Information and noise in stock markets: Evidence on the determinants and effects using new empirical measures

Thanh Huong Nguyen

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Finance

Finance Discipline Group, UTS Business School

University of Technology Sydney

July 2019

Certificate of original authorship

I, Thanh Huong Nguyen declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy in Finance, in the Business School at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree at any other academic institution except as fully acknowledged within the text.

This research is supported by the Australian Government Research Training Program.

Signature:

Production Note:
Signature removed prior to publication.

Thanh Huong Nguyen

Date: 16/02/2020

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my principal supervisor, Professor Talis Putnins, for his tremendous guidance, support, and encouragement. I would especially like to express my thanks for his patience and engagement throughout my research at UTS. The skills and knowledge that I have developed under his supervision will be of great value to my future academic career.

I am indeed fortunate to have Associate Professor Eliza Wu as my external supervisor. I would like to thank her for introducing me to UTS and providing me with a great opportunity to learn and work professionally at this university. Eliza also provided me with valuable comments for improving the materials in this thesis.

I am especially thankful to Dr Christina Sklibosios and Dr Adrian Lee for their helpful comments and encouragement. I also truly appreciate the help I have received from the administrative staff at the Finance Discipline Group: Andrea Greer, Mala Kapahi, Stephanie Ough, and Duncan Ford.

I gratefully acknowledge the Securities Industry Research Centre of Asia-Pacific (SIRCA), Thomson Reuters, and United Nations for providing access to data used in this thesis. I thank the Vietnam International Education Development (VIED) – UTS joint scholarship for financing my Ph.D. study, as well as colleagues from the University of Economics – The University of Danang for their great support when I was doing my Ph.D. overseas.

I extend my thanks to my friends and fellow research students for stimulating research discussions, and for the fun we have had over the last four years. A special thank you goes to Kate Ly’s family for making my stay in Sydney so enjoyable.

I am very grateful to my family and my husband, Phu Mai; and it is to them I dedicate my thesis. None of my successes and achievements would have been possible without their endless love and continuous support.

Preface

Some of the research presented in this thesis has also been incorporated in four working papers co-authored with Prof. Talis Putnins, Assoc. Prof. Eliza Wu, and Prof. Jonathan Brogaard. The four papers are titled:

1. "What moves stock prices? The role of news, noise, and information";
2. "Noisy prices and capital allocation efficiency";
3. "Time-varying gambling in stock markets and its effect on asset prices";
4. "Stock markets as casinos: International evidence on the drivers and effects of gambling in stock markets".

This research has benefited greatly from the thoughtful insights and feedback provided by Paul Besson, Jean-Philippe Bouchaud, Amit Goyal, Alok Kumar, Charles-Albert Lehalle, Joshua Mollner, Ryan Riordan, Elvira Sojli, and seminar and conference participants at the Market Microstructure Confronting Many Viewpoints #5 Conference, the Investment Management Research Conference, the Asian Finance Association Annual Meeting, the Financial Research Network (FIRN) Annual Meeting, FIRN Asset Pricing Workshop, the Australasian Finance and Banking Conference, the J.P. Morgan Quantference, Brigham Young University, Monash University, the University of Western Australia, Curtin University, the University of Technology Sydney, the University of Sydney, and the Queensland University of Technology.

The first working paper is scheduled for presentation at the American Finance Association (AFA) Annual Meeting in San Diego, California in January 2020.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

2SLS	Two-Stage Least Squares
AMEX	American Stock Exchange
CRSP	Center for Research in Security Prices
ETF	Exchange traded fund
EU	European Union
GBGC	Global Betting Gaming Consultants
GDP	Gross Domestic Product
IDV	Individualism versus Collectivism
IV	Instrumental Variable
IVR	Indulgence versus Restraint
NYSE	New York Stock Exchange
OLS	Ordinary Least Squares
SEC	Securities and Exchange Commission
SIC	Standard Industrial Classification
SIRCA	Securities Industry Research Centre of Asia-Pacific
VAR	Vector Auto-Regression
VMA	Vector Moving Average
UAI	Uncertainty Avoidance Index
US	United States of America

Abstract

This thesis comprises four studies relating to stock market efficiency, its measurement, its effects, and its determinants.

The first study proposes novel empirical measures that separate different types of information and noise as drivers of stock return variance. Specifically, the new methods disentangle four components: market-wide information, private firm-specific information revealed through trading, firm-specific information revealed through public sources, and noise. Overall, in US stocks, 31% of the return variance is attributable to noise, 37% to public firm-specific information, 24% to private firm-specific information, and 8% to market-wide information. Since the mid-1990s, there has been a dramatic decline in noise and an increase in firm-specific information, consistent with increasing market efficiency.

The second study examines how noise affects inference in existing empirical measures, such as idiosyncratic volatility (one minus the R^2 of a market model) and decompositions of cash flow and discount rate news. This thesis finds that after accounting for noise, cash flow information plays a considerably larger role in driving individual stock returns than previously believed and discount rate information plays a smaller role. Furthermore, the decrease in idiosyncratic volatility (increase of market model R^2) since 1997 is the result of a decrease in noise during this recent period. The evidence indicates that the market has become more efficient in the past two decades, contrary to what is implied by standard interpretations of R^2 as an inverse measure of efficiency.

In the third study, this thesis examines the real effects of stock market efficiency by analysing the relation between noise in stock prices and the efficiency of corporate investment and capital allocation at both the firm and industry levels. The analysis uses a long time-series of data from 1963, as well as a cross-section of 42 countries. Consistent with the notion that noise decreases investment efficiency, this research finds strong evidence that noise is negatively associated with the sensitivity of corporate investment to firms' growth opportunities and the sensitivity of industry-level investment to value added. These findings highlight the important real effects of

secondary market quality in determining firms' investment behaviour and the efficiency with which capital is allocated.

The forth essay provides evidence on how individual investors' behaviour, in particular investors' gambling activity in stocks, affects stock market efficiency. We develop novel measures of the amount of gambling in stock markets based on the turnover differences between lottery stocks and non-lottery stocks, and validate the measure. Using a global sample, we examine how much gambling occurs in different countries, what determines these levels, and how the gambling that occurs on stock markets affects a country's capital markets. We find that culture and economic factors are all significant drivers of a country's gambling propensity in both traditional venues and stock markets. Interestingly, we find a substitution effect—restrictions/bans on traditional gambling lead to a spillover of gambling onto stock market(s). Exploiting regulation of traditional gambling as an instrument, we find that increased gambling on stock markets makes them more liquid and efficient. Our findings have implications for using gambling regulation as a policy instrument to affect financial market quality.

Collectively, these studies contribute to our understanding of market efficiency, how to measure it, what drives its variation through time and across stocks, and how it affects resource allocation across companies and sectors.

Chapter 1: Introduction

This chapter presents the background and motivation of this research (Section 1.1), explores the research scope and key contributions (Section 1.2), and provides an outline of this thesis (Section 1.3).

1.1 Research background and motivation

The primary role of capital markets is allocation of scare resources and funds available for investment to their best uses. To perform this role well, prices in these markets need to provide accurate signals about firms' fundamentals to investors and corporate managers. In other words, informationally "efficient" markets drive efficient use of resources in an economy. For this reason, "market efficiency" has been the subject of considerable attention over recent decades. However, despite a large number of studies, important aspects of market efficiency remain poorly understood or remain as unresolved issues in the literature.

1.1.1 Definition and forms of market efficiency

Early work related to market efficiency is based on the random walk hypothesis. In an efficient market, when information arises, the news spreads very quickly and is impounded into the stock prices immediately. In other words, tomorrow's price changes reflect only tomorrow's news and are independent of the price changes today. In efficient markets, rational investors react instantaneously to any informational advantages they have and therefore trading on available information fails to generate abnormal profits. Fama, in his famous 1970 essay, summarises the random walk literature and proposes the concept of an efficient market that has been used in the finance literature ever since.

> A market in which prices always "fully reflect" available information is called "efficient" (Fama, 1970, p. 383).

An "efficient" market in the finance context refers to a market that is efficient in processing information and is thus *informationally* efficient. Fama (1970) classifies market efficiency into three categories: weak form, semi-strong form, and strong form, based on the information sets that are expected to affect stock prices. Under weak-form efficiency, the current prices reflect all information contained in historical prices, and it is therefore pointless to analyse past prices in an attempt to predict future prices. In semi-strong form efficient markets, stock prices reflect not only the information contained in past prices but also public information such as earnings and dividend announcements, stock splits, and economic and political news. Semi-strong form efficiency implies that the analysis of published financial statements, for example, does not allow one to systematically earn excess profits. A market is strong-form efficient if prices reflect all types of information whether available publicly or privately. Insiders who trade on private information cannot systematically make excess returns in a strong-form efficient market.

1.1.2 Tests for market efficiency: contentious evidence

Whether, and to what extent, financial markets are efficient is the subject of a longstanding debate in the finance literature, and more broadly in the finance profession. In fact, a large number of studies test the three forms of the efficient market hypothesis, producing a range of mixed findings including evidence for and against market efficiency.

On the one hand, the efficient market hypothesis is supported by a number of studies demonstrating the difficulty of beating the market (e.g., Jensen, 1968; Jensen, 1978; Malkiel, 2005), or the unpredictability of stock returns (e.g., Fama, 1965; Mandelbrot, 1966; Brealy et al., 2005). In addition, event studies look at the pattern of returns around a public event and show that stock prices adjust to new information immediately or over a period of a few days, consistent with a high degree of efficiency (e.g., Fama et al. 1969; Scholes, 1972; Keown and Pinkerton, 1981).[1]

---

[1] For illustration, Keown and Pinkerton (1981) use the event study methodology to evaluate the share price reactions to takeover announcements. They show that share prices start to rise prior to the announcement as the news is incorporated into the prices, then jumps on the date of announcement to reflect the takeover premium. This jump is not followed by a continued trend up or a reversal down, indicating that prices adjust to the public news instantaneously, consistent with the semi-strong form of the efficient market hypothesis.

On the other hand, there is an extensive body of research providing evidence inconsistent with market efficiency, including the large and growing behavioural finance literature. For example, Shiller (1981), in a seminal paper that spurred a long literature on "excess volatility", argues that the amount of volatility in stock markets is well beyond what could be expected if prices are equal to the present value of future cash flows. Henrikson (1984) suggests that fund managers have access to enough private information to generate abnormal returns in excess of expenses and management fees. Lo and MacKinlay (1988) find that stock price short-run serial correlations are not zero, rejecting the hypothesis that stock prices behave as true random walks. De Bondt and Thaler (1985) document that stock prices overreact to extreme bad or good news about firms, supporting substantial market inefficiencies. Evidence related to market anomalies implies that it is possible to predict future prices based on seasonal patterns or firm-specific characteristics such as firm size or firm value, or following earnings announcements, contradicting the semi-strong form of the efficient market hypothesis.[2]

Further, a growing literature seeks to explain these anomalies. Several studies demonstrate that the abnormal returns are caused by chance, the probability of over-reaction being approximately the same as the probability of under-reaction. The anomalies disappear when different models are used, an indication of the joint hypothesis problem.[3] There is also another perspective that tries to explain anomalies by behavioural approaches: investors suffer some cognitive limitations when they have to make decisions, causing irrational investment. Black (1986) introduces the concept of noise traders, namely those who trade on anything other than information.

[2] For instance, returns are found to be higher in January (e.g., Rozeff and Kinney, 1976; Haugen and Lakonishok, 1988), in small firms (e.g., Fama and French, 1993), in firms with low price to earnings and low price-to-book-value ratios (e.g., Basu, 1977; Campbell and Shiller, 1988b; Lakonishok et al., 1994), or in firms with high dividend yield (e.g., Rozeff, 1984; Fama and French, 1988) and so on. The tendency for stock prices to continue to drift upwards (downwards) following good (bad) earnings announcements (an effect known as the "post earnings announcement drift") is one of the earliest documented empirical "anomalies" and has remained a robust and puzzling empirical feature in the decades following its discovery by Ball and Brown (1968).

[3] Fama (1970) states that market efficiency must be tested jointly with an equilibrium-pricing model that provides expected returns to compare to the actual returns. However, the problem is that it is not certain if the model is imperfect or if the market is inefficient if efficiency is rejected. This join hypothesis problem makes it difficult, if not impossible, to definitively test the informational efficiency of a market.

The debate is still ongoing despite numerous attempts made in the literature to address it. Nevertheless, efficient markets in the strict sense are undoubtedly not possible in real financial markets, as perfect efficiency exists only in a competitive market with zero information and trading costs (Fama, 1970). Zhang (1999) states that "even the most competitive markets are not strictly efficient". Therefore, the concept of market efficiency has to accommodate various degrees of efficiency and admit that in practice, perfect efficiency is unlikely to be attainable.

1.1.3 Challenges to measuring stock market efficiency

Recently, empirical finance has become more devoted to assessing how efficient markets are instead of debating whether markets are efficient or not. There are several traditional efficiency measures that are largely based on the concept of weak-form market efficiency, including autocorrelations, variance ratios, the profitability of reversal strategies, information delay measures, profitability of momentum strategies, and intraday return predictability based on past order flow or past returns.

In addition, there are measures of the type or source of information in prices, which also shed some light on informational efficiency. For example, idiosyncratic volatility has been proposed as a measure of firm-specific information in stock returns and thus as a measure of market efficiency. It is commonly estimated as one minus the R^2 from the market model regression.[4] Campbell and Shiller (1988a, 1988b) and Campbell (1991) decompose the unexpected stock returns into cash flow news and discount rate news.

Measuring market efficiency is not an easy task and the existing metrics have several important limitations. For instance, measures like short-term reversals, momentum, and post-earnings announcement drift are estimated at the market level and therefore cannot gauge the efficiency of individual stock prices. Furthermore, the profitability of these "anomalies" may simply be a function of how well-known they are among market participants, rather than reflecting market efficiency more generally. By contrast, high-frequency measures such as autocorrelation and variance

[4] In this model, individual stock returns are regressed on market returns.

ratios can be estimated at the firm level. However, they measure departures from weak-form efficiency rather than capturing the amount of information in prices more generally.

The existing measures have even been shown to produce strange results that are inconsistent with the conventional wisdom when applied across country. For instance, Griffin, Kelly, and Nardari (2010) analyse several of the traditional measures in international markets, and find that emerging markets appear more efficient than developed markets according to the existing market efficiency metrics.[5] Griffin et al. (2010) argue that their results highlight crucial limitations of traditional weak-form efficiency measures and point to the importance of measuring informational aspects of efficiency: "one may need to focus on measuring the informational aspects of efficiency before making meaningful statements about relative efficiency, especially for settings with large disparities in the information environment" (Griffin et al., 2010, p. 3229).

Existing empirical methods that decompose stock return variance into various types of information also have their limitation and produce some puzzling results in recent times. Take the inverse of idiosyncratic volatility, the R^2 measure, as an example. Morck et al. (2013) document an upward trend in R^2 during the period 1997-2010. Under the standard interpretation that higher R^2 implies lower firm-specific information and thus a lower degree of efficiency, the time series trend suggests that market efficiency has been on the decline since the mid-1990s, which is at odds with much of the event study literature that suggests different market changes have generally improved market efficiency in recent years. The standard decomposition of cash flow and discount rate news as per Campbell and Shiller (1988a, 1988b) and Campbell (1991) can only be reliably performed using low-frequency data, which results in limited ability to examine time-series variation in the cash flow/discount rate components.

Another limitation is the inability of existing methods to disentangle information into more refined categories, which is important for understanding the

[5] Griffin et al. (2010) find that momentum trading strategies are more profitable in developed markets, prices deviate more from a random walk in developed markets, and prices in emerging markets incorporate past market returns more quickly than prices in developed markets.

implications of recent changes in markets. For example, some recent studies suggest the rapid growth in passive investing could harm the firm-specific information in stock prices (e.g., Cong and Xu, 2017). Similar concerns have been raised about the effects of high-frequency trading (e.g., Baldauf and Mollner, 2018). In addition, the partition into private and public information is crucial given substantial changes in the regulation of corporate information disclosure in recent decades (e.g., Regulation Fair Disclosure and the Sarbanes Oxley Act).

1.1.4 The presence of noise in stock prices

Noise is defined as the temporary deviations of prices from their equilibrium levels, which arise from microstructure frictions such as the bid-ask spread, as well as "price pressures" created by uninformed trading and imperfect liquidity (Asparouhova et al., 2013). Noise has a significant effect on returns not only at intraday horizons, but also at daily and monthly frequencies. For example, Blume and Stambaugh (1983) and Asparouhova et al. (2013) show that noise at daily frequencies causes an economically meaningful bias in returns. Hendershott et al. (2011) estimate that one-quarter of monthly return variance in NYSE stocks is due to transitory price changes explained by order imbalances and market-makers' inventories. Jegadeesh (1990) and Lehmann (1990) document significant reversals in stock returns at monthly and weekly horizons, respectively, also consistent with the notion that daily, weekly, and monthly returns contain substantial noise.

Despite the economically meaningful impacts of noise, the existing methods for performing stock return variance decompositions mostly ignore the role of noise in stock returns.

1.1.5 Implications/effects of informational efficiency

A growing number of empirical studies show that informational efficiency affects the real economy. For example, low information asymmetry between managers and outsiders, which occurs when stock prices are highly informative, encourages investors to provide more capital to firms in primary markets (Stiglitz and Weiss, 1981). Stock prices also affect corporate investment decisions by conveying new information to corporate managers. In fact, at the firm level, more firm-specific information in stock prices allows managers to improve their investment decisions and thus facilitates efficient corporate investment (Chen et al., 2007; Foucault and Fresard, 2012, 2014; Edman et al., 2017). Financial markets also help facilitate the identification of good investments, thereby improving capital allocation at the industry level and contributing to economic growth (Wurgler, 2000).

The analyses in the existing literature focus on the relation between the information component of stock prices and investment efficiency. There is little direct evidence on how mispricing/noise in stock prices impacts the efficiency of capital allocation.

There are studies of the determinants of market efficiency, especially the impacts of noise traders on market efficiency. These studies do not reach a consensus on the effects of noise traders. Behavioural efficiency supporters argue that more noise traders make the market itself noisier and prices therefore become less efficient. Kyle (1985) shows that more noise traders do not lead to greater inefficiency, and that instead, informed traders respond by trading more aggressively such that the same amount of information is incorporated into prices. Noise traders might even make price more informative by increasing the incentives for information acquisition and production (e.g., Kyle, 1989).

1.2 Research scope and contributions

This thesis aims to extend our understanding of the informational efficiency of markets by developing new measure the different types of information and noise in prices, analysing the determinants of stock market efficiency, and providing new evidence on the real effects of more efficient markets.

The first study (Chapter 2) proposes novel measures of stock market efficiency. Recognizing the limitations of existing measures, this thesis develops a return variance decomposition model that explicitly accounts for noise and simultaneously partitions information into various sources. This research disentangles four components: market-wide information, private firm-specific information revealed through trading, firm-specific information revealed through public sources, and noise. Overall, 31% of the return variance is from noise, 37% from public firm-specific information, 24% from private firm-specific information, and 8% from market-wide information. This approach enhances understanding of the specific sources of information that are impounded into stock prices, and allows the variance decompositions to be performed at the stock level and higher frequencies. These methods allow researchers to examine the time-series and cross-sectional variation in the components of stock return variance and have a number of future research applications.

The second study (Chapter 3) focuses on the effects of noise on existing empirical models, and how the variance decomposition (developed in Chapter 2) can be used to explain the puzzling trends in idiosyncratic volatility (and the R^2 of a market model) during the post-1997 period. This thesis finds that R^2 increases in the post-1997 period as a result of a decrease in noise during this period. The evidence indicates that the market has become more efficient in the past two decades, contrary to what is implied by standard interpretations of R^2 as an inverse measure of efficiency. Chapter 3 also explores how noise affects the standard cash flow/discount rate news decomposition by Campbell and Shiller (1988a, 1988b) and Campbell (1991). This thesis finds that, after accounting for noise in the return generating process, cash flow information plays a considerably larger role than discount rate information in driving individual stock returns. The difference in the importance of discount rate news compared to previous studies stems largely from the fact that noise influences return predictability.

