *EDUCATION* (correlation between education proxy and the natural logarithm of income is 0.88). Kumar et al. (2011) also omit the average income due to high correlation with education.

¹² We download demographic variables using NHGIS data finder, which is available at <https://data2.nhgis.org/main> (Manson et al., 2021).

by a market maker represents the maximum round-trip trading costs for investors and is basically determined by three factors: inventory holding costs, order processing costs¹³, and adverse selection costs (e.g., Bagehot, 1971; Attig et al., 2006; Lee and Chung, 2018; Stoll, 2000). Since our arguments on the relation between religiosity and liquidity are mainly based on informational aspects of trading costs, the bid-ask spread is a suitable measure of liquidity in our research setting. The bid-ask spread (*BAS*) represents the yearly average of daily bid-ask spreads calculated as $(Ask - Bid)/((Ask + Bid)/2)$.¹⁴ To mitigate the potential effect of data errors and outliers, we exclude all daily *BAS* that are negative, are greater than 50% of the midpoint, or are greater than \$5 (e.g., Chung et al, 2010; Chung and Zhang, 2014). *BAS* indicates illiquidity, thus larger *BAS* implies lower liquidity in the underlying stock. Data on daily bid prices and daily ask prices are retrieved from Refinitiv.¹⁵

3.4 Firm controls

Our aim is to isolate the effect of *REL* on *BAS*. Therefore, we consider a variety of control variables in our analyses that have been identified as relevant in the context of liquidity (e.g., Grullon et al., 2004; Loughran and Schultz, 2005; Hilary, 2006; Attig et al., 2006; Chung et al., 2010). These variables are firm size, stock price, number of analysts, Tobin's Q, capital expenditures, research and development, turnover, standard deviation of daily returns, leverage, cumulative returns, profitability, membership in the S&P 500, exchange listing, institutional ownership, and insider ownership. Table A1 in the Appendix provides detailed variable definitions.

¹³ We capture the order-processing costs by implementing a dummy variable (*NASD*). We explain this in more depth in Section 2.4.

¹⁴ Due to limited data availability, we were not able to retrieve intraday data for calculating the *BAS*. However, as shown by literature (see, e.g., Chung and Zhang, 2014; Fong et al., 2017), daily *BAS* are highly correlated with intraday-based spreads.

¹⁵ We also consider *AMIHUD_SQ* as well as *PIN* as dependent variable, respectively, to investigate our third hypothesis. Please refer to Table A1 in the Appendix for detailed variable descriptions.

Theory suggests that bigger firms reveal greater visibility and provide better information, thus reducing adverse selection risk (e.g., Brennan and Subrahmanyam, 1995; Easley et al., 1998; McInish and Van Ness, 2002; Kedia and Zhou, 2011). Therefore, we control for firm size with total assets (*SIZE*), the average share price (*PRICE*) as well as the number of analysts following a firm (*ANALYST*).¹⁶ Accordingly, high-growth firms may have higher stock liquidity due to higher attention from both media and investors (e.g., Chung et al., 2010; Gopalan et al., 2012). Growth options and investment opportunities are proxied by Tobin's Q (*Q*), capital expenditures (*CAPEX*), and research and development (*RnD*). We further include share turnover (*TURNOVER*) as we expect that increased investors' interest and high trading activity leads to lower spreads.¹⁷ Moreover, risk has an impact on liquidity since higher risk is related to higher spreads (e.g., Grullon et al., 2004). For this purpose, we include the yearly standard deviation of daily returns (*RISK*) and leverage (*LEVERAGE*) to proxy for risk. To control for the performance of firms, we further consider profitability (*ROA*) and cumulative stock returns (*CUMRET*) because investors are likely to be attracted by successful firms (Grullon et al., 2004).

Prior empirical work has also shown that the ownership structure, i.e., percentage of shares held by institutional and insider investors, of a company affects liquidity (e.g., Attig et al., 2006; Chung et al., 2010). For institutional ownership, at the end of each quarter, we identify all 13(f) institutions that are invested in the firm and calculate institutional ownership as the sum of all

¹⁶ Following previous literature (e.g., Chang et al., 2006; Guo et al., 2019) we set missing analyst firm years to zero. Thus, we implicitly assume that missing earnings forecasts is due to no analyst coverage (e.g., Chan and Hameed, 2006; Chang et al., 2006).

¹⁷ Turnover is defined as the annual average of total monthly number of shares traded divided by shares outstanding (Grullon et al., 2004). Following Lesmond (2005), we determine the shares outstanding and the respective adjustment factor annually, i.e., we keep the value constant throughout the year. Since our winsorized raw turnover measure is highly skewed, we use its natural logarithm in the following calculations (e.g., Jayaraman and Milbourn, 2012; Grullon et al., 2004; Chung and Charoenwong, 1998; Chordia et al., 2001). To ensure comparability between NYSE/NYSE AMEX and NASDAQ firms, we follow Gao and Ritter (2010) by adjusting the trading volume prior to 2004 for firms listed on NASDAQ. Please refer to Gao and Ritter (2010, p. 51-52) for a detailed explanation. Additionally, we also account for different regimes in market microstructure on NASDAQ and NYSE/AMEX by adding time-fixed effects.

holdings in the firm divided by the shares outstanding in the respective quarter. From this, we receive our yearly ownership variable (*OWN_INST*) by taking the average of the quarterly data over the calendar year. As pointed out by Baghdadi et al. (2018), focusing on the average level of institutional ownership reduces an impact by periodic sharp increases or decreases at a specific point in time. Following prior literature (e.g., Gompers and Metrick, 2001; Ferreira and Matos, 2008), we set *OWN_INST* to zero if a stock is not held by any institution, i.e., if institutional ownership is missing. In special cases, we observe, however, that our institutional ownership variable exceeds 100%.¹⁸ In such circumstances, we set the maximum ownership proportion of institutions to 100% (e.g., done in Lewellen, 2011; Striewe et al., 2016). We furthermore include shares held by closely related investors (*OWN_INSIDER*) to proxy for insider ownership (e.g., Ferreira and Matos, 2008 and 2010; Chung et al., 2010).

Moreover, following Hilary (2006), we include an indicator variable (*NASD*) to control for the exchange the firm is traded on. The dummy variable *NASD* takes the value of one if the respective firm is traded on NASDAQ and zero otherwise. We consider this variable to control for systematic microstructure differences between exchanges since past research reports that firms traded on NASDAQ are associated with higher spreads (e.g., Huang and Stoll, 1996; Hilary, 2006). This dummy variable also helps us to account for institutional features of the exchanges (e.g., Gao and Ritter, 2010).¹⁹

Lastly, to capture the effect of being included in a leading stock index and to control for possible industrial differences, we follow Agarwal (2007) and Chung et al. (2010) by

¹⁸ See Striewe et al. (2016) for a detailed discussion of possible reasons.

¹⁹ More recent research, however, find that the difference in NYSE and NASDAQ average spreads diminished after market reforms on NASDAQ started in 1997 (e.g., Weston, 2000; Dang et al., 2018). Further, our baseline result is also robust when we add an additional indicator for AMEX listed firms. Moreover, it is possible that companies voluntarily switch their trading location, e.g., from NYSE to NASDAQ (e.g., Dang et al., 2018). In a further test, to reduce concerns regarding liquidity effects stemming from changes in exchanges, we exclude all companies that experience a change in their trading location. The result of our baseline model does not alter with this specification.

considering a dummy variable for firms included in the S&P 500 index²⁰ (*SP500*) and dummy variables for two-digit SIC codes. We also include time dummies to control for time trends in liquidity, e.g., resulting from changes in minimum tick size, i.e., decimalization period. All presented control variables are collected from Refinitiv Eikon and Refinitiv Datastream, while ownership data comes from Refinitiv Ownership Profile (ROP).

3.5 Descriptive statistics

Table 1 provides descriptive statistics of our variables presented in Sections 3.1-3.4. Our liquidity measure (Section 3.3) and control variables (Section 3.4) are winsorized at the upper and lower 1-percentile by year to reduce the effect of outliers.²¹

[Insert Table 1 about here]

As Panel A in Table 1 shows, the mean value of *BAS* in our sample is 0.0077, indicating round trip trading costs of 77 basis points (bps) on average. The mean value and standard deviation (0.0134) are comparable to the quoted spread measures reported in existing studies (e.g., Chung et al., 2010; Gopalan et al., 2012).

REL reveals a mean value and standard deviation of 0.5163 and 0.1072, respectively, similar to the statistics documented in Callen and Fang (2015), and Jiang et al. (2018), among others. Unreported results show that the most religious state is Utah (*REL* equals 0.7558 with 50 companies headquartered here), while the least religious state is Maine (*REL* is 0.3189; number of companies is 7). We further observe that our companies are headquartered in 485 counties

²⁰ Each end of the year, we check the constituents list of S&P 500 companies. We then get a time series of yearly constituents, which we merge to our main dataset.

²¹ Implementing alternative methods of winsorizing, i.e., over the whole sample period or by year-industry, or omitting winsorization, do not affect our conclusions drawn from the baseline model.

with most of the companies being located in Santa Clara (298, California), followed by Middlesex (254, Massachusetts) and New York (220, New York). These counties reveal religiosity ratios of 0.4344, 0.6523, and 0.5325, respectively.

With regards to our controls, the average firm in our sample is covered by 7.70 analysts (untabulated) and its institutional ownership (insider ownership) is 0.6448 (0.1933).²² The average price of firms is \$42 with total assets of \$4.3 B (untabulated) and a mean Q of 2.30, indicating the coverage of rather larger firms (Agarwal, 2007). Taken together, our firm controls and demographic characteristics are mostly comparable to those reported in existing studies (e.g., Grullon et al., 2004; Chung et al., 2010; Kumar et al., 2011; Callen and Fang, 2015; Jiang et al., 2018; Zolotoy et al., 2019; Albuquerque et al., 2019).²³

A Pearson correlation matrix for the variables used in our baseline analysis (see Table IA1a, Panel A in the Internet Appendix (IA)) reveals that the correlation between BAS and REL is slightly positive (0.0207) and statistically significant at the 1% level, which contrasts with our first hypothesis. However, we should be cautious of overinterpreting this simple pairwise correlation since it does not control for the impact of other effects (e.g., the decimalization period). Furthermore, the level of pooled correlation between our controls is generally moderate (< 0.50), with some exceptions (see Table IA1a, Panel A in the IA). Further results of correlations among demographic variables (see Table IA1a, Panel B in the IA) indicate that REL is significantly correlated with the different demographic controls, with exception of $EDUCATION$ and $\ln(AGE)$. The correlations among demographic variables are reasonably moderate.²⁴

²² Note that the average analyst coverage (institutional ownership) is 8.14 (0.6644), when we omit companies with zero analysts following (zero institutional ownership). In our sample, 91.24% (99.61%) of firms are covered by at least one analyst (one institutional investor). Additionally, we document 6.97% of firm-years, in which institutional ownership exceeds 100%.

²³ We provide descriptive statistics for all additional variables used in further analyses in the Internet Appendix (Section IA1, Table IA1b).

²⁴ The average variance inflation factor value is 2.10.

4 The relation between religiosity and liquidity

4.1 Baseline regression results

To study the relation between religiosity and liquidity, we estimate our baseline model using ordinary least squares (OLS) regression technique with standard errors adjusted for heteroscedasticity and firm clustering.²⁵ Specifically, we employ the following empirical model (hereafter “baseline model”) to test our first hypothesis:

$$\begin{aligned}
 BAS_{i,t} = & \beta_0 + \beta_1 REL_{i,t} + \delta' FIRM_CONTROLS_{i,t} + \gamma' DEMO_{i,t} + \sum_j IND_j \\
 & + \sum_s YEAR_s + \varepsilon_{i,t}
 \end{aligned} \tag{1}$$

$BAS_{i,t}$ denotes the dependent variable representing our main liquidity measure, the bid-ask spread. $REL_{i,t}$ is the main variable of interest, computed as the sum of all adherents divided by the total population in the county of the firm’s headquarters location. Higher levels of $REL_{i,t}$ correspond to stronger religious norms. Based on our first hypothesis we expect a negative relation between $REL_{i,t}$ and $BAS_{i,t}$. The vector $FIRM_CONTROLS_{i,t}$ captures the firm attributes discussed in Section 3.4, $DEMO_{i,t}$ is a vector containing county level demographic factors (Section 3.2), $\sum IND_j$ denotes industry fixed effects based on the 2-digit SIC,²⁶ $\sum YEAR_s$ represents year fixed effects, and $\varepsilon_{i,t}$ is the error term.²⁷

²⁵ Our baseline results are robust if we adjust for year clustering, county level clustering, or year and county double clustering.

²⁶ The results are quantitatively and qualitatively the same when we rerun our baseline model using 1-digit or 4-digit SIC codes rather than 2-digit SIC codes.

²⁷ For the sake of brevity, we omit subscripts in the following.

The regression results are reported in Table 2. We estimate six specifications of the OLS regression. Column (1) reports the results of a reduced model, which is intended to minimize potential concerns arising from spuriously correlated independent variables (e.g., Hilary, 2006). In Column (2) we control for all firm characteristics presented in Section 3.4. In Column (3) we estimate our baseline model (see equation 1). Column (4) documents the results of a sample restricted to the two years for which we have direct survey data to mitigate bias due to linear interpolation. Finally, in column (5) and (6), we estimate our baseline model for the survey years separately.

[Insert Table 2 about here]

Across all specifications, the estimates on *REL* are negative and statistically significant. The point estimates are -0.0028, -0.0031, -0.0033, -0.0050, -0.0036 and -0.0054 for columns (1) to (6), respectively. Our results are consistent with hypothesis 1 that firms located in U.S. counties with higher level of religiosity reveal significant lower bid-ask spreads.²⁸

To address the economic significance of our results, we compare our economic magnitude of *REL* to that reported in Chung et al. (2010) for governance. We follow their way of quantifying the economic impact, thus moving from the first quartile (0.4369) to the third quartile (0.5945) of *REL*, the bid-ask spread decreases by 0.052% ($=0.0033 \times 0.1576 \times 100$). This is approximately 6.75% ($=0.00052/0.0077 \times 100$) of the mean spread for the average firm. Similar calculations in Chung et al. (2010) reveal that a raise in governance standards would

²⁸ The application of the alternative method of constructing *REL* and the vector of demographic variables by keeping values constant between the survey years (see Section 3.2, footnote 9) does not affect our conclusion. The coefficient slightly decreases to -0.0031 (t-statistic = -3.38).

decrease firms' quoted spreads by 4.5% of the mean quoted spread. The authors conclude an economically significant effect.²⁹

We next turn our focus to the control variables of column (3). The results are mostly consistent with our theoretical predictions and the effects documented in previous literature (e.g., Grullon et al., 2004; Hilary, 2006; Chung et al., 2010; Kedia and Zhou, 2011; Pham, 2020, among others). As expected, $\ln(TURNOVER)$, $\ln(SIZE)$, Q , RnD , $CAPEX$, ROA as well as $\ln(1+ANALYST)$ have negative coefficients. Furthermore, spreads are lower for firms with higher institutional ownership (OWN_INST), indicating that corporate monitoring is better. In turn, firms with higher risk ($RISK$), higher prices ($\ln(PRICE)$)³⁰, and leverage ($LEVERAGE$) tend to have higher spreads. Spreads are also higher, when the firm is listed in the S&P500 ($SP500$), which may be explained by poorer governance of these firms (Chung et al., 2010). Surprising is the lack of significance for the indicator variable for NASDAQ listings ($NASD$). However, since we already adjust for institutional features of the trading volume (Gao and Ritter, 2010), this may explain the insignificance of the dummy. Additionally, as shown by Weston (2000), differences in average spreads between NYSE and NASDAQ firms decreased after market reforms on NASDAQ started in 1997. Likewise, our variable on insider trading ($OWN_INSIDER$) is statistically insignificant, which may be explained by the information content being already captured by other variables (e.g., OWN_INST , $\ln(1+ANALYST)$) included in our baseline model (similarly in Chung et al., 2010). With regards to our set of demographic factors, we find that $\ln(TOTPOP)$ is negative and statistically significant related to BAS at the 10% level.

²⁹ Of course, this way of quantifying the economic significance is not limited to governance, but is also consistent with several studies on religiosity (e.g., Callen and Fang, 2015; and Jiang et al., 2018).

³⁰ We note that the findings with respect to $\ln(PRICE)$ among our different models are mixed. For example, we observe a negative and statistically significant sign in Model (5), while the coefficient is positive in our baseline regression. Narayan et al. (2015, p. 4497-4498) provide a detailed discussion of arguments, which motivates both, positive and negative effects of price on spreads. In addition, these findings may be also explained by pre- and post-decimalization effects (e.g., Gibson et al., 2003).

4.2 Additional control variables

In this section, we examine whether our main results are robust to the inclusion of supplemental control variables.³¹ We did not include these additional covariates in the main analysis because they mostly come at the cost of observations.

So far, our story relates to the effect of firms' religiosity on stock liquidity, mainly arising from trust in corporate behavior and quality in information flows. However, literature suggests that companies with good governance are likely to have enhanced financial and operational transparency, which reduces information asymmetry, and, consequently, improve stock liquidity (e.g., Chung et al., 2010).³² Although we capture some specific governance features in our main analysis (e.g. institutional ownership, number of analysts following a firm), we now explicitly include further aspects of (internal) corporate governance. Following previous studies (e.g., Kim et al., 2014; Callen and Fang, 2020; Hossain et al., 2021, among others), we consider Refinitiv's governance pillar score (*GOV*)³³, whether the CEO is a board member (*CEO_DUAL*), the board size (*BFSIZE*), or whether the firm has a CEO-Chairman separation (*BINDEP*). In addition, we rerun our baseline model by including an indicator variable, which is one if the respective firm is audited by a Big4 company (*BIG4*). Finally, we identify the first principal component extracted from the five above mentioned governance controls (e.g., Callen and Fang, 2020). Results are presented in models (1) to (6) in Table 3. We find that the coefficient on *REL* remains negative and statistically significant in all specifications, at least at

³¹ Due to the lack of readily available data, we are not able to include data based on CEO characteristics.

³² Likewise, Roy et al. (2022) document a liquidity-enhancing effect of the introduction of a mandated CSR regulation in India.

³³ For a detailed description of the categories and governance standards, which are included in the score, see Benz et al. (2020). Our results remain unchanged, when we use the overall environmental, social and governance (ESG) score of a firm instead of the isolated governance metric.

the 10% level, although the sample sizes are much smaller for the tests using governance variables, with exception of Model (5).³⁴

[Insert Table 3 about here]

Second, Grullon et al. (2004) detect that the degree of visibility, proxied by advertising expenses, has an impact on liquidity through the channel of information asymmetry. To further ensure that our results are not driven by this explanation, we rerun our baseline model including selling, general, and administrative scaled by total assets (*SG&A*) to proxy for advertising (e.g., Hawn and Ioannou, 2016). In Model (7), we show that our inferences on *REL* remain unchanged when controlling for advertising intensity.³⁵

Third, we add two balance sheet controls that are asset tangibility (*TANG*) and segment concentration of a firm (*COMPLEX*). Extant studies suggest that the asymmetric information problem is reduced for firms with more tangible assets, while stronger industry concentration may be associated with an increase in adverse selection risk (e.g., Grullon et al., 2004; Hilary, 2006; Chung et al., 2010). The coefficient on *REL* (Model 8) remains negative and statistically significant at the 1% level.

Finally, social capital as a measure of non-religious trustworthiness (*SOCIAL*) could constrain opportunistic firm behavior, thus probably impact the information environment of a firm. We also consider the abortion rate (*ABORT*), the per capita alcohol consumption (*ALC*),

³⁴ To be clear at this stage, we do not claim that second market liquidity of firms should be reduced to religiosity. For instance, the coefficient on the first principal component (Model (6)) is also negative and statistically significant at the 1% level. As we will show in Section 5, the impact of local religious norms is particularly evident for firms that operate in poor information environments, e.g., for small firms or firms with no analyst coverage. Since governance metrics from Refinitiv tend to be available for larger US companies to date (Refinitiv, 2022), it is a challenging task to finally judge the relevance of religiosity compared to corporate governance for the available set of companies.

³⁵ The unreported coefficient on our advertising coefficient is positive and statistically significant, which is in contrast with Grullon et al. (2004). This result may be explained by the different sample period and the fact that we already control for a variety of visibility measures (e.g., *SP500*, $\ln(\text{SIZE})$) in our baseline analysis.

and the state-GDP (*SGDP*) to further ensure that supplemental demographic factors do not drive our results (e.g. Hasan et al., 2017a; Jiang et al., 2018; Amin et al., 2021).³⁶ Results are documented in Model (9). Again, our coefficient on *REL* is statistically significant at the 1% level.

Even when we put all measures of alternative explanations together in one model along with the variables used in our baseline regression, the result holds (Model 10). Taken together, our results confirm the notion that religiosity probably fosters trust in corporate behavior and information flows, thus increasing stock liquidity.

4.3 Robustness tests

4.3.1 Different variable and model specifications

In this section, we examine whether our results are robust to different variable and model specifications of our baseline model. Table 4 reports the results.

Previous studies (e.g., Chung et al., 2010; Lee and Chung, 2018; Pham, 2020) approximate trading activity by using the natural logarithm of dollar trading volume (*TVOL*).³⁷ For this purpose, we replace in our baseline model $\ln(\textit{TURNOVER})$ by $\ln(\textit{TVOL})$. Additionally, Jayaraman and Milbourn (2012) point out that turnover itself may also be used as a measure of liquidity. Thus, we reestimate our baseline model by excluding $\ln(\textit{TURNOVER})$. Table 4, Models (1) and (2) present the results. We find that the coefficients on *REL* are still negative and statistically significant at the 1% level. Our next test addresses the concern about biases

³⁶ We obtain data on social capital from the Northeast Regional Center for Rural Development at Pennsylvania State University (NRCRD; <https://aese.psu.edu/nercrd/community/social-capital-resources>), while we receive the other measures from the National Institute on Alcohol Abuse and Alcoholism (<https://pubs.niaaa.nih.gov>), from the Guttmacher Institute (data.guttmacher.org; <https://osf.io/u58vf/>), and from the bureau of economic analysis (<https://www.bea.gov/data/gdp/gdp-state>), respectively.

³⁷ We do not use dollar trading volume in our baseline analysis since it captures the size effect, that is bigger companies have higher trading volume (e.g., Brennan et al., 2013). Consequently, our size variable is highly correlated with the dollar trading volume (correlation = 0.818).

stemming from our contemporaneous model. We rerun our baseline model by using lag one of continuous firm characteristics (e.g., Adhikari and Agrawal, 2016). The effect of religiosity is robust to these alternative specification (see Model 3), while the magnitude of the coefficient is close to that reported for the baseline regression. Finally, concerns may arise due to the replacement of missing values with zero for *RnD* and $\ln(1+ANALYST)$. In the same vein, the adjustment of our ownership variables (*OWN_INST* and *OWN_INSIDER*) may bias our results. For this purpose, we rerun our baseline model simply dropping all missing observations, and additionally, all observations above 100% for ownership variables (Model 4). Although the sample size is reduced to 22,166 observations, our overall conclusion drawn in Section 4.1 still holds.

[Insert Table 4 about here]

The next set of model specifications tackles concerns arising from geographical clustering of our sample. We start by excluding all companies, which are headquartered in the most conservative states since states such as Utah tend to be more religious and less diverse than other states (Cai and Shi, 2019).³⁸ Besides excluding the most conservative states, we also rerun our baseline model omitting the five most and five least religious counties. As documented in Section 3.5, most companies are headquartered in California, Texas, and New York. Thus, in our subsequent test, we exclude all companies, which are headquartered in these states to ensure that our results are not driven by these states. In a similar vein, we create a subsample, which omits the five largest counties in terms of number of observations. We also address the possibility that our results are driven by rural companies (e.g., Loughran and Schultz, 2005).

³⁸ Cai and Shi (2019) provide a list of the most conservative states, that are, Mississippi, Idaho, Alabama, Wyoming, Utah, South Dakota, Louisiana, North Dakota, South Carolina, Arkansas.

Therefore, we reestimate equation (1) using the ten largest metropolitan statistical areas only. Next, to capture differences in legal and social environment, we rerun our baseline model using state-level fixed effects along with the industry and year fixed effects (e.g., Hilary and Hui, 2009; Jiang et al., 2018; Cai and Shi, 2019). Finally, we follow Hilary and Hui (2009) and Cantrell and Yust (2018) by running a cross-sectional analysis on county level. We calculate the average value of each variable included in our baseline model for each county and reestimate equation (1) with standard errors adjusted for heteroscedasticity and county clustering. This test should minimize the risk that our results are driven by a small number of counties. Results of geographical subsampling are reported in Table 4, Models (5) to (11). We find that our results remain robust. Turning our focus to the state fixed effects regression in Model (10) as expected, the effect is much weaker in terms of statistical significance since the religiosity ratio is relatively stable over time. However, the coefficient is still statistically significant at the 10% level.³⁹

We now focus on temporal subsampling. Results are reported in Table 4, Models (12) to (16). We first divide the sample into two equal subperiods, i.e., from 1997-2008 and 2009-2020. The respective coefficients and significances for *REL* are reported in Models (12) and (13). Additionally, we rerun our baseline model considering (omitting) the financial crisis period from 2007 through 2009 (see Models 14 and 15). Finally, we conduct a cross-sectional regression using the Fama/MacBeth-procedure (Model 16) to rule out the possibility that the results are driven by cross-sectional correlation in a few years (e.g., Cantrell and Yust, 2018).

³⁹ Due to little within-firm time series variation in religiosity ratios, it is inappropriate to conduct a firm-fixed effects regression in this setting (e.g., Adhikari and Agrawal, 2016; and Zolotoy et al., 2019; for a discussion). Frijns et al. (2016) note that firm-fixed effects may lose statistical power in settings, in which the effects of variables differ mostly in the cross-section rather than over time. This is probably the reason that, with firm-fixed effects, *REL* turns insignificant. However, the negative sign persists. Moreover, as pointed out by Cantrell and Yust (2018), due to (technically induced) little yearly time series variation of *REL*, a full changes model seems also not appropriate in this setting. This could explain the insignificant coefficient on the first difference of *REL* (t-statistic = -1.59), when we estimate a full changes specification until the last survey year in 2010.

The coefficients on *REL* remain negative and statistically significant, at least, at the 5% level, thus supporting our first hypothesis.

4.3.2 Alternative measures of religiosity

Following McGuire et al. (2012) and Amin et al. (2021), we also consider an alternative measure of religiosity. We regress *REL* on the demographic controls, such as population, density, education, age, marriage, minority, election, social capital, alcohol consumption per capita, and abortion rate, respectively, and use the residuals as our revised measure of religiosity (*RES_REL*). This test is intended to purge religiosity from confounding effects of county characteristics. The reestimation of our baseline model produces similar results to those reported in Table 4 (see Model 17). Additionally, we replace *REL* by an indicator variable, which takes the value of one, if *REL* is greater than the yearly sample median, and zero otherwise (*HIGH_REL*). We replicate this analysis by using the sample tercile by year as an alternative threshold (*HIGH_REL1*). Results are reported in Model (18) and (19). The results reveal that alternative definitions of our religiosity variable do not alter our conclusion drawn in Section 4.1.

4.4 Religious subgroups

So far, we investigate whether the overall religious attitude has an impact on stock liquidity. However, an interesting question is related as to whether the type of religiosity also matters for liquidity. For this purpose, we replace the total religiosity ratio with catholic ratio (*CATH*), and protestant ratio (*PROT*), and in an additional analysis, we further decompose *PROT* into mainline protestants (*MPRT*) and evangelical protestants (*EVAN*). Jiang et al. (2018), for example, document that both protestant and catholic rate, respectively, have a significant impact

on the cost of debt, thus indicating that both subgroups matter in the same direction. Our analysis of religious subgroups reveals similar (see Section IA2, Table IA2) results, that is, both catholics and protestants negatively affect the bid-ask spread. A Wald-Test on the differences between coefficients of both religious groups furthermore indicate that the influence of *CATH* is not statistically significant different from *PROT*. When we additionally replace *PROT* by the protestant subgroups, i.e., *MPRT* and *EVAN*, we document a significant negative coefficient for *CATH* and for *EVAN* at the 5% level, while *MPRT* is insignificant. However, Wald-Tests show that the coefficients are not statistically significant different between religious subgroups (e.g., similar in Gao et al., 2017). Although catholics are more likely to be involved in gambling (e.g., Kumar et al., 2011), catholics and protestants probably share common characteristics of norms and values (e.g., solidarity and trust; Amin et al., 2021). Therefore, we posit that especially the overall strength of local religiosity impacts trust in corporate behavior and information production.

4.5 Endogeneity

We interpret religiosity as an exogenous antimanipulative factor that serves as a commitment device conveying honesty and credibility of firms' corporate actions and information flows. Our results so far suggest that companies headquartered in areas with stronger religiosity exhibit lower bid-ask spreads. Although we present several tests on omitted variables and model specifications in Sections 4.2 and 4.3, respectively, the relation between religiosity and liquidity may still be spurious due to reverse causality or further omitted correlated variables. Therefore, this section aims at establishing a potential causal link between religiosity and liquidity by conducting a set of causality tests (e.g., Adhikari and Agrawal, 2016; Jiang et al., 2018; Cai and Shi, 2019; Cai et al., 2019; Zolotoy et al., 2019). Of course, we acknowledge the fact that establishing causality in the setting of religiosity is a challenging and almost unsolvable task

(Cantrell and Yust, 2018). Besides numerous demographic county variables or other factors that may be correlated with religiosity, which we not aware of, the slow changing behavior of religious norms makes it difficult to, at least partially, rule out endogeneity since statistical techniques (especially fixed-effects regression) are not applicable in such settings (e.g., Zolotoy et al., 2019; Cantrell and Yust, 2018). The results are presented in Table 5.

4.5.1 Reverse causality and instrumental variable analysis

We start our endogeneity section by focusing on reverse causality, that is, the notion that the change in religiosity of the headquarter county is due to the firm-specific bid-ask spread. Numerous studies mentioned that relocations caused by changes in market outcomes are very unlikely (e.g., Callen and Fang, 2015; Cai and Shi, 2019). However, as pointed out by Cantrell and Yust (2018), firms may endogenously select headquarters that are common with their social religious norms, and it is possible that the information flows, and consequently, stock liquidity, could affect religiosity. For instance, bankrupt companies could harm local economy, resulting in job losses that force many employees to move. For this purpose, we conduct two additional analyses to mitigate concerns of biases arising from reverse causality.

First, following John et al. (2011), and Cai and Shi (2019), we rerun our baseline model using a subsample of firms, which operate in the manufacturing, mining, and agriculture sectors (SIC codes 100-3999). These firms are more likely to choose their headquarters location based on production and supply considerations rather than on religiosity. Second, following Cantrell and Yust (2018), and Callen and Fang (2020), we reestimate our baseline model using the fitted values of *REL* within a two-stage least squares approach (2SLS). These are estimated from the first stage regression of *REL* on our instrument, that is the county level religiosity in 1980⁴⁰,

⁴⁰ We do not use the religiosity ratio in 1971, since we use the adjusted longitudinal file in our main analysis (see Section 3.2).

including all other variables used in our baseline model (see equation 1). Since the religiosity ratios in 1980 are 17 years before our sample period starts, it seems very unlikely that the 1980s values of county religiosity are correlated with current firm liquidity (exclusion condition). However, as religiosity ratios tend to change very slowly over time, the values in 1980 should meet the relevance condition, i.e., should be sufficiently correlated with *REL*.⁴¹

Model (1) and (2) in Table 5 report the results. Consistent with our previous findings, the coefficient on *REL* remains negative and statistically significant, at least at the 10% level.

4.5.2 Placebo test

Our next attempt to address concerns of omitted variables is to conduct a placebo test (e.g., Zolotoy et al., 2019; Ma et al., 2021). This method is intended to further examine whether our main findings (see Section 4.1) are solely an artefact of omitted county variables. As described in Ma et al. (2021), we randomly shuffle religiosity ratios among each of our sample counties.⁴² For example, we replace the “true” religiosity ratios for years 1990, 2000, and 2010 of county A with the “false” ratios of county B; we then substitute the “true” ratios of county B with the “false” fractions of county C.⁴³ If there are omitted county level variables driving our results, we would expect that the coefficients on pseudo-*REL* are still negative and statistically significant, i.e., pseudo t-values would be smaller than -1.65, since the distribution of *REL* itself

⁴¹ The unreported first stage regression reveals that, as expected, our instrument is highly predictive of *REL*. The coefficient is 0.6375, significant at the 1% level, which is in line with results reported in Cantrell and Yust (2018). The first stage F-statistic (3,315.17) is also highly significant (p-value = 0.000), indicating that the weak instrument problem is not an issue. Moreover, the Hausman test (Hausman, 1978) fail to reject the exogeneity of *REL* (p-value = 0.638; similar results in Callen and Fang, 2020). In an unreported robustness test, we follow an alternative strategy proposed by Hilary and Hui (2009), and conducted in Callen and Fang (2015) and Jiang et al. (2018), among others, by taking lag 3 of *REL* and lag 3 of $\ln(TOTPOP)$ as our internal instruments. The second stage result of the fitted coefficient on *REL* is -0.0023, being significant at the 5% level (t-statistic = -2.10). The overidentification test in this setting shows that our instruments are valid.

⁴² In this test, our number of counties reduces to 483, since counties “Miami-Dade” and “Chesapeake City” reveal missing values for survey year 1990.

⁴³ We also apply this strategy to our demographic controls, except for *ELEC_REP*, since it is measured on state-level. However, the results still hold when we use pseudo values for *REL* only.

is not altered with this test. We repeat this procedure 500 times and plot the distribution of the pseudo t-values in Figure 1 (Ma et al., 2021).

[Insert Figure 1 about here]

Evidently, the pseudo t-values are distributed around zero, thus indicating that the coefficients of pseudo-*REL* are mostly not statistically significant different from zero. This finding is supported by a t-test on the pseudo t-values, which fails to reject the null hypothesis that the pseudo t-values are statistically significant different from zero (t-statistic = 0.86). Furthermore, the t-value of *REL* estimated in our baseline model (see Table 2, Column 3), represented by the vertical solid line, lies outside the distribution of the pseudo t-values. Taken together, the placebo test implies that our findings are unlikely to be driven by omitted county level variables.