The third study (Chapter 4) examines the relation between noise in stock prices and the efficiency of corporate investment and capital allocation at both the firm and industry levels using listed firms from 42 countries. Consistent with the notion that noise decreases investment efficiency, this research finds strong evidence that noise is negatively associated with the sensitivity of corporate investment to firms' growth opportunities. The research also finds that noise in stock prices decreases the elasticity of industry investment to value added, suggesting noise harms the efficiency of capital allocation across industries. These findings highlight the important real effects of secondary market quality in determining firms' investment behaviour and the efficiency with which capital is allocated.

The fourth study (Chapter 5) provides evidence regarding how individual investors' behaviour, in particular investors' propensity to gamble in stock markets, affects market efficiency. This research develops and validates a simple measure of the amount of gambling in stock markets based on the turnover differences between lottery stocks and non-lottery stocks. It contrasts the estimates of stock market gambling with “traditional” forms of gambling such as casinos and lotteries. The chapter finds that in the cross-section of countries, culture and economic factors are significant drivers of the amount of gambling in both traditional venues and stock markets. Interestingly, the chapter finds a substitution effect wherein restrictions/bans on traditional gambling lead to a spillover of gambling onto stock market(s). Exploiting these spillovers in the identification strategy, this research finds that increased gambling in stock markets makes them more liquid and makes prices reflect more information and less noise. These findings have implications for the use of gambling regulations as a policy instrument to affect financial market quality.

1.3 Thesis outline

This thesis is organized into six chapters. This introductory chapter provides the background and overview of the four studies that follow. Chapter 2 describes the new method developed in this thesis for decomposing stock return variance into various sources of information and noise. Chapter 3 focuses on the effects of noise on existing empirical models. Chapter 4 examines the relation between informational efficiency and capital allocation efficiency. Chapter 5 investigates how gambling in stock markets affects information and noise. Chapter 6 summarizes the conclusions that can be drawn from this thesis and describes some future research directions.

Chapter 2: Measuring the information and noise in prices

2.1 Introduction

The issue of what drives stock price movements is a fundamental question in finance, with implications for understanding risk, informational efficiency, and asset pricing. By understanding the stock return generating process, researchers can address questions such as whether private information is more important than public information, whether the role of market-wide information is increasing or decreasing over time, or how much noise is present in stock price movements. A methodology for measuring the return generating process is also useful for evaluating the impacts of recent phenomena such as the growth in passive investing and algorithmic trading, among others. Hence, this chapter develops a new tool that allows stock returns to be decomposed into various information components while simultaneously allowing for price changes to occur due to non-informational reasons.

There are currently two dominant approaches to decomposing the drivers of stock price changes. One approach exploits the canonical discounted cash flow valuation model to divide a return series into cash flow and discount rate related return components (Campbell and Shiller, 1988a, 1988b; Campbell, 1991). The second decomposes returns into market-wide news and firm-specific news using the R^2 from a regression of stock returns on market returns (Morck, Yeung, and Yu, 2000).

The existing decompositions have limitations, which we overcome with the variance decomposition developed in this chapter. To illustrate, Figure 2.1 shows the time series of the R^2 measure. The R^2 time series suggests that market efficiency has been on the decline since the mid-1990s (higher R^2 implies less firm-specific information is reflected in prices). This is at odds with much of the event study literature that suggests different market changes have generally improved market efficiency in recent years (e.g., Brogaard, Hendershott, and Riordan, 2014; Comerton-Forde and Putnins, 2015).

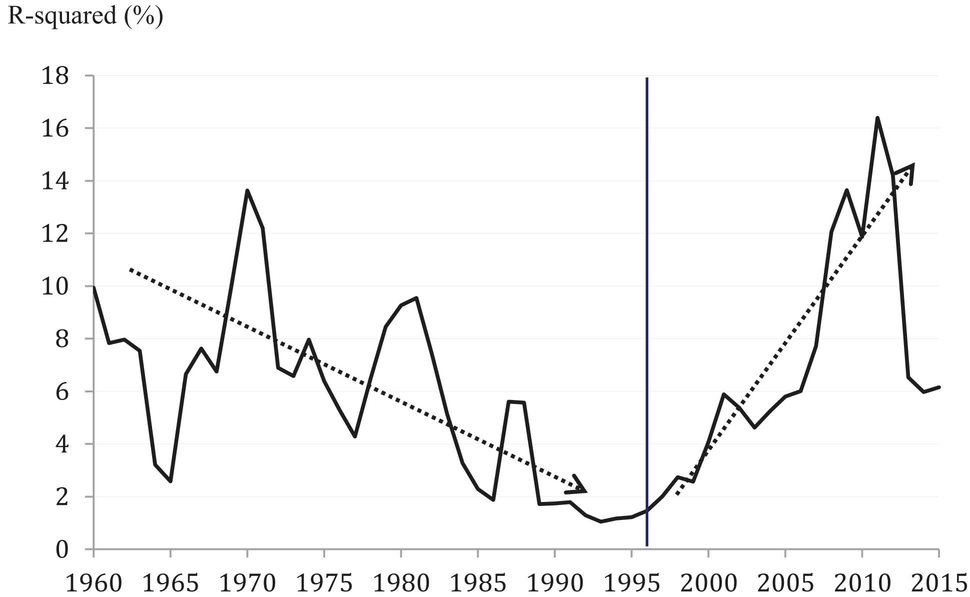


Figure 2.1. R^2 through time.
This figure shows the time-series trend in R^2 from 1960 to 2015. R^2 is calculated separately for each stock by regressing individual daily stock returns on daily market returns, and then averaging across stocks. The sample consists of stocks listed on NYSE, AMEX, and NASDAQ.

The first limitation of existing decompositions arises from ignoring the role of noise in stock returns. Temporary deviations of prices from their equilibrium levels are part of the return generating process (e.g., Hendershott and Menkveld, 2014; Asparouhova, Bessembinder, and Kalcheva, 2010; Asparouhova, Bessembinder, and Kalcheva, 2013). These price deviations, or "noise," arise from microstructure frictions such as bid-ask spreads, nonsynchronous trading, discrete price grids, and temporary price impacts of order imbalances, as well as changes in investor sentiment or other behavioural factors, in combination with limits to arbitrage (Asparouhova et al., 2013). Noise has a significant effect on returns at daily and monthly frequencies, not just intraday horizons. For example, Blume and Stambaugh (1983) and Asparouhova et al. (2013) show that noise at daily frequencies causes an economically meaningful bias in returns, equal to 50% or more of the corrected estimate and is able to explain the size effect. Jegadeesh (1990) and Lehmann (1990) document significant reversals in stock returns at monthly and weekly horizons, respectively, also consistent with the notion that daily, weekly, and monthly returns contain substantial noise.

Noise can distort the existing measures. For instance, an increase in R^2 is generally interpreted as indicating a relative decrease in the amount of firm-specific information in prices. However, if returns also contain noise, an alternative interpretation of a high R^2 is that there has been a decrease in the magnitude of idiosyncratic pricing errors and therefore an increase in R^2 does not necessarily indicate deteriorating informational efficiency.

The second limitation is the inability of existing methods to disentangle information into more refined categories. The partition into market-wide and firm-specific information has long been of interest in finance as firm-specific information is vital to efficient resource allocation across firms. Some recent studies suggest the rapid growth in passive investing could harm the firm-specific information in stock prices (e.g., Cong and Xu, 2017). Similar concerns have been raised about the effects of high-frequency trading (e.g., Baldauf and Mollner, 2018). Further, the partition into private and public information is important given substantial changes in the regulation of corporate information disclosure in recent decades (e.g., Regulation Fair Disclosure and the Sarbanes Oxley Act). Such regulations could result in better disclosure crowding out private information acquisition, with implications for the profitability of active investing and the levels of adverse selection. Having a variance decomposition methodology to shed light on cross-sectional variation and general time trends can aid research on these types of information topics.

This chapter proposes a new return variance decomposition model that explicitly accounts for noise and partitions information into various sources. Specifically, this study distinguishes between market-wide information, firm-specific information revealed through trading on private information, and firm-specific information revealed through public news. This approach allows for a more nuanced understanding of the specific sources of information that is impounded into stock prices along economically meaningful dimensions. It also allows the variance decompositions to be performed at higher frequencies (e.g., annual decompositions of daily returns) and therefore allows researchers to examine time-series variation in the components of stock return variance (most existing variance decompositions rely on low frequencies to mitigate the effects of noise, precluding time-series analysis).

The empirical return variance decomposition model draws on the market microstructure toolkit where separating temporary price movements driven by frictions from permanent price movements is commonplace. For example, Hasbrouck (1993) separates noise from information through a temporary-permanent component decomposition. Permanent innovations are those that affect the long-run expected value of the security, whereas temporary innovations affect prices in the short-run, but have no effect on the long-run expected value of the security. Permanent innovations in prices reflect innovations in the fundamental value driven by new information, whereas temporary innovations are pricing errors. These deviations from the fundamental value can arise from various well-documented market frictions including the bid-ask bounce, discrete price grids, temporary price pressures created by uninformed buying or selling, and non-synchronous trading.

This study builds on Hasbrouck's original decomposition by adapting the model to daily returns and separating innovations in the fundamental value into market-wide information, firm-specific information revealed through trading on private information, and firm-specific information revealed through other sources (public information). This gives rise to four components of return variance that map to the theoretical model. We estimate the model using daily returns on all common stocks listed on the NYSE, AMEX, and NASDAQ between 1960 and 2015, performing the variance decomposition separately for every stock in every year. This approach minimizes the issue of non-stationarity and allows us to examine how the variance components change in the cross-section and through time.

Intriguingly, this study finds that roughly 31% of daily return variance is noise. Firm-specific information accounts for the majority (61%) of stock return variance, with market-wide information accounting for the remaining 8% of variance in the full sample. We further partition firm-specific information and find that in the full sample, public firm-specific information plays a larger role than private firm-specific information that is impounded into prices through trading (37% and 24% of variance, respectively). While the estimates suggest that noise makes up an economically meaningful share of daily stock return variance, the estimate is substantially lower than estimates of noise at intra-day horizons (82%).[6] Our estimate is consistent with Hendershott and Menkveld (2014) who find that the ratio of "price pressure" (distortions of the midquote price from the efficient price) to the variance of efficient

[6] Extrapolating from Hasbrouck (1991b, 1993) for the year 1989.

prices is 33% in their sample of 697 NYSE stocks during 1994-2005. It is also consistent with Hendershott et al. (2011) who estimate that one-quarter of monthly return variance in NYSE stocks is due to transitory price changes explained by order imbalances and market-makers' inventories. Finally, the estimate is also consistent with the economically meaningful return reversals at daily, weekly, and monthly horizons, which reflect temporary departures from efficient prices due to imperfect liquidity (e.g., Jegadeesh, 1990; Lehmann, 1990; Avramov, Chordia, and Goyal, 2006; Nagel, 2012).

This study finds substantial time series variation in the components of variance. Some key trends stand out. First, noise increases from the 1970s to the mid-1990s, in particular around a period of collusion by dealers that widened bid-ask spreads. The subsequent decline in noise corresponds to a period with a general improvement in liquidity and exogenous decreases in tick sizes (minimum price increments). Separating the sample by firm size shows that the noise component decreases monotonically with firm size. Larger firms have less noise in their prices, as expected. When separating the sample by industry we observe only minor fluctuations in the different components of variance, suggesting that the findings are not specific to a particular industry, nor are they driven by a certain segment of the economy.

Second, the role of firm-specific information has increased through time, driven largely by increases in the amount of public firm-specific information that is reflected in prices. This trend is consistent with increasing informational efficiency through time, which one may expect given a variety of regulations such as the Regulation Fair Disclosure (2000) and Sarbanes Oxley Act (2002) that have increased the quality and quantity of corporate disclosure. Third, market-wide information has become a less important driver of stock returns over time. While market-wide information tends to spike during crises, it has generally declined from around 15% of variance to around 5-10% in recent years.

During the 55-year sample period, markets have changed dramatically. The number of exchanges has increased, bid-ask spreads have decreased, intermediaries have transitioned from dealers to market-makers to high-frequency traders, and investor horizons have declined. In addition, the quantity of information has ballooned and at the same time the cost and latency of accessing it has plummeted. These time varying changes can impact the various components of the price process differently.

Because we re-calculate the decomposition each year for each stock, our measure is able to vary along with these changes in the market.

We provide three detailed examples of how this decomposition can be useful for examining prominent questions in finance. First, we look at the role tick size has on the noise component. The tick size was reduced from eighths of a dollar to sixteenths of a dollar in 1997 and then from sixteenths to pennies in 2001. Chordia, Roll, and Subrahmanyam (2008) show that informational efficiency improved significantly around the change from quoting in eighths of a dollar to sixteenths, with the second event having a smaller effect. We exploit this natural experiment and examine how the noise share of variance is affected by the reduction in tick size from eighths of a dollar to sixteenths. We find that following the tick size reduction there is a substantial decrease in the noise in stock returns. Stocks experiencing a larger change in their relative tick size exhibit a larger decrease in noise.

Second, we examine exogenous shocks to analyst coverage. We find that a decline in analyst coverage harms the quality of the corporate information environment and reduces the proportion of public firm-specific information in prices. We use brokerage mergers and closures as plausible exogenous shocks to analyst coverage because the termination of coverage is not driven by the characteristics or behaviour of the firm (see Hong and Kacperczyk, 2010; Kelly and Ljungqvist, 2012; Brogaard et al., 2018). Additionally, exogenous decreases in analyst coverage are also associated with an increase in the relative level of noise in prices.

Third, we compare our variance components to two other measures of information in prices: the Hou and Moskowitz (2005) delay metric and the variance ratio (French and Roll, 1986; Chordia, Roll, and Subrahmanyam, 2011). The first measure focuses specifically on market-wide information, while the second measure focuses on information generally. We find a strong inverse relation between the delay metric and the share of variance that is attributable to market-wide information in our variance decomposition, whereas we find the variance ratio is positively related to all of the information components of variance and negatively related to the amount of noise in prices. The results of the decomposition reassuringly fit with what the existing, less flexible measures would predict.

Being able to accurately measure the amount and type of information in asset prices is vital to understanding the impacts of recent trends and innovations in finance.

This chapter makes a methodological contribution to the literature by developing a richer and more general variance decomposition that allows the separation of variance into multiple information and noise components.

This chapter proceeds as follows. Section 2.2 presents related literature, and Section 2.3 describes the empirical model for variance decomposition. Section 2.4 details further validation tests, and Section 2.5 concludes.

2.2 Related literature

This chapter relates to three main bodies of literature. The first is stock return variance decompositions. For example, Campbell and Shiller (1988a, 1988b) and Campbell (1991) decompose the variance of unexpected stock returns into two components: cash flow news and discount rate news. Chen, Da, and Zhao (2013) also decompose stock return variance into these components but with a method that directly incorporates cash flow forecasts. A different partition is used in Campbell et al. (2001), who decompose stock return variations into three components: a market-wide return, an industry-specific residual, and a firm-specific residual. A simple but widely used method for stock return decomposition is offered by Roll (1988), who simply distinguishes market-wide variations and firm-specific variations. Our study differs from these existing methods by using random walk variance decompositions, drawn from the market microstructure literature, to separate information from noise, and partition information into a more granular set of components. This separation of noise and information provides a measure related to market efficiency and decomposes the contributions of different information sources.

Second, our study is related to a growing body of literature on the significant impact of noise on asset prices and returns. For instance, Blume and Stambaugh (1983) show that zero-mean noise in prices leads to a positive bias in mean returns. Asparouhova et al. (2010) find that noisy prices lead to biases in intercept and slope coefficients obtained in any OLS regression using returns as the dependent variable. In addition, Asparouhova et al. (2013) find that correcting for the effects of noise in prices has significant effects on return premium estimates from monthly return data. Motivated by this strand of literature, we incorporate noise into our model as an important factor that might have impacts on stock price movements. In turn, by providing a convenient method to estimate noise in prices, our approach provides a tool that enables future research to more systematically examine the drivers and effects of noise.

Third, our study contributes to the extensive literature measuring market efficiency. For instance Bai, Philippon, and Savov (2016) use quarterly cross-sectional regressions of the extent to which market prices predict earnings as a measure of market-wide efficiency. In contrast, our method provides a more granular measure at the individual *stock* level, rather than at the market level, and does not require earnings information that may be subject to reporting bias.

There are several traditional efficiency measures that are largely based on the concept of weak-form market efficiency, including autocorrelations, variance ratios, reversal strategies, delay measures, post-earnings drift, profitability of momentum strategies, and intraday return predictability based on past order flow or past returns. Rosch, Subrahmanyam, and van Dijk (2017) examine the dynamics of market efficiency in the United States using the first principal component of four existing intra-day efficiency measures, while Griffin, Kelly, and Nardari (2010) analyse several of these traditional measures across international markets. Interestingly, Griffin et al. (2010) find that momentum trading strategies are more profitable in developed markets, prices deviate more from a random walk in developed markets, and prices in emerging markets incorporate past market returns more quickly than prices in developed markets. Thus, the existing measures produce results that are inconsistent with the conventional wisdom that emerging markets are less efficient than developed markets. Griffin et al. (2010) argue that their results highlight crucial limitations of traditional weak-form efficiency measures and point to the importance of measuring informational aspects of efficiency.

Our approach follows this call by focusing on decomposing the *information* in prices rather than contributing to the set of weak-form efficiency measures. Besides being based on the broader asset pricing methodology, our measure also differs from existing approaches, including Bai, Philippon, and Savov (2016), in that it not only measures efficiency in terms of information versus noise, but also decomposes the specific sources of information that is reflected in prices. The decomposition of information provides a more complete picture of the nature of market efficiency and how it evolves through time and in the cross-section.

2.3 Empirical model for variance decomposition

Our sample consists of all common stocks listed on the NYSE, AMEX, and NASDAQ. We use daily data on returns, prices, market capitalizations, volumes, and sectors for the period from 1960 to 2015 from the Center for Research in Security Prices (CRSP). Appendix 2.A contains a summary of variable definitions. Section 2.3.1 details construction of the baseline empirical model. Section 2.3.2 presents the estimated variance components in the full sample and compares the noise in daily returns to that of intraday returns. Section 2.3.3 characterizes how the variance components change through time and how they vary in the cross-section of stocks. Finally, to further validate the empirical model we examine how the variance components change around exogenous shocks to tick sizes and analyst coverage and how they relate to other measures of information and noise (Section 2.3.4).

2.3.1 Baseline variance decomposition model

We propose a variance decomposition model that separates noise and various sources of information. Our approach to separating noise from information builds on Hasbrouck (1993) who shows how temporary-permanent decompositions can be used on stock returns to separate temporary pricing errors from innovations in the fundamental value.[7] Permanent innovations in stock prices are those that affect the long-run expected value of the security, whereas temporary innovations affect prices in the short-run, but have no effect on the long-run expected value of the security. Permanent innovations in prices therefore reflect innovations in the fundamental value driven by new information being impounded into prices, whereas temporary innovations are pricing errors (deviations from fundamentals) generically referred to as “noise.” Noise is caused by many factors including the bid-ask spread, discrete price grids, illiquidity, temporary price pressures created by uninformed buying or selling, and non-synchronous trading.

Our variance decomposition departs from Hasbrouck (1993) in two important ways. First, we adapt the approach so that it can be applied at lower frequencies such as daily returns (Hasbrouck (1993) models intraday trade-to-trade returns). We

[7] Similar temporary-permanent decompositions are also used in empirical macroeconomics.

undertake an array of validation tests to verify that the variance decomposition produces reasonable estimates. Second, we push the variance decomposition further to separate the information into market-wide information, firm-specific information revealed through trading on private information, and firm-specific information revealed through public information. This gives rise to four components of variance that map to the theoretical model. These components are illustrated in Figure 2.2.

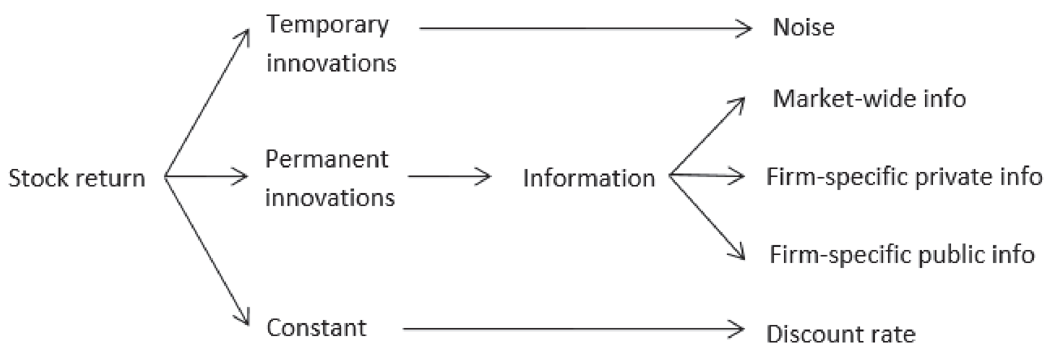


Figure 2.2. Stock return components.
Stock returns are decomposed into temporary innovations (noise), three types of information (permanent innovations), and a constant (discount rate). The first four of these are the variance components in the baseline model, while the fifth (the discount rate) does not contribute to variance in the model.

Consider p_t, the logarithm of the observed price at time t, as the sum of two components:

$$p_t = m_t + s_t \, , \tag{2.1}$$

where m_t is the efficient price and s_t is the pricing error. The pricing errors can have a temporary (short-run) effect on the price, but they do not affect price in the long run (no permanent effect). m_t follows a random walk with drift μ, and innovations w_t:

$$m_t = m_{t-1} + \mu + w_t \, . \tag{2.2}$$

The innovations reflect new information about the stock's fundamentals and are thus unpredictable, $\mathrm{E}_{t-1}[w_t] = 0$. The drift is the discount rate on the stock over the next period (day).[8] The stock return is therefore:

$$r_t = p_t - p_{t-1} = \mu + w_t + \Delta s_t \ . \tag{2.3}$$

Suppose that there are three sources of information impounded into stock prices: market-wide information, private firm-specific information incorporated through trading, and public firm-specific information such as firm-specific news disseminated in company announcements and by the media. The random-walk innovations, w_t, in (2.2) can then be decomposed into three parts:

$$w_t = \beta \varepsilon_{r_m,t} + \delta \varepsilon_{x,t} + u_t \ , \tag{2.4}$$

and thus

$$r_t = \mu + \beta \varepsilon_{r_m,t} + \delta \varepsilon_{x,t} + u_t + \Delta s_t \ , \tag{2.5}$$

where $\varepsilon_{r_m,t}$ is the unexpected innovation in the market return, $\beta \varepsilon_{r_m,t}$ reflects the market-wide information incorporated into stock prices, $\varepsilon_{x,t}$ is an unexpected innovation in signed dollar volume, $\delta \varepsilon_{x,t}$ is the firm-specific information revealed through trading on private information, and u_t is the remaining part of firm-specific information that is not captured by trading on private information. The separation of firm-specific information into private information associated with trading and public

[8] Later, in Chapter 3, when separating cash flow and discount rate news we allow the discount rate to be time-varying.

information not associated with trading follows Hasbrouck (1991a, 1991b). The pricing error, Δs_t, can be correlated with the innovations in the efficient price, w_t.

Equation (2.5) can be modeled as a vector auto-regression (VAR) to account for serial correlations in returns and other explanatory variables. We use a structural VAR with five lags to allow a full week of lagged effects:[9]

$$r_{m,t} = a_0 + \sum_{l=1}^{5} a_{1,l} r_{m,t-l} + \sum_{l=1}^{5} a_{2,l} x_{t-l} + \sum_{l=1}^{5} a_{3,l} r_{t-l} + \varepsilon_{r_m,t} \tag{2.6}$$

$$x_t = b_0 + \sum_{l=0}^{5} b_{1,l} r_{m,t-l} + \sum_{l=1}^{5} b_{2,l} x_{t-l} + \sum_{l=1}^{5} b_{3,l} r_{t-l} + \varepsilon_{x,t}$$

$$r_t = c_0 + \sum_{l=0}^{5} c_{1,l} r_{m,t-l} + \sum_{l=0}^{5} c_{2,l} x_{t-l} + \sum_{l=1}^{5} c_{3,l} r_{t-l} + \varepsilon_{r,t},$$

where $r_{m,t}$ is the market return, x_t is the signed dollar volume of trading in the given stock with positive values capturing net buying and negative values capturing net selling, and r_t is the stock return.[10] The lags of stock returns account for short-term momentum as well as reversals that can be driven by temporary price impacts from trading (e.g., Hendershott and Menkveld, 2014). The lags of signed dollar volume account for persistence in order flow (e.g., Hasbrouck, 1988). Finally, the lags of market returns account for first-order serial correlation in market returns due to non-

[9] There are two approaches to resolving the contemporaneous relations between variables in the VAR. One is to use a structural VAR that explicitly defines the assumptions about contemporaneous causality using economic arguments. The second is to use a reduced form VAR and apply Cholesky factorization to the variance-covariance matrix of reduced form innovations, which itself implies a recursive causal chain from the first to the last variable in the system. We take the former approach. The structural VAR approach has at least two advantages: (i) while both approaches must make assumptions about excluded contemporaneous relations to facilitate identification, the structural model allows the assumptions to be guided by economic reasoning rather than arbitrary allocation; and (ii) the structural approach produces a unique variance decomposition whereas the reduced form approach produces a decomposition that is sensitive to the ordering of variables in the model.

[10] Our measure of the market returns is the daily value-weighted market return excluding American Depository Receipts. We use market returns rather than a broader collection of factors such as market/size/value because we seek to identify the contribution of *market-wide* information to stock return variance. The variance decomposition could easily be extended to quantify the role of a variety of factor information in prices. We construct a proxy for the daily signed dollar volume of each stock as a product of price, volume, and sign of the return, similar to Pastor and Stambaugh (2003).

synchronous trading (e.g., Scholes and Williams, 1977) as well as delayed stock price reactions to market-wide information (e.g., Hou and Moskowitz, 2005).

The structural VAR above embeds contemporaneous relations between the variables. First, market-wide information can be reflected in stocks contemporaneously, but because each stock is a small part of the market index, individual stock returns and trades have a negligible contemporaneous impact on the market return. Second, trading activity in a stock can be contemporaneously caused by market returns and can contemporaneously cause changes in the stock price, but not vice versa. To the extent that returns can trigger trading activity contemporaneously (within the same day), such reverse causality would be attributed to trading activity having caused the returns. As a result of incorporating the contemporaneous relations between variables the structural VAR errors $\{\varepsilon_{r_m,t}, \varepsilon_{x,t}, \varepsilon_{r,t}\}$ are contemporaneously uncorrelated.

We separately estimate the VAR for every stock-year using daily data. Keeping the estimation windows to one-year periods mitigates concerns about non-stationarity or long-run trends in the model parameters and allows us to examine how the variance components change through time.

Before proceeding to the variance decomposition, Table 2.1 reports reduced-form VAR coefficient estimates averaged across the individual VAR models.[11] Below each average, in parentheses, Table 2.1 reports the percentage of negative statistically significant (at 5%) coefficients (first number in the parentheses) and the percentage of positive statistically significant (at 5%) coefficients (second number in the parentheses). The coefficients in the market return equation (Panel A) show a tendency for positive first-order serial correlation in market returns, consistent with the effects of non-synchronous trading (Scholes and Williams, 1977) and slow diffusion of market-wide information. In contrast, lags of other variables (trading in individual securities and individual stock returns) do not explain current market returns. The coefficients in the signed dollar volume equation (Panel B) show a

[11] The variance decomposition is based on the structural VAR in (2.6), but for practical reasons we follow the common practice of first estimating the reduced-form VAR and then using the coefficients and reduced form error covariances to compute the structural VAR impulse response functions and structural VMA. These structural VAR impulse response functions and structural VMA are used for the variance decomposition.

tendency for buying to follow positive market returns. They also show positive serial correlation in daily signed dollar volume, consistent with persistence in order flow (e.g., Hasbrouck (1988) and many subsequent studies). The coefficients in the individual stock return equation (Panel C) show that stock returns tend to be positively related to lagged market returns consistent with the known slow diffusion of market-wide information (e.g., Hou and Moskowitz, 2005). They also show a tendency in some stock-years for trading (innovations in signed dollar volume) to impact returns not only contemporaneously but also with a lag, suggesting that at times the information in trading takes more than one day to be fully reflected in prices. They also indicate the presence of negative serial correlation in stock returns at daily frequencies out to approximately four days, consistent with reversals of pricing errors due to price pressure (e.g., Hendershott and Menkveld, 2014). The VAR coefficient estimates support the use of five lags in the VAR because by the fifth lag very few coefficients are statistically different from zero (besides the 5% that would be expected by chance at the 95% confidence level).

In addition to the average coefficients, which reveal the lead-lag relations between variables, Table 2.1 also reports the average correlations of the reduced-form VAR residuals for pairs of variables. These correlations reveal the contemporaneous relations between innovations in the variables. Innovations in signed dollar volume are contemporaneously correlated with returns of individual stocks and with market returns. These correlations are consistent with buying pressure pushing prices up as well as positive returns inducing buying (and vice versa for negative returns). There is also a positive contemporaneous correlation between individual stock returns and market returns consistent with individual stocks contributing to the market return but also market returns reflecting market-wide information, which is impounded in individual stock prices.

Table 2.1. VAR coefficient estimates.

This table reports the mean coefficient estimates and the mean correlation between residuals for the baseline VAR model used to perform the variance decomposition. The VAR model is estimated separately for each stock in each year using daily observations. For the purpose of this table, each of the model coefficients is then averaged across stocks and years and reported in the table. Below each coefficient average, in parentheses we report the percentage of negative statistically significant (at 5%) coefficients (first number in the parentheses) and the percentage of positive statistically significant (at 5%) coefficients (second number in the parentheses). The correlations column is computed similarly, but rather than reporting coefficients it reports the correlations of the residuals for pairs of variables in the VAR. The variables used in the VAR are: daily market returns in basis points ($r_{m,t}$), daily signed dollar volume in \$ thousands (x_t), and daily stock returns in basis points (r_t). The columns $l = 1$ to $l = 5$ correspond to lags of the independent variables. The sample consists of stocks listed on NYSE, AMEX, and NASDAQ from 1960 to 2015.

Dependent variable	Independent variable	$l = 1$	$l = 2$	$l = 3$	$l = 4$	$l = 5$	Correlation
Panel A: Market return equation							
$r_{m,t}$	$r_{m,t-l}$	0.104	-0.032	0.018	-0.005	-0.010	
		(2.06%, 37.93%)	(8.13%, 0.34%)	(1.69%, 3.48%)	(5.05%, 1.81%)	(2.14%, 3.09%)	
	x_{t-l}	0.069	0.024	0.033	0.081	-0.034	0.134
		(2.36%, 2.83%)	(2.51%, 2.54%)	(2.41%, 2.48%)	(2.42%, 2.63%)	(2.40%, 2.44%)	(0.57%, 42.81%)
	r_{t-l}	-0.002	-0.0001	0.0006	0.0003	-0.001	0.229
		(3.76%, 3.07%)	(3.26%, 3.19%)	(3.04%, 3.14%)	(2.90%, 3.15%)	(3.22%, 2.83%)	(0.39%, 61.85%)
Panel B: Signed dollar volume equation							
x_t	$r_{m,t-l}$	-0.650	0.013	0.612	-0.529	-0.785	0.134
		(2.27%, 10.18%)	(2.42%, 2.91%)	(1.75%, 3.50%)	(2.05%, 2.85%)	(1.97%, 2.93%)	(0.57%, 42.81%)
	x_{t-l}	0.025	-0.016	0.003	-0.008	0.004	
		(8.69%, 14.74%)	(9.21%, 5.75%)	(5.45%, 6.14%)	(5.76%, 4.52%)	(4.26%, 5.15%)	
	r_{t-l}	1.736	0.631	0.196	-0.002	0.404	0.615
		(9.55%, 8.45%)	(4.66%, 4.83%)	(3.60%, 3.82%)	(3.01%, 3.50%)	(2.79%, 3.19%)	(0.05%, 99.42%)
Panel C: Stock return equation							
r_t	$r_{m,t-l}$	0.247	0.045	0.089	0.052	0.050	0.229
		(2.10%, 21.77%)	(2.93%, 4.78%)	(1.71%, 6.02%)	(2.46%, 4.37%)	(2.36%, 4.56%)	(0.39%, 61.85%)
	x_{t-l}	0.972	-0.077	-0.059	-0.150	-0.037	0.615
		(2.56%, 11.37%)	(3.06%, 3.97%)	(2.68%, 3.10%)	(2.80%, 2.66%)	(2.65%, 2.51%)	(0.05%, 99.42%)
	r_{t-l}	-0.112	-0.060	-0.030	-0.022	-0.007	
		(31.77%, 5.47%)	(16.40%, 2.61%)	(8.88%, 2.68%)	(6.27%, 2.58%)	(4.27%, 3.07%)	

Next we transform the VAR in Equation (2.6) to an infinite order structural vector moving average (VMA) model:

$$r_{m,t} = A_0 + \sum_{l=0}^{\infty} A_{1,l}\varepsilon_{r_m,t-l} + \sum_{l=1}^{\infty} A_{2,l}\varepsilon_{x,t-l} + \sum_{l=1}^{\infty} A_{3,l}\,\varepsilon_{r,t-l} \quad (2.7)$$

$$x_t = B_0 + \sum_{l=0}^{\infty} B_{1,l}\varepsilon_{r_m,t-l} + \sum_{l=0}^{\infty} B_{2,l}\varepsilon_{x,t-l} + \sum_{l=1}^{\infty} B_{3,l}\,\varepsilon_{r,t-l}$$

$$r_t = C_0 + \sum_{l=0}^{\infty} C_{1,l}\varepsilon_{r_m,t-l} + \sum_{l=0}^{\infty} C_{2,l}\varepsilon_{x,t-l} + \sum_{l=0}^{\infty} C_{3,l}\,\varepsilon_{r,t-l}.$$

The VMA form of the model is useful to glean the intuition for the temporary-permanent decomposition that separates noise from the innovations in the efficient price. The permanent effect on a stock's returns from a shock to the arrival of market-wide information, $r_{m,t}$ (unanticipated market returns), is given by $\theta_{r_m} = \sum_{l=0}^{\infty} C_{1,l}$. This is also the cumulative impulse response tracing time forward to the point where the response to the shock stabilizes. Similarly, the permanent effect on a stock's returns from trading activity, and from a shock to the stock's returns that is not due to market-wide information nor trading is $\theta_x = \sum_{l=0}^{\infty} C_{3,l}$, and $\theta_r = \sum_{l=0}^{\infty} C_{4,l}$, respectively. We estimate θ_{r_m}, θ_x, and θ_r from the impulse response functions of the structural model.

The information-driven innovation in the efficient price is given by $w_t = \theta_{r_m}\varepsilon_{r_m,t} + \theta_x\varepsilon_{x,t} + \theta_r\varepsilon_{r,t}$. The efficient price drift is given by $\mu = A_0$. The innovation in the pricing error is given by $\Delta s_t = r_t - \mu - w_t = r_t - A_0 - \theta_{r_m}\varepsilon_{r_m,t} - \theta_x\varepsilon_{x,t} - \theta_r\varepsilon_{r,t}$. It follows that the variance of information-driven innovations in the efficient price is $\sigma_w^2 = \theta_{r_m}^2\sigma_{\varepsilon_{r_m}}^2 + \theta_x^2\sigma_{\varepsilon_x}^2 + \theta_r^2\sigma_{\varepsilon_r}^2$. Recall, the structural model errors are contemporaneously uncorrelated by construction and therefore the covariance terms are all zero. The contribution to the efficient price variation from each of the information components is $\theta_{r_m}^2\sigma_{\varepsilon_{r_m}}^2$ (market-wide information), $\theta_x^2\sigma_{\varepsilon_x}^2$ (private firm-specific information), and $\theta_r^2\sigma_{\varepsilon_r}^2$ (public firm-specific information). The variance of noise, σ_s^2, is computed from the time-series of Δs_t.

To examine the contribution of each component in the total stock return variance, we construct two groups of new measures, the first being contributions to variance:

$$MktInfo = \theta_{r_m}^2 \sigma_{\varepsilon_{r_m}}^2$$

$$PrivateInfo = \theta_x^2 \sigma_{\varepsilon_x}^2$$

$$PublicInfo = \theta_r^2 \sigma_{\varepsilon_r}^2 \tag{2.8}$$

$$Noise = \sigma_s^2,$$

and the second being shares of variance:

$$\begin{gathered} MktInfoShare = \theta_{r_m}^2 \sigma_{\varepsilon_{r_m}}^2 / (\sigma_w^2 + \sigma_s^2) \\ PrivateInfoShare = \theta_x^2 \sigma_{\varepsilon_x}^2 / (\sigma_w^2 + \sigma_s^2) \\ PublicInfoShare = \theta_r^2 \sigma_{\varepsilon_r}^2 / (\sigma_w^2 + \sigma_s^2) \\ NoiseShare = \sigma_s^2 / (\sigma_w^2 + \sigma_s^2). \end{gathered} \tag{2.9}$$

Accordingly, $MktInfo$, $PrivateInfo$, $PublicInfo$, and $Noise$ are the variance contributions of market-wide information, trading on private firm-specific information, firm-specific information other than that revealed through trading, and noise, respectively. $PrivateInfoShare$, $PublicInfoShare$, and $MktInfoShare$ are corresponding shares of variance from those various sources of stock price movements.[12] These three shares sum to the contribution of the efficient price innovations (recall the covariances between these components are all zero by construction). Meanwhile, $NoiseShare$ reflects the relative importance of pricing errors due to illiquidity, price pressures, or other microstructure frictions.

[12] Rather than using the total variance of returns in the denominator of the variance shares, we instead take the sum of the efficient price innovations variance and the noise variance, which excludes the covariance between the efficient price component and Δs. In Appendix 2.B we show that ignoring this covariance term has a negligible effect on the overall estimates of variance shares.

As each stock may have different dynamics we perform the variance decomposition on each stock separately. In addition, because we examine a long time series in which dynamics and the drivers of return variance might change through time, we conduct the analysis separately year by year. Therefore, we estimate the variance decomposition detailed above separately for each stock-year.

2.3.2 Full sample estimates of variance components

Table 2.2 Panel A reports the estimated variance shares from the baseline model above for the full sample (all US stocks from 1960 to 2015). Recall the variance decomposition is performed separately for every stock-year. From the stock-year estimates we calculate variance-weighted averages of each component.[13] The results show that market-wide information is the smallest component and accounts for around 8% of stock return variance, while firm-specific information accounts for 61% (summing $PrivateInfoShare$ and $PublicInfoShare$). Most of the firm-specific information is impounded in prices through public information (37% of variance), while firm-specific private information that is impounded through trading accounts for around 24% of variance. Finally, noise accounts for a fairly substantial 31% of daily stock return variance overall.

[13] We use variance-weighted averages in our baseline results for comparability with Morck et al. (2000, 2013). The motivation is that when trying to understand the drivers of return variance, it makes sense to consider stocks weighted by their variance. Alternatively, using equal-weighting across stocks produces similar results (trends and levels of the components). Using value-weighting across stocks produces lower estimates of the level of noise (16% lower in the pooled sample) and higher estimates of the information components of variance, consistent with the notion that smaller stocks have noisier prices. More direct evidence on the cross-sectional differences in the variance components is presented in Section 2.3.4.

Table 2.2. Stock return variance components in the baseline model.
This table reports the mean variance shares (expressed as percentages of variance) for the period from 1960 to 2015. Stock return variance is decomposed into market-wide information ($MktInfoShare$), private firm-specific information ($PrivateInfoShare$), public firm-specific information ($PublicInfoShare$), and noise ($NoiseShare$). Panel A reports full sample averages. Panel B splits the sample into two sub-periods from 1960 to 1996, and from 1997 to 2015. Panel C groups stocks into quartiles by size (market capitalization) with quartiles formed separately each year. Panel D groups stocks into major industry groups: the *Consumer* group comprises the industries Consumer Durables, NonDurables, Wholesale, Retail, and some Services (Laundries, Repair Shops); the *Healthcare* group comprises the industries Healthcare, Medical Equipment, and Drugs; the *Manufact* group comprises the industries Manufacturing, Energy, and Utilities; the *HiTech* group comprises the industries Business Equipment, Telephone and Television Transmission; and the *Other* group comprises all other industries. The variance component shares are calculated separately for each stock in each year then averaged across stocks within the corresponding quartile or group. We also report the differences in means for the post-1997 period minus the pre-1997 period (Panel B) and quartile 1 minus quartile 4 (Panel C) and corresponding t-statistics in parentheses. ***, **, and * indicate statistically significant differences at the 1%, 5%, and 10% levels using standard errors clustered by stock and by year. The sample consists of stocks listed on NYSE, AMEX, and NASDAQ.