4.5.3 Entropy balancing and propensity score matching

To further improve identification between *REL* and *BAS*, we conduct entropy balancing (Hainmueller, 2012; and Hainmueller and Xu, 2013) like Jiang et al. (2018), Cai et al. (2019), and Mayberry (2020). We group firms into high- or low-religiosity strength by constructing an indicator variable (*HIGH_REL_DUMMY*), which is one if a firm's religiosity ratio is within the top tercile during the year (=treatment group), and zero if it is in the bottom tercile during the year (=control group). The implementation of entropy balancing is summarized as follows. First, we match our treatment and control group via the maximum-entropy reweighting scheme (Hainmueller and Xu, 2013) based on the first moment (Jiang et al., 2018; Cai et al., 2019) of all covariates used in our baseline model: firm characteristics (Section 3.4) and county attributes (Section 3.2). Specifically, entropy balancing computes weights for every control observation

such that the average equals those of the treatment observation. The reweighting scheme ensures that the first moment condition of the treatment group and the reweighted control group are virtually equal, and has the advantage that all control firms remain in the sample (e.g., Hainmueller, 2012; Hainmueller and Xu, 2013; Jiang et al., 2018; and Mayberry, 2020).⁴⁴ In the second step, we then use the weights obtained in the first step for the regression analysis with the treatment indicator (*HIGH_REL_DUMMY*) as the main explanatory variable, including all key covariates used in our baseline analysis.⁴⁵ Test diagnostics with respect to the means after the matching procedure reveal that we observe no statistically significant difference in the means of firm characteristics as well as county characteristics after the matching procedure (see Table IA3a, Panel A in the IA). Turning our focus to the results for the weighted regression (Table 5, Model 3a), we continue to find a negative and statistically significant effect, indicating that bid-ask spread decreases with the level of religiosity for firms that are virtually identical in other firm characteristics as well as county attributes.⁴⁶ In a subsequent test, we also verify that our results are not driven by extreme balancing weights (see, Hainmueller, 2012; Mayberry, 2020). For this purpose, we trim the extreme percentiles of our balancing weights, i.e., the 1% and 99% percentiles. As shown in Table 5 (Model 3b) our results are not affected by this adjustment.

For robustness purposes, we also conduct propensity score matching. Following Adhikari and Agrawal (2016), our first step is to estimate a logit model that regresses *HIGH_REL_DUMMY* on the set of firm characteristics described in Section 3.4, including

⁴⁴ See Hainmueller (2012), Hainmueller and Xu (2013), Jiang et al. (2018), and Mayberry (2020), for the main advantages of using entropy balancing over other conventional preprocessing schemes, such as propensity score matching.

⁴⁵ We use Stata's option [*pweight=_webal*] in combination with *reghdfe* for the regression analysis in the second step (see, Hainmueller and Xu, 2013, p. 13, for similar procedure with respect to the weight option).

⁴⁶ Our results are robust if we match on higher moments, i.e., variance or skewness (p-values < 0.10). However, when matching on the third moment, we excluded *MF_RATIO* from the matching scheme in the first step due to near collinearity with other demographic covariates (see IA, Table IA3a, Panel B)

time-fixed and industry-fixed effects.⁴⁷ We then use one firm located in a high religious area to match it to the closest firm headquartered in the low religious group without replacement and a caliper of 0.00001. Although the choice of the matching parameters comes heavily at the cost of losing observations, potential causal inferences can be only made when reasonable balancing is achieved. Our test diagnostics show that all covariates exhibit standardized biases lower than 3.1% after matching. Caliendo and Kopeinig (2005) interpret a standardized bias below 3% or 5% as sufficient. Moreover, all covariates are not statistically significant different for our treatment and control group after conducting propensity score matching (see Table IA3b, Panel B in the IA).⁴⁸ As shown in Table 5 (Model 4), we continue to find a negative relation between religiosity and liquidity for the propensity matched sample, although our sample size is severely reduced.⁴⁹ Our results are robust to the alternative matching procedure, thus not affecting our conclusions drawn from entropy balancing.

[Insert Table 5 about here]

4.5.4 Company relocation

In the context of local social norms, it is a challenging task to find an appropriate exogenous shock to provide additional evidence of potential causality (Amin et al., 2021). However, we

⁴⁷ Unlike with entropy balancing, we omit county attributes for propensity score matching in the first step to obtain reasonable balancing of the covariates (e.g., similarly done in Cai et al., 2019). Results of the first stage logit model can be found in the internet appendix (Table IA3b, Panel A).

⁴⁸ We use Stata's user-written command *pstest* (Leuven and Sianesi, 2003) to perform variables balance check before and after the matching procedure. For the second-stage regression, we use all treated observations that are on common support.

⁴⁹ We additionally carried out the following matching procedures: Matching with replacement, one nearest neighbor, and caliper of 0.0001 (e.g., Adhikari and Agrawal, 2016); matching without replacement, three nearest neighbors (e.g., Cai et al., 2019); and one-to-one matching with replacement, and caliper of 0.01 (e.g., Mayberry, 2020). As done in Mayberry (2020), in the auxiliary analyses with replacement, we employ frequency weights (Stata's option *[fweight=_weightl]*) to accommodate that one control firm can match with multiple treatment firms. In contrast, the matching procedure using three nearest neighbors requires the use of probability weights. Our Internet Appendix (Table IA3b, Panel C) reports the results.

argue that a headquarter (HQ) change may act as a sufficient exogenous event that drives changes in religiosity. For this purpose, we use historic HQ changes of a company as a quasi-exogenous shock to *REL* (e.g., Chhaochharia et al., 2012; and Hasan et al., 2017a, 2020; in alternative settings). To conserve space, Internet Appendix IA4 provides a detailed description of the initial sample construction.

We define a HQ relocation event when a firm reveals a change in the corporate headquarters' fips code located across two different states in two consecutive years (Chhaochharia et al., 2012). To be included at this stage, a company must satisfy the sampling criteria described in section 3.1, and, additionally, has been active at least two years before and after the relocation occurred (e.g., Hasan et al., 2017a).⁵⁰ Because the full sample covers the period from 1997-2020 and we require a two-year window around the relocation event, relocations are detected in the 1999-2018 period. Finally, we omit firms with multiple relocations to avoid any confounding events (Hasan et al., 2017a). With this strategy, we identify 164 relocation events in total, while the most take place in 2003, 2011, and 2017 (12 relocations), and the least in 1999 and 2001 (3 relocations).

In the next step, we measure the “strength” of changes in religiosity due to the HQ relocation as the difference between the *REL* in the post- and pre-relocation period. Specifically, it thus represents the difference in *REL* one year after the relocation and one year before the relocation. Subsequently, we create two dummy variables to pin headquarter changes involving a “measurable” change in religiosity (in a similar vein done in Adhikari and Agrawal, 2016): *ADH_INCR*, which equals one if the HQ change comes along with a change in religiosity that lies within the top quintile, zero otherwise; and *ADH_DECR*, which equals one if the HQ

⁵⁰ This choice is motivated as follows: First, a one-year threshold inherits the risk of considering companies that directly went inactive (e.g., due to mergers and acquisitions) in the year after the relocation occurred. Thus, the relocation would not represent a “true” picture of changes in social norms. Second, requiring more than two years before and after the relocation comes heavily at the cost of observations. Thus, two years seems appropriate in our setting.

change is accompanied by a change in religiosity that is in the bottom quintile, zero otherwise. We choose the top/bottom quintiles for two reasons. First, we do not expect that a change in *REL* that is close to zero has a “visible” effect on liquidity. Second, we do not choose more extreme percentiles to ensure a sufficient variation in our indicator variables. Finally, following Cai et al. (2019), we regress the changes of *BAS* on the indicator variables *ADH_INCR* and/or *ADH_DECR*, including the changes⁵¹ and the levels of all our control variables (see Sections 3.2 and 3.4) in the regression models. Since a firm appears only once in our sample, we replace within firm clustering by standard errors clustered at the industry level.⁵² Table 5 (Models 5a-5c) presents the results from the multivariate regressions based on HQ changes.

In Model 5a, we investigate HQ changes that comes along with a “measurable” increase of *REL*. Based on our hypothesis, we would expect that such HQ changes are related to a decrease in *BAS*. Indeed, we find a negative and statistically significant coefficient. In Model 5b, where we consider “measurable” decreases of *REL* only, we observe no association with *BAS*. When we consider both types together (Model 5c), we obtain similar results as with the solely considerations.

To further enhance validity, we test the sensitivity of our results regarding the classification of a “measurable” change in religiosity. For this purpose, we redefine the adherence increasing (decreasing) event, where *ADH_INCR* (*ADH_DECR*) takes the value of one if the changes in religiosity lies within the top (bottom) quartile or tercile, and zero otherwise. While the results for the quartile regressions are similar to those reported in Table 5, the results turn insignificant for the tercile regressions. The evidence therefore suggests that the trigger point of a “measurable” change in religiosity kicks in around the 75th percentile.

⁵¹ The change variables are defined in a similar way to the changes in *REL* over the same time.

⁵² Note that we observe ten singleton observations, which are dropped in the regression analysis. This results in 154 observations for our regression analyses. However, the inclusion of the singleton observations does not alter our conclusions. Likewise, results are robust if we cluster standard errors by year, or by industry and year, or when we exclude the vector of demographic variables (as done in Cai et al., 2019).

Taken together, changes in religiosity caused by headquarter relocations could contribute to changes in confidence in information flows of a firm, and consequently, in stock liquidity. Therefore, this analysis serves as another important component to further support our initial finding.

5 Information environment and religiosity

According to our second hypothesis, the importance of religiosity as an informal mechanism of trust in corporate behavior and information flows is probably more evident for stock liquidity, when firms operate in poor information environments, and liquidity providers have limited information about the firm. Therefore, we start with four proxies for firm level information asymmetry, and create two subsamples having high/low information problems (e.g., Callen and Fang, 2015; Jiang et al., 2018). Using separate regression models for each group is supported by previous research (e.g., Holgersson et al., 2014) and is in line with previous studies focusing on a similar purpose (e.g., Callen and Fang, 2015; Jiang et al., 2018; Pham, 2020; Bernile et al., 2015; Baik et al., 2010; Amin et al., 2021; among many others).

Our first proxy for information asymmetry is analyst coverage (e.g., Duarte et al., 2008; Jiang et al., 2018). Previous studies document that analysts play an effective role of external monitoring, thus reducing potential information problems because information is more widely distributed (e.g., Duarte et al., 2008; Bradley et al., 2021). A high information asymmetry firm is classified as one with no analyst coverage. The second proxy relates to the visibility of a firm. S&P 500 companies are more likely to be visible to market participants and media, thus having lower information asymmetry (e.g., Jiang et al., 2018). Based on this notion, non-S&P500 firms face higher information asymmetry. A closely related measure for visibility refers to company size. Literature suggests that bigger companies reveal improved corporate visibility (e.g., Dang

et al., 2018; Jiang et al., 2018). A firm is categorized as having high information asymmetry if its total assets are below the sample median by year (Jiang et al., 2018). The next proxy is based on firm location (e.g., Kedia and Zhou, 2011; El Ghouli et al., 2013). Loughran and Schultz (2005) find evidence that firms located in urban areas have a larger investor base since they are local stocks for many people. This, in turn, reduces information asymmetry. Therefore, if a firm is located more than 100 miles away from the nearest city center of the six financial centers (Boston, Chicago, Los Angeles, New York, Philadelphia, and San Francisco), we categorize this “far away” firm as one facing high information asymmetry (e.g., Loughran and Schultz, 2005; El Ghouli et al., 2013). Table 6 reports the results.

[Insert Table 6 about here]

Panel A of Table 6 shows that for each subsample of firms within the high information asymmetry cluster (Models (1), (3), (5), and (7), respectively), the coefficient on *REL* is always negative and statistically significant at the 1% level. Turning our focus to the low information problems firms (Models (2), (4), (6), and (8), respectively), coefficients are insignificant, except for the analyst coverage sample. In all models, the magnitude of the coefficients for each pair of subsamples is statistically significant different for the high information asymmetry group compared to the low information asymmetry group (see row “Differences in coefficients”), at least at the 10% level. The difference is especially pronounced for the analyst and S&P500 subsamples.⁵³

Taken together, these findings support our conjecture that the trust-based aspect of religiosity is particularly relevant when firms operate in settings, in which relatively less information about

⁵³ The differences between small and big companies gets even more pronounced when we shift our threshold from the median to terciles or quintiles.

the firm is otherwise available. In such circumstances, religiosity probably acts as an informal mechanism that reduces the relevance of information asymmetry, thus increasing stock liquidity.

6 Other aspects of liquidity and the trading environment of a firm

After finding empirical evidence for a potential causal link between *REL* and the *BAS*, that is more evident for firms operating in a poor information environment, we further apply three more measures as dependent variables to capture the multiple facets of liquidity and trading-specific information: the volatility of daily *BAS* (*VOLA_BAS*), the square root version of Amihud illiquidity measure (*AMIHUD_SQ*; Amihud, 2002; Gopalan et al., 2012) as a measure of price impact (e.g., Goyenko et al., 2009; Edmans et al., 2013), as well as the probability of information-based trading (*PIN*, see Easley et al., 1998; Easley et al., 2002; Brown and Hillegeist, 2007).⁵⁴ The latter two are used to test our third hypothesis. Detailed descriptions of the variable definitions are provided in Table A1. We regress *VOLA_BAS*, *AMIHUD_SQ*, and *PIN*, respectively, on *REL* using the same control variables as in our baseline model (see equation 1). Table 7 reports the results.

[Insert Table 7 about here]

Consistent with hypothesis 3, the estimates on our additional metrics are negative and statistically significant, at least at the 5% level. These results further indicate that firms located

⁵⁴ Since details on the estimation procedure of *PIN* are quite complex, and for brevity of this paper, the interested reader is referred to the original study by Easley et al. (1998, 2002), and Brown and Hillegeist (2007). We thank Steven Brown for sharing the estimates on the extended PINs. We retrieve these data from <https://terpconnect.umd.edu/~stephenb/>.

in more religious areas exhibit smaller price impacts of trades (see Model (2)).⁵⁵ This finding should be also reflected by smaller information-based trading (e.g., Chung et al., 2010). In line with this notion, our estimate on *PIN* is also negative and statistically significant at the 1% level (see Model (3)). From our perspective, *PIN* captures the probability of trading against a superiorly informed trader. This confirms the conjecture that religiosity impacts the trading and information environment of a firm positively in terms of reliance and trustworthiness (e.g., McGuire et al., 2012; Callen and Fang, 2015).⁵⁶ Consequently, since local religiosity probably fosters more credible and symmetric information flows, liquidity providers face less risk of trading against better (illegal) informed investors.⁵⁷ Taken together, companies headquartered in more religious areas exhibit smaller price impacts and lower probability of information-based trading.⁵⁸

7 Local religiosity and firm value

We document that local religious norms influence stock liquidity, particularly documented by a reduced bid-ask spread. In this context, extant research show that higher stock liquidity positively affects firm value (e.g., Fang et al., 2009).⁵⁹ This raises the natural question whether

⁵⁵ This effect is likely to be driven, at least partially, through the channel of lower bid-ask spreads and higher trading volume.

⁵⁶ Chung and Li (2003) show that the adverse selection component of the bid-ask spread is positive and significantly related to *PIN*.

⁵⁷ Unlike our view, Collin-Dufresne and Fos (2015) find that neither high-frequency nor low-frequency measures of stock liquidity increases when insiders trade. However, using data on illegal trading activity, Akey et al. (2022) provide further evidence on this issue. They show that liquidity providers respond to increase in informed order flow by posting higher spreads. Therefore, we argue that religiosity as a commitment device at least improves the trading environment in terms of illegal insider trading, when informed traders face criminal charges (Ahern, 2020).

⁵⁸ For *AMIHUD_SQ* as well as for *PIN*, we also test the channel of information asymmetry (see Section 5). Results are reported in the Internet Appendix Section IA5 (Table IA5). Overall, the results are qualitatively and mostly quantitatively similar.

⁵⁹ Prior research document that the risk-reducing effect of religiosity translates into lower cost of capital (e.g., El Ghouli et al., 2012; Jiang et al., 2018), suggesting an impact on religiosity on firm value.

the relation between local religiosity and stock liquidity has an implication for firm value (e.g., Amin et al., 2021).

From a theoretical perspective, religiosity is likely to enhance firm value through its risk reducing effect and the avoidance of corporate misconduct (e.g., Hilary and Hui, 2009; El Ghouli et al., 2012; McGuire et al., 2012; Jiang et al., 2018). However, religiosity induced risk aversion may also lead to refraining from risky investments with positive net present value, suggesting a negative effect on firm value (e.g., Hilary and Hui, 2009; Zolotoy et al., 2019). In addition, extant research failed to find a direct relationship between religiosity and firm value (e.g., Amin et al., 2021; Zolotoy et al., 2019). We briefly contribute to this ongoing debate.

In line with the literature, we do not find robust evidence that religiosity directly improves firm value.⁶⁰ However, consistent with Fang et al. (2009), our untabulated results show that firms tend to exhibit higher firm value when they reveal higher stock liquidity, which is, in turn, significantly influenced by religiosity. Like Amin et al. (2021), we therefore cautiously conjecture that there may be an indirect value-enhancing effect of religiosity via the channel of increased stock liquidity. Further research is thereby needed to disentangle value implications, if existing, caused by religiosity.

8 Conclusion

This study investigates the relation between religiosity and stock liquidity for a broad sample of U.S. listed firms. We find strong support that firms located in more religious areas tend to have lower bid-ask spreads. This negative relation remains statistically significant, even if we consider additional control variables, such as governance metrics, different model

⁶⁰ Despite the lack of significance in most circumstances, we throughout find a positive coefficient on religiosity when we regress firm value (Q) on REL and all baseline controls. Interestingly, when we add BAS , $AMIHUD_SQ$, or PIN , respectively, the coefficients on REL decrease, which could indicate an indirect firm value implication through the channel of stock liquidity.

specifications, or when we conduct endogeneity analyses. Further, our results support the notion that the impact of religiosity is especially emphasized when little information about the firm is conveyed. Also, firms headquartered in U.S. counties with a high level of religiosity reveal lower price impact and probability of information-based trading. Finally, we posit an indirect value-enhancing effect of religiosity through the channel of stock liquidity.

The present study contributes to and elaborates on existing literature on the effects of religiosity on the credibility of corporate behavior and information flows. Consistent with previous literature (e.g., Hilary and Hui, 2009; Jiang et al., 2018), religiosity as an important ethical factor not only matters for inside corporate behavior and culture, but particularly plays a crucial role of how a company is viewed by market participants. Admittedly, religiosity is a soft factor that fully deploy its impact when trust in corporate actions and information flows is of particular relevance, especially when little is known about the firm. We acknowledge that there may be further soft factors besides religiosity that affect corporations, which are worth recognizing and tackling in further studies (Callen and Fang, 2015).

Potential implications of our results arise for regulators and investors. Since religiosity represent an informal mechanism for reducing agency costs (e.g., McGuire et al., 2012; Leventis et al., 2018) and the relevance of information asymmetry, it possibly offers a new perspective to investment, monitoring and information searching strategies, especially for companies with a weak information environment. Taken together, our findings highlight the importance of local religious norms in enhancing financial market quality.

Appendix

Table A1 : Variable definition

Our download procedure in Datastream is summarized as follows. First, we choose as market “United States” and set the currency to “United States Dollars”. We then select all “Active” and “Dead” firms from all industries of type “Equity”, and consider “Major” shares with “Primary” codes only. Our base date is 2020, which means that all years equal 2020 and before are considered.

Variable	Definition
<i>Panel A: Main dependent variable</i>	
<i>BAS</i>	Adjusted ask price (Refinitiv Eikon item TR.ASKPRICE) minus adjusted bid price (Refinitiv Eikon item TR.BIDPRICE) divided by the spread midpoint, which is the sum of the adjusted ask price (Refinitiv Eikon item TR.ASKPRICE) and the adjusted bid price (Refinitiv Eikon item TR.BIDPRICE) divided by two ($(Ask - Bid)/((Ask + Bid)/2)$). We then take the average of daily figures to receive our yearly variable.
<i>Panel B: Variable of interest</i>	
<i>REL</i>	The number of religious adherents in a county divided by the county population in a year (ARDA). Religiosity is linearly interpolated between the survey years (https://www.thearda.com/Archive/Files/Descriptions/RCMSMGCY.asp).
<i>Panel C: Firm controls</i>	
$\ln(1+ANALYST)$	Number of analysts forecasting earnings per share for the following year (Datastream EPS1NE). We define the variable as the arithmetic mean number of monthly earnings forecasts during each calendar year. Firm-year in which a firm is not covered by any analysts, we set these values to zero (e.g., Chang et al., 2006; Bradley et al., 2021). Finally, we take the natural logarithm of one plus the annual value (e.g., similarly done in Chan and Hameed, 2006).

<i>CAPEX</i>	Capital Expenditures (Worldscope item 04601) divided by total assets (Worldscope item 02999).
<i>CUMRET</i>	Calculated as the cumulative firm-specific daily returns (calculated from Datastream item RI) during a calendar year.
<i>NASD</i>	Indicator variable that is one, if the firm is listed on NASDAQ, zero otherwise.
<i>OWN_INSIDER</i>	Number of shares held by insiders as a proportion of the number of shares outstanding (Worldscope item 08021). We set closely held shares to zero, if we observe a missing firm-year in our dataset. This variable has been used to proxy for insider holdings by researchers such as Ferreira and Matos (2008), and Ferreira et al. (2010).
<i>OWN_INST</i>	Institutional Ownership by all institutions (13(f) filings) (Refinitiv Ownership Profile item TR.FilingType) as a percentage of shares outstanding (Refinitiv Ownership Profile item TR.PctOfSharesOutHeld). All values above 100% are set to 100% (e.g., Lewellen, 2011). Also, we set institutional ownership to zero if a stock is not held by any institution (see Gompers and Metrick, 2001).
$\ln(\text{PRICE})$	The natural logarithm of the mean daily stock price (Datastream item P) during a calendar year (e.g., McInish and Van Ness, 2002).
<i>Q</i>	Total assets (Worldscope item 02999) plus market value of equity (Worldscope item 08001) minus book value of equity (Worldscope item 03995) divided by total assets (Worldscope item 02999).
<i>RISK</i>	The standard deviation of daily stock returns (calculated based on Datastream item RI) during a calendar year.
<i>RnD</i>	Research and development expenditures (Worldscope item 01201) divided by total assets (Worldscope item 02999). Following prior studies, we set missing values to zero (e.g. Chung et al., 2010).

<i>ROA</i>	Ratio of operating income (Worldscope item WC01551) to total assets (Worldscope item 02999).
<i>SP500</i>	Indicator variable, which equals one if a firm is a member of the S&P 500, zero otherwise. This dummy is generated from yearly constituents lists of the S&P 500 (e.g., Datastream item LS&PCOMP1220 for the constituents list as of December 2020). Empirically, each end of the year, we check the constituents list of S&P 500 companies. We then get a time series of yearly constituents, which we merge to our main dataset.
$\ln(\text{SIZE})$	The natural logarithm of annual total assets in thousands of dollars (Worldscope item 02999).
$\ln(\text{TURNOVER})$	Monthly share volume (Datastream item VO) divided by adjusted shares outstanding (Datastream items NOSH/AF; Ferreira and Matos, 2008). The shares outstanding are determined at the beginning of each year and kept constant for each day of the year (similarly done in Lesmond, 2005). We then take the mean from monthly turnover during a calendar year. At the end, we take the natural logarithm of the average turnover.
<i>Panel D: Demographic factors</i>	
$\ln(\text{AGE})$	The natural logarithm of the median age of the population in a county (U.S. Census Bureau). Age is determined by linear interpolation between the survey years.
<i>EDUCATION</i>	Education is defined as the fraction of county's population that is 25 years or older and hold a bachelor's degree or higher (U.S. Census Bureau). Education is determined by linear interpolation between the survey years.
$\ln(\text{DENSITY})$	The natural logarithm of the total population to the land area in the respective county. Density is determined by linear interpolation between the survey years (U.S. Census Bureau).
<i>MF_RATIO</i>	The ratio of male population to female population in a county (U.S. Census Bureau). Missing values

between survey years are obtained by linear interpolation.

MARRIAGE

The percentage of married households in a county (U.S. Census Bureau). Marriage is determined by linear interpolation between the survey years.

MINORITY

The percentage of non-white population in a county (U.S. Census Bureau). Missing values between survey years are obtained by linear interpolation.

ELEC_REP

The proportion of votes received by the republican candidate (<https://electionlab.mit.edu/data>).

$\ln(TOTPOP)$

The natural logarithm of the total population in a county (U.S. Census Bureau). Total population is determined by linear interpolation between the survey years.

Panel E: Variables used in auxiliary analyses

ABORT

ABORT is the abortion rate for women aged 15 to 44 (<https://osf.io/u58vf/>; accessed via <https://www.guttmacher.org/public-use-datasets>).

ADV

Advertising expenses are defined as selling, general & administrative expenses (Worldscope item WC01101) divided by total assets (Worldscope item WC02999); Hawn and Ioannou (2016).

ALC

ALC is defined as the per capita alcohol consumption rate (<https://pubs.niaaa.nih.gov>).

AMIHUDD_SQ

The square root of the absolute daily return calculated from total return indices (Datastream item RI) scaled by the adjusted daily dollar volume (Datastream items $RI*VO/(NOSH/AF)$) and multiplied by 1,000,000. We take the average of our daily measure to receive our yearly variable (e.g., Amihud, 2002; Gopalan et al., 2012).

BIG4

Indicator variable that is one, if a firm is audited by one of the big four audit companies, zero otherwise (Worldscope item WC07800).

BSIZE

BSIZE is number of board members at the end of the fiscal year (Refinitiv Datastream item CGBSDP060).

<i>BINDEP</i>	Indicator variable, which equals one if the company has a policy regarding the independence of its board, zero otherwise (Refinitiv Datastream item CGBSDP0012).
<i>CATH</i>	The number of catholic adherents in a county divided by the county population in a year (ARDA). Catholic adherence is linearly interpolated between the survey years.
<i>CEO_DUAL</i>	Indicator variable, which equals one if the CEO simultaneously chair the board or the chairman has been the CEO of the company, zero otherwise (Refinitiv Datastream item CGBSO09V).
<i>COMPLEX</i>	The Herfindahl-Hirschman Index based on production segments (calculated from Worldscope items WC19501-WC19591).
<i>EVAN</i>	The number of evangelical protestants including black protestant divided by the county population in a year (ARDA). <i>EVAN</i> is linearly interpolated between the survey years.
<i>GOV</i>	Weighted average relative rating of a company based on the reported governance information and the resulting three governance category scores (Refinitiv Datastream item CGSCORE).
<i>MPRT</i>	The number of mainline protestants in a county divided by the county population in a year (ARDA). Mainline protestants is linearly interpolated between the survey years.
<i>PIN</i>	Robust version of EKO PIN (Stephen Brown's website https://terpconnect.umd.edu/~stephenb/).
<i>PROT</i>	The number of protestant adherents in a county divided by the county population in a year (ARDA). Protestant adherence is linearly interpolated between the survey years.
<i>SOCIAL</i>	The first principal component from principal component analysis based on PVOTE, RESPN, NCCS, and ASSN (NRCRD at Pennsylvania State University; available at

<https://aese.psu.edu/nercrd/community/social-capital-resources>). Hasan et al. (2017a), among others, provide a comprehensive description of constructing the *SOCIAL* variable.

PVOTE = Percentage of voters who voted in presidential elections

RESPN = Response rate to the Census Bureau's decennial census

NCCS = Sum of tax-exempt nonprofit organizations divided by populations per 10,000

ASSN = Sum of social organizations divided by populations per 100,000

STATE_GDP

STATE_GDP represents the natural logarithm of the gross domestic product by state (Bureau of Economic Analysis).

TANG

Asset Tangibility is defined as property, plant, and equipment (Worldscope item WC02501) divided by total assets (Worldscope item WC02999).

$\ln(TVOL)$

The natural logarithm of daily price times (Datastream item P) times adjusted daily trading volume (Datastream items VO/(NOSH/AF)).

VOLA_BAS

The standard deviation of daily *BAS* during a calendar year (Refinitiv Eikon item).

Panel F: Supplemental data

COMPANY_IND

COMPANY_IND is based on two-digit standard industry classification (SIC) codes (Worldscope item 07021). We use the first SIC code, which is assigned to a company, i.e., this represent the business segment which provided most revenue.

Exchange Listing

Current exchange, on which the company is listed. Note that Datastream reports the old exchange name ("NYSE MKT") for AMEX firms. Therefore, we replace the exchange listing manually from "NYSE MKT" to "AMEX" for the respective companies.

fips

FIPS stands for "federal information processing standard". The 5-digit fips code is used by ARDA and

US Census Bureau to determine the location of a county, while the first two digits represent the state, where the county is located. ZIP and FIPS are used to merge company data with data on religiosity and demography.

ISIN

ISIN stands for “international securities identification number”. This is the main identifier of the companies in our sample.

ZIP

ZIP stands for “zone improvement plan”. It is used to determine the location of a company’s headquarter (Worldscope item 06025).

References

- Adhikari, B. K., Agrawal, A. (2016). Does religiosity matter for bank risk taking? *Journal of Corporate Finance* 38, 272-293. <http://dx.doi.org/10.1016/j.jcorpfin.2016.01.009>
- Agarwal, P. (2007). Institutional Ownership and Stock Liquidity. Working Paper, Johnson Graduate School of Management, Cornell University.
- Ahern, K. R. (2020). Do Proxies for Informed Trading Measure Informed Trading? Evidence from Illegal Insider Trades. *Review of Asset Pricing Studies* 10(3), 397-440. <https://doi.org/10.1093/rapstu/raaa004>
- Akey, P., Grégoire, V., Martineau, C. (2022). Price revelation from insider trading: Evidence from hacked earnings news. *Journal of Financial Economics* 143(3), 1162-1184. <https://doi.org/10.1016/j.jfineco.2021.12.006>
- Alam, Z. S., Chen, M. A., Ciccotello, C. S., Ryan, H. E. (2014). Does the Location of Directors Matter? Information Acquisition and Board Decisions. *Journal of Financial and Quantitative Analysis* 49(1), 131-164. <https://doi.org/10.1017/S002210901400012X>
- Albuquerque, R., Koskinen, Y., Zhang, C. (2019). Corporate Social Responsibility and Firm Risk: Theory and Empirical Evidence. *Management Science* 65(10), 4451-4469. <https://doi.org/10.1287/mnsc.2018.3043>
- Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time series effects. *Journal of Financial Markets* 5(1), 31-56. [https://doi.org/10.1016/S1386-4181\(01\)00024-6](https://doi.org/10.1016/S1386-4181(01)00024-6)
- Amin, M. R., Kim, I., Lee, S. (2021). Local religiosity, workplace safety, and firm value. *Journal of Corporate Finance* 70, 102093. <https://doi.org/10.1016/j.jcorpfin.2021.102093>
- Attig, N., Fong, W.-M., Gadhoun, Y., Lang, L. H. P. (2006). Effects of large shareholding on information asymmetry and liquidity. *Journal of Banking & Finance* 30, 2875-2892. <https://doi.org/10.1016/j.jbankfin.2005.12.002>
- Bagehot, W. (1971). The Only Game in Town. *Financial Analysts Journal* 27(2), 12-14. <https://doi.org/10.2469/faj.v27.n2.12>
- Baghdadi, G. A., Bhatti, I. M., Nguyen, L. H. G., Podolski, E. J. (2018). Skill or effort? Institutional ownership and managerial efficiency. *Journal of Banking & Finance* 91, 19-33. <https://doi.org/10.1016/j.jbankfin.2018.04.002>

- Baik, B., Kang, J.-K., Kim, J.-M. (2010). Skill or effort? Local institutional investors, information asymmetries, and equity returns. *Journal of Financial Economics* 97(1), 81-106. <https://doi.org/10.1016/j.jfineco.2010.03.006>
- Barro, R. J., McCleary, R. (2003). Religion and Economic Growth. NBER Working Paper Series 9682. <https://doi.org/10.3386/w9682>
- Benz, L., Paulus, S., Rohleder, M., Wilkens, M. (2020). Ownership comes with responsibility – The impact of ownership characteristics on CSR. Working Paper, University of Augsburg.
- Bernile, G., Korniotis, G., Kumar, A., Wang, Q. (2015). Local Business Cycles and Local Liquidity. *Journal of Financial and Quantitative Analysis* 50(5), 987-1010. <https://doi.org/10.1017/S0022109015000447>
- Bhattacharya, N., Desai, H., Venkataraman, K. (2013). Does Earnings Quality Affect Information Asymmetry? Evidence from Trading Costs. *Contemporary Accounting Research* 30(2), 482-516. <https://doi.org/10.1111/j.1911-3846.2012.01161.x>
- Blau, B. (2018). Does religiosity affect liquidity in financial markets. *Journal Behavioral and Experimental Finance* 19, 72-83. <https://doi.org/10.1016/j.jbef.2018.05.002>
- Bradley, D., Mao, C. X., Zhang, C. (2021). Does Analyst Coverage Affect Workplace Safety? *Management Science*, forthcoming. <https://doi.org/10.1287/mnsc.2021.4093>
- Brennan, M. J., Subrahmanyam, A. (1995). Investment analysis and price formation in securities markets. *Journal of Financial Economics* 38(3), 361-381. [https://doi.org/10.1016/0304-405X\(94\)00811-E](https://doi.org/10.1016/0304-405X(94)00811-E)
- Brennan, M. J., Huh, S.-W., Subrahmanyam, A. (2013). An Analysis of the Amihud Illiquidity Premium. *Review of Asset Pricing Studies* 3(1), 133-176. <https://doi.org/10.1093/rapstu/ras017>
- Brown, S., Hillegeist, S. A. (2007). How disclosure quality affects the level of information asymmetry. *Review of Accounting Studies* 12, 443-477. <https://doi.org/10.1007/s11142-007-9032-5>
- Buffett, W. (2011). Warren Buffet's Letter to Berkshire Shareholders. 26 February.
- Cantrell, B. W., Yust, C. G. (2018). The relation between religiosity and private bank outcomes. *Journal of Banking & Finance* 91, 86-105. <https://doi.org/10.1016/j.jbankfin.2018.04.009>
- Cai, J., Shi, G. (2019). Do Religious Norms Influence Corporate Debt Financing? *Journal of Business Ethics* 157, 159-182. <https://doi.org/10.1007/s10551-017-3701-5>

- Cai, J., Kim, Y., Li, S., Pan, C. (2019). Tone at the top: CEOs' religious beliefs and earnings management. *Journal of Banking & Finance* 106, 195-213. <https://doi.org/10.1016/j.jbankfin.2019.06.002>
- Caliendo, M., Kopeinig, S. (2005). Some Practical Guidance for the Implementation of Propensity Score Matching. Discussion Paper Series IZA DP No. 1588, Bonn.
- Callen, J. L., Fang, X. (2015). Religion and Stock Price Crash Risk. *Journal of Financial and Quantitative Analysis* 50(1/2), 169-195. <https://doi.org/10.1017/S0022109015000046>
- Callen, J. L., Fang, X. (2020). Local Gambling Norms and Audit Pricing. *Journal of Business Ethics* 164, 151-173. <https://doi.org/10.1007/s10551-018-4079-8>
- Chan, K., Hameed, A. (2006). Stock price synchronicity and analyst coverage in emerging markets. *Journal of Financial Economics* 80(1), 115-147. <https://doi.org/10.1016/j.jfineco.2005.03.010>
- Chang, X., Dasgupta, S., Hilary, G. (2006). Analyst Coverage and Financing Decisions. *Journal of Finance* 61(6), 3009-3048. <https://doi.org/10.1111/j.1540-6261.2006.01010.x>
- Chhaocchharia, V., Kumar, A., Niessen-Ruenzi, A. (2012). Local investors and corporate governance. *Journal of Accounting and Economics* 54, 42-67. <http://dx.doi.org/10.1016/j.jacceco.2012.03.002>
- Chircop, J., Johan, S., Tarsalewska, M. (2020). Does religiosity influence venture capital investment decisions? *Journal of Corporate Finance* 62, forthcoming. <https://doi.org/10.1016/j.jcorpfin.2020.101589>
- Chiswick, B. R. (1983). The Earnings and Human Capital of American Jews. *Journal of Human Resources* 18(3), 313-336. <https://doi.org/10.2307/145204>
- Chiswick, B. R. (1985). The Labour Market Status of American Jews: Patterns and Determinants. *The American Jewish Yearbook* 85, 131-153.
- Chordia, T., Subrahmanyam, A., Anshuman, V. R. (2001). Trading activity and expected stock returns. *Journal of Financial Economics* 59(1), 3-32. [https://doi.org/10.1016/S0304-405X\(00\)00080-5](https://doi.org/10.1016/S0304-405X(00)00080-5)
- Chung, K. H., Charoenwong, C. (1998). Insider Trading and the Bid-Ask Spread. *The Financial Review* 33(3), 1-20. <https://doi.org/10.1111/j.1540-6288.1998.tb01379.x>

- Chung, K. H., Elder, J., Kim, J.-C. (2010). Corporate Governance and Liquidity. *Journal of Financial and Quantitative Analysis* 45(2), 265-291. <https://doi.org/10.1017/S0022109010000104>
- Chung, K. H., Li, M. (2003). Adverse-Selection Costs and the Probability of Information-Based Trading. *Financial Review* 38, 257-272. <https://doi.org/10.1111/1540-6288.00045>
- Chung, K. H., Zhang, H. (2014). A simple approximation of intraday spreads using daily data. *Journal of Financial Markets* 17, 94-120. <https://dx.doi.org/10.1016/j.finmar.2013.02.004>
- Cialdini, R. B., Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In Gilbert, D. T., Fiske, S. T., Lindzey, G. (Eds.). *The handbook of social psychology*, 151-192. Oxford University Press Inc, New York
- Collin-Dufresne, P., Fos, V. (2015). Do Prices Reveal the Presence of Informed Trading? *Journal of Finance* 70(4), 1555-1582. <https://doi.org/10.1111/jofi.12260>
- Dang, V. A., Michayluk, D., Pham, T. P. (2018). The curious case of changes in trading dynamics: When firms switch from NYSE to NASDAQ. *Journal of Financial Markets* 41, 17-35. <https://doi.org/10.1016/j.finmar.2018.07.001>
- Diamond, D., W., Verrecchia, R. E. (1991). Disclosure, Liquidity, and the Cost of Capital. *Journal of Finance* 46(4), 1325-1359. <https://doi.org/10.1111/j.1540-6261.1991.tb04620.x>
- Duarte, J., Han, X., Harford, J., Young, L. (2008). Information asymmetry, information dissemination and the effect of regulation FD on the cost of capital. *Journal of Financial Economics* 87(1), 24-44. <https://doi.org/10.1016/j.jfineco.2006.12.005>
- Dyreng, S. D., Mayew, W. J., Williams, C. D. (2012). Religious Social Norms and Corporate Financial Reporting. *Journal of Business Finance & Accounting* 39(7/8), 845-875. <https://doi.org/10.1111/j.1468-5957.2012.02295.x>
- Easley, D., O'Hara, M. (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics* 19(1), 69-90. [https://doi.org/10.1016/0304-405X\(87\)90029-8](https://doi.org/10.1016/0304-405X(87)90029-8)
- Easley, D., O'Hara, M., Paperman, J. (1998). Financial analysts and information-based trade. *Journal of Financial Markets* 1(2), 175-201. [https://doi.org/10.1016/S1386-4181\(98\)00002-0](https://doi.org/10.1016/S1386-4181(98)00002-0)
- Easley, D., Hvidkjaer, S., O'Hara, M. (2002). Is Information Risk and Determinant of Asset Returns? *Journal of Finance* 62(5), 2185-2221. <https://doi.org/10.1111/1540-6261.00493>

- Edmans, A., Fang, V. W., Zur, E. (2013). The Effect of Liquidity on Governance. *Review of Financial Studies* 1(26), 1443-1482. <https://doi.org/10.1093/rfs/hht012>
- Eleswarapu, V. R., Venkataraman, K. (2006). The Impact of Legal and Political Institutions on Equity Trading Costs: A Cross-Country Analysis. *Review of Financial Studies* 19(3), 1081-1111. <https://doi.org/10.1093/rfs/hhj026>
- El Ghouli, S., Guedhami, O., Ni, Y., Pittman, J., Saadi, S. (2012). Does Religion Matter to Equity Pricing? *Journal of Business Ethics* 111, 491-518. <https://doi.org/10.1007/s10551-012-1213-x>
- El Ghouli, S., Guedhami, O., Ni, Y., Pittman, J., Saadi, S. (2013). Does Information Asymmetry Matter to Equity Pricing? Evidence from Firms' Geographic Location. *Contemporary Accounting Research* 30(1), 140-181. <https://doi.org/10.1111/j.1911-3846.2011.01147.x>
- Fang, V. W., Noe, T. H., Tice, S. (2009). Stock market liquidity and firm value. *Journal of Financial Economics* 94(1), 150-169. <https://doi.org/10.1016/j.jfineco.2008.08.007>
- Ferreira, M. A., Matos, P. (2008). The colors of investors' money: The role of institutional investors around the world. *Journal of Financial Economics* 88, 499-533. <https://doi.org/10.1016/j.jfineco.2007.07.003>
- Ferreira, M. A., Massa, M.; Matos, P. (2010). Shareholders at the Gate? Institutional Investors and Cross-Border Mergers and Acquisitions. *Review of Financial Studies* 23(2), 601-644. <https://doi.org/10.1093/rfs/hhp070>
- Fong, K. Y. L., Holden, C. W., Trzcinka, C. A. (2017). What Are the Best Liquidity Proxies for Global Research? *Review of Finance* 21(4), 1335-1401. <https://doi.org/10.1093/rof/rfx003>
- Francis, B., Hasan, I., Liu, L., Wu, Q., Zhao, Y. (2021). Financial analysts' career concerns and the cost of private debt. *Journal of Corporate Finance* 67, 101868. <https://doi.org/10.1016/j.jcorpfin.2020.101868>
- Frijns, B., Dodd, O., Cimerova, H. (2016). The impact of cultural diversity in corporate boards on firm performance. *Journal of Corporate Finance* 41, 521-541. <http://dx.doi.org/10.1016/j.jcorpfin.2016.07.014>
- Gao, X., Ritter, J. R. (2010). The marketing of seasoned equity offerings. *Journal of Financial Economics* 97(1), 33-52. <https://doi.org/10.1016/j.jfineco.2010.03.007>

- Gao, L., Wang, Y., Zhao, J. (2017). Does local religiosity affect organizational risk-taking? Evidence from the hedge fund industry. *Journal of Corporate Finance* 47, 1-22. <https://doi.org/10.1016/j.jcorpfin.2017.08.006>
- Gaspar, J.-M., Massa, M. (2007). Local ownership as private information: Evidence on the monitoring-liquidity trade-off. *Journal of Financial Economics* 83(3), 751-792. <https://doi.org/10.1016/j.jfineco.2005.10.010>
- Gibson, S., Singh, R., Yerramilli, V. (2003). The effect of decimalization on the components of the bid-ask spread. *Journal of Financial Intermediation* 12(2), 121-148. [https://doi.org/10.1016/S1042-9573\(03\)00017-2](https://doi.org/10.1016/S1042-9573(03)00017-2)
- Glosten, L. R., Harris, L. E. (1988). Estimating the components of the bid/ask spread. *Journal of Financial Economics* 21(1), 123-142. [https://doi.org/10.1016/0304-405X\(88\)90034-7](https://doi.org/10.1016/0304-405X(88)90034-7)
- Glosten, L. R., Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14(1), 71-100. [https://doi.org/10.1016/0304-405X\(85\)90044-3](https://doi.org/10.1016/0304-405X(85)90044-3)
- Gompers, P. A., Metrick, A. (2001). Institutional Investors and Equity Prices. *Quarterly Journal of Economics* 116(1), 229-259. <https://doi.org/10.1162/003355301556392>
- Gopalan, R., Kadan, O., Pevzner, M. (2012). Asset Liquidity and Stock Liquidity. *Journal of Financial and Quantitative Analysis* 47(2), 333-364. <https://doi.org/10.1017/S0022109012000130>
- Goyenko, R. Y., Holden, C. W., Trzcinka, C. A. (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics* 92, 153-181. <https://doi.org/10.1016/j.jfineco.2008.06.002>
- Goldstein, I., Spatt, C. S., Ye, M. (2021). Big Data in finance. *Review of Financial Studies* 34(7), 3213-3225. <https://doi.org/10.1093/rfs/hhab038>
- Graham, J. R., Harvey, C. R., Popadak, J., Rajgopal, S. (2017). Corporate Culture: Evidence from the Field. NBER Working Paper Series, No. 23255.
- Grullon, G., Kanatas, G., Weston, J. P. (2004). Advertising, Breadth of Ownership, and Liquidity. *Review of Financial Studies* 17(2), 439-461. <https://doi.org/10.1093/rfs/hhg039>
- Grullon, G., Kanatas, G., Weston, J. P. (2010). Religion and Corporate (Mis)Behavior. Working Paper, Rice University

- Guiso, L., Herrera, H., Morelli, M. (2016). Cultural Differences and Institutional Integration. *Journal of International Economics* 99, S97-S113. <https://doi.org/10.1016/j.jinteco.2015.11.005>
- Guo, B., Pérez-Castrillo, D., Toldrà-Simats, A. (2019). Firm's innovation strategy under the shadow of analyst coverage. *Journal of Financial Economics* 131(2), 456-483. <https://doi.org/10.1016/j.jfineco.2018.08.005>
- Hainmueller, J. (2012). Entropy Balancing for Causal Effects. A Multivariate Reweighting Method to Produce Balanced Samples in Observation Studies. *Political Analysis* 20, 25-46. <https://doi.org/10.1093/pan/mpr025>
- Hainmueller, J., Xu, Y. (2013). ebalance: A Stata Package for Entropy Balancing. *Journal of Statistical Software* 54(7), 1-18. <https://doi.org/10.18637/jss.v054.i07>
- Harris, L. E. (1994). Minimum Price Variations, Discrete Bid-Ask Spreads, and Quotation Size. *Review of Financial Studies* 7(1), 149-178. <https://doi.org/10.1093/rfs/7.1.149>
- Hasan, I., Hoi, C.-K., Wu, Q., Zhang, H. (2017a). Social Capital and Debt Contracting: Evidence from Bank Loans and Public Bonds. *Journal of Financial and Quantitative Analysis* 52(3), 1017-1047. <https://doi.org/10.1017/S0022109017000205>
- Hasan, I., Hoi, C.-K., Wu, Q., Zhang, H. (2017b). Does Social Capital Matter in Corporate Decisions? Evidence from Corporate Tax Avoidance. *Journal of Accounting Research* 55(3), 629-668. <https://doi.org/10.1111/1475-679X.12159>
- Hasan, I., Hoi, C.-K., Wu, Q., Zhang, H. (2020). Is social capital associated with corporate innovation? Evidence from publicly listed firms in the U.S. *Journal of Corporate Finance* 62, forthcoming. <https://doi.org/10.1016/j.jcorpfin.2020.101623>
- Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica* 46(6), 1251-1271. <https://doi.org/10.2307/1913827>
- Hawn, O., Ioannou, I. (2016). Mind the gap: the interplay between external and internal actions in the case of corporate social responsibility. *Strategic Management Journal* 37, 2569-2588. <https://doi.org/10.1002/smj.2464>
- Healy, P. M., Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics* 31(1-3), 405-440. [https://doi.org/10.1016/S0165-4101\(01\)00018-0](https://doi.org/10.1016/S0165-4101(01)00018-0)
- Hilary, G. (2006). Organized labor and information asymmetry in the financial markets. *Review of Accounting Studies* 11, 525-548. <https://doi.org/10.1007/s11142-006-9015-y>

- Hilary, G., Hui, K. W. (2009). Does religion matter in corporate decision making in America? *Journal of Financial Economics* 93, 455-473. <https://doi.org/10.1016/j.jfineco.2008.10.001>
- Hilary, G., Huang, S. (2021). Trust and Contracting: Evidence from Church Sex Scandals. *Journal of Business Ethics*, forthcoming. <https://doi.org/10.1007/s10551-021-04996-w>
- Hirshleifer, D. (2015). Behavioral Finance. *Annual Review of Financial Economics* 7, 133-159.
- Holgersson, H. E. T., Nordström, L., Öner, Ö. (2014). Dummy variables vs. category-wise models. *Journal of Applied Statistics*, 41(2), 233-241. <https://doi.org/10.1080/02664763.2013.838665>
- Huang, R. D., Stoll, H. R. (1996). Dealer versus auction markets: A paired comparison of execution costs on NASDAQ and the NYSE. *Journal of Financial Economics* 41, 313-357. [https://doi.org/10.1016/0304-405X\(95\)00867-E](https://doi.org/10.1016/0304-405X(95)00867-E)
- Huang, X., & Kang, J.-K. (2017). Geographic concentration of institutions, corporate governance, and firm value. *Journal of Corporate Finance* 47, 191-218. <https://doi.org/10.1016/j.jcorpfin.2017.09.016>
- Iannaccone, L. R. (1998). Introduction to the Economics of Religion. *Journal of Economic Literature* 36, 1465-1496. <http://www.jstor.org/stable/2564806>.
- Ince, O. S., Porter, R. B. (2006). Individual equity return data from Thomson Datastream: Handle with care! *The Journal of Financial Research* 29(4), 463-479. <https://doi.org/10.1111/j.1475-6803.2006.00189.x>
- Jayaraman, S., Milbourn, T. T. (2012). The Role of Stock Liquidity in Executive Compensation. *Accounting Review* 87(2), 537-563. <https://doi.org/10.2308/accr-10204>
- Jiang, F. W., John, K., Wei Li, C., Qian, Y. (2018). Earthly Reward to the Religious: Religiosity and the Costs of Public and Private Debt. *Journal of Financial and Quantitative Analysis* 53(5), 2131-2160. <https://doi.org/10.1017/S002210901800039X>
- John, K., Knyazeva, A., Nyazeva, D. (2011). Does geography matter? Firm location and corporate payout policy. *Journal of Financial Economics* 101, 533-551. <https://doi.org/10.1016/j.jfineco.2011.03.014>
- Kanagaretnam, K., Lim, C. Y., Lobo, G. J. (2014). Influence of National Culture on Accounting Conservatism and Risk-Taking in the Banking-Industry. *Accounting Review* 89(3), 1115-1149. <https://doi.org/10.2308/accr-50682>

- Kedia, S., Zhou, X. (2011). Local market makers, liquidity and market quality. *Journal of Financial Markets* 14(4), 540-567. <https://doi.org/10.1016/j.finmar.2011.02.002>
- Kumar, A., Page, J. K., Spalt, O. G. (2011). Religious beliefs, gambling attitudes, and financial markets outcomes. *Journal of Financial Economics* 102, 671-708. <https://doi.org/10.1016/j.jfineco.2011.07.001>
- Kurlat, P. (2018). Liquidity as Social Expertise. *Journal of Finance* 73(2), 619-656. <https://doi.org/10.1111/jofi.12606>
- Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica* 53(6), 1315-1335. <https://doi.org/10.2307/1913210>
- Landis, C., Skouras, S. (2021). Guidelines for asset pricing research using international equity data from Thomson Reuters Datastream. *Journal of Banking & Finance* 130, forthcoming. <https://doi.org/10.1016/j.jbankfin.2021.106128>
- Lang, M., Lins, K. V., Maffett, M. (2012). Transparency, Liquidity, and Valuation: International Evidence on When Transparency Matters Most. *Journal of Accounting Research* 50(3), 729-774. <https://doi.org/10.1111/j.1475-679X.2012.00442.x>
- Lee, J., Chung, K. H. (2018). Foreign Ownership and stock market liquidity. *International Review of Economics & Finance* 54, 311-325. <https://doi.org/10.1016/j.iref.2017.10.007>
- Lesmond, D. A. (2005). Liquidity of emerging markets. *Journal of Financial Economics* 77, 411-452. <https://doi.org/10.1016/j.jfineco.2004.01.005>
- Leuven, E., Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical Software Components S432001, Boston College Department of Economics, revised 01 Feb 2018.
- Leventis, S., Dedoulis, E., Abdelsalam, O. (2018). The Impact of Religiosity on Audit Pricing. *Journal of Business Ethics* 148, 53-78. <https://doi.org/10.1007/s10551-015-3001-x>
- Lewellen, J. (2011). Institutional investors and the limits of arbitrage. *Journal of Financial Economics* 102, 62-80. <https://doi.org/10.1016/j.jfineco.2011.05.012>
- Li, K., Mai, F., Shen, R., Yan, X. (2021). Measuring Corporate Culture Using Machine Learning. *Review of Financial Studies* 34(7), 3265-3315. <https://doi.org/10.1093/rfs/hhaa079>

- Loughran, T., Schultz, P. (2005). Liquidity: Urban versus rural firms. *Journal of Financial Economics* 78, 341-374. <https://doi.org/10.1016/j.jfineco.2004.10.008>
- Ma, L., Wang, X., Zhang, C. (2021). Does Religion Shape Corporate Cost Behavior? *Journal of Business Ethics* 170, 835-855. <https://doi.org/10.1007/s10551-019-04377-4>
- Manson, S., Schroeder, J., Van Riper, D., Kugler, T., Ruggles, S. (2021). IPUMS National Historical Geographic Information System: Version 16.0 [dataset]. Minneapolis, MN: IPUMS. <http://doi.org/10.18128/D050.V16.0>
- Mayberry, M. (2020). Good for managers, bad for society? Causal evidence on the association between risk-taking incentives and corporate social responsibility. *Journal of Business Finance and Accounting* 47(9-10), 1182-1214. <https://doi.org/10.1111/jbfa.12451>
- McGuire, S., T., Omer, T. C., Sharp, N. Y. (2012). The Impact of Religion on Financial Reporting Irregularities. *Accounting Review* 87(2), 645-673. <https://doi.org/10.2308/accr-10206>
- McInish, T. H., Van Ness, B. F. (2002). An Intraday Examination of the Components of the Bid-Ask Spread. *Financial Review* 37, 507-524. <https://doi.org/10.1111/1540-6288.00026>
- Narayan, P. K., Mishra, S., Narayan, S. (2015). New empirical evidence on the bid-ask spread. *Applied Economics* 47(42), 4484-4500. <https://doi.org/10.1080/00036846.2015.1031870>
- Omer, T. C., Sharp, N. Y., Wang, D. (2018). The Impact of Religion on the Going Concern Reporting Decisions of Local Audit Offices. *Journal of Business Ethics* 149, 811-831. <https://doi.org/10.1007/s10551-016-3045-6>
- Pirinsky, C., Wang, Q. (2006). Does Corporate Headquarters Location Matter for Stock Returns. *Journal of Finance* 61(4), 1991-2015. <https://doi.org/10.1111/j.1540-6261.2006.00895.x>
- Pham, M. H. (2020). In law we trust: Lawyer CEOs and stock liquidity. *Journal of Financial Markets* 50, forthcoming. <https://doi.org/10.1016/j.finmar.2020.100548>
- Refinitiv (2022). Environmental, Social and Governance Scores from Refinitiv. May 2022. https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf
- Roy, P. P., Rao, S., Zhu, M. (2022). Mandatory CSR expenditure and stock market liquidity. *Journal of Corporate Finance* 72, 102158. <https://doi.org/10.1016/j.jcorpfin.2022.102158>

- Shu, T., Sulaeman, J., Yeung, P. E. (2012). Local Religious Beliefs and Mutual Fund Risk-Taking Behaviors. *Management Science* 58(10), 1779-1796. <https://doi.org/10.1287/mnsc.1120.1525>
- Striewe, N., Rottke, N., Zietz, J. (2016). The Impact of Institutional Ownership on REIT Performance. *Journal of Real Estate Portfolio Management* 19(1), 17-30. <https://doi.org/10.1080/10835547.2013.12089939>
- Teo, M. (2009). The Geography of Hedge Funds. *Review of Financial Studies* 27(9), 3531-3561. <https://doi.org/10.1093/rfs/hhp007>
- Weston, J. P. (2000). Competition on the Nasdaq and the impact of Recent Market Reforms. *Journal of Finance* 60(6), 2565-2598. <https://doi.org/10.1111/0022-1082.00300>
- Zolotoy, L., O'Sullivan, D., Chen, Y. (2019). Local religious norms, corporate social responsibility, and firm value. *Journal of Banking & Finance* 100, 218-233. <https://doi.org/10.1016/j.jbankfin.2019.01.015>

Figures and Tables

Figure 1: Distribution of pseudo t-values of REL

This figure plots the t-values on pseudo-*REL* obtained from 500 regressions of our baseline model (see equation 1). In each regression, we replace *REL* and all demographic controls (with exception of *ELEC_REP*) by randomly assigned values of another county. The solid line represents the t-value of *REL* estimated in our baseline model (see Table 2, Column 3). The solid curve overlays the distribution represents the normal density curve.

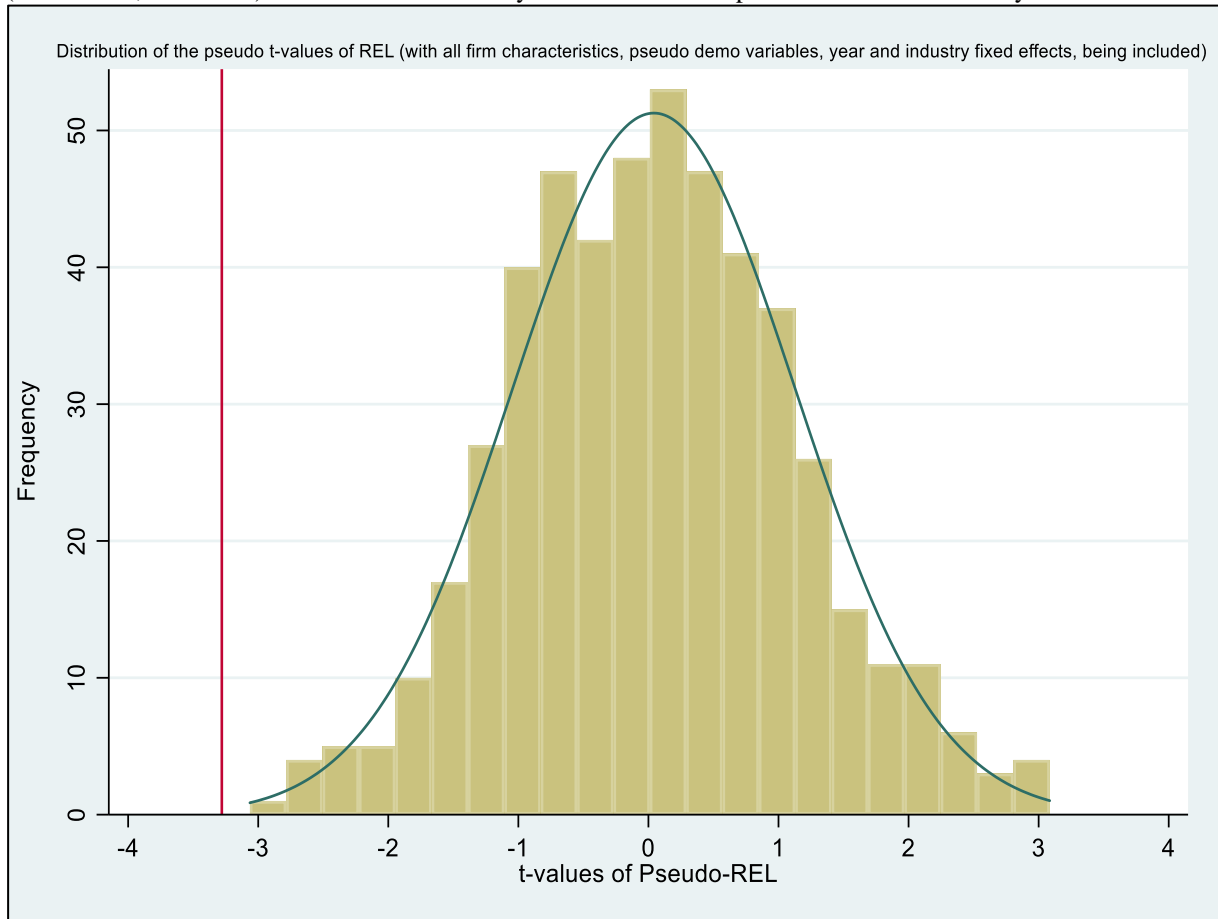


Table 1: Summary statistics

	<i>N</i>	Mean	Std.Dev.	Percentiles				
				5th. Pctl.	25th. Pctl.	Median	75th. Pctl.	95th. Pctl.
<i>Panel A: Bid-ask spread</i>								
BAS	46,201	0.0077	0.0134	0.0003	0.0009	0.0023	0.0083	0.0336
<i>Panel B: Variable of interest</i>								
REL	46,201	0.5163	0.1072	0.3464	0.4369	0.5177	0.5945	0.7014
<i>Panel C: Firm controls</i>								
ln(TURNOVER)	46,201	-2.0960	0.9780	-3.8780	-2.6335	-1.9918	-1.4590	-0.6821
ln(PRICE)	46,201	3.0901	0.9235	1.8047	2.4165	2.9959	3.6155	4.6859
NASD	46,201	0.5577	0.4967	0.0000	0.0000	1.0000	1.0000	1.0000
ln(SIZE)	46,201	6.5347	1.9032	3.5233	5.1947	6.45555	7.7877	9.8495
RISK	46,201	0.0316	0.0172	0.0132	0.0199	0.0273	0.0385	0.0638
CUMRET	46,201	0.1726	0.6739	-0.5964	-0.1894	0.0797	0.3775	1.1927
CAPEX	46,201	0.0504	0.0554	0.0042	0.0166	0.0329	0.0621	0.1628
RnD	46,201	0.0551	0.1131	0.0000	0.0000	0.0058	0.0664	0.2472
LEVERAGE	46,201	0.1974	0.1859	0.0000	0.0098	0.1692	0.3207	0.5496
Q	46,201	2.3029	1.9540	0.8834	1.2316	1.6898	2.6076	5.8232
ROA	46,201	0.0353	0.2198	-0.3460	0.0204	0.0773	0.1274	0.2318
ln(1+ANALYST)	46,201	1.7976	0.9179	0.0000	1.2040	1.8845	2.4918	3.1781
SP500	46,201	0.1607	0.3672	0.0000	0.0000	0.0000	0.0000	1.0000
OWN_INST	46,201	0.6448	0.2803	0.1057	0.4423	0.7110	0.8799	1.0000
OWN_INSIDER	46,201	0.1933	0.2084	0.0010	0.0209	0.1258	0.2959	0.6401
<i>Panel D: Demographic factors</i>								
ln(TOTPOP)	46,201	13.7593	1.0485	11.8495	13.2569	13.7602	14.3530	15.4631
ln(DENSITY)	46,201	6.3466	1.2837	4.3130	5.6488	6.4127	6.8317	8.5320
EDUCATION	46,201	0.3601	0.1025	0.2074	0.2819	0.3454	0.4400	0.5496
ln(AGE)	46,201	3.5774	0.0809	3.4530	3.5232	3.5774	3.6322	3.7040
MF_RATIO	46,201	0.9609	0.0354	0.8990	0.9375	0.9626	0.9869	1.0177
MARRIAGE	46,201	0.4144	0.0468	0.3244	0.3889	0.4213	0.4435	0.4838

Table 1 continued

MINORITY	46,201	0.3147	0.1430	0.0857	0.2000	0.3095	0.4335	0.5304
ELEC_REP	46,201	0.4423	0.0893	0.3132	0.3712	0.4436	0.4999	0.5930

This table reports descriptive statistics metrics of key variables used in our baseline analysis. The sample covers firm-year observations with nonmissing values for all variables from 1997 to 2020. All continuous variables are winsorized at the 1%/99% level. Appendix Table A1 provides a detailed description of the variables. The data are from Refinitiv Eikon, Refinitiv Datastream, ROP database, ARDA, U.S. Census Bureau, and MIT Election Lab.

Table 2: Baseline regressions

Variables	Dependent variable: BAS					
	(1) Reduced Model	(2) All Firm Controls	(3) Baseline Model	(4) Survey Years	(5) Yr 2000	(6) Yr 2010
<i>Variable of interest</i>						
REL	-0.0028*** (-2.59)	-0.0031*** (-3.32)	-0.0033*** (-3.28)	-0.0050*** (-3.23)	-0.0036* (-1.75)	-0.0054** (-2.37)
<i>Firm controls</i>						
ln(TURNOVER)	-0.0049*** (-23.18)	-0.0053*** (-22.74)	-0.0053*** (-22.73)	-0.0056*** (-12.82)	-0.0067*** (-11.46)	-0.0048*** (-7.06)
ln(PRICE)	0.0001 (0.36)	0.0006*** (4.08)	0.0006*** (4.01)	-0.0004* (-1.89)	-0.0014*** (-4.66)	0.0009*** (2.86)
NASD	0.0008** (2.47)	0.0004 (1.58)	0.0004 (1.60)	-0.0006 (-1.29)	-0.0011* (-1.81)	-0.0008 (-1.21)
ln(SIZE)	-0.0029*** (-25.67)	-0.0024*** (-18.74)	-0.0024*** (-18.69)	-0.0028*** (-13.51)	-0.0031*** (-11.94)	-0.0020*** (-5.69)
RISK		0.2289*** (18.77)	0.2289*** (18.74)	0.1921*** (7.56)	0.2503*** (7.01)	0.2498*** (3.83)
CUMRET		0.0005*** (5.61)	0.0005*** (5.57)	0.0004 (1.30)	0.0001 (0.31)	0.0002 (0.24)
CAPEX		-0.0058*** (-3.58)	-0.0057*** (-3.58)	-0.0038 (-1.13)	-0.0028 (-0.76)	-0.0093 (-1.46)
RnD		-0.0059*** (-3.99)	-0.0060*** (-4.01)	-0.0031 (-1.19)	-0.0106*** (-3.37)	-0.0044 (-1.06)
LEVERAGE		0.0060*** (10.08)	0.0060*** (10.06)	0.0096*** (7.63)	0.0102*** (7.18)	0.0051** (2.37)
Q		-0.0007*** (-13.90)	-0.0007*** (-13.91)	-0.0010*** (-11.08)	-0.0008*** (-8.11)	-0.0008*** (-3.62)
ROA		-0.0018** (-2.38)	-0.0018** (-2.40)	-0.0004 (-0.37)	0.0007 (0.61)	-0.0014 (-0.66)
ln(1+ANALYST)		-0.0008*** (-4.59)	-0.0008*** (-4.64)	-0.0016*** (-4.82)	-0.0024*** (-5.77)	-0.0008 (-1.47)
SP500		0.0053*** (17.10)	0.0052*** (16.91)	0.0069*** (12.42)	0.0067*** (10.06)	0.0055*** (6.86)
OWN_INST		-0.0050*** (-9.44)	-0.0050*** (-9.49)	-0.0056*** (-5.85)	-0.0023* (-1.87)	-0.0092*** (-6.16)
OWN_INSIDER		-0.0001 (-0.11)	-0.0000 (-0.08)	-0.0014 (-1.49)	-0.0010 (-0.92)	-0.0026 (-1.65)
<i>Demographic factors</i>						
ln(TOTPOP)			-0.0003* (-1.82)	-0.0001 (-0.27)	-0.0001 (-0.56)	0.0000 (0.04)
ln(DENSITY)			0.0003 (1.45)	0.0002 (0.63)	0.0003 (0.99)	-0.0000 (-0.02)
EDUCATION			-0.0001 (-0.07)	0.0018 (0.82)	0.0022 (0.80)	0.0034 (0.93)
ln(AGE)			-0.0026 (-1.25)	-0.0082*** (-2.58)	-0.0075** (-1.98)	-0.0028 (-0.60)
MF_RATIO			0.0033 (0.74)	-0.0087 (-1.37)	-0.0147* (-1.80)	0.0081 (0.69)
MARRIAGE			0.0019 (0.41)	0.0098 (1.47)	0.0055 (0.74)	0.0046 (0.41)
MINORITY			-0.0007 (-0.63)	-0.0006 (-0.33)	-0.0024 (-1.00)	0.0015 (0.71)
ELEC_REP			-0.0019 (-1.13)	-0.0054** (-2.09)	0.0004 (0.11)	-0.0026 (-0.71)
Constant	0.0173*** (16.82)	0.0096*** (7.08)	0.0183* (1.88)	0.0536*** (3.80)	0.0537*** (3.02)	0.0123 (0.55)
Year FE	YES	YES	YES	YES	NO	NO
Industry FE	YES	YES	YES	YES	YES	YES
Observations	46,200	46,200	46,200	3,776	1,940	1,830

Table 2 continued

R-squared	0.442	0.539	0.539	0.586	0.678	0.426
-----------	-------	-------	-------	-------	-------	-------

This table documents results of our baseline regressions. Model (1) and (2) includes controls for firm attributes. Model (1) is a reduced model, while Model (2) considers all firm controls. Model (3) is the baseline model presented in equation (1). It consists of firm characteristics, demographic controls as well as industry and year fixed effects. Finally, Model (4) reports the results for the survey years, i.e. 2000 and 2010, only, while Model (5) and Model (6) considers both survey years separately. Across all models, the dependent variable is *BAS*, which is calculated as $(Ask - Bid)/((Ask + Bid)/2)$. Appendix Table A1 defines all other variables in detail. Standard errors are adjusted for heteroscedasticity and within-firm clustering. t-statistics are reported in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 3: Additional control variables