	$MktInfoShare$ (%)	$PrivateInfoShare$ (%)	$PublicInfoShare$ (%)	$NoiseShare$ (%)
Panel A: Full sample				
	8.24	23.88	37.11	30.78
Panel B: Sub-periods				
1960-1996	7.00	21.89	35.63	35.47
1997-2015	9.58	26.03	38.71	25.69
Difference (Post-Pre 1997)	2.57	4.14	3.07	-9.78
	(1.73)*	(2.66)***	(3.37)***	(-5.42)***
Panel C: Quartiles by size (market capitalization)				
Q1=low	4.53	21.60	37.79	36.09
Q2	8.64	24.71	37.46	29.19
Q3	13.93	27.47	36.96	21.65
Q4=high	21.51	30.17	31.87	16.45
Difference (Q1-Q4)	-16.98	-8.58	5.92	19.64
	(-15.57)***	(-5.16)***	(6.16)***	(12.81)***
Panel D: Industry groups				
Consumer	7.69	23.24	36.49	32.59
Healthcare	7.32	28.57	37.39	26.72
HiTech	9.43	26.09	37.37	27.11
Manufact	9.15	24.29	35.02	31.55
Other	7.33	20.26	38.35	34.07

Before exploring the time-series and cross-sectional patterns in these variance shares, we consider how these estimates, in particular the noise in returns, compare to other estimates. The first comparison is with intraday returns (e.g., trade-to-trade), which is where similar temporary-permanent decompositions were first used to separate noise from information. Extrapolating from Hasbrouck (1991b, 1993), the implied noise share in intraday trade-to-trade returns is around 82% in US stocks in 1989.[14] For direct comparison, in the year 1989, our model estimates that the noise share in daily returns is around 35%. Therefore, the estimates of the level of noise in daily returns are considerably smaller than estimates of the noise in intraday trade-to-trade returns. One of the reasons for why there is less noise in daily returns than in trade-to-trade returns is that some sources of noise such as the bid-ask spread do not scale up when the return horizon is increased, yet fundamental volatility (the variance attributable to information) does scale up with the return horizon (this intuition is exploited in the Corwin and Schultz (2012) effective spread estimator). For example, a one-minute return between two successive trades can contain a whole bid-ask spread (if one trade occurs at the bid and the other at the offer) and one minute of fundamental volatility, while a one-day return can also contain a whole bid-ask spread (if one close occurs at the bid and the other at the offer) but a much larger 24 hours of fundamental volatility. Note, however, that bid-ask bounce is only one of several sources of noise in prices.

Another point of comparison is with the noise induced in daily returns by "price pressure", that is, temporary deviations from efficient prices due to risk-averse liquidity providers being unwilling to provide unlimited liquidity. Recently, Hendershott and Menkveld (2014), using data on New York Stock Exchange (NYSE) intermediaries, estimate that at daily frequencies the distortions in midquote prices caused by price pressure (i.e., separate from the effect of bid-ask-bounce) are economically large (0.49% on average) and have a half-life of 0.92 days. The ratio of price pressure (in the midquote) to the variance of the efficient midquote price is 0.33 or 33% in their sample of 697 NYSE stocks during 1994-2005. This ratio of one source of noise to the estimated efficient price volatility is similar in magnitude to the estimated noise share of variance in our model.

---

[14] Hasbrouck (1991, 1993) does not report a noise share comparable to ours but we are able to calculate one from his results as follows. The estimated variance of pricing errors in Hasbrouck (1993) is 10.89×10^{-6}, whereas the variance of random walk innovations in Hasbrouck (1991b) using the same sample is 4.7×10^{-6}. If we conservatively assume zero serial correlation in pricing errors (such an assumption will underestimate the noise variance in the following calculation), then the variance of changes in pricing errors is $2\times10.89\times10^{-6}$ (it is the variance of *changes* in pricing errors that adds noise to returns, not just variance of pricing errors). Now if we compute an implied noise share we get 82.17%.

Similarly, but at monthly frequencies and using a different approach, Hendershott et al. (2011) estimate that one-quarter of monthly return variance in NYSE stocks is due to transitory price changes that are themselves partially explained by cumulative order imbalances and market-makers' inventories (price pressure). Again, this is just one source of noise and in monthly returns, but it is also close in magnitude to our estimate.

Finally, our finding that a considerable proportion of the variance in daily returns is noise is consistent with studies such as Jegadeesh (1990) and Lehmann (1990) who document significant predictability (reversals) in stock returns at one month and one week horizons, respectively. Avramov, Chordia, and Goyal (2006) and Nagel (2012) show that the reversals reflect deviations from efficient prices. They find that non-informational demand generates price pressure that is reversed once liquidity suppliers react to potential profit opportunities and the uninformed demand for liquidity abates. While it is difficult to express the reversals documented by these studies as a percentage of variance to directly compare them to our estimates of noise, Jegadeesh (1990), Lehmann (1990), Avramov et al. (2006), and Nagel (2012) show that the price distortions involved in reversals are economically meaningful, consistent with the economically meaningful noise share estimated by our model. Similarly, Asparouhova et al. (2013) show that noise at daily frequencies causes an economically meaningful bias in returns, equal to 50% or more of the corrected estimate.

2.3.3 Variance components through time

Figure 2.3 shows how the stock return variance components change through time from 1960 to 2015. There are several noteworthy long-term trends. First, the amount of noise in prices has declined from around 40% of variance in the 1960s to around 20% of variance recently, although not monotonically. Noise rose through the 1990s, spiking in 1997, and has gradually declined since then. Table 2.2 Panel B confirms that stock returns after 1997 tend to have less noise and resultantly, higher information content. The differences between the two sub-periods are statistically significant as well as economically meaningful. For example, the average noise share decreases from 35.47% before 1997 to only 25.69% after 1997.

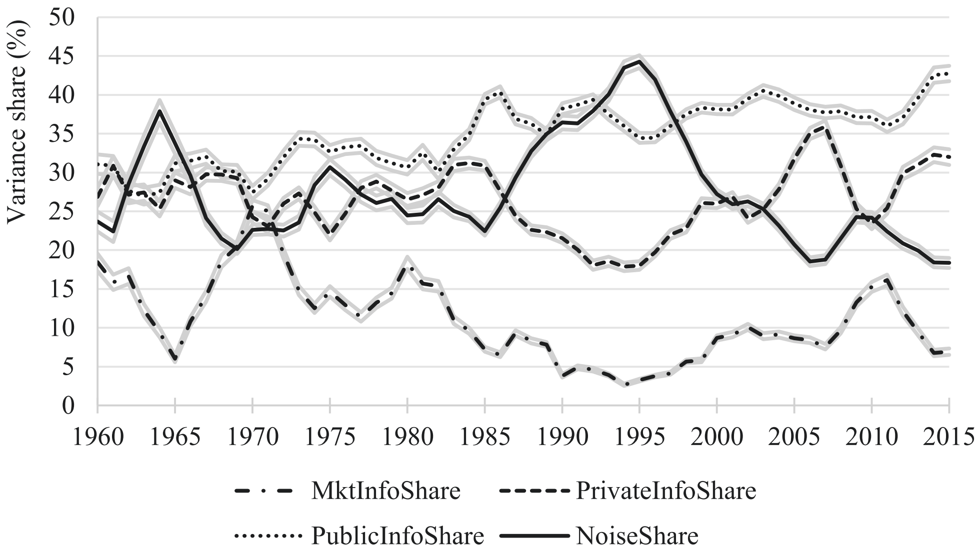


Figure 2.3. Stock return variance components for US stocks through time.
This figure shows the time-series trends in the percentage of stock price variance that is attributable to noise ($\boldsymbol{NoiseShare}$), market-wide information ($\boldsymbol{MktInfoShare}$), trading on private firm-specific information ($\boldsymbol{PrivateInfoShare}$), and public firm-specific information ($\boldsymbol{PublicInfoShare}$). The variance shares are calculated separately for each stock in each year based on a VAR model. For each variance shares in each year we then take an average across stocks. Light gray lines provide 99% confidence intervals. The sample consists of stocks listed on NYSE, AMEX, and NASDAQ from 1960 to 2015.

The high levels of noise in prices in the 1990s are at least partly driven by collusive behaviour of dealers during that period, which involved effectively widening the tick size by avoiding odd-eighth quotes and thereby increasing bid-ask bounce (Christie and Schultz, 1994). The post-1997 decline in noise to less than half of its peak levels is partly due to reductions in tick sizes starting in June 1997 as well as general improvements in liquidity and increases in turnover during the last two decades (e.g., Chordia, Roll, and Subrahmanyam, 2011). We provide more detailed analyses of these effects below (Section 2.4.1).

Second, while noise has declined through time, firm-specific information has become an increasingly important component of stock return variance. Together, the two firm-specific information components have increased from around 50% of variance in the early 1960s to above 70% of stock return variance in recent years. Table 2.2 Panel B confirms that this increase in firm-specific information is also statistically significant. The general trend is consistent with increasing informational efficiency through time. Interestingly, while public and private firm-specific information contribute approximately equally to stock return variance in the early 1960s, these components diverge through time, with publically available firm-specific information emerging as the dominant component, accounting for around 40% of stock return variance in recent years. This tendency is consistent with the objectives of a variety of regulations such as the Sarbanes Oxley Act (2002) and Regulation Fair Disclosure (2000) to increase both the quality and quantity of public disclosure by companies.

As a short aside, our estimates of the proportion of firm-specific information that is impounded in prices through trading (private information) compared to public information are similar to Hasbrouck's estimates of the role of trading in impounding new information. Using intraday data, Hasbrouck (1991b) estimates that in 1989 34.3% of the information in prices is impounded via individual trades. Despite differences in model, sample, and frequency, using daily data we estimate the fraction in 1989 to be around 39.1%.[15]

Third, while market-wide information tends to spike during crises, at other times it is generally not a substantial driver of individual stock returns. Throughout the sample period, market-wide information accounts for around 5-15% of stock return variance.

The broad trends illustrated in Figure 2.3 shed some light on recent issues concerning the information content of prices. For example, the concern that the growth in indexing and passive investing in recent years, and corresponding decline in active funds management, might harm the amount of firm-specific information in prices is not supported by the general trends in Figure 2.3. Similarly, suggestions that the increase in market model R^2 and stock correlations since the late 1990s reflects a deterioration in informational efficiency is also not supported by the data, which show declining noise levels and increasing dominance of informational components. We explore whether or not composition changes in the market contribute to some of the time-series trends in the next subsection after analyzing the cross-sectional variation.

[15] A further difference is that our estimate of 39.1% corresponds to the fraction of *firm-specific* information in prices that is impounded through trading, while the 34.3% corresponds to the fraction of *all* information impounded in prices.

2.3.4 *Variance components in the cross-section of stocks*

Table 2.2 Panels C and D report means of each of the variance components in size quartiles and industry groups with these groupings formed each year. The returns of large stocks tend to reflect more market-wide information and more private firm-specific information. The differences are particularly large for market-wide information, which accounts for 21.51% of variance in big stocks, but only 4.53% in small stocks. Large stocks also tend to have less noisy prices. Noise declines monotonically with size and the differences across stocks are large. For example, in small stocks, noise accounts for 36.09% of stock return variance, which is about twice that of big stocks at 16.44% of variance. The relatively low level of noise in large stocks is likely to be driven by a high level of liquidity, making their prices less susceptible to temporary deviations and price pressures. Panel D shows that there is considerably less variation across industry groups in what drives stock price movements than across size groups.

To test the cross-sectional determinants of the variance components in a multivariate setting, we estimate the following panel regressions of stock-year observations:

$$Share_{i,t} = \alpha + \gamma_1 D_t^{POST} + \gamma_2 lnP_{i,t} + \gamma_3 lnMC_{i,t} + \gamma_3 D_i^{Consumer} + \gamma_4 D_i^{Healthcare} + \gamma_5 D_i^{HiTech} + \gamma_6 D_i^{Manufact} + \varepsilon_{i,t} \ , \tag{2.10}$$

where $Share_{i,t}$ is one of the variance component shares ($MktInfoShare_{i,t}$, $PrivateInfoShare_{i,t}$, $PublicInfoShare_{i,t}$, $NoiseShare_{i,t}$) , D_t^{POST} is a dummy variable that takes the value of one after 1997 and zero before, $lnP_{i,t}$ is the log stock price, and $lnMC_{i,t}$ is the stock's log market capitalization. The dummy variables $D_i^{Consumer}$, $D_i^{Healthcare}$, D_i^{HiTech}, and $D_i^{Manufact}$ indicate the firm's industry group (the Other Industry group is the omitted category).

The regression results in Table 2.3 generally confirm the observations from the univariate analysis. Stock returns in the 1997-2015 part of the sample tend to contain significantly less noise and more public firm-specific information, even after controlling for other factors. Therefore, the time-series changes in noise and in firm-specific information are not driven simply by firms becoming larger through time. The returns of large stocks and high priced stocks are significantly more affected by market-wide information. Large stocks also tend to have less noisy prices and reflect relatively more firm-specific private information, controlling for other factors. Among the five industry groups, stocks in the Healthcare and Hi Tech sectors tend to have the highest levels of private firm-specific information and lowest levels of noise.

Table 2.3. Determinants of stock return variance components.
This table reports the results from panel regressions of stock-year observations in which the dependent variables are shares of stock return variance attributable to market-wide information ($MktInfoShare$), private firm-specific information ($PrivateInfoShare$), public firm-specific information ($PublicInfoShare$), and noise ($NoiseShare$). The explanatory variables are as follows. D_t^{POST} is a dummy variable that takes the value of one after 1997 and zero before. $lnP_{i,t}$ is the log price and $lnMC_{i,t}$ is the log market capitalization. $D_i^{Consumer}$, $D_i^{Healthcare}$, D_i^{HiTech}, and $D_i^{Manufact}$ are dummy variables that indicate the firm's industry group (the Other Industry grouping is the omitted category). T-statistics are in parentheses using standard errors clustered by stock and by year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample consists of stocks listed on NYSE, AMEX, and NASDAQ from 1960 to 2015.

Variable	$MktInfoShare$	$PrivateInfoShare$	$PublicInfoShare$	$NoiseShare$
Intercept	-1.50	14.48	39.42	47.59
	(-2.50)**	(10.13)***	(32.91)***	(24.35)***
D_t^{POST}	-0.28	1.38	3.67	-4.77
	(-0.21)	(0.97)	(4.51)***	(-2.83)***
$lnP_{i,t}$	2.00	-0.64	-1.09	-0.27
	(2.90)***	(-1.20)	(-2.64)***	(-0.39)
$lnMC_{i,t}$	1.86	1.93	-0.33	-3.46
	(4.59)***	(5.66)***	(-1.27)	(-9.77)***
$D_i^{Consumer}$	0.53	3.19	-1.81	-1.91
	(1.92)*	(7.08)***	(-2.92)***	(-2.71)***
$D_i^{Healthcare}$	-0.48	7.38	-1.45	-5.44
	(-0.91)	(8.74)***	(-2.36)**	(-6.62)***
D_i^{HiTech}	1.27	4.82	-1.30	-4.79
	(1.85)*	(6.38)***	(-2.04)**	(-7.92)***
$D_i^{Manufact}$	1.62	4.08	-2.92	-2.78
	(4.17)***	(7.51)***	(-4.17)***	(-3.55)***
$Adjusted\ R^2$	24.1%	5.9%	1.6%	14.7%

Given that the composition of stock return variance differs in the cross-section, in particular by size and to a lesser extent by industry, we examine to what extent the time-series patterns in variance components are due to market composition changes. The mix of industries in the market has changed through time and listed stocks have tended to become larger through time. Therefore, in Figures 2.4 and 2.5 we repeat the exercise of plotting the time-series of variance components, but this time by size group and by industry group. We form the size groups with respect to thresholds ($100 million and $1 billion in 2010 dollars) that are inflation adjusted through time, rather than size quartiles so as to keep the size groups relatively comparable through time even as the market composition changes. Time-series trends in the variance components within size or industry groups are less susceptible to compositional changes than the pooled time-series.

Figure 2.4 shows that all size groups have a similar trend with respect to market-wide information, including the peaks during crises. Large stocks consistently reflect more market-wide information through time than smaller stocks. All size groups show remarkably similar trends in private firm-specific information, except for a period of temporary divergence in the 1990s. The increase in public firm-specific information through time is driven mainly by smaller stocks, consistent with improvements in their disclosure. Finally, noise is consistently higher for smaller stocks and smaller stocks are largely responsible for the decline in noise through time, in particular since the mid-1990s.

Figure 2.5 shows that the variance components in different industry groups display remarkably similar time-series trends. Not only are the long-run trends in the types of information and noise similar across the industry groups, but so too are many of the year-to-year fluctuations. This result indicates that the time-series trends are not driven by changing industry composition in the market. Furthermore, it indicates that much of the variation in the information and noise shares is systematic and not just an artefact of estimation error or random fluctuations. Recall that the variance decomposition is performed separately (independently) for each stock in each year. The commonality in the variance component trends across groups of stocks (in this case industry groups) points to systematic drivers of the type of information and degree of noise in prices.

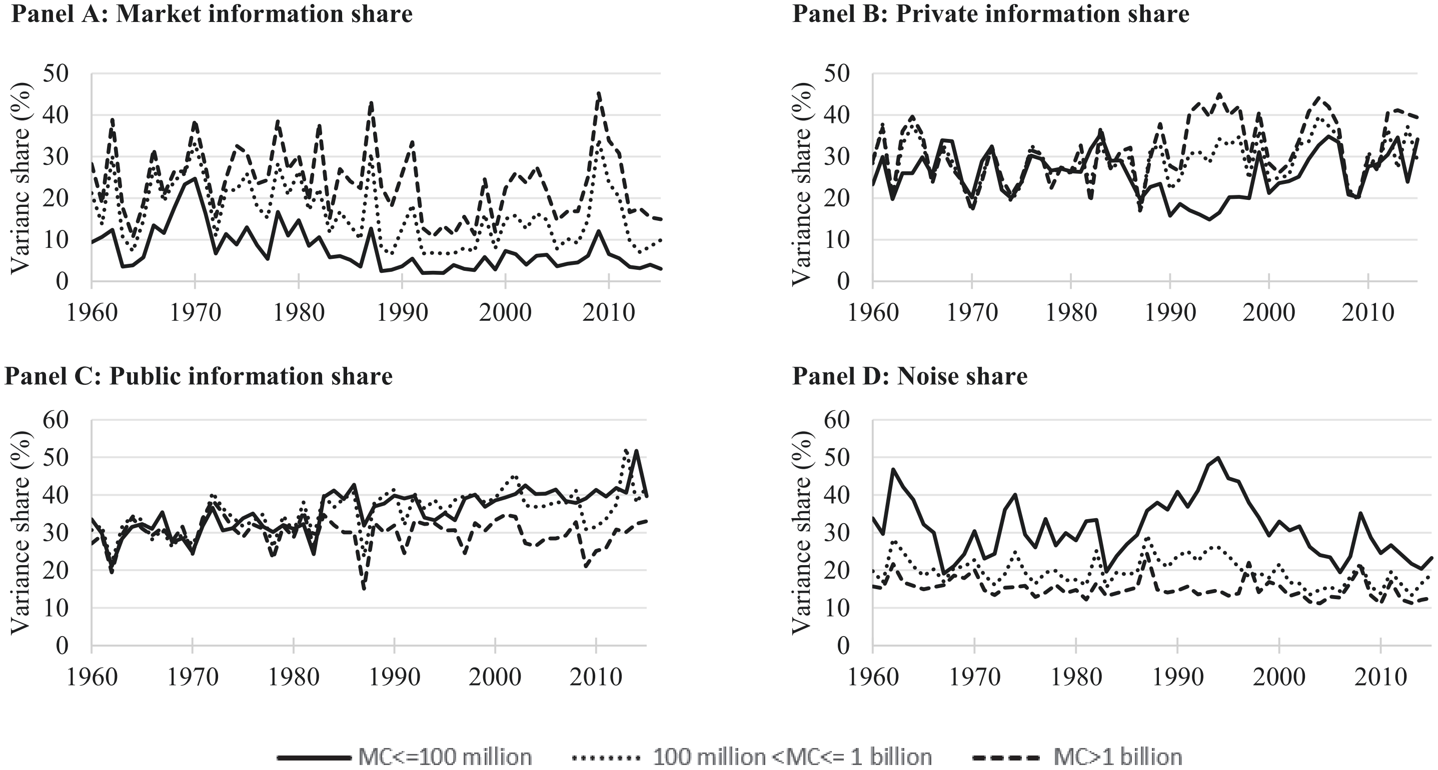


Figure 2.4. Variance components in size groups through time.
This figure shows the time-series trends in the percentage of stock return variance that is attributable to market-wide information (Panel A), private firm-specific information (Panel B), public firm-specific information (Panel C), and noise (Panel D) in three market capitalization groups: stocks with market capitalization less than $100 million, market capitalization between $100 million and $1 billion, and market capitalization greater than $1 billion. These breakpoint are in 2010 dollars and are adjusted for inflation forward and backward in time using the GDP price deflator. Each year stocks are assigned to one of the three groups based on their market capitalization at the start of the year. The variance component shares are calculated separately for each stock in each year then averaged for each size group in each year. The sample consists of stocks listed on NYSE, AMEX, and NASDAQ from 1960 to 2015.