Dimension	REL		N	R ²
	Coefficient	t-statistic		
<i>Corporate governance dimension</i>				
Model 1: Refinitiv Governance Pillar Score (<i>GOV</i>)	-0.0021**	-2.06	15,309	0.200
Model 2: CEO duality (<i>CEO_DUAL</i>)	-0.0021**	-2.12	15,289	0.191
Model 3: Board Size (<i>BSIZE</i>)	-0.0017*	-1.74	15,184	0.171
Model 4: Board Independence (<i>BINDEP</i>)	-0.0019*	-1.90	15,204	0.183
Model 5: Big4 (<i>BIG4</i>)	-0.0033***	-3.30	46,200	0.540
Model 6: First principal component of multiple governance variables	-0.0017*	-1.72	15,049	0.176
<i>Balance sheet dimension</i>				
Model 7: Advertising expenses (<i>SG&A</i>)	-0.0035***	-3.37	45,357	0.540
Model 8: Asset tangibility & segment concentration (<i>TANG & COMPLEX</i>)	-0.0032***	-3.13	45,586	0.544
<i>Demographic dimension</i>				
Model 9: Social Capital Index (<i>SOCIAL</i>), Abortion (<i>ABORT</i>), Alcoholism (<i>ALC</i>), state-GDP (<i>SGDP</i>)	-0.0033***	-3.19	46,200	0.540
Model 10: All variables together	-0.0017*	-1.79	14,525	0.161

This table documents results of considering additional control variables for different dimensions (Model 1 to Model 9). All models include the variables used in our baseline analysis. In Model 10, we put all presented additional control variables, i.e., the first principal component of multiple governance variables, advertising expenses, Herfindahl index, capital intensity, social capital, abortion, alcoholism, and state-GDP, together in one model along with the variables used in our baseline analysis. All models are estimated with year and industry fixed effects. Across all models, the dependent variable is *BAS*, which is calculated as $(Ask - Bid)/((Ask + Bid)/2)$. Appendix Table A1 defines all other variables used in detail. Standard errors are adjusted for heteroscedasticity and within-firm clustering. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4: Robustness tests

Dimension	<i>REL</i>		<i>N</i>	<i>R</i> ²
	Coefficient	t-statistic		
<i>Dimension: control variables</i>				
Model 1: Dollar trading volume (ln(<i>TVOL</i>))	-0.0026***	-2.66	46,200	0.552
Model 2: Omitting ln(<i>TURNOVER</i>)	-0.0029***	-2.63	46,200	0.484
Model 3: Lagged firm characteristics	-0.0030***	-2.83	39,339	0.486
Model 4: Exclude all replaced values	-0.0032***	-3.02	22,166	0.490
<i>Geographic dimension</i>				
Model 5: Excluding most conservative counties	-0.0030***	-2.82	44,802	0.539
Model 6: Excluding counties with 5 highest and lowest RELs	-0.0032***	-3.04	46,032	0.539
Model 7: Omitting CA, TX, and NY	-0.0040***	-3.56	33,546	0.551
Model 8: Excluding five largest counties (in terms of number of obs.)	-0.0035***	-2.99	37,300	0.547
Model 9: Only urban companies	-0.0046**	-2.48	17,668	0.515
Model 10: State-fixed effects	-0.0032*	-1.82	46,200	0.542
Model 11: County level estimation	-0.0056*	-1.89	485	0.687
<i>Temporal dimension</i>				
Model 12: period 1997-2008	-0.0035***	-2.90	23,820	0.586
Model 13: period 2009-2020	-0.0026**	-2.06	22,379	0.437
Model 14: Financial crisis period (2007-2009)	-0.0066***	-2.66	5,776	0.433
Model 15: Excluding financial crisis period	-0.0029***	-3.11	40,423	0.564
Model 16: Fama/MacBeth-procedure	-0.0030***	-4.84	46,201	0.543
<i>Dimension: variable of interest</i>				
Model 17: <i>RES_REL</i>	-0.0033***	-3.18	46,200	0.539
Model 18: <i>HIGH_REL</i>	-0.0005***	-2.68	46,200	0.539
Model 19: <i>HIGH_RELI</i>	-0.0007***	-2.74	30,612	0.538

This table documents results of alternative definitions of variables and model specifications for different dimensions. All models include the variables used in our baseline analysis, and are estimated with year and industry fixed effects, with exception of Model (9) and (10). Across all models, the dependent variable is *BAS*, which is calculated as $(Ask - Bid) / ((Ask + Bid) / 2)$. Appendix Table A1 defines all other variables used in detail. Standard errors are adjusted for heteroscedasticity and within-firm clustering, with exception of Model (11). *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 5: Estimates from reverse causality and endogeneity tests

Variables	Model (1) SIC 100-3999	Model (2) 2SLS	Model (3a) Entropy Balancing	Model (3b) Entropy Balancing	Model (4) Propensity Score Matching	Model (5a) HQ relocation	Model (5b) HQ relocation	Model (5c) HQ relocation
<i>Variable of Interest</i>								
REL	-0.0030** (-2.29)	-0.0028* (-1.85)						
HIGH_REL_DUMMY			-0.0007*** (-2.73)	-0.0009*** (-3.26)	-0.0009*** (-2.92)			
ADH_INCR						-0.0047*** (-2.87)		-0.0051*** (-3.41)
ADH_DECR							-0.0005 (-0.41)	-0.0013 (-0.78)
Baseline controls included	YES	YES	YES	YES	YES	YES	YES	YES
Changes in baseline controls included	NO	NO	NO	NO	NO	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES	YES	YES
Observations	26,822	46,179	30,612	29,998	9,671	154	154	154
R-squared	0.541	0.470	0.560	0.554	0.575	0.480	0.444	0.476

This table documents results from tests of reverse causality and endogeneity. Across models (1) to (4), the dependent variable is *BAS*, which is calculated as $(Ask - Bid)/(Ask + Bid)/2$. In Models (5a) to (5c) we use the change of *BAS* as the dependent variable, where the change is measured from $t-1$ to $t+1$. All models are estimated with year and industry fixed effects. Appendix Table A1 defines all variables used in detail. Standard errors are adjusted for heteroscedasticity and within-firm clustering, with exception of Models (5a) to (5c). t-statistics are reported in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 6: Information environment and religiosity

Variables	ANALYST		S&P500		SIZE		LOCATION	
	(1) NOT COVERED	(2) COVERED	(3) NON-S&P500	(4) S&P500	(5) BELOW MEDIAN	(6) ABOVE MEDIAN	(7) FAR AWAY	(8) CLOSE
<i>Variable(s) of Interest</i>								
REL	-0.0127*** (-3.02)	-0.0023*** (-2.75)	-0.0036*** (-3.19)	0.0000 (0.52)	-0.0040*** (-2.90)	-0.0013 (-1.21)	-0.0042*** (-2.92)	0.0006 (0.28)
Baseline controls included	YES	YES	YES	YES	YES	YES	YES	YES
Differences in coefficients (p-value)	0.01		0.00		0.09		0.07	
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES	YES	YES
Observations	4,044	42,156	38,773	7,424	23,106	23,093	29,448	16,749
R-squared	0.574	0.492	0.534	0.656	0.593	0.300	0.540	0.554

This table documents results from the effects of *REL* on subsamples, which are constructed based on measures of information asymmetry. Across all models, the dependent variable is *BAS*, which is calculated as $(Ask - Bid)/((Ask + Bid)/2)$. The left panels, i.e., models (1), (3), (5), and (7), respectively, represent the subsamples of firms, which face high information asymmetry. All models are estimated with year and industry fixed effects. Appendix Table A1 defines all variables used in detail. Standard errors are adjusted for heteroscedasticity and within-firm clustering. In the “Differences in coefficients” column, we test the null hypothesis of the equality between the coefficients of *REL* across the subsamples. t-statistics are reported in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 7: Alternative liquidity measures and the trading environment

Variables	Dependent Variable		
	(1) VOLA_BAS	(2) AMIHUQ_SQ	(3) PIN
<i>Variable of Interest</i>			
REL	-0.0026** (-2.39)	-0.0631** (-2.37)	-0.0202*** (-2.99)
Baseline controls included	YES	YES	YES
Year FE	YES	YES	YES
Industry FE	YES	YES	YES
Observations	46,195	46,200	24,834
Adj. R-squared	0.338	0.572	0.706

This table documents results of alternative measures of liquidity and private information. All models include the variables used in our baseline analysis, and are estimated with year and industry fixed effects. Appendix A1 defines all other variables used in detail. Standard errors are adjusted for heteroscedasticity and within-firm clustering. t-statistics are reported in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Internet Appendix to: Local Religiosity and Stock Liquidity

Oliver Entrop[†], Martin Rohleder[‡], Marco Seruset[#]

Abstract

This Internet Appendix (IA) contains additional results that supplements our main paper. It consists of six sections: The first section (Section IA1) provides a Pearson correlation matrix for the variables used in our baseline model as well as descriptive statistics of variables used in auxiliary analyses (Section 4.2 in the main paper). The second part (IA2) analyses religious subgroups, while the third part (IA3) contains test diagnostics and further results from Entropy Balancing (EB) and Propensity Score Matching (PSM), respectively. The fourth section (IA4) provides a detailed description of the sample construction for headquarter changes, while IA5 reports results for high and low information asymmetry firms when using alternative measures.

[†] Oliver Entrop, University of Passau, Chair of Finance and Banking, Innstraße 27, 94032 Passau, Germany, phone: +49 851 509 2460, E-mail: oliver.entrop@uni-passau.de

[‡] Martin Rohleder, University of Augsburg, Chair of Banking and Finance, Universitätsstraße 16, 86159 Augsburg, phone: +49 821 598 4120, email: martin.rohleder@uni-a.de

[#] Marco Seruset, University of Passau, Chair of Finance and Banking, Innstraße 27, 94032 Passau, Germany, phone: +49 851 509 2463, email: marco.seruset@uni-passau.de

Section IA1: Pearson correlation matrix and descriptive statistics

Table IA1a: Pearson correlation matrix of variables used in our baseline analysis

Panel A: Correlations among firm attributes used in our baseline analysis

	BAS	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
REL (2)	0.0207															
ln(TURNOVER) (3)	-0.4926	-0.0970														
ln(PRICE) (4)	-0.2512	-0.0198	0.2226													
NASD (5)	0.2121	-0.0916	-0.0364	-0.1679												
ln(SIZE) (6)	-0.5248	-0.0015*	0.2850	0.3772	-0.4450											
RISK (7)	0.3507	-0.0456	0.1783	-0.1790	0.2969	-0.4376										
CUMRET (8)	-0.0222	-0.0106	0.0495	-0.0243	0.0239	-0.0197	0.0297									
CAPEX (9)	0.0292	0.0121	-0.0299	0.0124	-0.0480	0.0200	0.0022*	-0.0397								
RnD (10)	0.0853	-0.0951	0.1235	-0.0078	0.2932	-0.3310	0.3223	-0.0327	-0.1511							
LEVERAGE (11)	-0.0822	0.0103	0.0367	0.0764	-0.2690	0.3853	-0.1262	-0.0344	0.0901	-0.2448						
Q (12)	-0.0701	-0.0476	0.1232	0.1781	0.1616	-0.1616	0.1058	0.4032	-0.0252	0.2885	-0.2122					
ROA (13)	-0.1944	0.0584	-0.0801	0.0281	-0.2208	0.3507	-0.4339	0.0919	0.1196	-0.6918	0.0954	-0.1097				
ln(1+ANALYST) (14)	-0.5256	-0.0563	0.4973	0.3243	-0.2191	0.7076	-0.2455	-0.0410	0.0669	-0.0480	0.1608	0.1031	0.1817			
SP500 (15)	-0.2156	0.0202	0.1105	0.2765	-0.2576	0.6167	-0.2526	-0.0198	-0.0083	-0.0961	0.1145	0.04171	0.1691	0.5119		
OWN_INST (16)	-0.5668	-0.0382	0.4988	0.2739	-0.1993	0.5348	-0.3518	-0.0289	-0.0624	-0.1555	0.1482	-0.0435	0.2508	0.5567	0.1757	
OWN_INSIDER	0.3263	0.0224	-0.3896	-0.2186	0.0864	-0.3001	0.1595	-0.0117	0.0782	-0.0392	-0.0412	-0.0351	-0.0223	-0.3496	-0.2406	-0.4596

Table IA1a continued

Panel B: Correlations among demographic variables used in our baseline analysis

	REL	ln(TOTPOP)	ln(DENSITY)	EDUCATION	ln(AGE)	MF_RATIO	MARRIAGE	MINORITY
ln(TOTPOP)	0.0358							
ln(DENSITY)	0.1755	0.5370						
EDUCATION	-0.0047*	0.1075	0.4351					
ln(AGE)	-0.0005*	-0.2612	-0.0917	0.1460				
MF_RATIO	-0.3048	0.0472	-0.4150	-0.0665	-0.3579			
MARRIAGE	-0.0105	-0.3593	-0.6042	-0.0185	0.3544	0.3113		
MINORITY	-0.1321	0.5632	0.5438	0.1539	-0.3439	0.0045*	-0.5808	
ELEC_REP	-0.0229	-0.1734	-0.3389	-0.2990	-0.2837	0.1897	0.0928	-0.0970

Panel A presents a Pearson correlation matrix of the firm controls, while Panel B documents correlations among demographic variables, used in our baseline analysis. Appendix Table A1 (Panel A to Panel D) in the main paper provides a detailed description of variables. * indicates p-values > 0.10.

Table IA1b: Summary statistics of variables used in auxiliary analyses

	<i>N</i>	Mean	Std.Dev.	Percentiles				
				5th. Pctl.	25th. Pctl.	Median	75th. Pctl.	95th. Pctl.
GOV	15,309	46.8073	22.1154	11.9100	28.8500	46.4700	64.3800	82.7000
CEO_DUAL	15,289	0.6414	0.4796	0.0000	0.0000	1.0000	1.0000	1.0000
BINDEP	15,204	0.8950	0.3065	0.0000	1.0000	1.0000	1.0000	1.0000
BSIZE	15,184	9.4394	2.2265	6.0000	8.0000	9.0000	11.0000	13.0000
BIG4	46,201	0.7786	0.4152	0.0000	1.0000	1.0000	1.0000	1.0000
ADV	45,358	0.3046	0.2782	0.0336	0.1229	0.2369	0.4022	0.7854
TANG	46,071	0.2419	0.2223	0.0187	0.0747	0.1674	0.3406	0.7492
COMPLEX	45,701	0.7114	0.2860	0.2558	0.4869	0.7411	1.0000	1.0000
SOCIAL	46,201	-0.5170	0.7624	-1.7536	-1.0882	-0.5017	-0.0449	0.5783
ABORT	46,201	18.9297	7.7692	8.7000	13.6000	17.3000	24.0000	34.8400
ALC	46,201	2.2954	0.3137	1.8900	2.1400	2.2600	2.3900	2.7700
SGDP	46,201	13.2121	0.9209	11.7153	12.5850	13.1522	13.9722	14.6636
ln(TVOL)	46,201	8.4726	2.3599	4.2915	6.8508	8.6321	10.1996	12.1564
RES_REL	46,201	0.0000	0.0935	-0.1453	-0.0654	0.0040	0.0637	0.1579
HIGH_REL	46,201	0.4943	0.5000	0.0000	0.0000	0.0000	1.0000	1.0000
HIGH_REL1	30,612	0.4940	0.5000	0.0000	0.0000	0.0000	1.0000	1.0000
CATH	46,201	0.2500	0.1311	0.0652	0.1533	0.2285	0.3507	0.4887
PROT	46,201	0.2036	0.1232	0.0747	0.1068	0.1644	0.2926	0.4578
MPRT	46,201	0.0736	0.0485	0.0225	0.0398	0.0638	0.0922	0.1728
EVAN	46,201	0.1300	0.0986	0.0275	0.0574	0.0989	0.1705	0.3215
VOLA_BAS	46,196	0.0060	0.0129	0.0001	0.0005	0.0016	0.0063	0.0245
AMIHUQ_SQ	46,201	0.1835	0.3458	0.0081	0.0245	0.0609	0.1781	0.8046
PIN	24,836	0.1681	0.0970	0.0604	0.1023	0.1415	0.2099	0.3636

This table reports descriptive statistics metrics of variables used in auxiliary analyses of Section 4.2, 4.3 and 6, respectively, in the main paper. The sample covers firm-year observations with nonmissing values for all variables from 1997 to 2020. All continuous variables are winsorized at the 1%/99% level. Appendix Table A1 (Panel E) in the main paper provides a detailed description of the variables. The data are from Refinitiv Eikon, Refinitiv Datastream, U.S. Census Bureau, ARDA, Stephen Brown's website, and Northeast Regional Center for Rural Development.

Section IA2: Religious subgroups

Table IA2: Analyses of religious subgroups

Variables	(1) CATH and PROT	(2) CATH, EVAN and MAIN
<i>Variables of Interest</i>		
CATH	-0.0021* (-1.82)	-0.0024** (-1.97)
PROT	-0.0028** (-2.14)	
MAIN		-0.0042** (-2.47)
EVAN		-0.0000 (-0.01)
Baseline Controls Included	YES	YES
CATH vs. PROT	0.628	
CATH vs. EVAN		0.445
CATH vs. MAIN		0.257
EVAN vs. MAIN		0.218
Year FE	YES	YES
Industry FE		
Observations	46,200	46,200
Adj. R-squared	0.540	0.540

This table documents results of our baseline regressions using religious subgroups. Model (1) considers catholics and protestants separately, while model (2), besides catholics, further separate protestants in evangelical and mainline protestants. Across both models, the dependent variable is *BAS*, which is calculated as $(Ask - Bid)/((Ask + Bid)/2)$. Appendix Table A1 defines all other variables in detail. Standard errors are adjusted for heteroscedasticity and within-firm clustering. t-statistics are reported in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Section IA3: Test diagnostics and further results

Table IA3a: Test diagnostics and further results from EB

Panel A: Difference in characteristics before and after entropy balancing (EB)						
Variables	Main Sample			EB Sample (after matching)		
	High REL	Low REL	Diff. (1-2)	High REL	Low REL	Diff. (1-2)
ln(TURNOVER)	-2.159	-1.967	-0.192***	-2.159	-2.159	0.000
ln(PRICE)	3.089	3.073	0.016	3.089	3.089	-0.000
NASD	0.509	0.659	-0.150***	0.509	0.509	0.000
ln(SIZE)	6.566	6.406	0.160***	6.566	6.566	-0.000
RISK	0.031	0.034	-0.003***	0.031	0.031	0.000
CUMRET	0.160	0.188	-0.028***	0.160	0.160	-0.000
CAPEX	0.050	0.048	0.002***	0.050	0.050	0.000
RnD	0.047	0.077	-0.030***	0.047	0.047	0.000
LEVERAGE	0.201	0.184	0.0167***	0.201	0.201	-0.000
Q	2.200	2.541	-0.341***	2.200	2.200	-0.000
ROA	0.042	0.011	0.031***	0.042	0.042	-0.000
ln(1+ANALYST)	1.758	1.851	-0.093***	1.758	1.759	-0.000
SP500	0.169	0.151	0.018***	0.169	0.169	-0.000
OWN_INST	0.645	0.641	0.004	0.645	0.645	-0.000
OWN_INSIDER	0.198	0.194	0.004*	0.198	0.198	-0.000
ln(TOTPOP)	13.837	13.787	0.050***	13.837	13.836	0.001
ln(DENSITY)	6.616	6.081	0.535***	6.616	6.615	0.001
EDUCATION	0.366	0.368	-0.002*	0.366	0.366	-0.000
ln(AGE)	3.582	3.578	0.004***	3.582	3.582	0.000
MF_RATIO	0.949	0.976	-0.027***	0.949	0.949	0.000
MARRIAGE	0.412	0.417	-0.005***	0.412	0.412	-0.000
MINORITY	0.296	0.338	-0.042***	0.296	0.296	0.000
ELEC_REP	0.436	0.430	0.006***	0.436	0.436	0.000
Number of treated units	15,122					
Number of control units	15,490					
Number of observations	30,612					

Panel B: Results from matching procedure on higher moments

Variables	EB on mean and variance	EB on mean, variance, and skewness (excluding MF_RATIO) from the matching scheme
<i>Variable of Interest</i>		
HIGH_REL	-0.0014* (-1.91)	-0.0014* (-1.68)
Baseline controls included	YES	YES
Year FE	YES	YES
Industry FE	YES	YES
Observations	30,612	29,998
R-squared	0.550	0.523

Panel A reports the difference in characteristics before and after the matching procedure for the treatment and control group. Panel B presents the regression results from matching on higher moment conditions, i.e. mean, variance, and skewness. t-statistics are reported in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table IA3b: First stage logit model, test diagnostics, and further results from PSM

Panel A: First stage logit model	
Variables	HIGH_REL
ln(TURNOVER)	-0.153*** (-3.431)
ln(PRICE)	0.00164 (0.0356)
NASD	-0.474*** (-4.466)
ln(SIZE)	-0.0466 (-1.069)
RISK	-3.144 (-1.512)
CUMRET	-0.00346 (-0.140)
CAPEX	-1.585** (-2.284)
RnD	-1.514*** (-3.775)
LEVERAGE	0.0387 (0.188)
Q	-0.0355** (-2.034)
ROA	-0.148 (-0.835)
ln(1+ANALYST)	-0.0987 (-1.543)
SP500	0.199 (1.392)
OWN_INST	0.316 (1.600)
OWN_INSIDER	0.0132 (0.0792)
INTERCEPT	-1.667 (-1.594)
Observations	30,541
Year FE	YES
Industry FE	YES
Area under ROC Curve	0.680

Panel B: Test diagnostics of PSM sample

Variables	%bias			Mean difference of PSM Sample		
	%bias	%bias	%reduction	High	Low	Diff.
	(unmatched)	(matched)	of bias	REL	REL	(1-2)
ln(TURNOVER)	-19.9	0.1	99.6	-2.080	-2.081	0.001
ln(PRICE)	1.8	-1.6	9.4	3.083	3.098	-0.015
NASD	-30.9	3.1	90.0	0.614	0.599	0.015
ln(SIZE)	8.4	2.1	75.4	6.467	6.427	0.040
RISK	-17.5	0.2	99.0	0.032	0.032	0.000
CUMRET	-3.9	0.7	82.5	0.157	0.152	0.005
CAPEX	3.0	1.7	42.7	0.044	0.043	0.001
RnD	-26.1	-1.5	94.1	0.057	0.059	-0.002
LEVERAGE	9.0	1.6	82.4	0.192	0.189	0.003
Q	-16.8	-1.3	92.4	2.307	2.333	-0.026
ROA	13.6	2.2	83.6	0.034	0.029	0.005
ln(1+ANALYST)	-10.1	-0.1	99.4	1.800	1.800	0.000
SP500	4.8	-0.3	93.8	0.158	0.159	-0.001
OWN_INST	1.4	-1.8	-29.2	0.645	0.650	-0.005
OWN_INSIDER	2.1	-0.8	63.7	0.190	0.192	-0.002

Panel C: Alternative matching parameters

Variables	Adhikari and Agrawal (2016)	Cai et al. (2019)	Mayberry (2020)
HIGH_REL_DUMMY	-0.0008*** (-2.63)	-0.0008*** (-2.90)	-0.0008*** (-2.62)
Baseline controls included	YES	YES	YES
Caliper	0.0001	NA	0.01
Replacement	YES	YES	YES
Neighbors	1	3	1
Highest %bias	3.6	6.8	6.3
Observations	27,511	25,215	30,523
R-squared	0.549	0.549	0.552

Panel A documents results from the first stage logit model, which is used to predict the probability of a firm being located in the top tercile by county level religiosity. In Panel B, we report results of the bias and the differences in characteristics before and after the matching procedure for the treatment and control group. We

use Stata's *pstest* after matching (Stata command *psmatch2*; Leuven and Sianesi, 2003) to receive test diagnostics for PSM sample. The matching procedure is based on matching with no replacement, one nearest neighbor, and a caliper of 0.00001. Panel C reports results from alternative matching parameters, which are based on previous studies. t-statistics are reported in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Section IA4: Detailed description of headquarter changes sample

As noted in the data description, Refinitiv reports the current location of a firm's headquarter only, not its historic headquarter. To remedy this issue, we follow recent studies (e.g., Hasan et al., 2017a and 2020; Chow et al., 2021) and extract historic HQ addresses from Securities and Exchange Commission (SEC) filings. We use the HQ relocations data of Loughran-McDonald augmented 10-X header data.⁴ The dataset captures all information in the header section of 10-K/Qs (and all variants) filed on EDGAR. In total, there are 1,417,141 filings for 44,582 firms with a unique Central Index Key (CIK) for the period from 1993 to 2021. We follow the procedure described in Garcia and Norli (2012) to further extract the filings and to obtain a time series of HQ locations for every company in our sample. We limit our dataset to a period starting in 1997 (first year of available observations on institutional ownership) and ending in 2020.

Another important issue of data preparation relates to the mapping of company identifiers between Refinitiv and SEC. Refinitiv provides a crosswalk for merging ISIN codes (our company identifier) with their respective CIK code (SEC company identifier). Unfortunately, the matching procedure leads to a significant loss of firms compared to our original sample (see Section 3.1). This may have at least two reasons. First, CIK number is generally not available in Refinitiv. Second, we are aware of the fact that not all companies that offer stocks must file electronically. For example, certain small companies are excluded from regular SEC reporting when they have less than \$10 million in assets, thus having no CIK number. In total, we cannot match 715 firms to their respective ISIN code. This is also the main reason, why we use the SEC dataset for the HQ changes test only. Nonetheless, we estimated our baseline model (see equation 1) with this reduced-firm dataset for the period 1997-2020. The results are robust (t -statistic = -2.17), while the economic significance is lower than reported in Section 4.1 in the

⁴ We thank Bill McDonald for sharing the headquarter relocation data, <https://sraf.nd.edu/data/augmented-10-x-header-data/>. We note that this file also contains historic SIC codes, which we utilize in this section rather than the static SIC codes provided by Refinitiv Datastream.

main paper (the *BAS* decreases by 0.0036% when moving from the first to the third quintile of *REL*, which is approximately 4.92% of the mean *BAS*). The partially loss of economic significance could be explained by the exclusion of smaller firms. For example, Jiang et al. (2018) congruently find that the effect of *REL* on the cost of private and public debt is much stronger for smaller firms, i.e., if its asset value is below the sample median.

Table IA5: Information environment and religiosity

Variables	ANALYST		S&P500		SIZE		LOCATION	
	(1) NOT COVERED	(2) COVERED	(3) NON-S&P500	(4) S&P500	(5) BELOW MEDIAN	(6) ABOVE MEDIAN	(7) FAR AWAY	(8) CLOSE
Panel A: Dependent variable: AMIHUQ_SQ								
REL	-0.2857** (-2.53)	-0.0430** (-2.47)	-0.0692** (-2.31)	-0.0082 (-1.13)	-0.0756* (-1.88)	-0.0182 (-1.23)	-0.0699* (-1.82)	0.0164 (0.25)
Baseline controls included	YES	YES	YES	YES	YES	YES	YES	YES
Differences in coefficients (p-value)	0.03		0.05		0.17		0.25	
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES	YES	YES
Observations	4,044	42,156	38,776	7,424	23,106	23,093	29,448	16,749
R-squared	0.652	0.552	0.570	0.115	0.613	0.463	0.583	0.573
Panel B: Dependent variable: PIN								
REL	-0.0806*** (-2.87)	-0.0125** (-2.19)	-0.0224*** (-3.06)	0.0025 (0.35)	-0.0294*** (-3.22)	-0.0073 (-0.91)	-0.0191** (-2.11)	-0.0120 (-0.73)
Baseline controls included	YES	YES	YES	YES	YES	YES	YES	YES
Differences in coefficients (p-value)	0.01		0.01		0.06		0.70	
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES	YES	YES
Observations	2,419	22,413	21,010	3,824	12,324	12,508	15,848	8,984
R-squared	0.481	0.694	0.673	0.491	0.658	0.636	0.698	0.721

This table documents results from the effects of *REL* on subsamples, which are constructed based on measures of information asymmetry. Panel A presents the results for *AMIHUQ_SQ*, while Panel B reports the results for *PIN* as dependent variable. The left panels, i.e., models (1), (3), (5), and (7), respectively, represent the subsamples of firms, which face high information asymmetry. All models are estimated with year and industry fixed effects. Appendix Table A1 defines all variables used in detail. Standard errors are adjusted for heteroscedasticity and within-firm clustering. In the “Differences in coefficients” column, we test the null hypothesis of the equality between the coefficients of *REL* across the subsamples. t-statistics are reported in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

References for Internet Appendix

- Adhikari, B. K., Agrawal, A. (2016). Does religiosity matter for bank risk taking? *Journal of Corporate Finance* 38, 272-293. <http://dx.doi.org/10.1016/j.jcorpfin.2016.01.009>
- Amin, M. R., Kim, I., Lee, S. (2021). Local religiosity, workplace safety, and firm value. *Journal of Corporate Finance* 70, 102093. <https://doi.org/10.1016/j.jcorpfin.2021.102093>
- Cai, J., Kim, Y., Li, S., Pan, C. (2019). Tone at the top: CEOs' religious beliefs and earnings management. *Journal of Banking & Finance* 106, 195-213. <https://doi.org/10.1016/j.jbankfin.2019.06.002>
- Chow, T., Huang, S., Klassen, K. J., Ng, J. (2021). The Influence of Corporate Income Taxes on Investment Location: Evidence from Corporate Headquarters Relocations. *Management Science*, forthcoming, 1-22. <https://doi.org/10.1287/mnsc.2020.3906>
- Fang, V. W., Noe, T. H., Tice, S. (2009). Stock market liquidity and firm value. *Journal of Financial Economics* 94(1), 150-169. <https://doi.org/10.1016/j.jfineco.2008.08.007>
- Fich, E. M., Harford, J., Tran, A. L. (2015). Motivated monitors: The importance of institutional investors' portfolio weights. *Journal of Financial Economics* 118(1), 21-48. <https://doi.org/10.1016/j.jfineco.2015.06.014>
- García, D., Norli, O. (2012). Geographic dispersion and stock returns. *Journal of Financial Economics* 106, 547-565. <http://dx.doi.org/10.1016/j.jfineco.2012.06.007>
- Hasan, I., Hoi, C.-K., Wu, Q., Zhang, H. (2017a). Social Capital and Debt Contracting: Evidence from Bank Loans and Public Bonds. *Journal of Financial and Quantitative Analysis* 52(3), 1017-1047. <https://doi.org/10.1017/S0022109017000205>
- Hasan, I., Hoi, C.-K., Wu, Q., Zhang, H. (2020). Is social capital associated with corporate innovation? Evidence from publicly listed firms in the U.S. *Journal of Corporate Finance* 62, forthcoming. <https://doi.org/10.1016/j.jcorpfin.2020.101623>
- Jiang, F. W., John, K., Wei Li, C., Qian, Y. (2018). Earthly Reward to the Religious: Religiosity and the Costs of Public and Private Debt. *Journal of Financial and Quantitative Analysis* 53(5), 2131-2160. <https://doi.org/10.1017/S002210901800039X>
- Leuven, E., Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical Software Components S432001, Boston College Department of Economics, revised 01 Feb 2018.

- Mayberry, M. (2020). Good for managers, bad for society? Causal evidence on the association between risk-taking incentives and corporate social responsibility. *Journal of Business Finance and Accounting* 47(9-10), 1182-1214. <https://doi.org/10.1111/jbfa.12451>
- Roy, P. P., Rao, S., Zhu, M. (2022). Mandatory CSR expenditure and stock market liquidity. *Journal of Corporate Finance* 72, 102158. <https://doi.org/10.1016/j.jcorpfin.2022.102158>
- Zolotoy, L., O'Sullivan, D., Chen, Y. (2019). Local religious norms, corporate social responsibility, and firm value. *Journal of Banking & Finance* 100, 218-233. <https://doi.org/10.1016/j.jbankfin.2019.01.015>

3 | Geographic Proximity and Stock Price Crash Risk: Evidence from Institutional Investors*

Marco Seruset[†]

Abstract

This study shows that higher physical distance to institutional shareholders is associated with higher stock price crash risk. Since monitoring costs increase with distance, the results are consistent with the monitoring theory of local institutional investors. Cross-sectional analyses show that the effect of proximity on crash risk is more pronounced for firms with weak internal governance structures. The significant relation between distance and crash risk still holds under the implementation of the Sarbanes-Oxley Act, however, to a lower extent. Also, the existence of the channel of bad news hoarding is confirmed. Finally, I show that there is heterogeneity in distance-induced monitoring activities of different types of institutions.

Keywords: Institutional investors; Stock price crash risk; Shareholder proximity; Monitoring; Bad news hoarding; Local bias

JEL classification: G02; G34

* I appreciate excellent comments from Miriam Kamper and Carl-Friederich Groesbrink. I am responsible for all remaining errors and omissions.

[†] Marco Seruset, University of Passau, Chair of Finance and Banking, Innstraße 27, 94032 Passau, Germany, phone: +49 851 509 2463, email: marco.seruset@uni-passau.de

1 Introduction

The impact of physical distance between firms and their institutional shareholders on financial outcomes is well documented in the literature (Quan and Zhang, 2021). Accordingly, geographical proximity fosters face-to-face communication, strengthens relational ties, increases knowledge acquisition, and enhances product outcomes (e.g., Kang and Kim, 2008; Chhaochharia et al., 2012; Huang and Kang, 2017; Quan and Zhang, 2021). Unlike remote institutions, proximity also facilitates access to formal and informal (soft) firm-specific information. It is likely that local institutions belong to the social networks of local firm managers, thus having a direct impact on their behavior (e.g., Chhaochharia et al., 2012; An et al., 2020). For instance, golf courses could increase the likelihood of network building among firm managers and institutional investors (Huang and Kang, 2017). Also, local media coverage in, amongst others, newspapers, radio, and television should reduce information asymmetries between local firms and local investors (e.g., Kang and Kim, 2008; Engelberg and Parsons, 2011; Chhaochharia et al., 2012; An et al., 2020).

One implication of my idea is that monitoring costs vary with distance. In the presence of local monitoring institutions, managers tend to voluntarily reduce opportunistic behavior, because local institutions are more likely to detect fraudulent activities. Put it differently, low monitoring activities promoted by greater distance between a firm and its institutional investors probably facilitates bad news hoarding of corporate managers. Literature argues that withholding bad news from investors causes future stock price crash risk (e.g., Hutton et al., 2009; Kim et al. 2011a and 2011b; Callen and Fang 2015a and 2015b). From this point of view, managers tend to withhold bad news for an extended period, allowing bad news to stockpile. If managers successfully block the flow of negative information into the stock market, the distribution of stock returns should be asymmetric. When the accumulation of bad news passes

a threshold and the news become public at once, this leads to a large negative drop in stock price (Hutton et al., 2009; Kothari et al., 2009). Based on the arguments that bad news hoarding causes stock price crashes and that more distant institutional shareholders monitor to a lower extent due to higher monitoring costs, I posit that more distant institutional shareholders increase the likelihood of bad news hoarding, thus increasing future stock price crash risk at the same time.