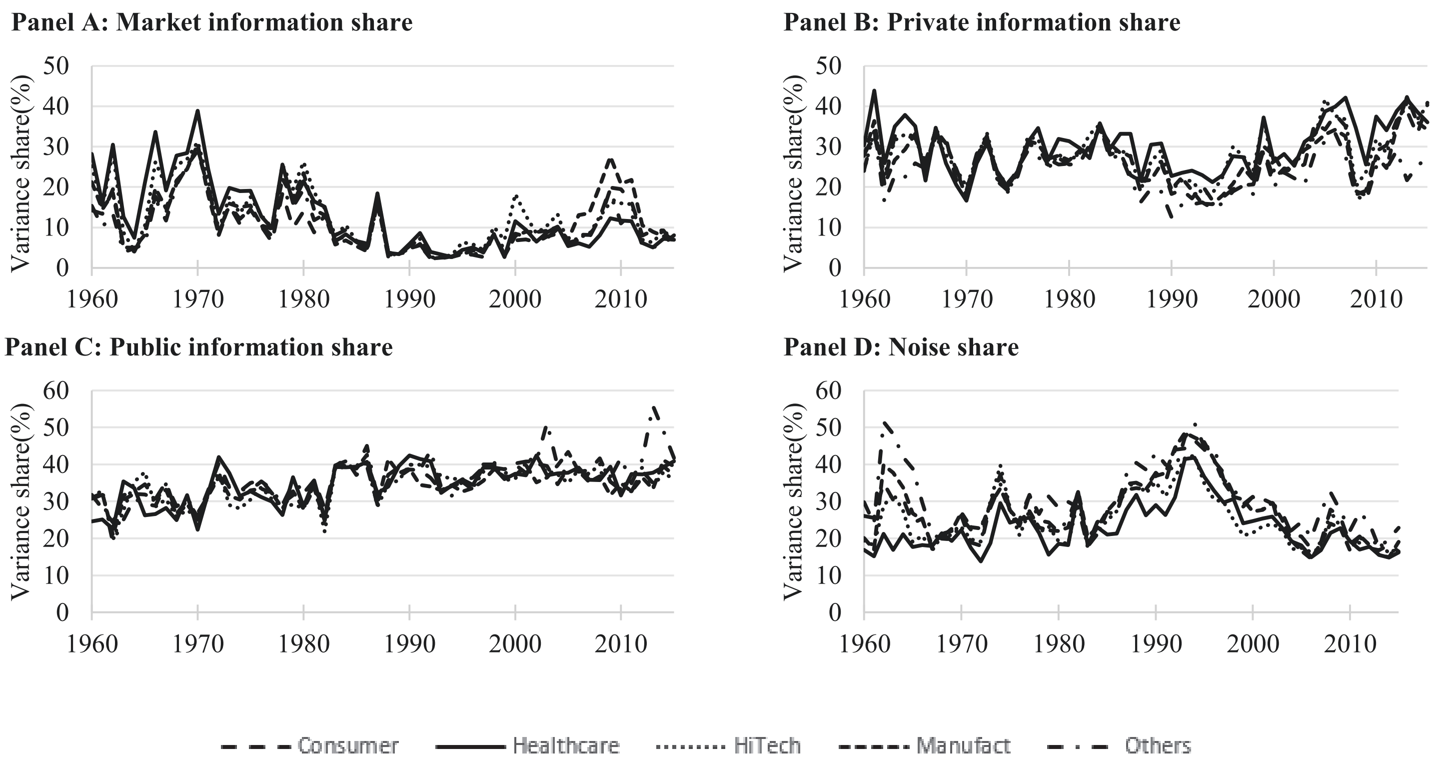


Figure 2.5. Variance components in major industry groups through time.
This figure shows the time-series trends in the percentage of stock return variance that is attributable to market-wide information (Panel A), private firm-specific information (Panel B), public firm-specific information (Panel C), and noise (Panel D) in five major industry groups. The *Consumer* group comprises the industries Consumer Durables, NonDurables, Wholesale, Retail, and some Services (Laundries, Repair Shops); the *Healthcare* group comprises the industries Healthcare, Medical Equipment, and Drugs; the *Manufact* group comprises the industries Manufacturing, Energy, and Utilities; the *HiTech* group comprises the industries Business Equipment, Telephone and Television Transmission; and the *Other* group comprises all other industries. The variance component shares are calculated separately for each stock in each year then averaged for each industry group in each year. The sample consists of stocks listed on NYSE, AMEX, and NASDAQ from 1960 to 2015.

2.4 Further validation tests

The preceding section provides some informal validation of the empirical variance decomposition model by showing that the variance components have reasonable time-series and cross-sectional properties, that they exhibit systematic variation, and that some of the estimated levels are consistent with estimates in other studies that use different approaches (usually capturing only one of the components). This section presents further validation tests that examine how the variance components respond to exogenous shocks to tick sizes (Section 2.4.1) and exogenous shocks to analyst coverage (Section 2.4.2). We also relate the variance components to other measures of noise and information (Section 2.4.3).

2.4.1 Exogenous shocks to tick sizes

Chordia, Roll, and Subrahmanyam (2008) show that decreases in tick sizes increase informational efficiency, implying a decrease in noise when tick sizes are smaller. During the sample period, the tick size was reduced from eighths of a dollar to sixteenths of a dollar on June 24, 1997, and then from sixteenths to pennies on January 29, 2001. This setting provides a natural experiment in which to test how the variance components, in particular noise, respond to the exogenous decreases in tick size and accompanying increase in informational efficiency.

If the estimated noise share of variance is indeed able to capture noise in prices, we would expect to see three patterns related to tick sizes and changes in tick sizes. First, stocks with larger relative tick sizes (tick size divided by price) should have noisier prices and thus a larger noise share. Given the tick size in dollars is the same for all stocks in the cross-section (at the time of the tick size changes) but price levels vary in the cross-section, we would expect lower priced stocks to have higher levels of noise because they have larger relative tick sizes. Second, we would expect that when the tick sizes are reduced, noise declines. Third, we would expect that the effects of tick size reductions are heterogeneous in the cross-section with lower priced stocks having a larger decline in noise because for such stocks the change in the relative tick size is larger. For example, the tick size reduction from eighths of a dollar to sixteenths is 1.25% of the price of a $5 stock, but only 0.125% of the price of a $50 stock.[16]

To exploit this natural experiment, we take a subsample of one year either side of the tick size reduction from eighths of a dollar to sixteenths of a dollar (i.e., we take the years 1996 and 1998), and estimate difference-in-differences models that exploit the cross-sectional heterogeneity in the treatment.[17] The highest priced quartile of stocks had the smallest change in relative tick size and therefore serves as a control group against which to measure the impact of the tick size reduction in other price quartiles:

$$\begin{aligned} NoiseShare_{it} = \alpha + \beta D^{POST}_{TickReduction,t} + \gamma_1 D^{POST}_{TickReduction,t} Q1_i \\ + \gamma_2 D^{POST}_{TickReduction,t} Q2_i + \gamma_3 D^{POST}_{TickReduction,t} Q3_i + \rho_1 Q1_i \\ + \rho_2 Q2_i + \rho_3 Q3_i + \varepsilon_{it} \end{aligned} \tag{2.11}$$

where $D^{POST}_{TickReduction,t}$ takes the value of one after the tick size reduction (1998) and zero otherwise. $Q1_i$, $Q2_i$, and $Q3_i$ are dummy variables that indicate the price quartile to which the firm belongs (the highest price quartile, $Q4_i$, is the omitted category). We also re-estimate the model (2.11) using the log price ($lnP_{i,t}$) instead of the price quartile indicators as a robustness test:

$$\begin{aligned} NoiseShare_{it} = \alpha + \beta D^{POST}_{TickReduction,t} + \gamma_1 D^{POST}_{TickReduction,t} lnP_{i,t} + \gamma_2 lnP_{i,t} \\ + \varepsilon_{it} \end{aligned} \tag{2.12}$$

---

16 Similarly, the tick-to-price, a measure of the pricing grid coarseness, decreases by a larger amount for lower priced stocks.

17 Chordia, Roll, and Subrahmanyam (2008) show that informational efficiency improved significantly around the change from eighths of a dollar to sixteenths and to a lesser extent from sixteenths to pennies given that the first change was relatively larger.

Table 2.4. **Effect of the tick size on noise.**
This table reports the results from panel regressions of stock-year observations in which the dependent variable is the share of stock return variance attributable to noise ($NoiseShare$). Models 1-4 examine how the noise share is affected by tick size reductions from eighths of a dollar to sixteens of a dollar on June 24, 1997 using two years of data around the change (1996, 1998). $D^{POST}_{TickReduction,t}$ is a dummy variable that takes the value of one after 1997 and zero before. $Q1_i$, $Q2_i$, and $Q3_i$ are dummy variables that indicate the price quartile to which the firm belongs (the highest price quartile, $Q4_i$, is the omitted category). $lnP_{i,t}$ is the log price. Model 5 examines how the collusion by NASDAQ dealers to avoid odd-eighth quotes impacts the noise in prices, using four years before and four years during the collusion (1987 to 1994). $D^{COLLUSION}_t$ takes the value of one in the collusion period (1991-1994) and zero otherwise. $NASDAQ_i$ is one for NASDAQ-listed stocks and zero otherwise. T-statistics are in parentheses using standard errors clustered by stock. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes all stocks listed on NYSE, AMEX, and NASDAQ.

Variable	(1)	(2)	(3)	(4)	(5)
Intercept	19.36	38.05	19.67	42.81	20.89
	(70.82)***	(103.54)***	(52.57)***	(76.74)***	(117.47)***
$D^{POST}_{TickReduction,t}$			-0.62	-8.49	
			(-1.54)	(-14.26)**	
$D^{POST}_{TickReduction,t} \times Q1_i$			-3.43		
			(-5.23)***		
$D^{POST}_{TickReduction,t} \times Q2_i$			-4.99		
			(-7.32)***		
$D^{POST}_{TickReduction,t} \times Q3_i$			-3.94		
			(-5.44)***		
$Q1_i$	14.30		16.02		
	(34.65)***		(27.26)***		
$Q2_i$	7.56		10.04		
	(17.67)***		(16.25)***		
$Q3_i$	5.51		7.48		
	(13.03)***		(11.92)***		
$D^{POST}_{TickReduction,t} \times lnP_{i,t}$				1.72	
				(8.38)***	
$lnP_{i,t}$		-4.86		-5.91	
		(-36.52)**		(-30.14)**	
$D^{COLLUSION}_t$					-2.45
					(-12.81)**
$D^{COLLUSION}_t \times NASDAQ_i$					8.32
					(24.89)***
$NASDAQ_i$					10.02
					(32.36)***
R^2	7.5%	7.5%	8.8%	9.0%	11.7%

In Table 2.4, Models 1 and 2 show a monotonic relation between a stock's price level and its level of noise. As expected, lower priced stocks have higher levels of noise consistent with the fact that they have larger relative tick sizes (tick size divided by price). The difference in the cross-section of stocks is economically meaningful: the highest priced quartile has a noise share of around 19%, while the next three price quartiles have noise shares that are 6%, 8%, and 14% higher with these differences being statistically significant as well. Models 3 and 4 show that noise tends to decline when the tick size is reduced consistent with our expectations. Furthermore, the decline in noise is larger for lower priced stocks consistent with the fact that for such stocks the change in the relative tick size is larger. Therefore, the analysis of how noise relates to tick sizes or price discreteness and how it changes around exogenous changes in tick sizes support the notion that the noise share is a useful measure of the amount of noise in prices and returns. We also examine a second natural experiment relating to the tick size. Christie and Schultz (1994) find evidence of collusive behaviour by NASDAQ dealers during a period from 1991 until the collusive behaviour was exposed a few years later. They show that NASDAQ dealers colluded to maintain artificially wide spreads by avoiding odd-eighth quotes. This behaviour increased the effective tick size and due to bid-ask bounce is expected to increase the noise in prices. Importantly, a natural control group is non-NASDAQ stocks.

To examine this natural experiment, we take a subsample of four years before and during the collusive behaviour and estimate a difference-in-differences model that uses non-NASDAQ stocks as a control group:

$$NoiseShare_{it} = \alpha + \beta D_t^{COLLUSION} + \gamma D_t^{COLLUSION} NASDAQ_i + \delta NASDAQ_i + \varepsilon_{it} \tag{2.13}$$

where $D_t^{COLLUSION}$ is a dummy variable that takes the value of one in the collusion period (1991-1994) and zero otherwise and the dummy variable $NASDAQ_i$ is one for NASDAQ-listed stocks and zero otherwise.

Model 5 in Table 2.4 shows that during the period of collusion by NASDAQ dealers, the returns of NASDAQ-listed stocks are significantly noisier, consistent with discreteness in price grids contributing to pricing errors and noise in returns. The magnitude of the effect is economically meaningful. The increase in noise for NASDAQ-listed stocks is estimated to be 8.32% of variance, which is large considering that the pooled sample mean noise share is around 30.78% of variance. The results therefore support the notion that collusion by NASDAQ dealers effectively widened the tick size and that the noise share from our variance decomposition model captures this increase in noise for the affected stocks.

2.4.2 *Exogenous shocks to analyst coverage*

Exogenous shocks to analyst coverage provide another natural experiment that changes the information environment for individual stocks. Given that analysts produce information about individual companies and disseminate this information to a variety of market participants, a reduction in analyst coverage is likely to reduce the amount of public firm-specific information in prices. As information in prices declines, the relative level of noise is likely to increase. Analyst coverage is expected to have little effect on market-wide information. The effects of analyst coverage on private firm-specific information is ambiguous: analyst-generated information that is made available to only some market participants might be impounded in prices through the course of those participants trading on the information (an increase in private information), but it might also crowd-out private information acquisition (a decrease in private information).

To test the impact of analyst coverage on the information and noise in prices, we use brokerage mergers/closures as a source of exogenous variations in analyst coverage. Broker mergers and closures are plausibly exogenous shocks because the termination of coverage is not driven by the characteristics or behaviour of the firm (see Hong and Kacperczyk, 2010; Kelly and Ljungqvist, 2012; Brogaard et al., 2018). We obtain a list of broker mergers and closures that combines the lists from Hong and Kacperczyk (2010), spanning 1984 to 2005, and Kelly and Ljungqvist (2012), spanning 2000 to 2008. Combining these lists, merging with CRSP and IBES (Institutional Brokers' Estimate System) data, and imposing the requirement that both the acquirer and target brokers must provide overlapping coverage for at least one firm before the broker merger (as per Kelly and Ljungqvist, 2012) we have 41 mergers/closures of brokers that occur during the period 1989-2009. Using the mergers/closures data, we calculate the number of exogenous analyst disappearances per stock-year. These mergers and closures result in exogenous coverage shocks to 4,546 firm-year observations.

Using the exogenous analyst coverage shocks, we estimate the following difference-in-differences model:

$$Share_{i,t} = \gamma_i + \delta_t + \beta_1 CoverageShock_{i,t} + \varepsilon_{i,t} \ , \tag{2.14}$$

where $Share_{i,t}$ is one of the variance component shares ($MktInfoShare_{i,t}$, $PrivateInfoShare_{i,t}$, $PublicInfoShare_{i,t}$, $FirmInfoShare_{i,t}$, $NoiseShare_{i,t}$) for stock i in year t, γ_i and δ_t are stock and time fixed effects, respectively, and $CoverageShock_{i,t}$ is the number of analyst disappearances due to mergers and closures of brokerage houses during the past two years.[18] We estimate the model above using the period from 1987 to 2011 given that the brokerage mergers and closures occur between 1989 and 2009 and we need to observe a two-year trend before and after the analyst disappearances.

The results in Table 2.5 show that exogenous decreases in analyst coverage are associated with a decline in public firm-specific information and an increase in the noise share of variance. These results are consistent with the notion that analysts produce firm-specific information that is made publically available and becomes reflected in prices. It also suggests the public firm-specific information component of variance from our variance decomposition model is able to detect this change in the information environment. The coefficient estimates indicate that the exogenous disappearance of each analyst is associated with a decline in public firm-specific information equal to around 1.59% of variance (for comparison, the pooled sample mean of public firm-specific information is around 37.11% of variance).

Shocks to analyst coverage have no significant effect on the amount of market-wide information in prices as expected. Neither do they have a significant impact on the amount of private firm-specific information in prices.

---

[18] In contrast to standard difference-in-differences models, here the "treatment" can have different magnitudes depending on how many analysts cease their coverage of a given stock. If the number of analyst disappearances is different in year $t-1$ and year $t-2$, we take the maximum of these two values.

Table 2.5. Effect of analyst coverage on variance components.

This table reports the results from difference-in-difference regressions of stock-year observations in which we examine the causal effect of an exogenous drop in analyst coverage (due to brokerage mergers and closures) on variance components. The dependent variables are shares of stock return variance attributable to market-wide information ($MktInfoShare$), private firm-specific information ($PrivateInfoShare$), public firm-specific information ($PublicInfoShare$), firm-specific information ($FirmInfoShare$), and noise ($NoiseShare$). $FirmInfoShare$ is the sum of $PrivateInfoShare$ and $PublicInfoShare$. The independent variable of interest is $CoverageShock_{i,t}$ which is the number of broker disappearances due to mergers and closures of brokerage houses during the past two years (max of the $t-1$ and $t-2$ values). T-statistics are in parentheses using standard errors clustered by stock and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes stocks listed on NYSE, AMEX, and NASDAQ from 1987 to 2011 (a period containing the analyst coverage shock events).

Variable	$MktInfoShare$	$PrivateInfoShare$	$PublicInfoShare$	$FirmInfoShare$	$NoiseShare$
Intercept	8.02	-3.96	-4.77	-8.73	0.71
	(122.82)***	(-66.20)***	(-110.76)*	(-107.62)*	(23.13)***
$CoverageShock_{i,t}$	0.43	0.57	-1.59	-1.03	0.59
	(0.60)	(0.87)	(-3.37)***	(-1.15)	(1.80)*
R^2	21.9%	6.9%	2.3%	9.2%	4.9%
Fixed Effects	Stock, Year	Stock, Year	Stock, Year	Stock, Year	Stock, Year

2.4.3 Relation between variance components and other measures of information and noise

In the next two validation tests, we examine the relation between the variance components and two other measures of information in prices: the Hou and Moskowitz (2005) delay metric and variance ratios as per French and Roll (1986) and Chordia, Roll, and Subrahmanyam (2011).