I examine the empirical link between institutional distance and future stock price crash risk with reference to firms headquartered in the U.S. I measure geographic proximity as the value-weighted distance between a firm and its institutional shareholders, taking into consideration that shareholders with large ownerships are more likely to monitor. By using a sample of publicly listed U.S. companies and over 39,000 firm-year observations, covering the period from 1997 through 2019, my baseline results show a significant positive relation between institutional distance and future stock price crash risk. This finding is also economically meaningful and comparable to existing studies such as Callen and Fang (2013; 2015a) or Hasan et al. (2021). Shifting my distance measure from the 25th to the 75th percentile of the distribution is associated with an estimated 13.45% to 14.77% increase in the 1-year ahead stock price crash risk. Concerns regarding endogeneity are addressed by several robustness tests, i.e., the use of additional control variables, firm and high-dimensional fixed effects, changes analyses, a two-stage least squares method, and a placebo test. Moreover, my cross-sectional analyses reveal that the association between proximity and future stock price crash risk is of particular importance when internal governance mechanisms are weak. In line with Chhaochharia et al. (2012), I also find that the effect of geographically closer located shareholders on stock price crash risk persists even after the introduction of the Sarbanes-Oxley Act (SOX) in 2002, albeit weakened. As expected, bad news hoarding behavior (Chang et al., 2017; Andreou et al., 2017; Al Mamun et al., 2020; Krishnamurti et al., 2021) of corporate

managers as an important mediating channel is confirmed. Moreover, the notion that every local institution monitors to the same extent may be inaccurate (e.g., Callen and Fang, 2013; An and Zhang, 2013; Andreou et al., 2016). Thus, motivated by recent findings of Heath et al. (2022), I finally shed light on the question as to whether local institutional investors with a “passive” orientation and an index-style investment behavior also appear as effective monitors. My results suggest that those investors are less involved in monitoring compared to local “active” institutions. Overall, my findings highlight the incremental impact of institutional distance on future stock price crash risk.

My study builds on three previous investigations. Callen and Fang (2013) examine the impact of institutional ownership on crash risk. Consistent with the monitoring theory, the authors provide evidence that stable institutional groups play an important part in reducing future stock price crash risk. Likewise, An and Zhang (2013) find that the presence of dedicated investors with large stake holdings and long-term investment horizons is negatively associated with the 1-year ahead stock price crash risk. In a more recent study, Andreou et al. (2016) confirm this evidence since they show that, besides CEO option incentives and the equity holdings of directors, transient institutional ownership increases crash risk, while insider ownership and several governance metrics reduce future crash risk.

To the best of my knowledge, there is, however, limited empirical evidence regarding the question on how physical distance of institutional shareholders affects future stock price crash risk. Thus, my examination differs from the presented studies in two important ways. First, while the above-mentioned insightful papers do not investigate the impact of distance of institutional investors on crash risk, my study focuses on the effect of physical distance as an important determinant of corporate monitoring activities. Furthermore, as pointed out by Heath et al. (2022), passively managed index funds accounts for over 30% of U.S. equity fund assets nowadays. Due to their significant economic impact, this raises the fundamental question of

their monitoring behavior. Thus, second, I extend the line of research, which investigates the monitoring activities among different types of institutional investors, and their impact on future crash risk, by considering their distance in addition to their holdings. In a broader setting, my study contributes to the ongoing debate as to whether big shareholders that follow a passive and index-oriented strategy act as effective monitors.

The remainder of this paper is organized as follows. Section 2 summarizes literature and develops testable hypotheses. In Section 3, I present the data as well as summary statistics and sorting results. Section 4 presents the main empirical tests, while Section 5 briefly concludes.

2 Related literature and hypothesis development

The large body of research on the determinants of firm-specific stock price crash risk initially focuses on managerial incentives for bad news hoarding. Hutton et al. (2009) show that firms with more opaque financial reporting are more prone to crash risk.¹ In a subsequent study, Kim et al. (2011a) provide evidence that complex tax shelters enable managers to act opportunistically and decrease corporate transparency. This leads to an abrupt stock crash when all unfavorable information is revealed at once. Moreover, Kim et al. (2011b) document that equity-based compensation induces CFOs to hide bad news, thus increasing future stock price crash risk. The authors explain this finding as evidence that CFO incentives are more influential in situations where financial expertise is of particular interest. With respect to opportunistic behavior of corporate managers, Kim et al. (2014) further show that an engagement in corporate social responsibility mitigates future stock price crash risk due to a higher standard of transparency, thus resulting in lower incentives of managers to hoard bad news. Similarly, Callen and Fang (2015b) document a positive relation between short interest and 1-year ahead

¹ Hutton et al. (2009) empirically test the model proposed by Jin and Myers (2006) on the determinants of crash risk.

stock crash risk, since short sellers are capable to detect bad news hoarding by managers. Likewise, even short-term lenders (may) act as effective monitors, thus reducing crash risk (Dang et al., 2018). Also, CEO inside debt holdings (He, 2015) and CEO centrality (Krishnamurti et al., 2021) decrease firm-specific stock price crash risk, whereas firms with overconfident (Kim et al., 2016), younger (Andreou et al., 2017), and more powerful CEOs (Al Mamun et al., 2020) reveal higher levels of crash risk. Likewise, stock price crash risk is increased when firms are led by CEOs with early-life disaster experience (Chen et al., 2021). Besides the managerial perspective on stock price crash risk, Callen and Fang (2015a) document that firms headquartered in counties with higher levels of religiosity exhibit lower levels of future stock price crash risk. Additionally, Kubick and Lockhardt (2016) find that firms located farther from the SEC reveal greater stock price crash risk, whereas Kim et al. (2019) show that a higher complexity of financial reports is associated with higher stock price crash risk. With respect to the location of a firm, Xu et al. (2020) provide empirical evidence that stock price crash risk is lower for locally important firms. Further informal determinants, such as political incentives (Piotrski et al., 2015), political connections (Lee and Wang, 2017), trust (Li et al., 2017), or internet searching (Xu et al., 2021) are also connected with stock price crash risk for the Chinese market. Most recently, Hasan et al. (2021) investigate the link between brand capital and future stock price crash risk. Consistent with agency theory, managerial opportunistic behavior is reduced by higher levels of brand capital, since firms are more exposed to investors and customers scrutiny. Regarding the impact of institutions' external monitoring on future stock price crash risk, the literature shows that monitoring activities by stable institutional investors attenuate managers' abilities to hide bad information, which is linked to lower future stock price crash risk (An and Zhang, 2013; Callen and Fang, 2013; Andreou et al., 2016).²

² A comprehensive literature review is provided by Habib et al. (2018).

A handful of research investigates the importance of location, as a proxy for the degree of monitoring activities, on firm outcomes.³ For instance, Kang and Kim (2008) document that geography plays an important role for M&A deals. Ayers et al. (2011) use local institutional ownership as a proxy for the cost of acquiring monitoring information and find that corporate managers are less involved in opportunistic financial reporting activities in the presence of local monitoring institutions. Moreover, Chhaochharia et al. (2012) show that monitoring effects of local institutional shareholders improve internal corporate governance, while these companies are also less likely to be involved in fraudulent activities and more profitable. This effect is even more pronounced when shareholders are classified as long-term, “dedicated” institutions. Also, Mazur et al. (2018) find that effective monitoring of local institutions reduces information asymmetries, relating to a lower litigation risk. While the mentioned studies focus on the distance between firms and their portfolio institutions, Huang and Kang (2017) additionally consider the geographic concentration among institutional investors. Their results indicate that even the distance among shareholders increases monitoring effectiveness, which improves corporate governance and increases firm value.

From my perspective, I posit that geographically distant investors monitor corporate behavior less effectively than their local peers, because monitoring costs increase with distance (e.g., Coval and Moskowitz, 1999; Lin and Png, 2003; Degryse and Ongena, 2005; Chhaochharia et al., 2012). I predict that firms with more distant institutional shareholders reveal higher levels of future stock price crash risk, since a lower level of monitoring activities by distant shareholders decreases information transparency, thereby increasing managers’ possibilities to hoard bad news. Formally, I propose the following hypothesis:

³ There are also several studies focusing on an information-based explanation of local institutional ownership, e.g., Coval and Moskowitz, 1999; Gaspar and Massa, 2007; Baik et al., 2010; Bernile et al., 2015.

Hypothesis 1: Firms with more distant shareholders reveal higher levels of future stock price crash risk.⁴

I also address the question as to whether the increased monitoring activities of local institutional shareholders differ according to the quality of internal corporate governance (e.g., Chhaochharia et al., 2012; Andreou et al., 2016). Considering that firms with high local institutional ownership are subject to better internal corporate governance, monitoring aspects are likely to be of particular interest when internal governance is weak. Another purpose of this paper is to examine to which extent the distance to institutional shareholders plays an important role after the introduction of the SOX. Multiple studies document improvements in different facets of corporate behavior after the SOX (e.g., Cohen et al., 2008; Hutton et al., 2009; Callen and Fang, 2017), which is likely to be associated with lower bad news hoarding. Based on these findings, I expect that the importance of institutional distance is more pronounced for firms with weak internal governance and before the introduction of the SOX.

Hypothesis 2: The relation between geographic proximity and future stock price crash risk is more emphasized when (a) internal governance is weak and (b) before the introduction of the SOX.

Closely related to the governance channel, I also shed light on the channel of bad news hoarding behavior of corporate managers. Based on the argumentation provided earlier, I expect

⁴ Due to the construction of my main distance variable (see Section 3.2), I take the perspective of distant shareholders for the formulation of my hypotheses to match them with the sign of the estimated coefficients. However, some explanations are from the view of local institutions. Thus, for instance, re-expressing Hypothesis 1 from a geographic proximate ownership perspective leads to the following: Firms with more local institutions have lower levels of future stock price crash risk.

that withholding unfavorable information is increased when a firm reveals more distant shareholders.

Hypothesis 3: Firms with more distant shareholders expose higher levels of bad news hoarding.

Finally, prior research documents that monitoring incentives of institutional shareholders may differ between various types of shareholders (e.g., Callen and Fang, 2013; An and Zhang, 2013; Andreou et al., 2016). Also, Heath et al. (2022) provide empirical evidence that index funds are less effective monitors relative to their “active” peers. Consequently, I expect that the monitoring activities of local “active” shareholders are stronger, thus having an impact on future stock price crash risk. From the perspective of distant shareholders, my final hypothesis is:

Hypothesis 4: Firms with more distant “active” institutions show higher levels of crash risk, while there is no significant effect for their distant “passive” counterparts.

3 Data and univariate analyses

3.1 Firm level sample

My data is retrieved from four sources. I begin with all active and dead U.S. companies traded on NYSE, AMEX (formerly “NYSE MKT”), or NASDAQ, that are covered by Refinitiv Datastream (RD) from 1973 through 2020 (22,420 companies). Restricting the sample to U.S. headquartered firms with common equity and available company identifier (e.g., Porter and Ince, 2006; Landis and Skouras, 2021), 11,039 companies remain for the RD sample.

While RD reports the current location of a company only, I extract historic headquarters' locations from the SEC's EDGAR platform.⁵ I identify the institutional location by using the longitude and latitude that are connected to the respective zip code of the companies' headquarters. To receive a time series of historic addresses, I follow the procedure proposed by Garcia and Norli (2012). The SEC uses the Central Index Key (CIK) as the unique identifier for the firms, while the main identifier for the RD sample is represented by the International Securities Identification Number (ISIN). I map ISIN codes to their respective CIK codes using Refinitiv's crosswalk. After merging the RD dataset with the historical headquarter (HQ) information from the SEC, excluding companies with missing or unknown SIC code, I am able to match 8,007 firms and 96,156 forms for the period from 1993 through 2020.

The calculation of the distance measures (see Section 3.2) requires headquarter addresses and common stock holdings information of 13(f) institutions. These are compiled by Refinitiv Ownership Profile (ROP) and are available from 1997 through 2020. For each investor, Refinitiv reports the current business address along with the respective zip code. I use the Gazetteer File from the U.S. Census Bureau to match zip codes to the latitude and longitude coordinates for US-based investors. Missing information on longitude and latitude, and coordinates for non-US shareholders are retrieved manually.⁶

Ultimately, financial data are obtained from Refinitiv Datastream (RD). Following common practice in literature (e.g., Hutton et al., 2009; Kim et al., 2014), I exclude financials (2-digit SIC codes 60-69) and firms from the utility sector (2-digit SIC code 49) due to their regulated nature. I also delete all firm-years if I observe firms with year-end share prices below \$1, or those with fewer than 26 weeks of stock return data. This ensures that my results are not driven

⁵ I thank Bill McDonald for sharing historical headquarter data on his website, <https://sraf.nd.edu/data/augmented-10-x-header-data/>. This file contains also historic SIC codes, which I use for my analysis rather than SIC codes provided by Refinitiv Datastream.

⁶ The Gazetteer File from the U.S. Census Bureau is available at https://www2.census.gov/geo/docs/maps-data/data/gazetteer/2021_Gazetteer. Please refer to Section 3.2 for a detailed description of calculating distance measures.

by low liquid stocks (Kim et al., 2014). Requiring all firm-years to have full availability⁷ of baseline variables for the sample period (1997-2019) yields a final sample of 39,652 firm-year observations, which represents 4,100 firms.⁸ All of the continuous variables are winsorized at the 1% and 99% levels (e.g., Chhaochharia et al., 2012; Kim et al., 2011a and 2011b; Xu et al., 2021).

3.2 Measures for shareholder proximity

My main variable of interest to measure shareholder proximity is the value-weighted distance between a firm's headquarter (Pirinsky and Wang, 2006) and all its institutional investors, which is labelled as $DIST_VAL_ALL_t$.⁹ The weights are determined based on the equity ownership of an institution in the firm. This measure not only captures the distance to institutions, but also considers that large shareholders with high ownership have stronger incentives to monitor corporate behavior (Chhaochharia et al., 2012; Huang and Kang, 2017). It also avoids having to set a threshold for defining the number of relevant shareholders. For robustness, I also compute the percentage of shareholders that are located within a 100-mile radius ($PERC_LOCAL100_t$) as well as the equal-weighted distance ($DIST_EQ_10_t$) to its ten largest shareholders (Chhaochharia et al., 2012).¹⁰

⁷ Following Kothari et al. (2005), I also eliminate all observations where there are fewer than ten observations in a 2-digit industry code for a given year, when estimating the accrual measures.

⁸ Since I compute 1-year ahead stock price crash risk measures, the year 2020 is omitted from the sample. Note that, when using $DUVOL_{t+1}$ as dependent variable, the sample size reduces to 39,639 firm-years.

⁹ As noted earlier, my dataset contains the latitude and longitude for each firm and for each institutional investor that has a holding in that firm. Using the Great Distance Formula (Stata package *geodist*; Picard, 2010), I then receive the distance in miles between the company and each shareholder. See Coval and Moskowitz (1999) for a detailed description.

¹⁰ Note that the investor sample is not limited to US-based investors to receive a stronger evidence for geographic proximity. Considering US-based institutions only, does, however, not alter my overall conclusions, since almost 97% of institutional investors are headquartered in the United States (Coval and Moskowitz, 1999).

3.3 Measures of stock price crash risk

I employ two commonly used measures to proxy crash risk, following Chen et al. (2001).¹¹ Both measures rely on firm-specific weekly returns, which are estimated from the expanded market model for each firm i and for each year t (e.g., Jin and Myers, 2006; Hutton et al., 2009):

$$r_{i,\tau} = \alpha_i + \beta_{1i}r_{m,\tau-2} + \beta_{2i}r_{m,\tau-1} + \beta_{3i}r_{m,\tau} + \beta_{4i}r_{m,\tau+1} + \beta_{5i}r_{m,\tau+2} + \varepsilon_{i,\tau}, \quad (1)$$

where $r_{i,\tau}$ is the logarithmic return of firm i in week τ , and $r_{m,\tau}$ is the logarithmic return of the Wilshire 5000 market-cap weighted market index in week τ . The lead- and lag-terms are included to control for non-synchronous trading (Dimson, 1979). As common in literature (e.g., Hutton et al., 2009) I then define the firm-specific weekly returns for firm i in week τ ($W_{i,\tau}$) as the natural logarithm of one plus the residual return from equation (1), i.e., $W_{i,\tau} = \ln(1 + \varepsilon_{i,\tau})$.

My first measure of stock price crash risk, the negative conditional skewness (*NSKEW*), is calculated as the negative of the raw third central moment of firm-specific weekly returns divided by the sample variance of firm-specific weekly returns to the power 3/2 (An and Zhang, 2013; Utz, 2018; Balachandran et al., 2020).¹² I obtain *NSKEW* for each firm i in year t as follows:

$$NSKEW_{i,t} = - \frac{\left[n(n-1)^{\frac{3}{2}} \sum_{\tau=1}^n (W_{i,\tau,t} - \bar{W}_{i,t})^3 \right]}{\left[(n-1)(n-2) \left(\sum_{\tau=1}^n (W_{i,\tau,t} - \bar{W}_{i,t})^2 \right)^{\frac{3}{2}} \right]}, \quad (2)$$

¹¹ The subsequent studies also follow Chen et al. (2001) and use these proxies: Kim et al. (2011a and 2011b), Kim et al., (2014), Callen and Fang (2015a and 2015b), Xu et al. (2021), as well as Hasan et al. (2021).

¹² I will present results for alternative measures of crash risk in my robustness tests (see Section 4.3.2).

where $W_{i,\tau,t}$ is the firm-specific weekly return calculated from equation (1). $\bar{W}_{i,t}$ is the average firm-specific weekly return in year t , and n is the number of observations in the respective year (An and Zhang, 2013; Utz, 2018; Balachandran et al., 2020). As outlined by An and Zhang (2013), Callen and Fang (2015b), and Utz (2018), the cubed standard deviation (the denominator) is a standard normalization factor in order to compare skewness across returns with different variances.¹³ The minus sign in front of the adjusted skewness indicates that an increase in *NSKEW* corresponds to a firm being more “crash prone” (e.g., An and Zhang, 2013; Callen and Fang, 2015a and 2015b).

My second measure of crash risk is the “down-to-up volatility” (*DUVOL*), which is computed following Chen et al. (2001), Dang et al. (2019), and Hu et al. (2020b). For each stock, I separate all weeks with firm-specific returns above (“up weeks”) or below (“down weeks”) the sample mean of the year, respectively. Next, I compute the empirical standard deviation for each of these subsamples separately. Finally, I receive my second measure of crash risk by taking the natural logarithm of the standard deviation of down weeks divided by the standard deviation of up weeks. As with *NSKEW*, higher values of *DUVOL* correspond to a firm of being more “crash prone” (Callen and Fang, 2015b; Xu et al., 2021).¹⁴ Since this second measure does not account for the third moment of firm-specific weekly returns, it is less likely to be affected by a small number of extreme returns (Callen and Fang, 2015b).

¹³ Technically, I adjust Stata’s empirical skewness measure by $(n(n-1))^{0.5}/(n-2)$ to receive equation (2).

¹⁴ In particular, I estimate *DUVOL* for each firm i in year t by $\ln \left[\frac{\sqrt{\frac{\sum_{DOWN} (W_{i,\tau,t} - \bar{W}_{i,t})^2}{(n_d-1)}}}{\sqrt{\frac{\sum_{UP} (W_{i,\tau,t} - \bar{W}_{i,t})^2}{(n_u-1)}}} \right]$, where n_d and n_u are the number of down and up weeks, respectively.

3.4 Empirical model

To study the relation between institutional distance and future stock price crash risk, I estimate my baseline model using ordinary least squares (OLS) regression technique with standard errors adjusted for heteroscedasticity and firm clustering. Specifically, I employ the following empirical model (hereafter “baseline model”) to test my first hypothesis:

$$\begin{aligned}
 \text{CRASH_RISK}_{i,t+1} &= \beta_0 + \beta_1 \ln(\text{DIST_VAL_ALL}_{i,t}) + \delta' \text{FIRM_CONTROLS}_{i,t} + \sum_j \text{IND}_j \\
 &+ \sum_s \text{YEAR}_s + \varepsilon_{i,t+1},
 \end{aligned} \tag{3}$$

where $\text{CRASH_RISK}_{i,t+1}$ is a scalar containing either NSKEW_{t+1} or DUVOL_{t+1} in year $t+1$, respectively (see Section 3.3). My main variable of interest is $\ln(\text{DIST_VAL_ALL}_t)$, which is the natural logarithm of the value-weighted distance in year t , as discussed in Section 3.2. Like in prior studies (e.g., Kim et al., 2014; Callen and Fang, 2015a and 2015b; Xu et al., 2021; Hasan et al., 2021), there is a one-year lag between the dependent variable and all independent variables to predict crash risk and to reduce concerns stemming from reverse causality.

I control for a set of variables ($\text{FIRM_CONTROLS}_{i,t}$) that have been shown to determine future stock price crash risk (e.g., Chen et al., 2001; Hutton et al., 2009; Kim et al., 2011a and 2011b; Kim et al., 2014; An and Zhang, 2013; Callen and Fang, 2015a and 2015b; Xu et al., 2021). This set includes NSKEW_t or DUVOL_t , respectively, DTURN_t , SIGMA_t , CUMRET_t , SIZE_t , MB_t , LEV_t , ROA_t , ACCM_t , and OWN_INST_t .

To account for potential serial correlation in my crash risk measures (NSKEW_{t+1} or DUVOL_{t+1} , respectively) for my sample firms, I include the lagged dependent variables as

controls (e.g., Kim et al., 2014; Hasan et al., 2021).¹⁵ Next, Chen et al. (2001) show that heterogenous opinions among investors cause greater crashes. Thus, I consider the detrended average monthly stock trading volume ($DTURN_t$) to cater for investor heterogeneity beliefs. $DTURN_t$ is calculated as the average monthly share turnover in year t minus the average monthly share turnover in year $t-1$. $SIGMA_t$ is defined as the standard deviation of firm-specific weekly returns in year t . Stocks with higher volatility are more likely to experience a crash (Chen et al., 2001; Kim et al., 2011). Likewise, Chen et al. (2001) show that stocks with high past stock returns are more “crash prone”. For this purpose, I include the cumulative firm-specific weekly returns ($CUMRET$). Borrowing the findings of previous studies (e.g., Chen et al., 2001; Hutton et al., 2009), I consider total assets ($SIZE_t$) as well as the market-to-book ratio (MB_t). Larger firms and firms with high growth opportunities are shown to reveal greater future stock price crash risk. Moreover, based on the findings of Hutton et al. (2009), I include two further regressors: leverage (LEV_t), defined as the long-term debt divided by total assets, and return on assets (ROA_t), calculated as the net income before extraordinary items to total assets. Hutton et al. (2009) show that companies with higher leverage and better financial performance are less likely to suffer a future crash. Again, exploiting the findings of Hutton et al. (2009), I consider accrual manipulation ($ACCM_t$), computed as the three-year moving sum of the absolute value of annual performance-adjusted discretionary accruals (Kothari et al., 2005; Callen and Fang, 2015a and 2015b) to capture effects of financial reporting opacity. To ensure

¹⁵ By including the lagged dependent variables as controls, my baseline model is basically of dynamic nature. It is noteworthy, especially in the case of firm fixed effects regression (see Section 4.2.3), that only if $T \rightarrow \infty$, the within estimator of the independent variables is consistent, as emphasized by Baltagi (2021, p. 188). However, based on Monte Carlo simulations, Flannery and Hankins (2013) state: “[...] FE often is the most accurate with respect to the exogenous variables, but exhibits much higher errors for the lag.” To avoid dynamic panel biases caused by the inclusion of a lagged dependent variable (Nickell, 1981), I follow Kim et al. (2019) and run all my models by excluding the lagged crash risk measure. In any case, the results are quantitatively and qualitatively similar. Moreover, Dang et al. (2018) propose the application of a system GMM, thus estimating equation (3) in both levels and first-differences using appropriate instruments for the two potential endogeneous variables, crash risk ($NSKEW_t$ or $DUVOL_t$, respectively) and $\ln(DIST_VAL_ALL_t)$. Although the coefficients on distance remain statistically significant, the null hypothesis of the Hansen test is rejected, thus indicating a weak instrument problem. This is the main reason why I do not present results of the dynamic model. Ultimately, Gaspar and Massa (2007), and Kang and Kim (2008) additionally point out that local ownership is likely to be exogenous.

that the impact of value-weighted distance on future stock price crash risk is not solely driven by the ownership structure (An and Zhang, 2013), I include institutional ownership (OWN_INST_t). I start the construction by identifying all 13(f) institutions that have a holding in a company at the end of each quarter. I then divide the total institutional holdings by the firm's total number of shares outstanding. OWN_INST_t is the average percentage of shares held over the four quarters over the year. Lastly, to capture industrial differences and time trends, I include dummy variables for 2-digit SIC codes as well as time dummies. All presented control variables are collected from Refinitiv Datastream, while data on institutional ownership are retrieved from Refinitiv Ownership Profile. Finally, SIC codes come from the SEC file. Appendix A1 provides comprehensive variable descriptions.¹⁶

[Insert Table 1 about here]

3.5 Descriptive statistics

Panel A of Table 1 presents descriptive statistics for the variables used in my baseline regression model from 1997 through 2019 for the sample firms. The mean values of $NSKEW_{t+1}$ ($NSKEW_t$) and $DUVOL_{t+1}$ ($DUVOL_t$) are 0.238 (0.254) and 0.088 (0.095), respectively, while the median values are lower, amounting to 0.111 (0.118) and 0.062 (0.066) for $NSKEW_{t+1}$ ($NSKEW_t$) and $DUVOL_{t+1}$ ($DUVOL_t$), respectively. These estimates are higher than those reported in existing studies with reference to institutional ownership, such as An and Zhang (2013), or Andreou et al. (2016).¹⁷ Possible reasons for the deviations are (Al Mamun et al., 2020): (a) a difference in

¹⁶ I consider additional variables that are used either as additional control variables or as partitioning variables. A detailed description of these measures is presented in Appendix A1, Section D.

¹⁷ The mean (median) values for $NSKEW$ reported in An and Zhang (2013) are -0.165 (-0.195), and 0.101 (0.046) for Andreou et al. (2016), whereas, the mean (median) $DUVOL$ is -0.100 (-0.114) for An and Zhang (2013), and -0.000 (-0.011) for Andreou et al. (2016). Note that I do not compare my statistics to those reported in Callen and Fang (2013), since the authors use daily return data to calculate measures of crash risk.

the use of the market index to estimate firm-specific weekly returns: My study employs the Wilshire 5000 market index rather than the CRSP value-weighted market index¹⁸; (b) a difference in sample period: While my study focuses on the period between 1997 through 2019, the study period for An and Zhang (2013) is 1987-2010, and 2002-2013 for Andreou et al. (2016), respectively; and (c) a difference in defining sample selection criteria: I exclude all stocks with year-end share prices below \$1 (following Kim et al., 2014), whereas Andreou et al. (2016) set a price filter of \$2.5. An and Zhang (2013) do not discuss as to whether they exclude small priced stocks. Ultimately, Chen et al. (2021) also report higher values than those documented in other studies. For instance, the authors find a mean value for $DUVOL_{t+1}$ of 0.086, which is close to my observed value. According to the authors, this discrepancy to other studies may be explained by the use of rather larger firms.¹⁹

Further, my sample firms reveal an average value-weighted distance to their institutional shareholders ($DIST_VAL_ALL_t$) of 1,238 miles, which is similar to, however slightly higher than, the reported mean in Chhaochharia et al. (2012).²⁰ Also consistent with Chhaochharia et al. (2012), $DIST_VAL_ALL_t$ has a reasonably high level of variation in the amount of 535.112 miles, while most of the variation is cross-sectional in nature. Untabulated calculations of the absolute within group changes of $DIST_VAL_ALL_t$ reveal a mean (standard deviation) of 13.388 (184.500), indicating a modest temporal component (e.g., Hilary and Hui, 2009, for similar calculations in a different setting).²¹

With regards to my controls, the average change in monthly turnover is 0.007. The average firm in my sample reveals a firm-specific return of -0.183, while $SIGMA_t$ is 0.059. Moreover,

¹⁸ Due to data restrictions, I am not able to use the CRSP value-weighted market index for estimating firm-specific weekly returns.

¹⁹ It is likely that my sample also covers mostly large firms due to the matching procedure of CIK and ISIN.

²⁰ Chhaochharia et al. (2012) document a mean value-weighted distance of 1,132 miles. My slightly higher value is likely to be caused by the inclusion of non-US based investors. When I restrict the variable to U.S. investors only, I observe a mean value of 1,127 miles, which is almost identical to Chhaochharia et al. (2012).

²¹ Since investor addresses are static, the within group time variation can stem from two sources only: either from headquarter relocations of sample firms, or from changes in the ownership structure.

the sample firms have, on average, a 60 percent institutional ownership, and their measure of discretionary accruals is 0.185. Taken together, descriptive figures of my firm controls are mostly in line with prior studies, which investigate the determinants of future stock price crash risk (e.g., Kim et al., 2014; Callen and Fang, 2015a and 2015b; Al Mamun et al., 2020; Hasan et al., 2021). In addition to the variables used in my baseline analysis, I report the summary statistics of the variables used as additional controls or as partitioning variables (Panel B of Table 1). For instance, the average values of $CRASH_{D_{t+1}}$ and $CRASH_{EXTR_{t+1}}$ are 0.310 and 2.830, respectively, which are higher than the means reported in Hasan et al. (2021), whereas these higher values are consistent with the larger values for $NSKEW_{t+1}$ and $DUVOL_{t+1}$, respectively, documented in my study. Finally, $\ln(DIST_{EQ_{10t}})$ and $PERC_{LOCAL100t}$ reveal mean values of 7.009 miles and 0.090, respectively. All variables are defined in Appendix A1.

[Insert Table 2 about here]

3.6 Correlation analysis

Turning the focus to Table 2, $\ln(DIST_{VAL_ALL_t})$ is positively correlated with both $NSKEW_{t+1}$ (0.04; t-statistic = 7.94) and $DUVOL_{t+1}$ (0.04; t-statistic = 8.19), respectively, while the correlation coefficients are statistically significant at the 1% level. These univariate tests are consistent with my first hypothesis, indicating a possible link between shareholder proximity and future crash risk. Moreover, as expected, the correlation coefficient between the two crash risk measures is positive and statistically significant (0.96, t-statistic = 649.55), which is in line with prior literature (e.g., Kim et al., 2014; Hasan et al., 2021). Lastly, my measures of crash risk (i.e., $NSKEW_{t+1}$ and $DUVOL_{t+1}$) are positively correlated with lagged crash risk, share turnover, standard deviation, market-to-book ratio, size (significant at the 5% level for

$NSKEW_{t+1}$), discretionary accruals, and institutional ownership, whereas crash risk is negatively correlated with lagged cumulative return, leverage, and return on assets (insignificant for $DUVOL_{t+1}$). These results are largely consistent with previous studies, with exception of the negative correlation between cumulative return and future stock price crash risk.

3.7 Sorting results

Borrowing the idea of Chhaochharia et al. (2012), I additionally present sorting results based on annual distance quintiles²² to examine the distribution of crash risk and selected firm characteristics. Specifically, observations with the lowest (highest) levels of the value-weighted distance fall into the group of firms that lie in Q1 (Q5). They are labelled as “Local” (“non-L”). Moreover, to provide first empirical insights on the heterogeneity of institutional investors, I also form distance portfolios, which consider an institutions’ investment style and investment orientation.²³ To construct these subsamples, I split the sample in “active” and “passive” institutions and calculate their value-weighted distance separately.²⁴ As for the overall sample, I then compute annual distance quintiles for each subsample (i.e., “active” and “passive” institutions) and compare the means of future crash risk and firm characteristics between the Q1 and Q5 groups. I present the results in Table 3.

[Insert Table 3 about here]

²² The results also hold when I split the sample based on terciles (e.g., Hasan et al., 2021).

²³ Prior research for the U.S. typically relies on the classification of institutional investors using Bushee’s classification scheme (e.g., Bushee, 1998; Bushee and Noe, 2000; Bushee, 2001). Due to data limitation, this classification scheme is not available.

²⁴ This classification is motivated by the findings of Heath et al. (2022), which I present in Section 4.7 in more detail. To save space in this section, the definitions of the physical distance between firms, and “active” ($DIST_VAL_ALL_Ind_i$) and “passive” ($DIST_VAL_ALL_exInd_i$) institutions, respectively, are captured in Section 4.7 and in Appendix A1 (Panel D).

First, in columns (1)-(3) of Table 3, I find that, compared to firms with high distance (Q5 = “Non-L”) to their institutional investors, firms with low distance (Q1 = “Local”) to their institutional investors have lower future stock price crash risk. They also tend to have lower stock return volatility, market-to-book ratio, discretionary accruals, and institutional ownership. Firms, which are closer to their institutional investors also have higher return, leverage, and financial performance. The sorting results based on the subsamples further show that the difference in crash risk between firms in the highest (Q5 = “Non-L”) and lowest (Q1 = “Local”) distance quintiles (see column (6) compared to column (9)) are much more pronounced, when local investors are defined as “active” institutions in contrast to “passive” shareholders, which is consistent with the findings of Chhaochharia et al. (2012) and Heath et al. (2022). I interpret this result as an indication that local “actively”-oriented investors are more likely to monitor corporate behavior than their local “passive” counterparts. Overall, these results are in line with my expectations and theoretical predictions (see Hypothesis 1 and 4).

4 The relation between shareholder proximity and stock price crash risk

4.1 Baseline regression results

The regression results are reported in Table 4. The main measures for crash risk are $NSKEW_{t+1}$ and $DUVOL_{t+1}$, respectively, and my main variable of interest is $\ln(DIST_VAL_ALL_t)$. The main dependent variable in columns (1) to (3) is $NSKEW_{t+1}$, while columns (4) to (6) refer to $DUVOL_{t+1}$. More specifically, in columns (1) and (4), I control for year and industry fixed effects only, whereas columns (2) and (5) add all firm characteristics presented in Section 3.4, with exception of the lagged crash risk. This is intended to minimize potential statistical

concerns arising from the inclusion of the lagged crash risk in my baseline model (see footnote 15). Finally, columns (3) and (6) report the results of my baseline model (see equation (3)).

[Insert Table 4 about here]

Across all specifications, the estimates on $\ln(DIST_VAL_ALL_t)$ are positive and statistically significant at the 1% level. For instance, the point estimates for my baseline regressions in column (3) and (6) are 0.049 (t-statistics = 4.24) and 0.020 (t-statistics = 4.17), respectively, indicating that firms with more distant shareholders reveal significant higher levels of future stock price crash risk. In other words, higher levels of local ownership reduce crash risk.²⁵

The results are also significant in economic terms. To place my results in the context of prior studies on stock price crash risk, I estimate the economic magnitude of $\ln(DIST_VAL_ALL_t)$ using the interquartile range (Callen and Fang, 2015a; Hasan et al., 2021). Thus, moving from the first quartile (6.734) to the third quartile (7.380) of $\ln(DIST_VAL_ALL_t)$, the future crash risk ($NSKEW_{t+1}$) increases by 0.032 (= 0.049×0.65), which is roughly 13.45% (= 0.032/0.238×100%) of the mean $NSKEW_{t+1}$ for the average firm.²⁶ The economic significance of $DUVOL_{t+1}$ is comparable, albeit slightly larger. For $DUVOL_{t+1}$, I observe an economic impact of 14.77% of the sample mean, when setting $\ln(DIST_VAL_ALL_t)$ to their 25th and 75th

²⁵ Due to the very high correlation between $CUMRET_t$ and $SIGMA_t$, I omit the firm-specific stock return volatility from the regression model. Likewise, in a subsequent test, I also exclude $\ln(SIZE_t)$, $CUMRET_t$ and OWN_INST_t from my baseline model one at a time (similarly done in e.g., Callen and Fang, 2013). Untabulated results are quantitatively and qualitatively similar to those reported in Table 4. Additionally, while I find a positive and significant result for $\ln(DIST_VAL_ALL_t)$, the results may also be driven by the value-weighted distance among the top ten investors (e.g., Huang and Kang, 2017), or by the percentage held by block holders. The inclusion of these variables does not affect my results. Finally, I add the changes in institutional ownership ($\Delta OWN_INST_t = OWN_INST_t - OWN_INST_{t-1}$) to my baseline regression to alleviate concerns that the value-weighted distance is solely driven by changes of total institutional ownership in the period $t-1$. As before, the coefficients on $\ln(DIST_VAL_ALL_t)$ are positive and statistically significant at the 1% level.

²⁶ Callen and Fang (2015a) report that a raise in religiosity would decrease firms' future stock price crash risk by 6.34%, whereas Hasan et al. (2021) document that an increase in brand capital from the 25th to the 75th percentile reduces stock price crash risk by an average of 8.51%.

percentile values and holding all other variables at their mean values. Taken together, these results are economically meaningful.

I now turn the focus to the control variables of columns (3) and (6), respectively. Signs and statistical significances are mostly consistent with my expectations and the findings documented in previous literature (e.g., Kim et al., 2014; Callen and Fang, 2015a; Andreou et al., 2016; Hasan et al., 2021). For example, high $DTURN_t$, $SIGMA_t$, $CUMRET_t$, MB_t , and $ACCM_t$ increase future stock price crash risk (e.g., Kim et al., 2014; Callen and Fang, 2015a; Hasan et al., 2021). Moreover, firms with higher leverage reveal lower future crash risk (e.g., Hutton et al., 2009, Callen and Fang, 2015a, Hasan et al., 2021). Furthermore, future stock price crash risk is higher for firms with higher institutional ownership, which is in line with An and Zhang (2013). The authors argue that the impact of transient institutional investors overwhelms that of dedicated investors, which explains the positive coefficient.²⁷ Surprising is the lack of statistical significance for $\ln(SIZE_t)$ and ROA_t . It is likely that OWN_INST_t , at least partly, captures the information content inheriting size and financial performance. In an (untabulated) subsequent analysis, I test this notion by excluding OWN_INST from my baseline analysis. Results for $\ln(SIZE_t)$ turn significant with the expected coefficients, while ROA_t remains insignificant.²⁸

Overall, my results are consistent with Hypothesis 1 that firms with more distant institutional shareholders reveal higher levels of future stock price crash risk.

²⁷ When I decompose OWN_INST_t into percentage held by long- and short-term investors (Bourveau et al., 2022), I find similar results for short-term investors as An and Zhang (2013) (see, e.g., Ramalingegowda and Yu, 2012).

²⁸ For example, Kim et al. (2011a) also document an insignificant coefficient on ROA in their OLS regression results. Moreover, I additionally estimate the economic magnitude of the controls, when moving from the first quartile to the third quartile of the respective variable. Using $NSKEW_{t+j}$ as dependent variable, the economic impact of $\ln(DIST_VAL_ALL_t)$ (0.032) is lower (in absolute terms) than that of $SIGMA_t$ (0.166), OWN_INST_t (0.155), $CUMRET_t$ (0.105), and LEV_t (-0.034), while higher than the rest of the control variables. Results for $DUVOL_{t+j}$ are qualitatively similar.

4.2 Endogeneity

My results so far suggest that companies with more distant investors exhibit higher levels of future stock price crash risk. However, a potentially endogenous relation between my distance measure and future crash risk is a concern in my analysis. Endogeneity may potentially arise due to correlated omitted variables as well as reverse causality. Therefore, this section aims at providing further robust evidence on the results by establishing a potential causal link between distance and future crash risk. It is of course a challenge to consider all possible sources of endogeneity. Thus, although the results survive a battery of endogeneity tests, the findings should be interpreted cautiously.

4.2.1 Additional control variables

Like prior studies (e.g., Kim et al., 2014; Callen and Fang, 2013, 2015a and 2015b; Hasan et al., 2021), I examine the impact of additional control variables that might be correlated with both the institutional distance and the future crash risk.²⁹

First, as pointed out by Hasan et al. (2021) and shown by previous studies (e.g., Kim et al., 2011a and 2011b; Chang et al., 2017; Hasan et al., 2021; Andreou et al., 2021), bad news hoarding could be dependent on information asymmetry stemming from stock liquidity, financial reporting opacity, analyst coverage, firms' distress risk, and tax avoidance. Thus, according to this evidence, I reestimate my baseline model adding liquidity ($AMIHUD_t$), real earnings management (REM_t), the number of analysts following a firm (NUM_ANALY_t), a distance-to-default measure (DTD_t)³⁰, and a measure of tax avoidance (BTD_t ; Frank et al., 2009)

²⁹ Note that I do not include these control variables in my main regression, since they either come (heavily) at the cost of observations or are not commonly used in literature.

³⁰ Data on distance-to-default are obtained from the Credit Research Initiative (CRI, 2022), maintained by the Risk Management Institute (RMI) of the National University of Singapore. Data are available at <https://nuscri.org/en/> (CRI, 2022). Please refer to https://nuscri.org/en/white_paper/ and https://nuscri.org/en/technical_document/ for more information.

as control variables. As columns (1) and (5) of Table 5 reveal, I continue to find a positive relation between institutional distance and future crash risk with coefficients being significant at the 1% level.

[Insert Table 5 about here]

Second, controlling for soft determinants of future stock price crash risk and demographic factors, by means of e.g., an indicator capturing whether the firm is proximate, i.e. within a 250-mile radius, to a SEC office (*DUM_DIST_SEC_t*; Kubick and Lockhardt, 2016), religiosity (*REL_t*; Callen and Fang, 2015a), and the total population and the median income in a county ($\ln(TOTPOP_t)$ and $\ln(INCOME_t)$; Xu et al., 2020), does not alter my main results (see columns (2) and (6)). Next, since prior studies show that future crash risk is related to governance mechanisms, I include external and internal governance metrics (e.g., Chhaochharia et al., 2012; Andreou et al., 2016; Chang et al., 2017; Hasan et al., 2021). Specifically, I employ CEO-Chair duality (*DUAL_t*), an indicator for board independence (*INDEP_t*), the board size (*BSIZE_t*), and an indicator capturing whether the company is audited by a Big 4 company (*BIG4_t*). Additionally, to capture the overall governance performance, I include Refinitiv's governance pillar score (*GOV_t*).³¹ Columns (3) and (7) present the results. I continue to observe a statistically significant relation between shareholder proximity and crash risk, indicating that my results are not sensitive to governance factors.

Lastly, I re-run my baseline equation using all presented factors in one model. As shown in columns (4) and (8), statistical significances persist. Taken together, these results deliver a

³¹ For a detailed description of the categories and governance standards, which are included in the score, see Benz et al. (2020). Also, my results remain unchanged, when I use the overall ESG score rather than the governance score.

contribution to the establishment of a potential causal relation between crash risk and distance measures.³²

4.2.2 Entropy balancing

To ensure that my results are driven by differences in shareholder distance, not other factors, I employ the entropy balancing matching scheme introduced by Hainmueller (2012) and Hainmueller and Xu (2013) and applied in various recent studies, such as Jiang et al. (2018) and Cai et al. (2019). For this purpose, I group firms into high- or low-distant firms to their institutional shareholders by constructing an indicator variable, $HIGH_VAL_ALL_t$. This indicator takes the value of one if the firm's value-weighted distance is above the 75th percentile (=treatment group), and zero if it is below the 25th percentile (=control group).³³ Treatment and control group are matched by means of an entropy reweighting scheme based on the first moment of all covariates used in my baseline analysis. The computed weights ensure that treatment and control group are virtually identical regarding their firm characteristics.³⁴ I then employ $HIGH_VAL_ALL_t$ as the main variable of interest along with the weights estimated from the entropy balancing procedure to rerun my regression model. Additionally, I repeat this exercise by excluding extreme balancing weights (Hainmueller, 2012; Mayberry, 2020). Table 6 presents the results.

[Insert Table 6 about here]

³² A few recent studies explicitly control for CEO characteristics, such as CEO age, gender, or overconfidence (e.g., Kim et al., 2011a; Andreou et al., 2016; Hasan et al., 2021). I am not able to control for these additional factors due to data restrictions. I acknowledge the fact that this is a potential limitation of my study.

³³ A detailed description of the implementation of entropy balancing is provided in Hainmueller (2012), Hainmueller and Xu (2013), Jiang et al. (2018), and Mayberry (2020). The results are similar when I use the median value as partitioning threshold.

³⁴ Untabulated test diagnostics reveal that I observe no statistically significant difference in the means of firm characteristics after the application of entropy balancing.

I find that the coefficients on $HIGH_VAL_ALL_t$ are significant at the 1% level in all regressions, indicating that the crash risk increases for firms with distant shareholders similar in firm characteristics. I also find qualitatively similar results when matching is conducted based on higher moments, i.e., variance or skewness (untabulated to save space).

4.2.3 Firm and high dimensional fixed effects regression

Another common way in literature to reduce endogeneity concerns stemming from omitted time-invariant firm characteristics is to employ a firm fixed effects regression specification (e.g., Chhaochharia et al., 2012; Chang et al., 2017; Hasan et al., 2021). However, the estimation of a firm fixed effects regression may lose its statistical power in my setting due to the little within-firm variation of the distance measures, which could make identification difficult (e.g., Chhaochharia et al., 2012; Frijns et al., 2016).³⁵

[Insert Table 7 about here]

Despite this fact, I continue to observe a positive and statistically significant coefficient in all specifications (see Table 7, columns (1), (2), (4), and (5), respectively). In addition, closely following Hasan et al. (2021), I also estimate my baseline model using high-dimensional fixed effects.³⁶ I replace my year dummies by year×industry indicators, which allows for time-varying industry specific unobserved heterogeneity (Klasa et al., 2018; Hasan et al., 2021). Table 7, columns (3) and (6) present the results. Across all specifications, I find that the

³⁵ As noted earlier, to avoid concerns regarding biased estimates on my distance variable (Nickell, 1981), I also estimate both models without lagged dependent variable.

³⁶ I use Stata's command *reghdfe* in combination with the option *absorb(ID year#Industry FE)* for estimating the high dimensional fixed effects models.

coefficients on my distance measure are positive and statistically significant at least at the 5% level. Overall, this evidence corroborates my first hypothesis.

4.2.4 Change analysis

To further establish a potential causal link between institutional distance and future stock price crash risk, I conduct a change analysis. According to Wooldridge (2006, p. 475), “differencing panel data over time, in order to eliminate a time-constant unobserved effect, is a valuable method for obtaining causal effects.” Chhaochharia et al. (2012) point out that, although distance measures are relatively stable over time (see Section 3.6), the relative importance of local monitors can change over time, when the holdings of non-local institutions vary. Hence, a relative decrease in distance to potential monitors enhances the monitoring effect.

Following past research (e.g., Chhaochharia et al., 2012; Kim et al., 2019; Hasan et al., 2021), I regress changes in future stock price crash risk ($\Delta NSKEW_{t+1}$ and $\Delta DUVOL_{t+1}$) on changes in the value weighted distance measure ($\Delta DIST_VAL_ALL_t$), while I also include the changes of other firm characteristics (see Section 3.3).³⁷ In a subsequent analysis, borrowing the idea of Aggarwal et al. (2011), I regress changes of institutional distance ($\Delta DIST_VAL_ALL_{t+1} = DIST_VAL_ALL_{t+1} - DIST_VAL_ALL_t$) on changes of crash risk ($\Delta NSKEW_t$ or $\Delta DUVOL_t$), along with the changes in other firm characteristics. This test is intended to mitigate concerns stemming from reverse causality (similarly applied in Aggarwal et al., 2011; Chhaochharia et al., 2012).³⁸ If the direction is from institutional distance to future

³⁷ The change is defined as $\Delta NSKEW_{t+1} = NSKEW_{t+1} - NSKEW_t$. The same method is applied for $\Delta DUVOL_{t+1}$, while the change in the value-weighted distance is $\Delta DIST_VAL_ALL_t = DIST_VAL_ALL_t - DIST_VAL_ALL_{t-1}$. Likewise, the changes for the control variables are defined from t to $t-1$. Note that I also include the changes in $NSKEW_t$ or $DUVOL_t$ as control variables. However, results are robust when I omit the changes in the lagged dependent variables.

³⁸ Reverse causality in my setting is the notion that proximity between an institutional shareholder and its portfolio firm may provide the institution with information that is associated with crash risk, justifying its investment or deinvestment in a firm (Mazur et al., 2018). A first attempt to minimize such concerns is (a) the lead-lag setting, which I employ in my baseline analysis. In addition, (b) I consistently control for the lag of the

crash risk solely, lower levels of stock price crash risk would not bring institutional owners closer to the firm (Chhaochharia et al., 2012). Table 8 reports the results.

[Insert Table 8 about here]

I find that changes in institutional distance are positively and statistically significant related to changes in future stock price crash risk with coefficients being significant at the 1% level (see columns (1) and (2)). Conversely, the reverse causality test (columns (3) and (4)) reveals statistically insignificant results. Taken together, this evidence is consistent with the view that the direction is stronger from institutional distance to crash risk than the reverse potential causal link, again supporting my first hypothesis.³⁹

4.2.5 Instrumental variable regression

To further shed light on whether there is an endogenous relation between institutional distance and crash risk, I estimate a two-stage least-squares (2SLS) regression (e.g. Kim et al., 2014; Jiraporn et al., 2014; Hollander and Verriest, 2016; Dimitrescu and Zakriya, 2021; Chang et al., 2021). According to the referenced articles, I use the state-level and industry-level average of $\ln(DIST_VAL_ALL_t)$ as instruments. In particular, I compute the state-level instrument as the average of $\ln(DIST_VAL_ALL_t)$ for all other firms that are located in the same state in the same

crash risk measure to account for persistence in crash risk documented in previous studies. Following Andreou et al. (2021), (c) I additionally re-estimate my baseline model by including the three most recent lagged values of the dependent variable as additional controls, i.e., the crash risk measures in year t , $t-1$, and $t-2$, respectively. Although the sample size is reduced by including more lags, my results show that the significances of the coefficient estimates remain persistent, and more importantly, economic significances are as high as reported in Section 4.1. Thus, these findings along with the results presented in Section 4.2.4 indicate that my main inferences should not be confounded by reverse causality.

³⁹ In Table 8, the number of observations is lower in columns (1) and (2), where the dependent variables are the future stock price crash risk measures, compared to columns (3) and (4), where the dependent variable is the future distance measure. This is due to a lack of data on institutional distance before 1997, whereas crash risk measures are available in 1996. However, the results are similar when I run the regressions with the same sample size.

year but do not share the same industry. Likewise, the industry-level instrument is calculated as the average of $\ln(DIST_VAL_ALL_t)$ of all other firms in the same industry in the same year but located in a different state. Ultimately, further tests are conducted to assess the validity of these two instruments. I expect the current value of $\ln(DIST_VAL_ALL_t)$ to be significantly correlated with a state- and industry-level factor, respectively, which satisfy the relevance criterion.⁴⁰ Because the contribution of the focal firm is excluded in both cases, it is also unlikely that the state- and industry-level averages affect a firm's crash risk (= exclusion criterion). Moreover, Kubick and Lockhart (2016) find an association between future stock price crash risk and the firm's proximity to the nearest SEC office. Thus, following Ayers et al. (2011), I include the distance between the firm's headquarters and the nearest SEC office as an additional control variable ($\ln(FIN_DIST_SEC_t)$). Results are reported in Table 9.

[Insert Table 9 about here]

Columns (1) and (3) document the second-stage fitted regressions for both measures of crash risk without the additional control variable, whereas columns (2) and (4) include $\ln(FIN_DIST_SEC_t)$. Consistent with my first hypothesis, I continue to find a positive and statistically significant relation between $\ln(DIST_VAL_ALL_t)$ and stock price crash risk in columns (1) to (4). Equally important, the test diagnostics reveal the desired outcomes. First, the Kleibergen-Paap rk Wald F -statistic is higher than the critical values suggested by the Stock-Yogo test, indicating that my instruments are valid. Second, I observe p-values > 0.10 for the Hansen J -statistic, which is away from rejection of the null hypothesis, thus suggesting that the instruments are exogenous. Finally, the Hausman test (Hausman, 1978) fails to reject

⁴⁰ Untabulated results of the first-stage regressions show that both instruments are statistically significant at the 1% level.

the exogeneity of $\ln(DIST_VAL_ALL_t)$. Overall, the positive relation between institutional distance and future crash risk holds, thus further attenuating concerns regarding endogeneity.

4.2.6 Placebo test

Following Xu et al. (2020) and Chen et al. (2021), I complement my endogeneity section by conducting a placebo test. This method is intended to reduce concerns that the main findings are driven by spurious correlations related to my sample firms. As described in Xu et al. (2020), I randomly assign coordinates to each firm in my sample. Subsequently, I recompute the value-weighted distance to the location of all actual institutional investors and construct a pseudo- $DIST_VAL_ALL_t$ variable. I estimate my baseline model using the $\ln(\text{pseudo-}DIST_VAL_ALL_t)$ variable instead of the true $\ln(DIST_VAL_ALL_t)$ variable. I repeat this exercise 500 times, thus generating 500 coefficient estimates of the $\ln(\text{pseudo-}DIST_VAL_ALL_t)$ variable for each of the two crash risk measures used in my analysis. Following Chen et al. (2021), I employ these estimates and construct an empirical distribution of the pseudo variable.

[Insert Table 10 about here]

The results of this procedure (see Table 10) show that the actual coefficient estimates reported in Table 4 lie at the extreme upper tail of the empirical distribution of $\ln(\text{pseudo-}DIST_VAL_ALL_t)$ coefficients, suggesting that my baseline results are unlikely to be driven by spurious correlations (Chen et al., 2021).

4.3 Alternative proxies and additional analysis

4.3.1 Alternative definition of local monitors

Besides using $\ln(DIST_VAL_ALL_t)$ as my main independent variable, I further employ two alternative distance measures (see Section 3.2) to capture multiple facets of geographic proximity, and present the results in Table 11.

[Insert Table 11 about here]

In columns (1) and (2), I use the number of institutions located within a 100-mile radius ($PERC_LOCAL100_t$) as a proxy for geographic proximity. Chhaochharia et al. (2012) point out that this measure captures potential nonlinear effects of the distance between a firm and its institutional shareholders. Moreover, in columns (3) and (4), I regress $NSKEW_{t+1}$ and $DUVOL_{t+1}$, respectively, on the equal-weighted distance ($\ln(DIST_EQ_10_t)$) to the top ten institutional investors.

The coefficient estimates on $PERC_LOCAL100_t$ are negative and statistically significant at the 1% level, while I observe positive and statistically significant results for $\ln(DIST_EQ_10_t)$, supporting my main results. This evidence shows that using alternative measures of institutional distance does not alter my conclusions drawn in Section 4.1.

4.3.2 Alternative measures of future stock price crash risk

Following Hutton et al. (2009), Kim et al. (2011a and 2011b), and Callen and Fang (2015a and 2015b), I employ three alternative measures of firm-specific future stock price crash risk. First, I compute an indicator variable, $CRASH_3.09_t$, which is one if one or more firm-specific weekly returns fall 3.09 standard deviations below the mean weekly firm-specific return during a year,

and zero otherwise.⁴¹ Second, following Andreou et al. (2017) and Hasan et al. (2021), I consider extreme sigma ($CRASH_EXTR_t$), which is calculated as the negative of the worst deviation of firm-specific weekly returns from the average firm-specific weekly returns, scaled by the standard deviation of firm-specific weekly returns during a year. Specifically, I compute $CRASH_EXTR_t$ as follows:

$$CRASH_EXTR_{i,t} = -Min\left[\frac{W_{i,t,t} - \bar{W}_{i,t}}{\sigma_{r,t}}\right], \quad (4)$$

whereas higher values of $CRASH_EXTR_t$ indicate that a firm is more “crash prone”. Finally, I borrow the measure used by Callen and Fang (2015a and 2015b), $CRASH_COUNT_t$, which is the difference between the number of firm-specific weekly returns exceeding 3.09 standard deviations below the average firm-specific weekly return and the number of firm-specific weekly returns exceeding 3.09 standard deviations above the firm-specific weekly return during the year. Results are presented in Table 12.

[Insert Table 12 about here]

I observe that $\ln(DIST_VAL_ALL_t)$ is positively related to all alternative measures of firm-specific stock price crash risk, with coefficients being significant at least at the 5% level. Again, these results support my first hypothesis.

⁴¹ I estimate this model using logit regression technique with standard errors adjusted for heteroscedasticity and firm clustering, while Table 12 reports marginal effects. The choice of 3.09 standard deviations follows Hutton et al. (2009). The 0.1% cutoff of the normal distribution is merely used as a method to receive a reasonable reference point for an extreme event, although the weekly firm-specific returns are not normally distributed (Hutton et al., 2009). Like Hutton et al. (2009), I observe a greater frequency of crashes than the benchmark would imply.

4.3.3 Estimation of the expanded market model using daily returns

As described in Section 3.3, I regress weekly firm returns on the past, current and future weekly returns of the value-weighted market index (Wilshire 5000) to receive firm-specific weekly returns (e.g., Kim et al., 2011a and 2011b; Kim et al., 2014; Xu et al., 2021; Hasan et al., 2021; see equation (1)). However, some recent crash risk-related studies (e.g., Callen and Fang 2015a and 2015b) use daily returns as an alternative. Thus, following Hu et al. (2020a), I test the robustness of my results by using daily returns rather than weekly returns for estimating the expanded market model (see equation (1)).⁴² Subsequently, I employ the firm-specific daily returns to calculate the crash risk measures, $NSKEW_DAILY_{t+1}$ and $DUVOL_DAILY_{t+1}$, respectively. In addition, I replace my control variables $SIGMA_t$ and $CUMRET_t$, respectively, with their daily counterparts, i.e., by $SIGMA_DAILY_t$ and $CUMRET_DAILY_t$.⁴³ With these modifications, I then reestimate my baseline model (see equation (3)). Results are reported in Table 13.

[Insert Table 13 about here]

I observe that the coefficients of my distance measures continue to be positive and statistically significant at the 1% level for columns (1) to (4). Thus, employing daily returns rather than weekly returns does not alter my conclusions.

⁴² According to Hu et al. (2020a), a possible advantage of using daily rather than weekly returns stems from the fact that the non-synchronicity can be captured more rigorously. However, taking daily returns may also introduce noise to the firm-specific daily returns.

⁴³ Descriptive statistics for these additional variables are provided in Table 1, Panel B.

4.3.4 Other robustness tests

I conduct further robustness tests, which are untabulated to conserve space. Accordingly, my main findings are robust to several specifications: (a) using Russell 3000 instead of the Wilshire 5000 as the value-weighted market index; (b) omitting winsorization of all continuous variables; (c) clustering standard errors by year or by industry; (d) adding state-level fixed effects along with year and industry fixed effects; (e) estimating my baseline model with 1-digit or 4-digit SIC codes rather than 2-digit SIC codes; (f) estimating equation (3) with Fama-MacBeth procedure; (g) re-running my baseline model (equation (3)) by excluding firms that are located within 100 miles of New York City to further ensure that my results of local monitoring do not simply capture the fact that a company is headquartered in or near New York (Coval and Moskowitz, 1999; El Ghouli et al. 2013). I still observe qualitatively similar results.

4.4 Monitoring and crash risk: the role of corporate governance

Chhaochharia et al. (2012) find that firms with high local institutional ownership have better internal corporate governance mechanisms. Moreover, prior studies also show that weak governance mechanisms result in a low-quality information environment, which increases managers' incentives to hoard bad news (e.g., Ryan and Wiggins, 2004; Bae et al., 2006; Kim et al., 2011a; Hossain et al., 2022). Based on this evidence, I expect the role of local institutions as effective monitors to be more pronounced for firms with weak internal governance mechanisms. To test this hypothesis, I adopt two internal governance proxies that are used in prior literature (e.g., Kim et al., 2014; Hasan et al., 2021; Hossain et al., 2022).

[Insert Table 14 about here]

First, following Kim et al. (2014), I employ an overall governance index to proxy for the effectiveness of corporate governance. This measure is based on the Refinitiv's governance ratings score (GOV_t), which is computed as the weighted average relative rating of a company based on the reported governance information and the resulting three governance category scores (Refinitiv, 2022). Following Hasan et al. (2021), I subgroup the sample using the yearly median value of GOV_t (Table 14, Panel A). As expected, the coefficient for the low-quality internal governance group is statistically significant at the 5% level, whereas the coefficients on my distance measure are insignificant for the above-median group.

Second, previous studies document that the proportion of independent board members ($PROP_BINDEP_t$) serves as an effective proxy for the quality of firms' internal governance, since independent board members limit bad news hoarding and opportunistic behavior by managers, assess board decisions in an objective manner, and act in the best interest of shareholders (e.g., Xie et al., 2003; Nguyen and Nielsen, 2010; Hasan et al., 2021). As before, I expect that the effect of local monitors is more prominent for the below-median group, i.e., for the subsample with a low proportion of independent board members. Panel B of Table 14 confirms this notion, since the coefficients for my distance measure are always positive and statistically significant at conventional levels in the below-median group. Overall, although the differences in coefficients are statistically significant in one case only⁴⁴, the results suggest that the distance-induced monitoring effect is of particular importance when internal corporate governance is weak, which is consistent with Hypothesis 2. Weak governance probably facilitates bad news hoarding behavior by corporate managers, causing stock crashes.⁴⁵

⁴⁴ For instance, Chen et al. (2021) also document two insignificant results when analyzing the cross-sectional relation between governance and stock price crash risk.

⁴⁵ Like Chhaochharia et al. (2012), I additionally analyze whether monitoring effects of local institutions are more evident among firms with low information asymmetry, i.e., firms with low analyst coverage and small firms. In line with the findings of Chhaochharia et al. (2012), I do not find any differences between firms with high and low information asymmetry, which is inconsistent with an information-based view of my results. Also, following Hasan et al. (2021), I test various other channels, such as the role of tax avoidance or earnings management.

4.5 Effect of SOX issued in 2003 and the information advantage of local institutions

Motivated by Chhaochharia et al. (2012), I use the implementation of the SOX as a natural experiment to further shed light on the story on the governance channel in my research setting. Congress passed the SOX in July 2002 in response to several corporate scandals starting in late 2001. The restrictions that accompany the new rules are: (a) a majority of board members on a single board being independent and (b) the members of compensation, audit, and nominating committees being independent (Chhaochharia et al., 2012). Chhaochharia et al. (2012) provide evidence that the effect of local monitoring is indeed more pronounced before 2003. Thus, the new exchange regulation leads to an overall improved corporate governance, which possibly attenuates the effect of monitoring activities by local investors. To test this conjecture, I create two subsamples for the period before and after 2003, respectively, while I investigate the effect of $\ln(DIST_VAL_ALL_t)$ on future crash risk for both samples separately (similarly applied in Chhaochharia et al., 2012). Table 15 presents the results of this analysis.

[Insert Table 15 about here]

I observe that the coefficients in both subsamples are significant at the 1% level. Moreover, the test of differences in coefficients are significant at the 10% level, indicating that the exchange regulations attenuate the effect of distance, which is in line with Chhaochharia et al. (2012) and with Hypothesis 2. Overall, these findings are consistent with the conjecture that the SOX improves internal corporate governance, which again reduces bad news hoarding and

Untabulated results are also inconsistent with an explanation based on tax avoidance or earnings management, thus supporting my monitoring hypothesis.

subsequently future stock price crash risk, while the effect of local monitors persists, albeit to a lower extent than before 2003.

4.6 Bad news hoarding tests

To provide additional evidence on potential mechanisms between institutional distance and crash risk, I shed further light on the channel of bad news hoarding. As before, from my perspective, the mechanism underpinning the relationship between institutional distance and stock price crash risk is the monitoring behavior of institutional shareholders. If, however, firms have more distant shareholders, managers can more easily hoard bad news, since monitoring costs increase with distance.

First, following Chang et al. (2017) and Krishnamurti et al. (2021), I estimate the sudden release of unexpected earnings ($SURP_UE_t$). $SURP_UE_t$ is an indicator variable that takes the value of one if the firm's unexpected earnings in the current year are in the bottom tercile of the distribution and its unexpected earnings in the previous year are non-negative, and zero otherwise. Unexpected earnings are defined as the ratio of the change in income before extraordinary items over year t and $t-1$ and total assets at year $t-1$ (Kothari et al., 2006, Krishnamurti et al., 2021). As pointed out by Chang et al. (2017), the sudden release of very bad news is not expected based on the previous year's performance disclosure, which is surprising for the market, causing stock price crashes.

[Insert Table 16 about here]

Second, I estimate two further variables, $C_STR_BREAK1_t$ and $C_STR_BREAK2_t$, following Andreou et al. (2017) and Krishnamurti et al. (2021). The idea behind these variables is to identify longer strings of consecutive earnings increases, which may result from bad news

hoarding (Myers et al., 2007; Hutton et al., 2009). Thus, a break in strings that causes a stock crash is another channel to investigate as to whether this mechanism is more pronounced for firms with more distant shareholders. In particular, $C_STR_BREAK1_t$ is an indicator variable that is one if a firm experiences a stock price crash in the current year⁴⁶ and a reduction of earnings in the current year, but an increase in earnings in the previous year, and zero otherwise. The construction of $C_STR_BREAK2_t$ follows the same approach, though the firm reports a consecutive increase in earnings during the last two years (Andreou et al., 2017; Krishnamurti et al., 2021). Table 16 reports the results for the tests of bad news hoarding.⁴⁷

I observe statistically significantly positive coefficients in columns (1) to (3), supporting my expectations (Hypothesis 3) that firms with more distant shareholders are more likely to hoard bad news due to the increasing costs of monitoring that comes along with distance. In other words, withholding bad information by corporate managers is attenuated when a firm is surrounded by more local institutions.

4.7 Do all proximate investors monitor?

So far, my analysis implicitly assumes that all institutional investors are potential monitors. However, previous studies document heterogeneity in investors' engagement in monitoring. For example, non-transient (dedicated/quasi-indexer)⁴⁸ investors are more likely to monitor corporate management (e.g., Callen and Fang, 2013; Chang et al., 2017; Huang and Kang, 2017). This is congruent with the view that increasingly large positions held by index funds result in strong incentives to monitor corporate behavior. However, a recent study by Heath et al. (2022) investigates the monitoring behavior of passively managed index funds. Although

⁴⁶ To identify stock price crashes, I employ $CRASH_3.09$, as the corresponding variable.

⁴⁷ Again, I estimate these models using logit regression technique with standard errors adjusted for heteroscedasticity and firm clustering, while Table 16 reports marginal effects.

⁴⁸ This classification scheme follows Bushee (1998), Bushee and Noe (2000), and Bushee (2001).

index funds pursue long-term investments with low turnover, the authors provide comprehensive and robust evidence that, compared to actively managed funds, index funds are less effective monitors. This is likely due to weakened corporate governance in firms following an increase in index fund holdings. Also, Heath et al. (2022) argue that, due to their low-cost structure, passively managed index funds have limited resources to invest in monitoring. Based on these empirical insights, I further examine as to whether the type of institutional shareholders matters for the relation between institutional distance and future crash risk.

Empirically, I divide the sample into two groups (see Section 3.7). The first group of investors contains all passively-oriented institutions (investor sub-type “investment advisor” or “investment advisor/hedge fund”)⁴⁹ with investment style “index” and orientation “passive” in the Refinitiv Ownership Profile. Based on the findings of Heath et al. (2022), I conjecture that these shareholders are involved in monitoring activities to a lower extent (“passive”). I further classify all other investors as potential “active” monitoring institutions. Subsequently, I construct the value-weighted distance measures for “active” ($DIST_VAL_ALL_exInd_t$) and “passive” ($DIST_VAL_ALL_Ind_t$) institutional shareholders separately, in the same way as described in Section 3.2. I then re-estimate my baseline regressions using the revised distance measures. Table 17 reports the regression results.⁵⁰

[Insert Table 17 about here]

Columns (1) and (4) show that the coefficients on the distance measures are positive and statistically significant at the 1% level, indicating that firms with more distant “active”

⁴⁹ For example, “BlackRock Institutional Trust Company” is classified as “investment advisor” with investment style “Index” and orientation “passive”, whereas another big player, “The Vanguard Group, Inc.”, is categorized as “investment advisor/hedge fund” with investment style “Index” and orientation “passive”.

⁵⁰ Following Heath et al. (2022), I also restrict the sample period to 2004 through 2019, because this period mainly captures the dramatic increase in index investing. My results are robust to this specification.

shareholders exhibit higher levels of future stock price crash risk. In columns (2) and (5), I conduct this analysis for the distance measures using the value-weighted distance to “passive” institutional investors only. I observe that the coefficients on the distance measures are insignificant. These results also hold, when I combine the distance measures of active and passive shareholders in one regression model (see columns (3) and (6)). Moreover, in columns (3) and (6), Wald-Tests show that the coefficients are statistically significantly different between the investor subgroups.

Taken together, these results are consistent with Heath et al. (2022) and my fourth hypothesis, supporting the conjecture that geographically proximate “active” shareholders (i.e., institutions that are not classified as passively managed index funds) with substantial ownership have stronger incentives to monitor rather than geographically concentrated “passive” investors. Ultimately, stronger monitoring of proximate “active” institutions reduces future stock price crash risk.