The Hou and Moskowitz (2005) delay metric is a measure of how efficiently market-wide information is reflected in individual stock prices. It is constructed in each stock-year by estimating a regression of daily stock returns ($r_{i,t}$) on daily market returns ($r_{m.t}$) and ten lags of daily market returns:

$$r_{i,t} = \alpha_i + \beta_i r_{m.t} + \sum_k^{10} \delta_{i,k} r_{m,t-k} + \varepsilon_{i,t} \, , \tag{2.15}$$

Regression (2.15) is estimated once with all of the lags of market returns (unconstrained) saving the regression R^2 ($R^2_{unconstrained}$) and once without the lags of market returns (constraining all $\delta_{i,k}$ to zero) also saving the R^2 ($R^2_{constrained}$). The delay metric is then constructed from the two R^2 as follows to measure the incremental explanatory power of the lagged market returns:

$$Delay_{i,t} = 1 - \frac{R^2_{constrained}}{R^2_{unconstrained}} \quad , \tag{2.16}$$

If market-wide information is perfectly and instantly reflected in the stock's prices, the two R^2 are equal and $Delay_{i,t} = 0$, but if stock prices are sluggish in reflecting market-wide information, then $R^2_{unconstrained} > R^2_{constrained}$ and $Delay_{i,t} > 0$. We therefore expect that higher values of $Delay_{i,t}$ should be associated with less market-wide information in prices (lower $MktInfoShare$) and less efficient, noisier prices (higher $NoiseShare$).

We estimate the relation between each of the variance components ($Share_{i,t}$) and the delay metric using the following panel regressions, including stock and year fixed effects (γ_i and δ_t):

$$Share_{i,t} = \gamma_i + \delta_t + \beta_1 Delay_{i,t} + \varepsilon_{i,t} \ . \tag{2.17}$$

Table 2.6 Panel A reports the result. There is a very strong inverse relation between the delay metric and the share of variance that is attributable to market-wide information by our variance decomposition. An increase in $Delay_{i,t}$ from zero (full efficiency) to 0.5 (half way to the maximum value of $Delay_{i,t}$) is associated with a reduction of $MktInfoShare$ by 9.17% of variance (a large effect, considering the pooled sample mean of $MktInfoShare$ is 8.24%). The results therefore suggest that $MktInfoShare$ reflects the efficiency with which market-wide information is reflected in prices.

Next we examine the variance ratios measure (French and Roll, 1986; Chordia et al., 2011), defined as the variance of returns during trading hours (variance of open-to-close returns, ${}^1/_n \sum_1^n r_{intraday,i,d}^2$) divided by the variance of overnight returns (variance of close-to-open returns, ${}^1/_n \sum_1^n r_{overnight,i,d}^2$). We measure the variance ratio for each stock i in each year t using daily, d, observations, excluding weekends:

$$VarianceRatio_{i,t} = \frac{{}^1/_n \sum_1^n r_{intraday,i,d}^2}{{}^1/_n \sum_1^n r_{overnight,i,d}^2} \ . \tag{2.18}$$

Table 2.6. Relation between variance components and other measures of information in prices.
This table reports the results from panel regressions of stock-year observations in which the dependent variables are shares of stock return variance attributable to market-wide information ($MktInfoShare$), private firm-specific information ($PrivateInfoShare$), public firm-specific information ($PublicInfoShare$), firm-specific information ($FirmInfoShare$), and noise ($NoiseShare$). $FirmInfoShare$ is the sum of $PrivateInfoShare$ and $PublicInfoShare$. Panel A reports the relation between the variance components and a measure of the delay with which a stock's prices respond to market-wide information ($Delay_{i,t}$). Panel B reports the relation between the variance components and a measure of the amount of information impounded in prices during the trading session ($VarianceRatio_{i,t}$ is the volatility of open to close returns divided by the volatility of overnight (close to open) returns). The sample includes stocks listed on NYSE, AMEX, and NASDAQ from 1960 to 2015 (Panel A) and 1992 to 2015 (Panel B). T-statistics are in parentheses using standard errors clustered by stock and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels.

Panel A: Delay metric

Variable	$MktInfoShare$	$PrivateInfoShare$	$PublicInfoShare$	$FirmInfoShare$	$NoiseShare$
Intercept	-2.31	2.19	1.60	3.79	-1.48
	(-18.11)**	(10.12)***	(7.55)***	(19.09)***	(-9.84)***
$Delay_{i,t}$	-18.34	-1.05	4.57	3.52	14.83
	(-175.60)**	(-6.45)***	(26.08)***	(17.47)***	(87.67)***
R^2	34.1%	7.3%	4.1%	13.4%	11.2%
Fixed Effects	Stock, Year	Stock, Year	Stock, Year	Stock, Year	Stock, Year

Panel B: Variance ratio

Variable	$MktInfoShare$	$PrivateInfoShare$	$PublicInfoShare$	$FirmInfoShare$	$NoiseShare$
Intercept	-0.62	1.97	1.33	3.30	-2.68
	(-16.60)***	(49.96)***	(50.29)***	(60.44)***	(-40.76)**
$VarianceRatio_{i,t}$	0.44	0.18	0.27	0.45	-0.89
	(4.77)***	(1.83)*	(4.01)***	(3.40)***	(-5.74)***
R^2	17.5%	6.2%	1.9%	8.9%	6.4%
Fixed Effects	Stock, Year	Stock, Year	Stock, Year	Stock, Year	Stock, Year

French and Roll (1986) and Chordia et al. (2011) show that the variance ratio is related to the amount of information reflected in prices, in particular through trading, and they use this ratio as a measure of informational efficiency. Under this interpretation, we would expect the variance ratio to be positively related to most or all of the information components of variance and negatively related to the amount of noise in prices.

We estimate the relation between each of the variance components and the variance ratio metric using similar panel regressions as before in (2.17) including stock and year fixed effects. The results in Table 2.6 Panel B show that this is the case. The variance ratio is positively related to all of the information components of variance and negatively related to the amount of noise in prices. This result supports the interpretation of the variance ratio as an efficiency measure as well as the ability of the variance decomposition to separate out information from noise.

2.5 Conclusion

This study decomposes stock return variance to better understand the roles of different types of information and noise in driving stock price movements. We find that a substantial proportion of return variance, 31%, is noise. Firm-specific information accounts for the majority (61%) of stock return variance, with market-wide information accounting for the remaining 8% of variance in the full sample. We further partition firm-specific information and find that in the full sample, public firm-specific information plays a larger role than private firm-specific information that is impounded into prices through trading.

There is substantial time-series variation in the components of variance, with some key trends standing out. First, noise increases from the 1970s to the mid-1990s, in particular around a period of collusion by dealers to effectively widen bid-ask spreads, and has substantially declined since then. The decline in noise is attributable in part to narrower tick sizes, which reduces bid-ask bounce, and a general improvement in liquidity and increase in turnover. We show that the recent decrease in noise due to improved liquidity is largely responsible for the increasing R^2 of a market model in the past two decades. An important implication is that a lower R^2 is not necessarily associated with more informationally efficient prices, in contrast to the interpretation of R^2 in prior studies.

Second, the role of firm-specific information has increased through time, driven largely by increases in the amount of public firm-specific information that is reflected in prices. This trend is consistent with increasing informational efficiency through time. The increasing importance of public firm-specific information in stock prices is also consistent with a variety of regulatory reforms such as the Sarbanes Oxley Act (2002) and Regulation Fair Disclosure (2000) aimed at improving both the quality and quantity of corporate disclosure.

Third, market-wide information has over time become a less important driver of stock returns. While market-wide information tends to spike during crises it has generally declined from around 15% of variance to around 5-10% in recent years.

Overall, the broad trends in the components of stock return variance shed light on recent issues concerning the information content of prices. For example, the concern that the growth in indexing and passive investing in recent years and corresponding decline in active funds management might harm the amount of firm-specific information in prices is not intuitively supported by the observed time trends. Similarly, suggestions that the increase in market model R^2 and stock correlations in recent years reflects a deterioration in informational efficiency is also not supported by the data, which show declining noise levels and increases in informational components.

While our results provide some new insights about these issues, we leave a more detailed examination of each of these issues to future research. This study's contribution is largely methodological. The framework for variance decomposition developed in this chapter can be applied to analyzing each of these issues and others, in particular due to: (i) its ability to isolate noise from information, which is crucial for correctly characterizing the information in prices; and (ii) the ability to obtain higher frequency estimates of variance components, which is important in analyzing effects that vary through time and recent phenomena that require current estimates of the information/noise components of variance.

Appendix 2.A. Variable definitions

The table below provides descriptions and notation for the variables that are components of stock return variance. Each variable is estimated separately for each stock in each year using daily observations. When aggregating across stocks, we take variance-weighted averages (as per Morck et al., 2000, 2013).

Variable	Notation	Description
Stock return co-movement	R^2	R^2 is estimated by regressing individual daily stock returns on daily market return.
Noise share	$NoiseShare$	The share of stock return variance that is attributable to noise.
Market-wide information share	$MktInfoShare$	The share of stock return variance that is attributable to market-wide information.
Private firm-specific information share	$PrivateInfoShare$	The share of stock return variance that is attributable to trading on private firm-specific information.
Public firm-specific information share	$PublicInfoShare$	The share of stock return variance that is attributable to public firm-specific information.
Firm-specific information share	$FirmInfoShare$	The share of stock return variance that is attributable to firm-specific information (sum of $PrivateInfoShare$ and $PublicInfoShare$).

Appendix 2.B. Effect of including the covariance between noise and information

In computing the variance shares in Equation (2.9), we ignore the covariance between information (innovations in the efficient price) and noise (changes in the pricing error). Here we show that accounting for this covariance has little effect on our estimates of the variance shares.

One way to account for the covariance term is to distribute it between the information components of variance and the noise component of variance in the same proportions as the variances of these components and then recompute the variance shares from the covariance-adjusted components using the total return variance as the normalizing variable. In this approach, we allocate a fraction α of $2cov(w_t, \Delta s_t)$ to the information variance and a fraction $(1-\alpha)$ to the noise variance, where $\alpha = \frac{\sigma_w^2}{\sigma_w^2+\sigma_s^2}$. Consequently the information and noise shares of variance become:

$$InfoShare = \left(\sigma_w^2 + \frac{\sigma_w^2}{\sigma_w^2+\sigma_s^2} 2cov(w_t, \Delta s_t)\right) / \sigma_r^2$$

$$= \sigma_w^2 \left(1 + \frac{2cov(w_t, \Delta s_t)}{\sigma_w^2+\sigma_s^2}\right) / \sigma_r^2$$

$$= \sigma_w^2 \left(\frac{\sigma_w^2 + 2cov(w_t, \Delta s_t) + \sigma_s^2}{\sigma_w^2+\sigma_s^2}\right) / \sigma_r^2$$

$$= \sigma_w^2 \left(\frac{\sigma_r^2}{\sigma_w^2+\sigma_s^2}\right) / \sigma_r^2 = \sigma_w^2 / (\sigma_w^2+\sigma_s^2) \tag{2.B.1}$$

$$NoiseShare = \left(\sigma_s^2 + \frac{\sigma_s^2}{\sigma_w^2+\sigma_s^2} 2cov(w_t, \Delta s_t)\right) / \sigma_r^2$$

$$= \sigma_s^2 \left(1 + \frac{2cov(w_t, \Delta s_t)}{\sigma_w^2+\sigma_s^2}\right) / \sigma_r^2 = \sigma_s^2 \left(\frac{\sigma_r^2}{\sigma_w^2+\sigma_s^2}\right) / \sigma_r^2$$

$$= \sigma_s^2 / (\sigma_w^2+\sigma_s^2)$$

Equation (2.B.1) shows that after the distribution of the covariance term into information and noise components, we have exactly the same variance shares as in the baseline (Equation (2.9)).

An alternative way of distributing the covariance term is to add it entirely to either the information component of variance or to the noise component of variance thereby producing upper and lower bounds on the variance shares. Applying this approach we find that the upper and lower bounds are extremely narrow (e.g., the noise share has a lower bound of 25.73% and an upper bound of 27.86%, whereas the information share has a lower bound of 72.14% and an upper bound of 74.27%). Therefore, ignoring the covariance term in the baseline variance shares has little effect on the results.

Chapter 3: The effects of noise on existing empirical models

3.1 Introduction

Given the significant effect of noise on asset pricing/returns documented in the finance literature, the previous chapter proposes a new return variance decomposition model that separates noise and various sources of information. In this chapter, we examine how noise distorts inference in existing empirical variance decomposition models and how those models can be adjusted to account for the effects of noise.

We begin with the variance decomposition of Campbell and Shiller (1988a, 1988b) and Campbell (1991), which partitions stock return variance into two parts: cash flow news and discount rate news. The foundation for this decomposition is the discounted cash flow model of asset pricing. If the value of a stock is the present value of its future cash flows, then fluctuations in the stock's value, returns, can either come from changes in the expected future cash flows ("cash flow news") or from changes in the discount rates applied by investors when evaluating the present values of the cash flows ("discount rate news").

In the Campbell and Shiller (1988a, 1988b) and Campbell (1991) approach, returns are assumed to be information-driven, and there is no allowance for noise in this standard model. Because the standard forms of these models do not account for noise, they are typically only applied using low-frequency data to minimize the effects of noise. For example, a typical application would be to use a decade of monthly returns, resulting in one estimate for the proportions of cash flow and discount rate news for that decade. Being restricted to such low frequencies severely limits the ability to examine time-series variation in the cash flow/discount rate components. In this chapter, we show how to solve this limitation by explicitly accounting for noise, thereby enabling a more accurate and granular re-examination of the roles of cash flow and discount rate news.

After accounting for noise in the return generating process, we find a considerably smaller role for discount rate news and a much larger role for cash flow news. In fact, without accounting for noise, the ratio of cash flow news to discount rate news for individual stocks is around five, similar to Vuolteenaho (2002). After accounting for noise, the ratio increases to around 25 because at the individual stock level the discount rate news accounts for only a small (3%) fraction of the stock return variance. The difference in the importance of discount rate news compared to previous studies stems largely from the fact that noise induces return predictability. For example, returns have negative first-order serial correlation at many frequencies due to pricing errors and price pressure (e.g., Jegadeesh, 1990; Roll, 1984). Without accounting for noise, the noise-induced variation in predicted returns is attributed to variation in the discount rate thereby overstating the role of discount rate news.

In an extension of the variance decomposition developed in Chapter 2, we partition each of the information components in our baseline model into a cash flow and discount rate part. We find that the ratio of cash flow news to discount rate news is higher in firm-specific information than in market-wide information. Therefore, although cash flow news is generally responsible for a larger share of stock return variation than discount rate news, cash flow news tends to be more idiosyncratic and discount rate news more systematic. This result helps reconcile existing studies in which cash flow news is found to be more important in stock-level analyses (e.g., Vuolteenaho, 2002), while discount rate news plays a larger role in portfolio-level analyses in which much of the idiosyncratic variation is removed (e.g., Campbell, 1991).

In this chapter we also re-visit the debate on the interpretation of idiosyncratic volatility, or one minus the R^2 from a regression of stock returns on market returns (a "market model"), as a measure of market efficiency. Roll (1988) proposes a decomposition of stock return variance into market-wide variations and firm-specific variations using the R^2 from a market model regression. Under the assumption that variation in prices reflects the arrival of new information, R^2 is often used as a measure of the relative amount of market-wide information in prices, and $1 - R^2$ by extension, is often used a measure of the amount of firm-specific information and thus informational efficiency (e.g., Morck et al., 2000; Wurgler, 2000; DeLisle et al., 2015; Lin et al., 2015).

However, this interpretation is challenged in the literature by a contrasting view based on the notion that variance in prices reflects information *and* noise arising from illiquidity, microstructure frictions, non-synchronous trading and so on. High idiosyncratic volatility could be driven by high levels of noise in stock prices; therefore, a low R^2 is likely to indicate a high amount of noise and *inefficiency* in prices (e.g., Dasgupta et al., 2010; Kelly, 2014; Bramante et al., 2013). Much of the disagreement about how to interpret the variance decomposition implied by R^2 comes down to whether noise is a significant driver of stock price movements and whether it has time-series or cross-sectional variations that can systematically distort the interpretation of R^2.

To illustrate the challenges that arise in reconciling standard variance decompositions with the interpretations of R^2, we consider the salient case of the long-term time-series trends in R^2 documented for the US markets. As Figure 2.1 in Chapter 2 shows, in the US stock market R^2 has increased markedly during the past 20 years, reversing the previous trend of declining R^2 from 1960 to the mid-1990s. Under the classic interpretation, this increase in the R^2 would imply a decrease in the informational efficiency of prices in recent years. However, if returns also contain noise, an alternative interpretation of the trend in the R^2 is that the amount of firm-specific noise in returns has decreased and so R^2 has increased, reflecting *improved* informational efficiency. Put simply, accounting for noise and other drivers of returns can lead to completely different interpretations of the evolving nature of the information in prices. In this chapter, we use the variance decomposition from Chapter 2 to help to resolve this puzzle.

To understand the drivers of R^2, we first derive theoretical relations between R^2 and various information components and noise. We provide an extension of the model of Jin and Myers (2006), allowing for noise traders and pricing errors. We allow returns to be driven by noise, firm-specific information revealed by trading on private information or by other sources such as public news, and market-wide information. The theory shows that R^2 is affected by both firm-specific information and noise. If any of the firm-specific information components of variance are high then R^2 will be low. Similarly, if noise is high R^2 will be low. Therefore, R^2 is not unambiguously associated with more or less informationally efficient prices.

Results of our empirical tests show that the noise component is an important driver of the R^2 trends through time. Decreases in noise during the last two decades explain a large part of the increase in R^2 during the same period. In summary, the recently upward trending R^2 is *not* symptomatic of deteriorating informational efficiency (as would be implied by common interpretations of the R^2 as an inverse indicator of informational efficiency); on the contrary, the increasing R^2 is driven by decreasing noise, which implies more informative prices.

This chapter proceeds as follows. Section 3.2 discusses related literature, and Section 3.3 presents the theoretical model. Section 3.4 explores extensions incorporating cash flow and discount rate news, and Section 3.5 examines the empirical relation between R^2 and the variance components. Section 3.6 concludes the chapter.

3.2 Related literature

This chapter relates to two major strands of literature. The first is stock return variance decompositions. For example, Campbell and Shiller (1988a, 1988b) and Campbell (1991) decompose the variance of unexpected stock returns into two components: cash flow news and discount rate news. Chen, Da, and Zhao (2013) also decompose stock return variance into these components but with a method that directly incorporates cash flow forecasts. A different partition is used in Campbell et al. (2001), who decompose stock return variations into three components: a market-wide return, an industry-specific residual, and a firm-specific residual. Roll (1988) offers a simple but widely used method for stock return decomposition, which simply distinguishes market-wide variations and firm-specific variations. In the previous chapter, we also propose a random walk variance decomposition model drawn from the market microstructure literature to separate information from noise, and partition information into a more granular set of components. The current chapter makes a methodological contribution to the literature by providing a cleaner measure of cash flow and discount rate news components in prices.

Second, our study is related to an ongoing debate regarding whether firm-specific return variations capture firm-specific information or noise. A long-standing interpretation of a higher R^2 is that it implies a lower amount of firm-specific information is reflected in stock prices, and therefore higher R^2 indicates price inefficiency. Morck et al. (2000) document low R^2s in developed countries but high R^2s in emerging and less developed countries. They argue that in poor information environments with weak property rights, stock prices will be less efficient in capturing firm-specific information and will therefore be driven primarily by market-wide information, causing high R^2s. Along similar lines, other studies relate lower R^2 to a better allocation of capital (Wurgler, 2000), higher economic growth (Durnev et al., 2004), and more information about firms' future earnings (DeLisle et al., 2015).

Contrary to the aforementioned view, another body of literature provides evidence to indicate that greater R^2 is associated with less pricing errors, and hence could be used as a price efficiency indicator. For instance, Teoh et al. (2007) find that a low R^2 is associated with high uncertainty about firm fundamentals, which may involve greater mispricing. Under this interpretation, higher R^2s signal more efficient stock markets. Dasgupta et al. (2010) also provide theoretical and empirical evidence to show that R^2 can increase when transparency improves, as stock prices are more informative about future events. Similarly, Kelly (2014) shows that R^2 is positively associated with the quality of the information environment. He finds that a lower R^2 is associated with greater information costs, greater trading costs, and lower liquidity. Gassen et al. (2015) argue that one should control for illiquidity when testing the relationship between R^2 and transparency as R^2 is positively related to liquidity. In contrast, Morck et al. (2013) find a significantly smaller R^2 for US stocks as compared to those in Russia and China, suggesting an inverse relation between R^2 and liquidity. Thus, the issue of whether R^2 is an inverse measure of firm-specific information or an inverse measure of noise remains a puzzle. Our baseline variance decomposition accounting for multiple factors including information and noise allows us to reconcile a longstanding debate in the R^2 literature about how this measure relates to market efficiency.