5 Conclusion

I study the relationship between geographic proximity and firm-specific stock price crash risk for a broad U.S. sample covering the period from 1997 through 2019. Collectively, I find strong support that firms with more distant shareholders are more prone to crash risk. Importantly, I also document that my main results are not driven by endogeneity problems. The relation between institutional distance and crash risk remains statistically significant, even if I employ alternative measures of institutional distance and crash risk, crash risk measures based on daily data, and alternative regression specifications. The cross-sectional analysis reveals that the main effect may be more distinct for firms with weak internal governance structures. Moreover, although the introduction of the SOX in 2002 attenuates the effect of local monitoring, it still

persists from 2003 onwards. My tests of bad news hoarding confirm the notion that firms with more distant institutional shareholders are more likely to hoard bad news, which increases future stock price crash risk. Finally, the findings of the heterogeneity of investor types reveal that institutional investors, which are not classified as index-oriented shareholders tend to monitor corporate behavior, thus having an impact on the distance-crash risk relation.

My study extends a growing strand of research examining the influence of geographic proximity on corporate outcomes and practices. It also elaborates on the literature examining determinants of stock price crash risk by documenting the importance of physical distance in the formation of stock price crashes. Furthermore, it provides new evidence on how geographic proximity to institutional investors affects firm-specific crash risk when considering the heterogeneity of institutional investors. Overall, potential implications of my findings arise for (equity) investors and regulators. Since the monitoring activity also depends on the location of institutional investors, it possibly offers an additional perspective in predicting the likelihood of bad news hoarding, and thus of futures crashes (e.g., Callen and Fang, 2013).

Appendix

Table A1 : Variable definition

Variable	Definition
<i>Panel A: Main dependent variable</i>	
<i>DUVOL</i>	A measure of stock price crash risk, which is calculated as the natural logarithm of the standard deviation of firm-specific weekly returns for the “down-week” sample divided by the standard deviation of firm-specific weekly returns for the “up-week” sample over the year. Source: Refinitiv Datastream.
<i>NSKEW</i>	A measure of stock price crash risk, which is the negative coefficient of skewness of firm-specific weekly returns over the year. Source: Refinitiv Datastream.
<i>Panel B: Variable of interest</i>	
$\ln(DIST_VAL_ALL)$	The natural logarithm of the value-weighted distance between a firm and its institutional shareholders, where the weights are determined by the percentage held by the institution in the firm. Based on the holdings data, I identify all institutions that have a stake in a particular firm. Zip codes of US-headquartered firms and institutions are then converted to latitude and longitude using the Gazetteer File provided by the U.S. Census Bureau. Data for missing or non-US-headquartered institutions are retrieved manually via internet searching. I then calculate the distance between a portfolio firm and an institution employing the great distance formula (Stata package: <i>geodist</i> ; Picard, 2010) (e.g., Coval and Moskowitz, 1999; Chhaochharia et al., 2012). Source: Refinitiv Ownership Profile and EDGAR file.
<i>Panel C: Firm controls</i>	
<i>ACCM</i>	A measure of discretionary accruals. Total accruals are calculated as the change in total current assets

(Worldscope item: 02201), minus the change in cash and cash equivalents (Worldscope item: 02001), minus the change in total current liabilities (Worldscope item: 03101), plus the change in short-term debt included in current liabilities (Worldscope item: 03051), minus depreciation and amortization expenses (Worldscope item: 01151), divided by total assets (Worldscope item: 02999). (Kothari et al., 2005; Fernandes and Ferreira, 2008; Dechow and Dichev, 2002). Using the modified Jones model (formula (6) in Kothari et al., 2005, p. 173), I then receive the discretionary accruals. Finally, following Callen and Fang (2015a and 2015b), I calculate the three-year moving sum of the absolute value of discretionary accruals using years $t-2$ to t to receive ACCM. Source: Refinitiv Datastream.

CUMRET

The cumulative firm-specific weekly returns over the year. Source: Refinitiv Datastream.

DTURN

The detrended trading volume, which is calculated as the average share turnover (Refinitiv Datastream item: *VO*) in each month in year t minus the average monthly share turnover over the previous year ($t-1$). Monthly share turnover is defined as the monthly share trading volume divided by the monthly number of shares outstanding (Refinitiv Datastream item: *NOSH*). As per Ferreira and Matos (2008), I adjust the number of shares outstanding by an adjustment factor (Refinitiv Datastream item: *AF*). Finally, following Gao and Ritter (2010), I adjust NASDAQ trading volume prior to 2004. Source: Refinitiv Datastream.

LEV

The total debt (Worldscope item: 03255) divided by total assets (Worldscope item: 02999). Source: Refinitiv Datastream.

MB

The market value of equity (Worldscope item: 08001) divided by the book value of equity (Worldscope item: 03501). Source: Refinitiv Datastream.

OWN_INST

Institutional ownership by all institutions (13(f) filings) as a percentage of shares outstanding. All

values above 100% are set to 100% (e.g., Lewellen, 2011). Source: Refinitiv Ownership Profile.

ROA The net income before extraordinary items (Worldscope item: 01551) divided by total assets (Worldscope item: 02999). Source: Refinitiv Datastream.

$\ln(\text{SIZE})$ The natural logarithm of a firm's total assets (Worldscope item: 02999). Source: Refinitiv Datastream.

SIGMA The standard deviation of firm-specific weekly return during the year. Source: Refinitiv Datastream.

Panel D: Further variables used in auxiliary analyses

AMIHUD A stock illiquidity measure proposed by Amihud (2002). It is the absolute daily return calculated from adjusted daily prices (Refinitiv Datastream item: *P*) scaled by the adjusted daily dollar volume (Refinitiv Datastream items $P*VO/(NOSH/AF)$) and multiplied by 1,000,000. The average of daily figures is calculated during a calendar year. Source: Refinitiv Datastream.

BIG4 Indicator variable that is one, if a firm is audited by one of the Big 4 audit companies, zero otherwise (Worldscope item: 07800). Source: Refinitiv Datastream.

BOARD_INDEP Indicator variable, which equals one if the company has a policy regarding the independence of its board, zero otherwise (Refinitiv Datastream item *CGBSDP0012*). Source: Refinitiv Datastream.

BOARD_SIZE *BFSIZE* is number of board members at the end of the fiscal year (Refinitiv Datastream item *CGBSDP060*). Source: Refinitiv Datastream.

BTD A measure of tax avoidance following Frank et al. (2009). It is calculated as the pre-tax income (Worldscope item: 01401) minus income taxes (Worldscope item: 01451) divided by the statutory tax rate in the U.S., divided by lagged total assets (Worldscope item: 02999). Statutory tax rate is 35%

until 2017, while it decreases to 21% from 2018-2020 (<https://www.taxpolicycenter.org/statistics/corporate-top-tax-rate-and-bracket>). Source: Refinitiv Datastream.

C_STR_BREAK1

An indicator variable that takes the value of one when a firm experiences stock price crash risk and firm earnings decreased in the current year but increased in the previous year, zero otherwise (Andreou et al., 2017). Source: Refinitiv Datastream.

C_STR_BREAK2

An indicator variable that takes the value of one when a firm experiences stock price crash risk and firm earnings decreased in the current year but increased in the previous two years, zero otherwise (Andreou et al., 2017). Source: Refinitiv Datastream.

CEO_DUALITY

Indicator variable, which equals one if the CEO simultaneously chairs the board or the chairman has been the CEO of the company, zero otherwise (Refinitiv Datastream item *CGBSO09V*). Source: Refinitiv Datastream

CRASH_COUNT

The difference between the number of firm-specific weekly returns exceeding 3.09 standard deviations below the mean firm-specific weekly returns over the year and the number of firm-specific weekly returns exceeding 3.09 standard deviations above the mean firm-specific weekly returns over the year. The choice of 3.09 relates to the frequencies of 0.1% in the normal distribution. Source: Refinitiv Datastream.

CRASH_D

An indicator variable that is one for a firm-year with at least one crash week, and zero otherwise. Crash weeks are defined as those weeks where firm-specific weekly returns are 3.09 standard deviations below the mean firm-specific weekly return over the year. 3.09 is chosen to generate frequencies of 0.1% in the normal distribution (Hutton et al., 2009; Callen and Fang, 2015b). Source: Refinitiv Datastream.

CRASH_EXTR

Negative of the worst deviation of firm-specific weekly returns from the mean firm-specific weekly return scaled by the standard deviation of firm-specific weekly returns. Source: Refinitiv Datastream.

<i>CUMRET_DAILY</i>	The cumulative firm-specific daily returns over the year. Source: Refinitiv Datastream.
$\ln(\text{DEMO_TOTPOP})$	The natural logarithm of the total population in a county. Total population is determined by linear interpolation between the survey years. Source: U.S. Census Bureau.
$\ln(\text{DIST_EQ_10})$	The equal-weighted distance to the ten largest shareholders of a firm. Source: Refinitiv Ownership Profile and EDGAR file.
$\ln(\text{DIST_VAL_ALL_exInd})$	The natural logarithm of the value-weighted distance between a firm and its institutional shareholders, where the weights are determined by the stake of the institution in the firm. For this variable, I exclude all institutional investors that are classified as investor sub-type “investment advisor” or “investment advisor/hedge fund” with investment style “index” and orientation “passive” in the Refinitiv Ownership Profile. Source: Refinitiv Ownership Profile and EDGAR file.
$\ln(\text{DIST_VAL_ALL_Ind})$	The natural logarithm of the value-weighted distance between a firm and its institutional shareholders, where the weights are determined by the stake of the institution in the firm. For this variable, I consider institutional investors that are classified as investor sub-type “investment advisor” or “investment advisor/hedge fund” with investment style “index” and orientation “passive” in the Refinitiv Ownership Profile. Source: Refinitiv Ownership Profile and EDGAR file.
<i>DTD</i>	A measure of credit risk. It is the distance-to-default, which measures the distance between the default point and the expected value of a firm’s assets. Source: Risk Management Institute at the National University of Singapore (CRI, 2020).
<i>DUM_DIST_SEC</i>	An indicator variable that is one for companies that are located within a 250-mile radius to the nearest SEC office, zero otherwise. Source: Refinitiv Ownership Profile and https://www.sec.gov/page/sec-regional-offices .

<i>DUVOL_DAILY</i>	A measure of stock price crash risk, which is calculated as the natural logarithm of the standard deviation of firm-specific daily returns for the “down-week” sample divided by the standard deviation of firm-specific daily returns for the “up-week” sample over the year. Source: Refinitiv Datastream.
<i>GOV</i>	Weighted average relative rating of a company based on the reported governance information and the resulting three governance category scores (Refinitiv Datastream item <i>CGSCORE</i>). Source: Refinitiv Datastream.
$\ln(\text{FIN_DIST_SEC})$	The natural logarithm of the distance between a firm and the closest SEC office. Source: Refinitiv Ownership Profile and https://www.sec.gov/page/sec-regional-offices .
$\ln(\text{INCOME})$	The natural logarithm of the per-capita income in a county. Income is determined by linear interpolation between the survey years. Source: U.S. Census Bureau.
<i>NSKEW_DAILY</i>	A measure of stock price crash risk, which is the negative coefficient of skewness of firm-specific daily returns over the year. Source: Refinitiv Datastream.
<i>NUM_ANALY</i>	The number of analysts following a firm during the calendar year (Refinitiv Datastream item: <i>EPSINE</i>). Source: Refinitiv Datastream.
<i>PROP_BINDEP</i>	Percentage of independent board members (Refinitiv Datastream item <i>CGBS07V</i>). Source: Refinitiv Datastream.
<i>PROP_LOCAL100</i>	Proportion of total institutional ownership by institutional investors located within a 100-mile radius of a firm. Source: Refinitiv Ownership Profile and EDGAR file.
<i>REL</i>	Number of adherents in the county divided by the total population in the county. Source: ARDA.
<i>REM</i>	The three-year moving sum of abnormal cash flow from operations (multiplied by negative one) and abnormal production costs. Abnormal cash flow from

operations and abnormal production costs are calculated following Roychowdhury (2006). Source: Refinitiv Datastream.

SIGMA_DAILY

The standard deviation of firm-specific daily returns during the year. Source: Refinitiv Datastream.

SURP_UE

Indicator variable equal to one if the unexpected earnings are in the lowest decile for the current year and non-negative in the previous year, zero otherwise (Chang et al., 2017). Source: Refinitiv Datastream.

References

- Aggarwal, R., Erel, I., Ferreira, M., Matos, P. (2011). Does governance travel around the world? Evidence from institutional investors. *Journal of Financial Economics* 100(1), 154-181. <https://doi.org/10.1016/j.jfineco.2010.10.018>
- Al Mamun, M., Balanchandran, B., Nhan Duong, H., (2020). Powerful CEOs and stock price crash risk. *Journal of Corporate Finance* 62, 101582. <https://doi.org/10.1016/j.jcorpfin.2020.101582>
- An, H., & Zhang, T. (2013). Stock price synchronicity, crash risk, and institutional investors. *Journal of Corporate Finance* 21, 1-15. <https://doi.org/10.1016/j.jcorpfin.2013.01.001>
- An, Z., Chen, C., Naiker, V., Wang, J. (2020). Does media coverage deter firms from withholding bad news? Evidence from stock price crash risk. *Journal of Corporate Finance* 64, 101664. <https://doi.org/10.1016/j.jcorpfin.2020.101664>
- Andreou, P. C., Antoniou, C., Horton, J., Louca, C. (2016). Corporate Governance and Firm-specific Stock Price Crashes. *European Financial Management* 22(5), 916-956. <https://doi.org/10.1111/eufm.12084>
- Andreou, P. C., Louca, C., Petrou, A. P. (2017). CEO Age and Stock Price Crash Risk. *Review of Finance* 21(3), 1287-1325. <https://doi.org/10.1093/rof/rfw056>
- Andreou, C. K., Andreou, P. C., Lambertides, N. (2021). Financial distress risk and stock price crashes. *Journal of Corporate Finance* 67, 101870. <https://doi.org/10.1016/j.jcorpfin.2020.101870>
- Ayers, B. C., Ramalingegowda, S., Yeung, P. E. (2011). Hometown advantage: The effects of monitoring institution location on financial reporting discretion. *Journal of Accounting and Economics* 52, 41-61. <https://doi.org/10.1016/j.jacceco.2011.03.003>
- Bae, K.-H., Lim, C., Wei, K. C. J. (2006). Corporate Governance and Conditional Skewness in the World's Stock Markets. *Journal of Business* 79(6), 2999-3028. <https://doi.org/10.1086/508006>
- Baik, B., Kang, J.-K., Kim, J.-M. (2010). Local institutional investors, information asymmetry, and equity returns. *Journal of Financial Economics* 97(1), 81-160. <https://doi.org/10.1016/j.jfineco.2010.03.006>

- Balachandran, B., Duong, H. N., Luong, H., Nguyen, L. (2020). Does takeover activity affect stock price crash risk? Evidence from international M&A laws. *Journal of Corporate Finance* 64, 101697. <https://doi.org/10.1016/j.jcorpfin.2020.101697>
- Baltagi, B. H. (2021). *Econometric Analysis of Panel Data*. Springer Texts in Business and Economics, Cham, Switzerland, Sixth Edition. <https://doi.org/10.1007/978-3-030-53953-5>
- Benz, L., Paulus, S., Rohleder, M., Wilkens, M. (2020). Ownership comes with responsibility – The impact of ownership characteristics on CSR. Working Paper, University of Augsburg.
- Bernile, G., Kumar, A., Sulaeman, J. (2015). Home away from Home: Geography of information and Local Investors. *Review of Financial Studies* 28(7), 2009-2049. <https://doi.org/10.1093/rfs/hhv004>
- Bourveau, T., Brochet, F., Garel, A. (2022). The Capital Market Consequences of Tenure-Based Voting Rights: Evidence from Florange Act. *Management Science* 0(0). <https://doi.org/10.1287/mnsc.2022.4320>
- Bushee, B. J. (1998). The influence of institutional investors on myopic R&D investment behavior. *Accounting Review* 73(3), 305-333.
- Bushee, B. J., & Noe, C. (2000). Corporate Disclosure Practices, Institutional Investors, and Stock Return Volatility. *Journal of Accounting Research* 38, 171-202. <http://dx.doi.org/10.2307/2672914>
- Bushee, B. J. (2001). Do Institutional Investors Prefer Near-Term Earnings over Long-Run Value? *Contemporary Accounting Research* 18(2), 207-246. <https://doi.org/10.1506/J4GU-BHWH-8HME-LE0X>
- Cai, J., Kim, Y., Li, S., Pan, C. (2019). Tone at the top: CEOs' religious beliefs and earnings management. *Journal of Banking & Finance* 106, 195-213. <https://doi.org/10.1016/j.jbankfin.2019.06.002>
- Callen, J. L., & Fang, X. (2013). Institutional investor stability and crash risk: Monitoring versus short-termism? *Journal of Banking & Finance* 37(8), 3047-3063. <https://doi.org/10.1016/j.jbankfin.2013.02.018>
- Callen, J. L., & Fang, X. (2015a). Religion and Stock Price Crash Risk. *Journal of Financial and Quantitative Analysis* 50(1-2), 169-195. <https://doi.org/10.1017/S0022109015000046>
- Callen, J. L., & Fang, X. (2015b). Short interest and stock price crash risk. *Journal of Banking & Finance* 60, 181-194. <https://doi.org/10.1016/j.jbankfin.2015.08.009>

- Callen, J. L., & Fang, X. (2017). Crash Risk and the Auditor-Client Relationship. *Contemporary Accounting Research* 34(3), 1715-1750. <https://doi.org/10.1111/1911-3846.12311>
- Chang, X., Chen, Y., Zolotoy, L. (2017). Stock Liquidity and Stock Price Crash Risk. *Journal of Financial and Quantitative Finance* 52(4), 1065-1637. <https://doi.org/10.1017/S0022109017000473>
- Chang, K., Kabongo, J., Li, Y. (2021). Geographic proximity, long-term institutional ownership, and corporate social responsibility. *Review of Quantitative Finance and Accounting* 56, 297-328. <https://doi.org/10.1007/s11156-020-00895-9>
- Chhaochharia, V., Kumar, A., Niessen-Ruenzi, A. (2012). Local investors and corporate governance. *Journal of Accounting and Economics* 54(1), 42-67. <https://doi.org/10.1016/j.jacceco.2012.03.002>
- Chen, J., Hong, H., Stein, J. C. (2001). Forecasting crashes: trading volume, past returns, and conditional skewness in stock prices. *Journal of Financial Economics* 61(3), 345-381. [https://doi.org/10.1016/S0304-405X\(01\)00066-6](https://doi.org/10.1016/S0304-405X(01)00066-6)
- Chen, Y., Fan, Q., Yang, X., Zolotoy, L. (2021). CEO early-life disaster experience and stock price crash risk. *Journal of Corporate Finance* 68, 101928. <https://doi.org/10.1016/j.jcorpfin.2021.101928>
- Cohen, D. A., Dey, A., Lys, T. Z. (2008). Real and Accrual-Based Earnings Management in the Pre- and Post-Sarbanes-Oxley Periods. *Accounting Review* 83(3), 757-787. <https://doi.org/10.2308/accr.2008.83.3.757>
- Coval, J., & Moskowitz, T. (1999). Home bias at home: local equity preference in domestic portfolios. *Journal of Finance* 54(6), 2045-2073. <https://doi.org/10.1111/0022-1082.00181>
- CRI (2022). Credit research initiative of the national university of Singapore. <https://nuscri.org/en/>.
- Dang, V. A., Lee, E., Liu, Y., Zeng, C. (2018). Corporate debt maturity and stock price crash risk. *European Financial Management* 24, 451-484. <https://doi.org/10.1111/eufm.12134>
- Dang, T. L., Faff, R., Luong, H., Nguyen, L. (2019). Individualistic cultures and crash risk. *European Financial Management* 25(3), 622-654. <https://doi.org/10.1111/eufm.12180>
- Dechow, P. M., & Dichev, I. D. (2002). The Quality of Accruals and Earnings: The Role of Accrual Estimation Errors. *Accounting Review* 77, 35-59. <https://doi.org/10.2308/accr.2002.77.s-1.35>

- Degryse, H., & Ongena, S. (2005). Distance, lending relationships, and competition. *Journal of Finance* 60(1), 231-266. <https://doi.org/10.1111/j.1540-6261.2005.00729.x>
- Dimson, E. (1979). Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics* 7, 197-226. [https://doi.org/10.1016/0304-405X\(79\)90013-8](https://doi.org/10.1016/0304-405X(79)90013-8)
- Dumitrescu, A., & Zakriya, M. (2021). Stockholdings and the stock price crash risk: What matters in corporate social performance. *Journal of Corporate Finance* 67, 101871. <https://doi.org/10.1016/j.jcorpfin.2020.101871>
- El Ghouli, S., Guedhami, O., Ni, Y., Pittman, J., Saadi, S. (2013). Does Information Asymmetry Matter to Equity Pricing? Evidence from Firms' Geographic Location. *Contemporary Accounting Research* 30(1), 140-181. <https://doi.org/10.1111/j.1911-3846.2011.01147.x>
- Engelberg, J. E., & Parsons, C. A. (2011). The Causal Impact of Media in Financial Markets. *Journal of Finance* 66(1), 67-97. <https://doi.org/10.1111/j.1540-6261.2010.01626.x>
- Fernandes, N., & Ferreira, M. A. (2008). Does international cross-listing improve the information environment. *Journal of Financial Economics* 88(2), 216-244. <https://doi.org/10.1016/j.jfineco.2007.06.002>
- Ferreira, M. A., & Matos, P. (2008). The colors of investors' money: The role of institutional investors around the world. *Journal of Financial Economics* 88, 499-533. <https://doi.org/10.1016/j.jfineco.2007.07.003>
- Flannery, M. J., & Hankins, K. W. (2013). Estimating dynamic panel models in corporate finance. *Journal of Corporate Finance* 19, 1-19. <https://doi.org/10.1016/j.jcorpfin.2012.09.004>
- Frank, M. M., Lynch, L. J., Rego, S. O. (2009). Tax Reporting Aggressiveness and Its Relation to Aggressive Financial Reporting. *Accounting Review* 84(2), 467-496. <https://doi.org/10.2308/accr.2009.84.2.467>
- Frijns, B., Dodd, O., Cimerova, H. (2016). The impact of cultural diversity in corporate boards on firm performance. *Journal of Corporate Finance* 41, 521-541. <http://dx.doi.org/10.1016/j.jcorpfin.2016.07.014>
- Gao, X., & Ritter, J. R. (2010). The marketing of seasoned equity offerings. *Journal of Financial Economics* 97(1), 33-52. <https://doi.org/10.1016/j.jfineco.2010.03.007>
- García, D., & Norli, O. (2012). Geographic dispersion and stock returns. *Journal of Financial Economics* 106, 547-565. <http://dx.doi.org/10.1016/j.jfineco.2012.06.007>

- Gaspar, J.-M., & Massa, M. (2007). Local ownership as private information: Evidence on the monitoring-liquidity trade-off. *Journal of Financial Economics* 83(3), 751-792. <https://doi.org/10.1016/j.jfineco.2005.10.010>
- Habib, A., Hasan, M. M., Jiang, H. (2018). Stock price crash risk: review of the empirical literature. *Accounting and Finance* 58(51), 211-251. <https://doi.org/10.1111/acfi.12278>
- Hainmueller, J. (2012). Entropy Balancing for Causal Effects. A Multivariate Reweighting Method to Produce Balanced Samples in Observation Studies. *Political Analysis* 20, 25-46. <https://doi.org/10.1093/pan/mpr025>
- Hainmueller, J., & Xu, Y. (2013). ebalance: A Stata Package for Entropy Balancing. *Journal of Statistical Software* 54(7), 1-18. <https://doi.org/10.18637/jss.v054.i07>
- Hasan, I., Hoi, C.-K., Wu, Q., Zhang, H. (2017). Social Capital and Debt Contracting: Evidence from Bank Loans and Public Bonds. *Journal of Financial and Quantitative Analysis* 52(3), 1017-1047. <https://doi.org/10.1017/S0022109017000205>
- Hasan, M. M., Taylor, G., Richardson, G. (2021). Brand Capital and Stock Price Crash Risk. *Management Science* 0(0). <https://doi.org/10.1287/mnsc.2021.4197>
- Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica* 46(6), 1251-1271. <https://doi.org/10.2307/1913827>
- He, G. (2014). The effect of CEO inside debt holdings on financial reporting quality. *Review of Accounting Studies* 20, 501-536. <https://doi.org/10.1007/s11142-014-9305-8>
- Heath, D., Macciocchi, D., Michaely, R., Ringgenberg, M. C. (2022). Do Index Funds Monitor? *Review of Financial Studies* 35(1), 91-131. <https://doi.org/10.1093/rfs/hhab023>
- Hilary, G., & Hui, K. W. (2009). Does religion matter in corporate decision making in America? *Journal of Financial Economics* 93(3), 455-473. <https://doi.org/10.1016/j.jfineco.2008.10.001>
- Hollander, S., & Verriest, A. (2016). Bridging the gap: the design of bank loan contracts and distance. *Journal of Financial Economics* 119(2), 399-419. <https://doi.org/10.1016/j.jfineco.2015.09.006>
- Hu, J. (B.), Li, X., Duncan, K., Xu, J. (J.) (2020a). Corporate relationship spending and stock price crash risk: Evidence from China's anti-corruption campaign. *Journal of Banking & Finance* 113, 105758. <https://doi.org/10.1016/j.jbankfin.2020.105758>

- Hu, J., Li, S., Taboado, A. G., Zhang, F. (2020b). Corporate board reforms around the world and stock price crash risk. *Journal of Corporate Finance* 62, 101557. <https://doi.org/10.1016/j.jcorpfin.2020.101557>
- Huang, X., & Kang, J.-K. (2017). Geographic concentration of institutions, corporate governance, and firm value. *Journal of Corporate Finance* 47, 191-218. <https://doi.org/10.1016/j.jcorpfin.2017.09.016>
- Hutton, A. P., Marcus, A. J., Tehranian, H. (2009). Opaque financial reports, R², and crash risk. *Journal of Financial Economics* 94(1), 67-86. <https://doi.org/10.1016/j.jfineco.2008.10.003>
- Hossain, M. M., Mammadov, B., Vakilzadeh, H. (2022). Wisdom of the crowd and stock price crash risk: evidence from social media. *Review of Quantitative Finance and Accounting* 58, 709-742. <https://doi.org/10.1007/s11156-021-01007-x>
- Ince, O. S., & Porter, R. B. (2006). Individual equity return data from Thomson Datastream: Handle with care! *The Journal of Financial Research* 29(4), 463-479. <https://doi.org/10.1111/j.1475-6803.2006.00189.x>
- Jiang, F., John, K., Wei Li C., Qian, Y. (2018). Earthly Reward to the Religious: Religiosity and the Costs of Public and Private Debt. *Journal of Financial and Quantitative Analysts* 53(5), 2131-2160. <https://doi.org/10.1017/S002210901800039X>
- Jin, L., & Myers, S. C. (2006). R² around the world: New theory and new tests. *Journal of Financial Economics* 76(2), 257-292. <https://doi.org/10.1016/j.jfineco.2004.11.003>
- Jiraporn, P., Jiraporn, N., Boepsert, A., Chang, K. (2014). Does Corporate Social Responsibility (CSR) Improve Credit Ratings? Evidence from Geographic Identification. *Financial Management* 43(3), 505-531. <https://doi.org/10.1111/fima.12044>
- Kang, J.-K., & Kim, J.-M. (2008). The Geography of Block Acquisitions. *Journal of Finance* 63(6), 2817-2858. <https://doi.org/10.1111/j.1540-6261.2008.01414.x>
- Kim, J.-B., Li, Y., Z., Zhang, L. (2011a). Corporate tax avoidance and stock price crash risk: Firm-level analysis. *Journal of Financial Economics* 100(3), 639-662. <https://doi.org/10.1016/j.jfineco.2010.07.007>
- Kim, J.-B., Li, Y., Z., Zhang, L. (2011b). CFO vs. CEO: equity incentives and crashes. *Journal of Financial Economics* 101(3), 713-730. <https://doi.org/10.1016/j.jfineco.2011.03.013>
- Kim, Y., Li, H., Li, S. (2014). Corporate social responsibility and stock price crash risk. *Journal of Banking & Finance* 43, 1-13. <http://dx.doi.org/10.1016/j.jbankfin.2014.02.013>

- Kim, J.-B., & Zhang, L. (2016). Accounting Conservatism and Stock Price Crash Risk: Firm-level Evidence. *Contemporary Accounting Research* 33(1), 412-441. <https://doi.org/10.1111/1911-3846.12112>
- Kim, J.-B., Wang, Z., Zhang, L. (2016). CEO Overconfidence and Stock Price Crash Risk. *Contemporary Accounting Research* 33(4), 1720-1749. <https://doi.org/10.1111/1911-3846.12217>
- Kim, C. (F.), Wang, K., Zhang, L. (2019). Readability of 10-K Reports and Stock Price Crash Risk. *Contemporary Accounting Research* 36(2), 1184-1216. <https://doi.org/10.1111/1911-3846.12452>
- Klasa, S., Ortiz-Molina, H., Serfling, M., Srinivasan, S. (2018). Protection of trade secrets and capital structure decisions. *Journal of Financial Economics* 128(2), 266-286. <https://doi.org/10.1016/j.jfineco.2018.02.008>
- Kothari, S. P., Leone, A. J., Wasley, C. E. (2005). Performance matched discretionary accrual measures. *Journal of Accounting and Economics* 39(1), 163-197. <https://doi.org/10.1016/j.jacceco.2004.11.002>
- Kothari, S. P., Shu, S., Wysocki, P. D. (2009). Do Managers Withhold Bad News? *Journal of Accounting Research* 47(1), 241-276. <https://doi.org/10.1111/j.1475-679X.2008.00318.x>
- Krishnamurti, C., Chowdhury, H., Han, H. D. (2021). CEO centrality and stock price crash risk. *Journal of Behavioral and Experimental Finance* 31, 100551. <https://doi.org/10.1016/j.jbef.2021.100551>
- Kubick, T. R., & Lockhart, G. B. (2016). Proximity to the SEC and Stock Price Crash Risk. *Financial Management* 45(2), 341-367. <https://doi.org/10.1111/fima.12122>
- Landis, C., & Skouras, S. (2021). Guidelines for asset pricing research using international equity data from Thomson Reuters Datastream. *Journal of Banking & Finance* 130, forthcoming. <https://doi.org/10.1016/j.jbankfin.2021.106128>
- Lee, W., & Wang, L. (2017). Do political connections affect stock price crash risk? Firm-level evidence from China. *Review of Quantitative Finance and Accounting* 48, 643-676. <https://doi.org/10.1007/s11156-016-0563-3>
- Lewellen, J. (2011). Institutional investors and the limits of arbitrage. *Journal of Financial Economics* 102, 62-80. <https://doi.org/10.1016/j.jfineco.2011.05.012>
- Li, X., Wang, S. S., Wang, X. (2017). Trust and stock price crash risk: Evidence from China. *Journal of Banking & Finance* 76, 74-91. <https://doi.org/10.1016/j.jbankfin.2016.12.003>

- Lin, C.-C., & Png, I. (2003). Monitoring costs and the mode of international investment. *Journal of Economic Geography* 3(3), 261-274. <https://doi.org/10.1093/jeg/3.3.261>
- Mazur, M., Salganik-Shoshan, G., Walker, T., Wang, J. (2018). Proximity and litigation: Evidence from the geographic location of institutional investors. *Journal of Financial Markets* 40, 60-74. <https://doi.org/10.1016/j.finmar.2018.05.002>
- Mayberry, M. (2020). Good for managers, bad for society? Causal evidence on the association between risk-taking incentives and corporate social responsibility. *Journal of Business Finance and Accounting* 47(9-10), 1182-1214. <https://doi.org/10.1111/jbfa.12451>
- Nguyen, B. D., & Nielsen, K. M. (2010). The value of independent directors: Evidence from sudden deaths. *Journal of Financial Economics* 98(3), 550-567. <https://doi.org/10.1016/j.jfineco.2010.07.004>
- Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica* 49(6), 1417-1426. <https://doi.org/10.2307/1911408>
- Picard, R. (2010). GEODIST: Stata module to compute geographical distances. Statistical Software Components S457147, Boston College Department of Economics, revised 24 Jun 2019.
- Piotroski, J. D., Wong, T. J., Zhang, T. (2015). Political Incentives to Suppress Negative Information: Evidence from Chinese Listed Firms. *Journal of Accounting Research* 53(2), 405-459. <https://doi.org/10.1111/1475-679X.12071>
- Pirinsky, C., & Wang, Q. (2006). Does Corporate Headquarters Location Matter for Stock Returns. *Journal of Finance* 61(4), 1991-2015. <https://doi.org/10.1111/j.1540-6261.2006.00895.x>
- Quan, Y., & Zhang, W. (2021). Geographic distance and board monitoring: Evidence from the relocation of independent directors. *Journal of Corporate Finance* 66, 101802. <https://doi.org/10.1016/j.jcorpfin.2020.101802>
- Ramalingegowda, S., & Yu, Y. (2012). Institutional ownership and conservatism. *Journal of Accounting and Economics* 53(1-2), 98-114. <https://doi.org/10.1016/j.jacceco.2011.06.004>
- Refinitiv (2022). Environmental, Social and Governance Scores from Refinitiv. May 2022. https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf
- Roychowdhury, S. (2006). Earnings management through real activities manipulation. *Journal of Accounting and Economics* 42(3), 335-370. <https://doi.org/10.1016/j.jacceco.2006.01.002>

- Ryan Jr., H. E., & Wiggins III, R. A. (2004). Who is in whose pocket? Director compensation, board independence, and barriers to effective monitoring. *Journal of Financial Economics* 73(3), 497-524. <https://doi.org/10.1016/j.jfineco.2003.11.002>
- Strauss-Kahn, V., & Vives, X. (2009). Why and where do headquarters move? *Regional Science and Urban Economics* 39(2), 168-186. <https://doi.org/10.1016/j.regsciurbeco.2008.07.001>
- Utz, S. (2018). Over-investment or risk mitigation? Corporate social responsibility in Asia-Pacific, Europe, Japan, and the United States. *Review of Financial Economics* 36, 167-193. <https://doi.org/10.1016/j.rfe.2017.10.001>
- Wooldridge, J. M. (2006). *Introductory Econometrics: A Modern Approach*. Thomson South-Western, Mason, OH.
- Xie, B., Davidson III, W. N., DaDalt, P. J. (2003). Earnings management and corporate governance: the role of the board and the audit committee. *Journal of Corporate Finance* 9(3), 295-316. [https://doi.org/10.1016/S0929-1199\(02\)00006-8](https://doi.org/10.1016/S0929-1199(02)00006-8)
- Xu, L., Yu, C.-F. (J.), Zurbrugg, R. (2020). The benefit of being a local leader: Evidence from firm-specific stock price crash risk. *Journal of Corporate Finance* 65, 101752. <https://doi.org/10.1016/j.jcorpfin.2020.101752>
- Xu, Y., Xuan, Y., Zheng, G. (2021). Internet searching and stock price crash risk: Evidence from a quasi-natural experiment. *Journal of Financial Economics* 141(1), 255-275. <https://doi.org/10.1016/j.jfineco.2021.03.003>

Tables

Table 1: Summary statistics

	<i>N</i>	Mean	Std.Dev.	Percentiles				
				5th.	25th.	Median	75th.	95th.
<i>Panel A: Variables used in the main analysis</i>								
NSKEW _{t+1}	39,652	0.238	0.977	-1.096	-0.319	0.111	0.619	2.104
DUVOL _{t+1}	39,639	0.088	0.411	-0.541	-0.182	0.062	0.325	0.822
DIST_VAL_ALL _t	39,652	1,237.934	535.112	554.706	840.790	1,097.252	1,603.300	2,225.106
ln(DIST_VAL_ALL _t)	39,652	7.023	0.460	6.318	6.734	7.001	7.380	7.708
NSKEW _t	39,652	0.254	0.970	-1.032	-0.307	0.118	0.628	2.123
DUVOL _t	39,652	0.095	0.404	-0.517	-0.175	0.066	0.327	0.823
DTURN _t	39,652	0.007	0.102	-0.118	-0.023	0.002	0.031	0.146
SIGMA _t	39,652	0.059	0.035	0.021	0.035	0.050	0.073	0.125
CUMRET _t	39,652	-0.183	0.186	-0.569	-0.239	-0.119	-0.059	-0.022
ln(SIZE _t)	39,652	6.219	2.000	3.023	4.778	6.153	7.587	9.686
MB _t	39,652	3.967	6.079	0.693	1.409	2.342	4.086	11.789
LEV _t	39,652	0.163	0.172	0.000	0.001	0.122	0.273	0.496
ROA _t	39,652	-0.020	0.248	-0.450	-0.016	0.040	0.081	0.164
ACCM _t	39,652	0.185	0.236	0.026	0.069	0.121	0.209	0.529
OWN_INST _t	39,652	0.602	0.300	0.063	0.356	0.665	0.862	1.000
<i>Panel B: Variables used in the auxiliary analyses</i>								
PROP_LOCAL100 _t	39,652	0.090	0.128	0.000	0.011	0.037	0.100	0.378
ln(DIST_EQ_10 _t)	39,652	7.009	0.430	6.311	6.724	7.003	7.334	7.677
ln(DIST_VAL_ALL_exln _t)	39,636	6.968	0.582	6.005	6.631	6.983	7.415	7.768

ln(DIST_VAL_ALL_In _t)	36,513	7.140	0.283	6.727	7.017	7.154	7.308	7.539
CRASH_D _{t+1}	39,652	0.310	0.463	0.000	0.000	0.000	1.000	1.000
CRASH_EXTR _{t+1}	39,652	2.830	0.898	1.746	2.190	2.617	3.264	4.676
CRASH_COUNT _{t+1}	39,652	0.155	0.678	-1.000	0.000	0.000	1.000	1.000
NSKEW_D _{t+1}	39,612	0.389	1.848	-1.800	-0.460	0.059	0.777	3.993
NSKEW_D _t	39,612	0.432	1.820	-1.597	-0.431	0.062	0.767	4.071
DUVOL_D _{t+1}	39,612	0.046	0.361	-0.485	-0.174	0.013	0.228	0.713
DUVOL_D _t	39,612	0.053	0.351	-0.451	-0.167	0.015	0.227	0.715
SIGMA_D _t	39,612	0.030	0.018	0.011	0.017	0.025	0.037	0.064
CUMRET_D _t	39,612	-0.217	0.205	0.659	-0.293	-0.145	-0.073	-0.028
AMIHU _D _t	39,149	0.871	5.242	0.000	0.001	0.007	0.098	3.369
REM _t	35,492	0.060	1.336	-2.034	-0.418	0.089	0.594	1.812
NUM_ANALY _t	37,359	7.404	7.097	0.000	2.000	5.167	10.667	22.500
BTD _t	39,407	-0.043	0.881	-1.307	-0.047	0.132	0.272	0.596
DTD _t	38,683	4.709	2.633	1.297	2.802	4.210	6.114	9.936
REL _t	39,727	0.521	0.110	0.348	0.437	0.524	0.598	0.715
DUM_DIST_SEC _t	39,652	0.543	0.498	0.000	0.000	1.000	1.000	1.000
ln(FIN_DIST_SEC _t)	39,652	4.135	1.721	0.822	3.016	4.565	5.623	5.961
ln(DEMO_TOTPOP _t)	39,606	13.717	1.101	11.689	13.204	13.755	14.336	15.463
ln(INCOME _t)	39,606	10.956	0.264	1.0554	10.759	10.973	11.153	11.393
BIG4 _t	39,652	0.733	0.442	0.000	0.000	1.000	1.000	1.000
CEO_DUALITY _t	11,504	0.652	0.476	0.000	0.000	1.000	1.000	1.000
BOARD_INDEP _t	11,454	0.899	0.301	0.000	1.000	1.000	1.000	1.000
PROP_BINDEP _t	11,419	0.791	0.132	0.556	0.737	0.818	0.889	0.917
BOARD_SIZE _t	11,451	9.546	2.239	6.000	8.000	9.000	11.000	13.000
GOV _t	11,501	0.471	0.219	0.121	0.296	0.469	0.645	0.826

SURP_UE _{t+1}	37,354	0.048	0.214	0.000	0.000	0.000	0.000	0.000
C_STR_BREAK1 _{t+1}	34,559	0.092	0.289	0.000	0.000	0.000	0.000	1.000
C_STR_BREAK2 _{t+1}	34,532	0.051	0.000	0.000	0.000	0.000	0.000	1.000

This table presents descriptive statistics for measures of stock price crash risk, distance variables, as well as control variables used in my baseline analysis (Panel A). Panel B reports summary statistics for variables either used as additional controls or as partitioning variables. The sample covers non-missing firm-year observations from 1997 through 2019. See Appendix A for detailed variable definitions.

Table 2: Pearson correlation results

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1) NSKEW _{t+1}	1.00													
(2) DUVOL _{t+1}	0.96*	1.00												
(3) ln(DIST_VAL_ALL _t)	0.04*	0.04*	1.00											
(4) NSKEW _t	0.04*	0.03*	0.05*	1.00										
(5) DUVOL _t	0.04*	0.04*	0.05*	0.96*	1.00									
(6) DTURN _t	0.05*	0.05*	0.00	0.05*	0.05*	1.00								
(7) SIGMA _t	0.04*	0.03*	0.00	0.29*	0.25*	0.20*	1.00							
(8) CUMRET _t	-0.03*	-0.02*	0.00	-0.32*	-0.27*	-0.18*	-0.98*	1.00						
(9) MB _t	0.05*	0.05*	0.07*	-0.04*	-0.05*	0.04*	0.03*	-0.03*	1.00					
(10) LEV _t	-0.02*	-0.02*	-0.03*	0.00	0.00	0.01	-0.14*	0.13*	0.12*	1.00				
(11) ROA _t	-0.02*	-0.01	-0.03*	-0.07*	-0.06*	-0.03*	-0.44*	0.42*	-0.16*	0.08*	1.00			
(12) ln(SIZE) _t	0.01	0.03*	0.14*	0.02*	0.04*	-0.02*	-0.50*	0.46*	-0.02*	0.39*	0.34*	1.00		
(13) ACCM _t	0.04*	0.04*	0.02*	0.03*	0.02*	0.02*	0.20*	-0.19*	0.11*	-0.07*	-0.23*	-0.19*	1.00*	
(14) OWN_INST _t	0.07*	0.08*	0.14*	0.08*	0.10*	0.00	-0.41*	0.39*	0.01*	0.20*	0.25*	0.64*	-0.13*	1.00

This table presents a Pearson correlation matrix of key variables for the sample of firms included in my study. The sample covers non-missing firm-year observations from 1997 through 2019. * indicates statistical significance at the 1% level. Appendix A1 contains detailed variable descriptions.

Table 3: Sorting results

Variable	Partitioning variable: Value-weighted distance								
	Full sample			“Active” institutions			“Passive” institutions		
	Local (1)	Non-L (2)	L-NL (3)	Local (4)	Non-L (5)	L-NL (6)	Local (7)	Non-L (8)	L-NL (9)
NSKEW _{t+1}	0.192	0.293	-0.101***	0.177	0.286	-0.109***	0.188	0.249	-0.061***
DUVOL _{t+1}	0.065	0.111	-0.047***	0.058	0.108	-0.050***	0.068	0.092	-0.024***
NSKEW _t	0.202	0.313	-0.110***	0.193	0.304	-0.111***	0.183	0.278	-0.095***
DUVOL _t	0.070	0.121	-0.051***	0.066	0.116	-0.050***	0.064	0.106	-0.042***
DTURN _t	0.006	0.007	-0.001	0.006	0.006	0.000	0.007	0.008	-0.001
SIGMA _t	0.060	0.065	-0.005***	0.061	0.065	-0.004***	0.060	0.060	0.000
CUMRET _t	-0.191	-0.217	0.026***	-0.194	-0.217	0.023***	-0.194	-0.191	-0.003
MB _t	3.721	.4839	-1.119***	3.487	4.785	-1.299***	3.859	4.626	-0.766***
LEV _t	0.161	0.129	0.031***	0.154	0.129	0.025***	0.170	0.147	0.024***
ROA _t	-0.039	-0.064	0.025***	-0.043	-0.066	0.023***	-0.026	-0.037	0.011**
ln(SIZE _t)	5.640	6.134	-0.494***	5.498	6.093	-0.595***	5.924	6.443	-0.519***
ACCM _t	0.199	0.206	-0.007*	0.200	0.207	-0.008*	0.192	0.194	-0.002
OWN_INST _t	0.552	0.612	-0.064***	0.527	0.601	-0.074***	0.511	0.609	-0.098***

This table documents the mean estimates of stock price crash risk and main firm characteristics in Q5 (= “Non-L”) and Q1 (= “Local”) sorted by $\ln(\text{DIST_VAL_ALL}_t)$ for columns 1 to 3, $\ln(\text{DIST_VAL_ALL_Ind}_t)$ for columns 4 to 6, and $\ln(\text{DIST_VAL_ALL_exInd}_t)$ for columns 7 to 9. Columns (1)-(3) report sorting results for the full sample, while Columns (4)-(6) and Columns (7)-(9) present results for the subsample of “active” and “passive” investors, respectively. For columns (4) to (9) investors are classified based on the investor category combined with their investment style and orientation. Finally, columns (3), (6), and (9) report the difference in means between “Local” and “Non-L”. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. All variables are defined in Appendix A1.

Table 4: Baseline regression results

Dependent Variable	NSKEW _{t+1}			DUVOL _{t+1}		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Variable of Interest</i>						
ln(DIST_VAL_ALL _t)	0.074*** (6.25)	0.049*** (4.21)	0.049*** (4.24)	0.033*** (6.59)	0.020*** (4.15)	0.020*** (4.17)
<i>Control Variables</i>						
NSKEW _t			0.018*** (2.74)			
DUVOL _t						0.017*** (2.94)
DTURN _t		0.282*** (4.91)	0.287*** (4.99)		0.128*** (5.59)	0.130*** (5.66)
SIGMA _t		4.230*** (4.70)	4.355*** (4.81)		1.886*** (5.14)	1.916*** (5.21)
CUMRET _t		0.517*** (3.20)	0.585*** (3.55)		0.240*** (3.66)	0.262*** (3.95)
MB _t		0.005*** (5.35)	0.005*** (5.54)		0.003*** (6.88)	0.003*** (7.06)
LEV _t		-0.127*** (-3.72)	-0.126*** (-3.72)		-0.071*** (-4.95)	-0.070*** (-4.96)
ROA _t		-0.008 (-0.26)	-0.011 (-0.38)		0.014 (1.22)	0.013 (1.15)
ln(SIZE _t)		0.001 (0.15)	-0.000 (-0.09)		0.003 (1.53)	0.002 (1.32)
ACCM _t		0.076** (2.51)	0.077** (2.57)		0.035*** (2.92)	0.035*** (2.97)
OWN_INST _t		0.003*** (13.04)	0.003*** (12.66)		0.001*** (13.58)	0.001*** (13.24)
Constant	-0.282*** (-3.39)	-0.470*** (-5.59)	-0.459*** (-5.49)	-0.142*** (-4.06)	-0.226*** (-6.40)	-0.220*** (-6.29)
Year FE	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES
Observations	39,652	39,652	39,652	39,639	39,639	39,639
Adj. R ²	0.019	0.029	0.029	0.021	0.033	0.033

This table documents the results of my baseline regression model. In Columns (1) and (4), I control for year and industry fixed effects only, whereas I include all firm characteristics (except the lagged dependent variable) in Columns (2) and (5). Finally, Columns (3) and (6) report the results of the baseline model as presented in equation (3). Across columns (1)-(3), the dependent variable is NSKEW_{t+1}, while columns (4)-(6) report the results for DUVOL_{t+1}. Each model includes year and industry fixed effects. Standard errors are adjusted for heteroscedasticity and within-firm clustering. t-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 5: Regression results with additional control variables

Dependent variable	NSKEW _{t+1}				DUVOL _{t+1}			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Variable of Interest</i>								
ln(DIST_VAL_ALL _t)	0.038*** (2.91)	0.049*** (3.66)	0.052* (1.87)	0.076** (2.13)	0.017*** (3.06)	0.019*** (3.39)	0.023** (2.01)	0.036** (2.48)
<i>Additional Controls</i>								
AMIHU _t	-0.011*** (-8.70)			0.011 (0.44)	-0.005*** (-8.61)			0.014 (1.09)
REM _t	-0.001 (-0.29)			-0.016* (-1.77)	-0.001 (-0.69)			-0.008* (-1.95)
NUM_ANALY _t	0.003** (2.23)			0.003 (1.40)	0.001** (2.57)			0.002* (1.77)
DTD _t	-0.009*** (-2.89)			-0.010* (-1.85)	-0.003** (-2.32)			-0.003 (-1.27)
BTD _t	-0.008 (-0.56)			-0.017 (-0.59)	-0.003 (-0.59)			-0.004 (-0.34)
FIN_DIST_SEC _t		0.010 (0.79)		-0.029 (-1.15)		0.006 (1.18)		-0.014 (-1.36)
REL _t		-0.011 (-0.21)		0.058 (0.55)		-0.006 (-0.28)		0.035 (0.79)
ln(TOTPOP _t)		-0.003 (-0.51)		-0.013 (-1.31)		-0.001 (-0.37)		-0.006 (-1.46)
ln(INCOME _t)		-0.038 (-1.19)		0.035 (0.80)		-0.013 (-1.37)		0.015 (0.84)
DUAL _t			-0.041** (-2.02)	-0.054*** (-2.58)			-0.014 (-1.63)	-0.020** (-2.27)
INDEP _t			0.042 (1.42)	0.026 (0.86)			0.020 (1.60)	0.013 (0.98)
BSIZE _t			0.003 (0.54)	0.002 (0.43)			0.001 (0.58)	0.001 (0.52)
GOV _t			0.000 (0.13)	-0.000 (-0.12)			0.000 (0.70)	0.000 (0.38)
BIG4 _t			-0.023 (-0.58)	-0.016 (-0.39)			-0.008 (-0.46)	-0.002 (-0.14)
Baseline Controls	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES	YES	YES
Observations	35,406	39,272	11,428	10,769	35,397	39,260	11,428	10,769
Adj. R ²	0.028	0.031	0.016	0.025	0.034	0.035	0.023	0.016

This table reports regression results of the relation between shareholder proximity and future stock price crash risk after including additional controls. Columns (1)-(4) present results for NSKEW_{t+1} as the dependent variable, while columns (5)-(8) report the results for DUVOL_{t+1}. Each model includes year and industry fixed effects. Standard errors are adjusted for heteroscedasticity and within-firm clustering. t-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 6: Entropy balancing

Dependent Variable	NSKEW _{t+1}		DUVOL _{t+1}	
	(1)	(2)	(3)	(4)
<i>Variable of Interest</i>				
HIGH_VAL_ALL _t	0.045*** (3.26)	0.040*** (2.91)	0.018*** (3.19)	0.017*** (3.02)
Baseline Controls	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Observations	26,434	25,904	26,425	25,895
Adj. R ²	0.028	0.029	0.031	0.033

This table presents results from the regressions for the entropy matched sample. Columns (1) and (3) report the regression results using all weights, while columns (2) and (4) exclude extreme weights, i.e. the 1% and 99% percentiles. Each model includes year and industry fixed effects. Standard errors are adjusted for heteroscedasticity and within-firm clustering. t-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 7: Firm/high-dimensional fixed effects regression results

Dependent Variable	NSKEW _{t+1}			DUVOL _{t+1}		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Variable of Interest</i>						
ln(DIST_VAL_ALL _t)	0.069*** (2.87)	0.073*** (2.97)	0.058** (2.35)	0.034*** (3.29)	0.036*** (3.40)	0.029*** (2.70)
<i>Control Variables</i>						
NSKEW _t		-0.074*** (-11.13)				
DUVOL _t					-0.070*** (-11.55)	
DTURN _t	0.411*** (7.01)	0.379*** (6.52)	0.441*** (7.26)	0.175*** (7.42)	0.163*** (6.99)	0.188*** (7.71)
SIGMA _t	2.787*** (2.75)	2.446** (2.45)	2.568** (2.47)	1.131*** (2.77)	1.062*** (2.62)	1.015** (2.42)
CUMRET _t	0.838*** (4.69)	0.558*** (3.13)	0.830*** (4.53)	0.324*** (4.55)	0.238*** (3.35)	0.319*** (4.36)
MB _t	0.010*** (7.47)	0.009*** (6.87)	0.009*** (7.00)	0.005*** (8.66)	0.005*** (8.08)	0.005*** (8.28)
LEV _t	-0.223*** (-3.85)	-0.221*** (-3.74)	-0.227*** (-3.84)	-0.113*** (-4.65)	-0.112*** (-4.52)	-0.115*** (-4.68)
ROA _t	0.083** (2.19)	0.086** (2.25)	0.079** (2.03)	0.048*** (3.23)	0.048*** (3.14)	0.047*** (3.05)
ln(SIZE _t)	0.131*** (10.17)	0.139*** (10.40)	0.137*** (10.12)	0.060*** (11.20)	0.063*** (11.35)	0.062*** (11.26)
ACCM _t	0.018 (0.52)	0.014 (0.39)	0.005 (0.11)	0.009 (0.65)	0.008 (0.54)	0.004 (0.21)
OWN_INST _t	0.003*** (5.58)	0.003*** (6.44)	0.003*** (5.56)	0.001*** (6.00)	0.002*** (6.82)	0.001*** (5.90)
Constant	-1.052*** (-5.85)	-1.161*** (-6.31)	-1.186*** (-6.32)	-0.521*** (-6.84)	-0.568*** (-7.26)	-0.581*** (-7.28)
Year FE	YES	YES	NO	YES	YES	NO
Firm FE	YES	YES	YES	YES	YES	YES
Year×Industry FE	NO	NO	YES	NO	NO	YES
Observations	39,652	39,652	39,235	39,639	39,639	39,222
Adj. R ²	0.028	0.032	0.065	0.034	0.038	0.066

This table reports regression results using firm fixed effects and high-dimensional fixed effects. Columns (1), (2), (4), and (5) present the regression estimates for firm fixed effects specification, whereas columns (3) and (6) employ high-dimensional fixed effects. Standard errors are adjusted for heteroscedasticity and within-firm clustering. t-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 8: Regression results from change analysis

Dep. Variable =	Direction crash-distance		Direction distance-crash	
	(1) Δ NSKEW _{t+1}	(2) Δ DUVOL _{t+1}	(3) Δ DIST_VAL_ALL _{t+1}	(4) Δ DIST_VAL_ALL _{t+1}
<i>Variables of Interest</i>				
Δ DIST_VAL_ALL _t	0.103*** (2.87)	0.045*** (3.09)		
Δ NSKEW _t			0.001 (1.28)	
Δ DUVOL _t				0.003 (1.28)
Baseline Controls (Δ)	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Observations	34,581	34,570	36,780	36,780
Adj. R ²	0.267	0.269	0.002	0.002

This table presents regression results for the relationship between institutional distance and stock price crash risk applying change analysis. In columns (1) and (2), I regress the changes in future stock price crash risk on changes in value-weighted distance, whereas in columns (3) and (4), changes in future institutional distance are regressed on changes in stock price crash risk. Across all models, I consider changes in control variables. Standard errors are adjusted for heteroscedasticity and within-firm clustering. t-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 9: Regression results from 2SLS instrumental variable approach

Dependent Variable	NSKEW _{t+1}		DUVOL _{t+1}	
	(1)	(2)	(3)	(4)
<i>Variables of Interest</i>				
Instrumented ln(DIST_VAL_ALL _t)	0.043*** (2.82)	0.040** (2.47)	0.018*** (2.90)	0.015** (2.29)
<i>Control Variables</i>				
NSKEW _t	0.017*** (2.63)	0.017*** (2.63)		
DUVOL _t			0.017*** (2.84)	0.017*** (2.82)
DTURN _t	0.289*** (5.01)	0.289*** (5.00)	0.130*** (5.67)	0.130*** (5.67)
SIGMA _t	4.352*** (4.79)	4.358*** (4.80)	1.914*** (5.19)	1.920*** (5.21)
CUMRET _t	0.583*** (3.53)	0.583*** (3.53)	0.261*** (3.93)	0.261*** (3.93)
MB _t	0.005*** (5.51)	0.005*** (5.52)	0.003*** (7.04)	0.003*** (7.07)
LEV _t	-0.125*** (-3.68)	-0.126*** (-3.71)	-0.070*** (-4.90)	-0.071*** (-4.97)
ROA _t	-0.001 (-0.33)	-0.001 (-0.33)	0.013 (1.18)	0.013 (1.18)
ln(SIZE _t)	0.000 (0.00)	0.000 (0.04)	0.002 (1.37)	0.002 (1.45)
ACCM _t	0.077*** (2.58)	0.077** (2.57)	0.035*** (2.98)	0.035*** (2.96)
OWN_INST _t	0.003*** (12.61)	0.003*** (12.61)	0.001*** (13.17)	0.001*** (13.17)
ln(DIST_SEC _t)		0.002 (0.48)		0.002 (1.18)
Year FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Observations	39,528	39,528	39,515	39,515
Adj. R ²	0.010	0.010	0.011	0.011
<i>Test diagnostics</i>				
Kleibergen-Paap rk Wald F-statistic	7,148.55	6,383.46	7,143.22	6,377.54
Hansen J-statistic (p- value)	0.155	0.157	0.365	0.374
Hausman test (p- value)	0.331	0.298	0.428	0.313

This table presents regression results for the second-stage regressions. The first instrument is the state average of ln(DIST_VAL_ALL_t), where the contribution of the focal firm and the firms in the same industry are excluded. The second instrument is the industry average of ln(DIST_VAL_ALL_t), where the contribution of the focal firm and the firms located in the same state are excluded. In columns (1) and (3), I use all baseline control variables, whereas columns (2) and (4) add the distance to the nearest SEC office as an additional control variable in both stages. Standard errors are adjusted for heteroscedasticity and within-firm clustering. t-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 10: Distribution of betas after placebo tests

Dependent variable	NSKEW _{t+1} (1)	DUVOL _{t+1} (2)
Mean β ln(pseudo_DIST_VAL_ALL _t)	0.009	0.004
Min β ln(pseudo_DIST_VAL_ALL _t)	-0.029	-0.011
1% perc. β ln(pseudo_DIST_VAL_ALL _t)	-0.022	-0.009
5% perc. β ln(pseudo_DIST_VAL_ALL _t)	-0.011	-0.004
25% perc. β ln(pseudo_DIST_VAL_ALL _t)	0.001	0.001
Median β ln(pseudo_DIST_VAL_ALL _t)	0.009	0.004
75% perc. β ln(pseudo_DIST_VAL_ALL _t)	0.017	0.007
95% perc. β ln(pseudo_DIST_VAL_ALL _t)	0.028	0.012
99% perc. β ln(pseudo_DIST_VAL_ALL _t)	0.038	0.016
Max β ln(pseudo_DIST_VAL_ALL _t)	0.041	0.017
Coefficient of actual ln(DIST_VAL_ALL _t) as reported in Table 4	0.049	0.020

This table presents the results of the placebo tests. I randomly assign coordinates to each sample firm, generating a pseudo-DIST_VAL_ALL_t variable by measuring the value-weighted distance between the pseudo location of a company and the location of its actual institutional shareholders. I then use ln(pseudo-DIST_VAL_ALL_t) to estimate my baseline model, whereby I repeat this procedure 500 times. The presented empirical distribution is based on the 500 coefficient estimates. For comparison, I also report the coefficient estimates of the actual value of ln(DIST_VAL_ALL_t) for each of the two crash risk measures.

Table 11: Robustness tests: alternative measures of geographic proximity

Dependent Variable	NSKEW _{t+1}		DUVOL _{t+1}	
	(1)	(2)	(3)	(4)
<i>Variables of Interest</i>				
PERC_LOCAL100 _t	-0.141*** (-3.55)		-0.059*** (-3.57)	
ln(DIST_EQ_10 _t)		0.048*** (3.91)		0.021*** (4.07)
<i>Control Variables</i>				
NSKEW _t	0.017*** (2.70)	0.018*** (2.74)		
DUVOL _t			0.017*** (2.92)	0.017*** (2.93)
DTURN _t	0.282*** (4.90)	0.286*** (4.98)	0.127*** (5.57)	0.130*** (5.66)
SIGMA _t	4.465*** (4.94)	4.359*** (4.81)	1.962*** (5.35)	1.916*** (5.21)
CUMRET _t	0.592*** (3.60)	0.585*** (3.55)	0.265*** (4.00)	0.261*** (3.94)
MB _t	0.005*** (5.82)	0.005*** (5.59)	0.003*** (7.34)	0.003*** (7.10)
LEV _t	-0.139*** (-4.13)	-0.129*** (-3.84)	-0.075*** (-5.37)	-0.071*** (-5.06)
ROA _t	-0.014 (-0.48)	-0.009 (-0.33)	0.012 (1.04)	0.014 (1.21)
ln(SIZE _t)	0.002 (0.41)	0.001 (0.21)	0.003* (1.83)	0.003 (1.60)
ACCM _t	0.077** (2.55)	0.077** (2.55)	0.035*** (2.95)	0.035*** (2.96)
OWN_INST _t	0.003*** (12.63)	0.003*** (12.65)	0.001*** (13.21)	0.001*** (13.23)
Constant	-0.115*** (-2.99)	-0.456*** (-5.16)	-0.078*** (-4.95)	-0.225*** (-6.14)
Year FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Observations	39,652	39,652	39,639	39,639
Adj. R ²	0.029	0.029	0.033	0.033

This table documents the results of my baseline regression model using alternative measures of institutional distance. Each model includes year and industry fixed effects. Standard errors are adjusted for heteroscedasticity and within-firm clustering. t-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 12: Robustness tests: alternative measures of crash risk

Dependent Variable	(1) CRASH_3.09 _{t+1}	(2) CRASH_EXTR _{t+1}	(3) CRASH_COUNT _{t+1}
<i>Variable of Interest</i>			
ln(DIST_VAL_ALL _t)	0.017*** (3.08)	0.045*** (4.23)	0.018** (2.12)
Baseline Controls	YES	YES	YES
Year FE	YES	YES	YES
Industry FE	YES	YES	YES
Observations	39,652	39,652	39,652
Pseudo or adj. R ²	0.020	0.029	0.016

This table presents the results of my baseline regression model using alternative measures of firm-specific stock price crash risk. Each model includes year and industry fixed effects. Standard errors are adjusted for heteroscedasticity and within-firm clustering. t-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 13: Robustness tests: daily data

Dependent Variable	NSKEW_ D_{t+1}		DUVOL_ D_{t+1}	
	(1) Univariate	(2) Including controls	(3) Univariate	(4) Including controls
<i>Variable of Interest</i>				
ln(DIST_VAL_ALL $_t$)	0.123*** (5.69)	0.081*** (3.83)	0.022*** (5.31)	0.013*** (3.21)
<i>Control Variables</i>				
NSKEW_ D_t		0.018*** (2.65)		
DUVOL_ D_t				0.020*** (3.16)
DTURN $_t$		0.466*** (4.25)		0.100*** (4.74)
SIGMA_ D_t		7.231** (2.46)		2.032*** (3.22)
CUMRET_ D_t		0.406 (1.59)		0.136** (2.57)
MB $_t$		0.010*** (5.21)		0.002*** (6.75)
LEV $_t$		-0.194*** (-2.97)		-0.059*** (-4.64)
ROA $_t$		0.008 (0.15)		0.037*** (3.69)
ln(SIZE $_t$)		-0.008 (-1.03)		0.001 (0.72)
ACCM $_t$		0.136** (2.52)		0.034*** (3.26)
OWN_INST $_t$		0.006*** (13.73)		0.001*** (12.19)
Constant	-0.475*** (-3.15)	-0.682*** (-4.45)	-0.109*** (-3.74)	-0.157*** (-5.16)
Year FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Observations	39,612	39,612	39,612	39,612
Adj. R ²	0.017	0.026	0.031	0.041

This table documents the results of my baseline regression model, while using daily firm-specific returns rather than weekly firm-specific returns to construct variables of crash risk. In Columns (1) and (3), I control for year and industry fixed effects only, whereas I include all firm characteristics in Columns (2) and (4). Across columns (1)-(2), the dependent variable is NSKEW $_{t+1}$, while columns (3)-(4) report the results for DUVOL $_{t+1}$. Each model includes year and industry fixed effects. Standard errors are adjusted for heteroscedasticity and within-firm clustering. t-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 14: Moderating role of internal corporate governance

Dependent Variable	NSKEW _{t+1}		DUVOL _{t+1}	
Panel A: Refinitiv ESG corporate governance scores				
	(1) Strong <i>GOV_t</i>	(2) Weak <i>GOV_t</i>	(3) Strong <i>GOV_t</i>	(4) Weak <i>GOV_t</i>
<i>Variable of Interest</i>				
ln(DIST_VAL_ALL _t)	0.009 (0.22)	0.077** (2.06)	0.013 (0.76)	0.028* (1.84)
Differences in coefficients (p-value)		0.21		0.49
Baseline Controls	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Observations	5,746	5,755	5,746	5,755
Adj. R ²	0.018	0.020	0.017	0.019
Panel B: Proportion Independent Board Members				
	(1) High <i>PROP_BINDEP_t</i>	(2) Low <i>PROP_BINDEP_t</i>	(3) High <i>PROP_BINDEP_t</i>	(4) Low <i>PROP_BINDEP_t</i>
<i>Variable of Interest</i>				
ln(DIST_VAL_ALL _t)	-0.009 (-0.22)	0.093** (2.51)	0.009 (0.56)	0.032** (2.09)
Differences in coefficients (p-value)		0.06		0.31
Baseline Controls	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Observations	5,366	6,051	5,366	6,051
Adj. R ²	0.019	0.014	0.019	0.012

This table presents cross-sectional regression results for the moderating effect of internal corporate governance proxies on the relationship between institutional distance and stock price crash risk. In columns (1) and (2) of Panel A and B, respectively, *NSKEW_t* is the dependent variable, whereas in columns (3) and (4), *DUVOL_t* serves as the dependent variable. Standard errors are adjusted for heteroscedasticity and within-firm clustering. t-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 15: SOX and monitoring

Dependent Variable	NSKEW _{t+1}		DUVOL _{t+1}	
	(1) Subsample: Year ≤ 2003	(2) Subsample: Year > 2003	(3) Subsample: Year ≤ 2003	(4) Subsample: Year > 2003
<i>Variable of Interest</i>				
ln(DIST_VAL_ALL _t)	0.080*** (4.47)	0.039*** (2.67)	0.032*** (4.22)	0.016*** (2.72)
Differences in coefficients (p-value)		0.07		0.09
Baseline Controls	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Observations	12,103	27,548	12,097	27,541
Adj. R ²	0.061	0.023	0.067	0.025

This table presents the estimates from subsample analyses before and after 2003 and investigates the effect of institutional distance on future crash risk measures. Standard errors are adjusted for heteroscedasticity and within-firm clustering. t-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 16: Bad news hoarding tests

Dependent Variable	(1) SURP_UE _{t+1}	(2) C_STR_BREAK_1 _{t+1}	(3) C_STR_BREAK_2 _{t+1}
Variable of Interest			
ln(DIST_VAL_ALL _t)	0.009*** (4.33)	0.013*** (4.05)	0.006*** (3.20)
Baseline Controls	YES	YES	YES
Year FE	YES	YES	YES
Industry FE	YES	YES	YES
Observations	37,354	34,559	34,532
Pseudo R ²	0.116	0.041	0.063

This table presents the estimates from logit regressions where the dependent variable is one of the revised crash risk measures, which are based on earnings information. Coefficients are reported as marginal effects. Standard errors are adjusted for heteroscedasticity and within-firm clustering. t-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 17: Do all investors monitor?

Dependent Variable	NSKEW _{t+1}			DUVOL _{t+1}		
	(1) Active	(2) Passive	(3) Together	(4) Active	(5) Passive	(6) Together
<i>Variables of Interest</i>						
ln(DIST_VAL_ALL_ex_Ind _t)	0.054*** (5.94)		0.058*** (6.04)	0.028*** (5.91)		0.025*** (6.05)
ln(DIST_VAL_ALL_Ind _t)		0.005 (0.25)	-0.016 (-0.85)		-0.000 (-0.04)	-0.009 (-1.16)
Wald-Test of difference in coefficients (p-value)			0.00			0.00
Baseline Controls	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES
Observations	39,636	38,660	38,644	39,623	38,647	38,631
Adj. R ²	0.029	0.028	0.029	0.033	0.032	0.033

This table presents the estimates from subsample analyses based on investor classification. Specifically, columns (1) and (4) consider the distance to investors that are classified as “active” monitors. Moreover, columns (2) and (5) report the results for their “passive” counterparts, i.e., the proximity to investors with subtype “investment advisor” or “investment advisor/hedge fund” and investment style “index” and orientation “passive” in Refinitiv Ownership Profile. Columns (3) and (6) combine both measures of institutional distance. Standard errors are adjusted for heteroscedasticity and within-firm clustering. t-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.