3.3 Theoretical model

To understand the drivers of R^2, we derive a modified version of the Jin and Myers (2006) model in which we account for noise and also partition firm-specific information into a component revealed by trading on private information and a component revealed through public information such as company announcements and news. The theory provides guidance on how empirical quantities, such as the R^2, will be affected by different types of information and noise.

The model starts with the firm's cash flow generating process:

$$C_t = K_0 X_t, \tag{3.1}$$

where C_t is the future cash flow at time t, K_0 is the initial investment, and X_t captures the random shocks to the cash flow process.

Departing from the Jin and Myers (2006) model, in which the focus is solely on information innovations, we suppose that X_t, as it is perceived by investors ($\widetilde{X_t}$), is driven by four components, namely three information-driven components and noise:

$$\widetilde{X_t} = \theta_{1,t} + \theta_{2,t} + \theta_{3,t} + \theta_{4,t}. \tag{3.2}$$

The term $\theta_{1,t}$ reflects market-wide cash flow information, while $\theta_{2,t}$ and $\theta_{3,t}$ reflect firm-specific cash flow information, and $\theta_{4,t}$ reflects noise. The distinction between $\theta_{2,t}$ and $\theta_{3,t}$ is that $\theta_{2,t}$ is firm-specific information revealed through trading on private information, while $\theta_{3,t}$ is firm-specific information revealed through public information such as company announcements and news. Firm-specific information ($\theta_{2,t}$ and $\theta_{3,t}$) is orthogonal to market-wide information ($\theta_{1,t}$) and noise is assumed to be uncorrelated with information.

The noise ($\theta_{4,t}$) is a reduced form way to account for many sources of noise in prices. It accounts for the fact that prices set by investors following their beliefs about $\widetilde{X_t}$ can deviate from intrinsic values that are based solely on true information. The addition of $\theta_{4,t}$ to the model is in the spirit of Black's (1986) notion that "noise trading is trading on noise as if it were information", i.e., investors believe $\theta_{4,t}$ to be information about cash flows, causing prices to depart from their efficient intrinsic values. There are of course other sources of noise in prices such as a discrete pricing grid, non-synchronous trading and so on, but for the sake of simplicity we stick to one noise term as that is sufficient to characterize the effect of noise on stock returns.

Similar to Jin and Myers (2006), we assume that all of the information and noise components follow stationary AR(1) processes driven by a set of random shocks ($\varepsilon_{1,t}$, $\varepsilon_{2,t}$, $\varepsilon_{3,t}$, and $\varepsilon_{4,t}$).[19] That is, $\theta_{i,t+1} = \theta_{i,0} + \varphi\theta_{i,t} + \varepsilon_{i,t+1}$, where $0 < \varphi < 1$. We define r as the discount rate and K_t as the investors' valuation of the firm, which is the present value of future cash flows (assuming each period's cash flow is paid out), conditional on the information that the investors have at date t about $\widetilde{X_t}$:

$$K_t = PV\{E(C_{t+1}|I_t), E(C_{t+2}|I_t), \dots; r\}. \tag{3.3}$$

For investors $I_t = \{\theta_{1,t} + \theta_{2,t} + \theta_{3,t} + \theta_{4,t}\}$ and therefore the value of K_t is:

$$K_t = \frac{K_0 X_0}{r(1-\varphi)} - \frac{K_0 X_0 \varphi}{(1+r-\varphi)(1-\varphi)} + \frac{\varphi}{1+r-\varphi} K_0(\theta_{1,t} + \theta_{2,t} + \theta_{3,t} + \theta_{4,t}). \tag{3.4}$$

Let $\tilde{r}_{t+1}$ be the total realized return on the firm's shares in the period $t+1$. The return $\tilde{r}_{t+1}$ is calculated as the change in the investors' valuation of the firm from one period to the next plus that period's cash flow, which is paid out, and is therefore a function of the shocks to the investors' information about the cash flow process:

$$\tilde{r}_{t+1} = r + b_t\left(\varepsilon_{1,t+1} + \varepsilon_{2,t+1} + \varepsilon_{3,t+1} + \varepsilon_{4,t+1}\right), \tag{3.5}$$

where

$$b_t = \frac{(1+r)}{\frac{X_0(1+r)}{r} + \varphi(\theta_{1,t} + \theta_{2,t} + \theta_{3,t} + \theta_{4,t})}. \tag{3.6}$$

---

[19] This and other assumptions made in this section are for the convenience of solving the motivating theoretical model (they mainly follow from Jin and Myers (2006)). We do not rely on these assumptions in the empirical model in the following section.

The derivation of Equations (3.5-3.6) is provided in Appendix 3.B (Proof 1). Equation (3.6) implies that the random component of realized stock returns, $b_t(\varepsilon_{1,t+1} + \varepsilon_{2,t+1} + \varepsilon_{3,t+1} + \varepsilon_{4,t+1})$, is driven by shocks to the various types of information (market-wide information, firm-specific information revealed through trading on private information, and public firm-specific information) as well as innovations in noise.

The model suggests stock return variance can be decomposed into four distinct sources of variation. Performing a variance decomposition on the realized returns in this theoretical framework serves as a guide for the empirical decomposition in the next section. Because of the independence between the components of realized returns, the variance of realized returns is equal to the sum of the contributions from each of the components. This allows us to define a set of variance shares as:

$$\eta_j = \frac{Var(\varepsilon_{j,t})}{Var(\varepsilon_{1,t} + \varepsilon_{2,t} + \varepsilon_{3,t} + \varepsilon_{4,t})}, \tag{3.7}$$

with subscript $j = \{1, 2, 3, 4\}$ denoting each return component. The first of these variance shares, η_1, is the contribution of market-wide information to an individual stock's variance. The second (η_2) is the contribution of firm-specific information that is revealed through trading (private firm-specific information), while the third (η_3) is the contribution of public firm-specific information. The last component (η_4) is the effect of noise on stock return variance. We also return to the theoretical model later in the paper (Section 3.5) to analyze what drives trends in the R^2 of a market model regression.

3.4 Extensions incorporating cash flow and discount rate news

This section extends the variance decomposition proposed in Chapter 2 by separating each of the information components of variance into cash flow and discount rate parts. One reason for doing so is that by accounting for noise, decompositions of cash flow/discount rate news can be performed at higher frequencies (traditionally, monthly returns are used to minimize concerns about noise), which allows examination of the time-series trends in those information components.

Our sample consists of all common stocks listed on the NYSE, AMEX, and NASDAQ. We use daily data on returns, prices, market capitalizations, volumes, and sectors for the period from 1960 to 2015 from the Center for Research in Security Prices (CRSP). Appendix 3.A contains a summary of variable definitions. First, we review the standard approach for separating cash flow and discount rate news, developed by Campbell and Shiller (1988a, 1988b) and Campbell (1991) and subsequently used in many papers (Section 3.4.1). We then extend the standard approach by accounting for noise, noting how noise impacts the estimated cash flow and discount rate news (Section 3.4.2). Finally, we use cash flow/discount rate decompositions to produce an extended version of our variance decomposition (Section 3.4.3).

3.4.1 The standard approach to separating cash flow and discount rate news

Campbell and Shiller (1988a, 1988b) and Campbell (1991) show, without having to make behavioural or preference assumptions, that an unexpected stock return, $\varepsilon_{r_{t+1}}$ is equal to two parts:

$$
\begin{aligned}
\varepsilon_{r_{t+1}} &= (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} - (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} \\
&= \varepsilon_{CASHFLOWS_{t+1}} + \varepsilon_{DISCOUNT_{t+1}},
\end{aligned} \tag{3.8}
$$

where $\varepsilon_{CASHFLOWS_{t+1}} = (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j}$ is cash flow news and $\varepsilon_{DISCOUNT_{t+1}} = -(\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j}$ is discount rate news, d_t is the log dividend at time t, r_t is the log holding period return at time t, and $\rho \approx 0.96$ is a constant.

The terms in Equation (3.8) can be estimated from a VAR in which one of the variables is the log stock return.[20] The typical approach is to use the VAR to estimate discount rate news because that does not require information on dividends, and then obtain the cash flow news as the difference between the unexpected stock return and the discount rate news, $\varepsilon_{CASHFLOWS_{t+1}} = \varepsilon_{r_{t+1}} - \varepsilon_{DISCOUNT_{t+1}}$. The importance of cash flow news and discount rate news can be quantified by the variance or standard deviation of the two time-series: $\varepsilon_{CASHFLOWS_t}$ and $\varepsilon_{DISCOUNT_t}$.

3.4.2 Accounting for noise when separating cash flow and discount rate news

A limitation of the standard approach (summarized above) for separating cash flow and discount rate news is that it does not account for the noise in stock returns. Without accounting for noise, the cash flow/discount rate decomposition can only be reliably performed using low-frequency data such as monthly returns so that the ratio of noise to information remains within acceptable error tolerances. Therefore, the standard approach has limited ability to examine time-series variation in the cash flow/discount rate components. For example, with monthly returns and a minimum of say 120 time-series observations in the VAR, one can obtain a single value of cash flow and discount rate variance every ten years. Accounting for noise, however, allows us to apply the decomposition to daily data and thereby estimate cash flow and discount rate news variances every year. This frequency reveals time-series trends in cash flow and discount rate news and also enables us to further partition the information components in our baseline model.

To understand how noise manifests in a standard cash flow/discount rate decomposition and therefore how to approach the task of isolating noise in the decomposition, consider Figure 3.1 Panel A. A stock return is composed of a discount rate that captures the required or expected rate of return, noise, and information. Noise has an expected and an unexpected component. The expected component arises from reversals of pricing errors. For example, a positive pricing error is expected to reverse

---

[20] For example, once the VAR is estimated, one can obtain the time t expectations of returns at $t+2$, $t+3$ and so on (multi-step forecasts from the VAR) from which one can compute $\sum_{j=1}^{\infty} \rho^j \mathrm{E}_t[r_{t+1+j}]$. Repeating this process at time $t+1$ one obtains $\sum_{j=1}^{\infty} \rho^j \mathrm{E}_{t+1}[r_{t+1+j}]$. The difference gives the discount rate news at time $t+1$, i.e., $\varepsilon_{DISCOUNT_{t+1}} = -(\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} = \sum_{j=1}^{\infty} \rho^j \mathrm{E}_t\left[r_{t+1+j}\right] - \sum_{j=1}^{\infty} \rho^j \mathrm{E}_{t+1}[r_{t+1+j}]$.

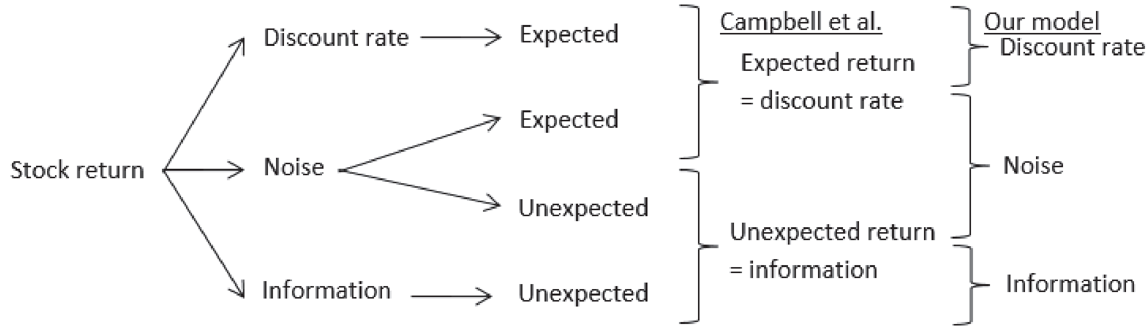
resulting in an expected negative return component.[21] The unexpected component of noise reflects random changes to the pricing errors. Thus, the expected return is made up of the discount rate and the return from the expected change in the pricing error.

The unexpected return is driven by information arrivals and shocks to pricing errors (unexpected noise). Therefore, in the standard cash flow/discount rate decomposition, noise contaminates the estimated discount rate news because the expected return reflects the discount rate *and* noise. Noise also contaminates the estimated cash flow news component because: (i) cash flow news is usually calculated as the difference between the unexpected stock return and the discount rate news, which is contaminated by noise; and (ii) part of the unexpected return, which goes into the cash flow news calculation, is noise. To resolve these issues, our modified cash flow/discount rate decomposition first removes noise from both the expected and unexpected returns, resulting in a method that is suitable for higher frequency data.

---

[21] There are several reasons why pricing errors can be inferred from past returns and their reversals are somewhat predictable. At the most basic level, bid-ask bounce (trade prices oscillating between the bid and the ask or offer quotes) creates negative serial correlation in returns and therefore a predictable "noise" component of returns (e.g., Roll, 1984). For example, if a stock's closing price is at the bid quote, its next close could be at the bid or the ask/offer and therefore, merely on the basis of bid-ask bounce, in expectation the next closing price will be higher, i.e., there is an expected positive noise return. Return predictability due to pricing errors goes well beyond the bid-ask bounce effect. Negative serial correlation is also found in midquote returns of individual stocks (suggesting quoted prices also suffer from temporary mean-reverting pricing errors) and at longer horizons such as weekly and monthly returns (e.g., Jegadeesh, 1990; Lehmann, 1990; Hendershott and Menkveld, 2014). The economically meaningful reversals in returns at daily through to monthly horizons (which has been validated in many studies) is linked to imperfect liquidity and the inability for the market to absorb order imbalances without temporarily deviating from efficient prices (e.g., Avramov et al., 2006; Hendershott et al., 2011; Nagel, 2012). The existence of predictable reversals in returns due to temporary price distortions from efficient prices is also supported by market microstructure theory. For example, classic inventory control models of liquidity supply show that when risk averse liquidity providers receive many buy orders, they "shade" their subsequent quoted prices upward (above the efficient price) to attract sellers and thereby revert their inventory towards zero, and vice versa when they receive many sell orders (e.g., Stoll, 1978, Ho and Stoll, 1981, and many subsequent models). Similarly, return reversals due to distortions from efficient prices arise in models of adverse selection (e.g., Kyle, 1985; Glosten and Milgrom, 1985) if liquidity providers are risk averse or less than perfectly competitive (e.g., Subrahmanyam, 1991; Nagel, 2012).

Panel A: Adjusting a standard cash flow/discount rate decomposition to account for noise



Panel B: Extended variance decomposition

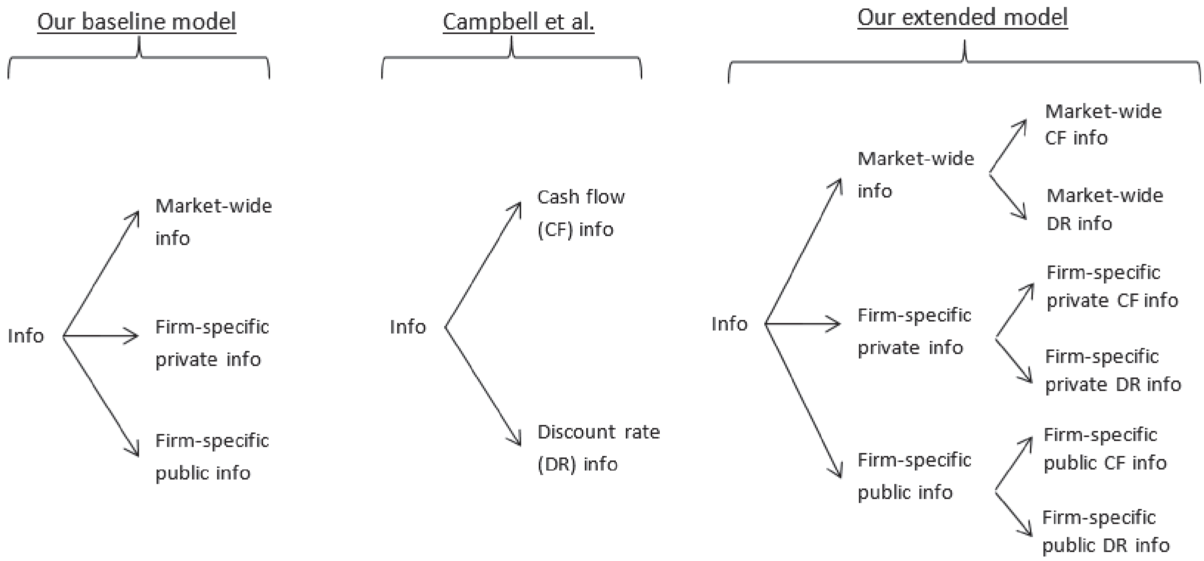


Figure 3.1. Extension of variance decomposition to cash flow and discount rate information.

Panel A shows how noise is dealt with in a standard cash flow/discount rate news decomposition (e.g., Campbell, 1991) and in our modified cash flow/discount rate news decomposition. In the standard decomposition, the expected changes in pricing errors contaminate the discount rate (expected return) and the unexpected changes in pricing errors contaminate the cash flow news. In our modified decomposition, noise is removed from both the discount rate and cash flow news. Panel B shows how our baseline variance decomposition is extended by splitting each of the baseline model's information components into a cash flow and discount rate part.

First we modify our baseline model to allow for a time-varying discount rate. The efficient price from Equation (2.2) in Chapter 2 becomes:

$$m_t = m_{t-1} + \mu_t + w_t \,, \tag{3.9}$$

and the stock return from Equation (2.3) becomes:

$$r_t = p_t - p_{t-1} = \mu_t + w_t + \Delta s_t \,, \tag{3.10}$$

where the time-varying drift, μ_t, is the discount rate on the stock over the time t period, w_t is an innovation that reflects new information about the stock's fundamentals, and Δs_t is the change in pricing error. Noise has an expected component ($\mathrm{E}_{t-1}[\Delta s_t]$) and an unexpected component (ε_{s_t}), $\Delta s_t = \mathrm{E}_{t-1}[\Delta s_t] + \varepsilon_{s_t}$. The expected component comes from the fact that pricing errors are temporary and therefore tend to reverse as discussed above. Consequently, the expected return ($\mathrm{E}_{t-1}[r_t]$) is made up of the discount rate and the expected change in the pricing error, $\mathrm{E}_{t-1}[r_t] = \mu_t + \mathrm{E}_{t-1}[\Delta s_t]$. Similarly, the unexpected return ($\varepsilon_{r_t} = r_t - \mathrm{E}_{t-1}[r_t]$) is made up of new information about the stock's fundamentals and unexpected changes in the pricing error (noise), $\varepsilon_{r_t} = w_t + \varepsilon_{s_t}$.

The information-driven innovation in the efficient price is the same as in our baseline model and is estimated from the VAR/VMA: $w_t = \theta_{r_m}\varepsilon_{r_m,,t} + \theta_x\varepsilon_{x,t} + \theta_r\varepsilon_{r,t}$. We estimate the expected return on the stock over the next period, $\mathrm{E}_{t-1}[r_t]$, as the one-period-ahead forecast of the return from the VAR, in the spirit of Campbell (1991). We isolate the expected noise part of the expected return by considering what part of the expected return is predicted by past unexpected changes in the pricing error, $\mathrm{E}_{t-1}[\Delta s_t] = \frac{Cov(\mathrm{E}_{t-1}[r_t],\varepsilon_{s_{t-1}})}{Var(\varepsilon_{s_{t-1}})}\varepsilon_{s_{t-1}}$.[22] The other part of the expected return is the discount rate, $\mu_t = \mathrm{E}_{t-1}[r_t] - \mathrm{E}_{t-1}[\Delta s_t]$. We obtain the unexpected innovations in the pricing error from $\varepsilon_{s_t} = r_t - \mathrm{E}_{t-1}[r_t] - w_t = r_t - \mathrm{E}_{t-1}[\Delta s_t] - \mu_t - w_t$ and

[22] This approach is equivalent to estimating the predictive regression, $\mathrm{E}_{t-1}[r_t] = a + b\varepsilon_{s_{t-1}} + e_{t-1}$, where the estimate of the coefficient b is given by $\hat{b} = \frac{Cov(\mathrm{E}_{t-1}[r_t],\varepsilon_{s_{t-1}})}{Var(\varepsilon_{s_{t-1}})}$ and the part of $\mathrm{E}_{t-1}[r_t]$ that is explained by $\varepsilon_{s_{t-1}}$ is $\hat{b}\varepsilon_{s_{t-1}}$. This approach picks up the first-order negative serial correlation in returns that occurs at daily frequencies due to bid-ask bounce and price pressures.

consequently the total change in the pricing error (sum of expected and unexpected parts) is $\Delta s_t = \mathrm{E}_{t-1}[\Delta s_t] + \varepsilon_{s_t} = r_t - \mu_t - w_t$.

A simple schematic of what is going on in the process above is shown in Figure 3.1 Panel A. We break noise into expected and unexpected parts. Subtracting expected noise from the expected return gives the “clean” discount rate. The clean discount rate is similar to the discount rate in Campbell (1991) but purged of noise. Subtracting unexpected noise from the unexpected return gives the “clean” information. The clean information is similar to the cash flow and discount rate information in Campbell (1991) but purged of noise.

Next we apply a cash flow/discount rate decomposition similar to Campbell (1991), but using the clean discount rate and the clean information. Using the de-noised expected return ($\mathrm{E}_t[\mu_{t+1}]$) in place of the standard expected return ($\mathrm{E}_t[r_{t+1}]$), we estimate discount rate news using the Campbell (1991) approach:

$$\begin{aligned}\varepsilon_{DISCOUNT_{t+1}} &= -(\mathrm{E}_{t+1} - \mathrm{E}_t)\sum_{j=1}^{\infty}\rho^j \mu_{t+1+j} \\ &= \sum_{j=1}^{\infty}\rho^j \mathrm{E}_t\left[\mu_{t+1+j}\right] - \sum_{j=1}^{\infty}\rho^j \mathrm{E}_{t+1}\left[\mu_{t+1+j}\right].\end{aligned} \tag{3.11}$$

Also following Campbell (1991), but using the de-noised unexpected return instead of the standard unexpected return, we estimate the cash flow news at time $t+1$ as the informational part of the return that is not associated with discount rate news:

$$\varepsilon_{CASHFLOWS_{t+1}} = w_{t+1} - \varepsilon_{DISCOUNT_{t+1}}. \tag{3.12}$$

From the time-series of the cash flow and discount rate news, we compute the variances $Var(\varepsilon_{CASHFLOWS_t})$ and $Var(\varepsilon_{DISCOUNT_t})$. We also compute the variance of noise, $Var(\Delta s_t)$.[23] We then plot the cash flow news, the discount rate news, and the noise as shares of variance.

Figure 3.2 plots the time-series of the cash flow news, discount rate news, and noise, expressed as shares of stock return variance.[24] Panel A reports results from the standard model that does not account for noise as represented in Equation (3.8), while Panel B is the model that accounts for noise and is described in Equations (3.11-3.12). In the model that does not account for noise, cash flow news is estimated to account for around 75% of stock return variance, while discount rate information makes up around 10%. The remaining variation is attributable to time-series variation in the discount rate itself (15%), which is different from discount rate news.[25] These results are consistent with Vuolteenaho (2002) who also performs a variance decomposition on individual stocks without accounting for noise and finds similar estimates.[26]

---

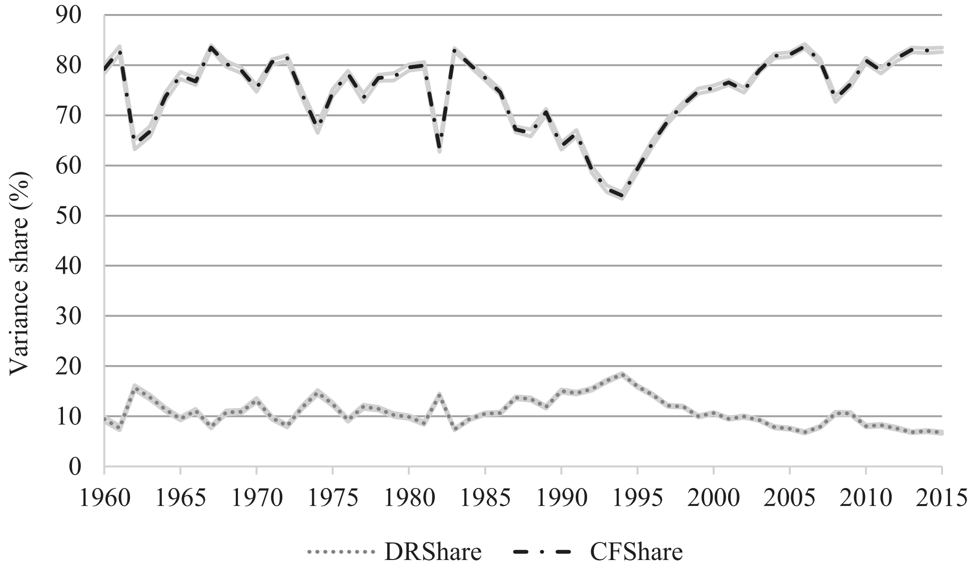
[23] The variance of noise differs slightly from our baseline model because we allow for a time-varying discount rate.

[24] In expressing the variance components as "shares" of variance, to make the results comparable to other models in the paper, we must also consider the covariance between cash flow and discount rate news. Given the total information in this model is the same as in the baseline model, to ensure the sum of the information component variances in this model equal the variance of information in the baseline model, we allocate a fraction α of $2Cov(\varepsilon_{DISCOUNT_t}, \varepsilon_{CASHFLOWS_t})$ to the cash flow news variance and a fraction $(1-\alpha)$ to the discount rate news variance, where $\alpha = \frac{Var(\varepsilon_{DISCOUNT_t})}{Var(\varepsilon_{DISCOUNT_t})+Var(\varepsilon_{CASHFLOWS_t})}$. Doing so does not change the ratio of cash flow news to discount rate news and, for consistency, we apply this covariance attribution to both the models that account for noise and those that do not.

[25] The time-varying discount rate, $\mathrm{E}_t[r_{t+1}]$ in the model that does not account for noise and μ_t in the model that does account for noise, gives rise to variation in returns directly by determining the average rates of return in different periods, whereas the discount rate *news* captures price changes that occur when expectations of the discount rate change and the stock is re-priced accordingly. Given our focus on information and noise, we do not report the time-varying discount rate variance share in the plots.

[26] To better compare with Vuolteenaho (2002), we also calculate the ratio of cash flow news variance to discount rate news variance over the period from 1960 to 1996. Despite differences in data frequency and VAR model used, the ratio of cash flow news variance to discount rate news variance is about five times in our model, which is very similar to the ratio reported in Vuolteenaho (2002) for the same period of time.

Panel A: Cash flow / discount rate decomposition not accounting for noise



Panel B: Cash flow / discount rate decomposition accounting for noise

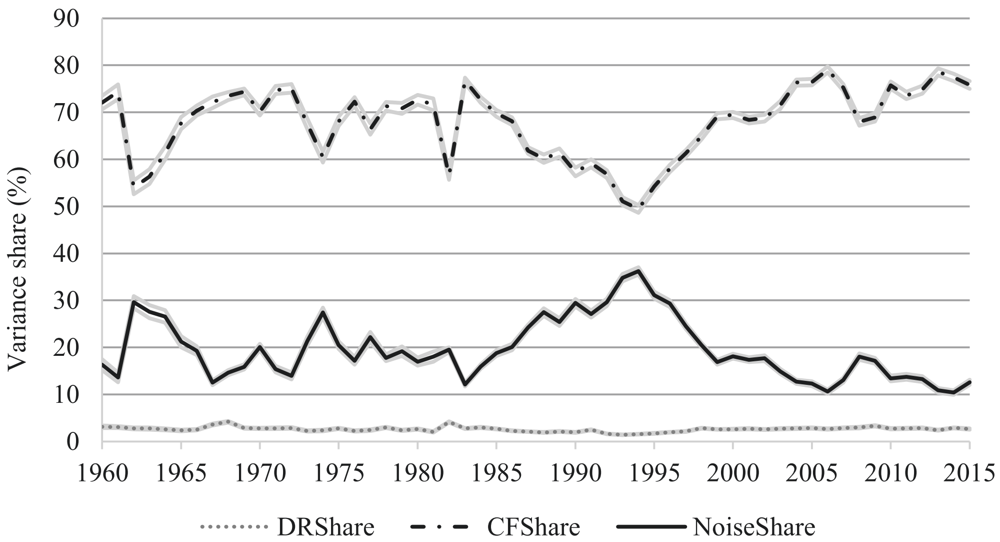


Figure 3.2. Cash flow news, discount rate news, and noise through time.
This figure shows the time-series trends in the percentage of stock return variance that is attributable to time-variation in the cash flow news ($CFShare$), discount rate news ($DRShare$), and noise ($NoiseShare$) from 1960 to 2015. Panel A shows the components estimated from a standard cash flow / discount rate news decomposition that does not account for noise. Panel B shows the components estimated from our modified cash flow / discount rate news decomposition that does account for noise. The variance components are calculated separately for each stock each year, and then averaged across stocks each year. Light gray lines provide 99% confidence intervals. The sample consists of stocks listed on NYSE, AMEX, and NASDAQ.

Other studies have performed similar decompositions on portfolios of stocks rather than individual stocks (e.g., Campbell, 1991; Campbell and Ammer, 1993). In portfolios, discount rate news plays a larger role, suggesting that cash flow news is more idiosyncratic than discount rate news. The dominance of cash flow information in our stock-level variance decomposition and the fact that cash flow information tends to be relatively idiosyncratic is also consistent with our baseline decomposition, which shows that idiosyncratic information is a far more important driver of individual stock returns than market-wide information.

Figure 3.2, Panel B adjusts the standard cash flow/discount rate decomposition for noise and reveals some interesting differences. A striking result is that almost all of the stock price variations associated with information is driven by cash flow news, with very little variation attributed to discount rate news. In fact, cash flow news is responsible for 68% of stock return variance in the full sample, whereas discount rate news accounts for less than 3%. It is natural to expect that accounting for noise would decrease both of these information components as some of the variation labelled as information in the standard models is noise. The interesting observation is that they do not decrease by a similar amount. The decrease in estimated discount rate news is far greater, resulting in a substantial increase in the estimated ratio of cash flow news to discount rate news when accounting for noise.

The results suggest that much of what is usually labelled as discount rate news is actually noise. Why? The primary reason is that noise creates considerable return predictability, so expected returns are not good measures of discount rates. Noise creates return predictability because pricing errors are stationary, mean reverting processes. Prices are drawn towards fundamental values in the long run, so a positive noise-driven return shock in one period leads to a negative expected return component over the next period and vice versa. The empirical consequence of pricing error reversals is the widely documented negative serial correlation in returns, which is observed at a wide range of frequencies from the classic monthly reversals anomaly (e.g., Jegadeesh, 1990) to weekly, daily, and intraday horizons (e.g., Roll, 1984). Without accounting for noise, variation in the discount rate is overestimated when the expected/forecast return is taken as an estimate of the discount rate, leading to a substantial overestimation of the discount rate news component.

Estimates of cash flow news are also affected by explicitly accounting for noise, but to a lesser extent due to two opposing effects. These effects are best illustrated by recognizing that cash flow news is the difference between estimated information and estimated discount rate news: $\varepsilon_{CASHFLOWS_t} = w_t - \varepsilon_{DISCOUNT_t}$. First, removing noise shrinks the estimated information shocks (w_t), which tends to decrease cash flow news. But, second, as explained above, the estimated discount rate news ($\varepsilon_{DISCOUNT_t}$) is considerably smaller after accounting for noise and this effect tends to increase the estimated cash flow news. The opposing effects explain why the estimated magnitude of cash flow news is less affected by accounting for noise than the estimated magnitude of discount rate news.

An advantage of isolating noise is the ability to apply the decomposition over relatively short windows using high-frequency data. Unlike previous studies, this allows us to examine the time-series variation in the cash flow and discount rate news. Figure 3.1 shows that since the late 1990s, there has been a notable increase in the proportion of stock returns that are attributable to cash flow news, mirroring the decrease in noise during the same period. This trend matches our earlier decomposition that shows firm-specific information has become an increasingly important component of stock returns during the past two decades, consistent with the widely held view that financial markets are now more informationally efficient than in previous decades.

3.4.3 Extended variance decomposition

Armed with a method to separate cash flow and discount rate news at the daily frequency purged of noise, we further extend our baseline variance decomposition by splitting each information component into a cash flow part and a discount rate part. This extended decomposition of information is illustrated in Figure 3.1 Panel B. Note that the noise and time-varying discount rate components are not shown.

The six information components in the extended decomposition are obtained from the following regressions of cash flow and discount rate news on each of the information components from our variance decomposition:

$$\varepsilon_{DISCOUNT_t} = \beta_1 r_{A,t} + \beta_2 r_{B,t} + \beta_3 r_{C,t}$$

$$\varepsilon_{CASHFLOWS_t} = \gamma_1 r_{A,t} + \gamma_2 r_{B,t} + \gamma_3 r_{C,t}, \tag{3.13}$$

where the information components are market-wide information ($r_{A,t} = \theta_{r_m} \varepsilon_{r_m,t}$), firm-specific private information ($r_{B,t} = \theta_x \varepsilon_{x,t}$), and firm-specific public information ($r_{C,t} = \theta_r \varepsilon_{r,t}$).[27] From the fitted values we obtain six sources of variance: market-wide discount rate and cash flow news, $\widehat{\beta_1} r_{A,t}$ and $\widehat{\gamma_1} r_{A,t}$, firm-specific discount rate and cash flow news incorporated through trading on private information, $\widehat{\beta_2} r_{B,t}$ and $\widehat{\gamma_2} r_{B,t}$, and firm-specific discount rate and cash flow news incorporated through public information, $\widehat{\beta_3} r_{C,t}$ and $\widehat{\gamma_3} r_{C,t}$, respectively. In expressing the variance components as variance shares, we add back the covariance between cash flow and discount rate news as before, preserving the total variance attributable to information.

Table 3.1 reports means of the seven variance components from our extended decomposition, namely six information components and a noise component, expressed as percentages of variance.[28] The pooled sample results are presented in Panel A. Panel B shows the results separately for the two sub-periods, before and after 1997. Results for size, price, and industry subgroups are presented in Panel C, D, and E, respectively. Consistent with our earlier observation corroborating Chen et al. (2013) that cash flow news is a much larger driver of individual stock returns than discount rate news, we also find that the cash flow parts of the market-wide and firm-specific information components are much larger than the corresponding discount rate parts. Overall, firm-specific cash flow information comprises the largest contribution to individual stock return variance, accounting for 57% of variance (the sum of the CF columns for $PrivateInfoShare$ and $PublicInfoShare$ in Table 3.1 Panel A).

[27] In the regression, $\beta_i + \gamma_i = 1$, thereby preserving the total amount of each information type.
[28] For conciseness, we do not report the share of variance attributable to time-variation in the discount rate (μ_t), which is why the seven reported components sum to slightly less than 100%.

Table 3.1. Stock return variance components in the extended decomposition model.

This table reports mean variance shares (expressed as percentages of variance). Using an extended decomposition model, stock return variance is decomposed into market-wide information ($MktInfoShare$), private firm-specific information ($PrivateInfoShare$), public firm-specific information ($PublicInfoShare$), and noise ($NoiseShare$). The three information components are further decomposed into discount rate (DR) and cash flow (CF) related components. Panel A reports full sample averages. Panel B splits the sample into two sub-periods from 1960 to 1996, and from 1997 to 2015. Panels C and D group stocks into quartiles by price and size (market capitalization), respectively, with quartiles formed separately each year. Panel E groups stocks into major industry groups: the *Consumer* group comprises the industries Consumer Durables, NonDurables, Wholesale, Retail, and some Services (Laundries, Repair Shops); the *Healthcare* group comprises the industries Healthcare, Medical Equipment, and Drugs; the *Manufact* group comprises the industries Manufacturing, Energy, and Utilities; the *HiTech* group comprises the industries Business Equipment, Telephone and Television Transmission; and the *Other* group comprises all other industries. The variance components are calculated separately for each stock in each year and then averaged across stocks within the corresponding quartile or group. We also report the differences in means for the post-1997 period minus the pre-1997 period (Panel B) and quartile 1 minus quartile 4 (Panels C and D) and corresponding t-statistics in parentheses. ***, **, and * indicate statistically significant differences at the 1%, 5%, and 10% levels using standard errors clustered by stock and by year. The sample consists of stocks listed on NYSE, AMEX, and NASDAQ from 1960 to 2015.

	$MktInfoShare$ (%)		$PrivateInfoShare$ (%)		$PublicInfoShare$ (%)		$NoiseShare$
	DR	CF	DR	CF	DR	CF	(%)
Panel A: Full sample							
	0.43	7.61	0.66	22.60	1.35	34.79	22.07
Panel B: Sub-periods							
1960-1996	0.34	6.45	0.58	20.71	1.24	33.36	26.61
1997-2015	0.52	8.85	0.75	24.63	1.48	36.34	17.16
Difference	0.18	2.40	0.17	3.92	0.24	2.99	-9.45
(Post-Pre 1997)	(2.45)**	(1.70)*	(2.78)***	(2.66)***	(3.49)***	(3.32)***	(-6.06)***
Panel C: Quartiles by Price							
Q1=low	0.30	4.56	0.64	21.50	1.36	35.36	24.66
Q2	0.47	9.00	0.67	22.61	1.41	35.30	21.27
Q3	0.69	13.75	0.73	25.23	1.29	33.71	16.25
Q4=high	0.90	19.12	0.80	27.45	1.19	30.43	12.14
Difference	-0.60	-14.56	-0.17	-5.95	0.17	4.94	12.52
(Q1-Q4)	(-4.92)***	(-15.96)***	(-3.35)***	(-4.71)***	(2.47)**	(6.22)***	(10.90)***
Panel D: Quartiles by size (market capitalization)							
Q1=low	0.28	4.05	0.59	20.29	1.31	35.28	26.29
Q2	0.44	8.04	0.69	23.49	1.42	35.22	21.25
Q3	0.67	13.02	0.78	26.22	1.42	34.90	14.58
Q4=high	0.92	20.30	0.87	28.89	1.30	30.13	9.85
Difference	-0.64	-16.25	-0.27	-8.61	0.01	5.15	16.44
(Q1-Q4)	(-5.10)***	(-15.13)***	(-4.69)***	(-5.49)***	(0.14)	(5.42)***	(11.85)***
Panel E: Industry groups							
Consumer	0.40	7.08	0.66	21.98	1.31	34.14	23.77
Healthcare	0.41	6.69	0.74	27.20	1.46	35.11	18.30
HiTech	0.47	8.74	0.71	24.68	1.37	35.14	18.75
Manufact	0.45	8.49	0.70	23.00	1.32	32.86	23.04
Other	0.40	6.73	0.59	19.12	1.34	35.87	24.90

What is perhaps more interesting is that the ratio of cash flow to discount rate news differs across the three information components. The differences are consistent with the notion that cash flow news tends to be more idiosyncratic than discount rate news. For example, the ratio of cash flow news to discount rate news in the firm-specific information component is around 29 times, whereas in market-wide information it is around 18 times. We observe this relation in all price and size quartiles as well as industry groups. This finding helps reconcile differing results in the literature: when variance decompositions are performed on portfolios of stocks (e.g., Campbell, 1991; Campbell and Ammer, 1993), in which most of the firm-specific variation is cancelled out through diversification, leaving predominantly market-wide information, discount rate news tends to be more important than when variance decompositions are performed on individual stocks (e.g., Vuolteenaho, 2002; Chen et al., 2013). We extend these findings by showing that once we isolate the noise component of returns, the importance of cash flow news relative to discount rate news becomes even more apparent.

3.5 The empirical relation between R^2 and the variance components

Drawing on the theoretical model in Section 3.3, R^2 can be written as a function of the four components of stock return variance or the four variance shares:[29]

$$R^2 = \frac{Var(\varepsilon_{1,t})}{Var(\varepsilon_{1,t} + \varepsilon_{2,t} + \varepsilon_{3,t} + \varepsilon_{4,t})} = \frac{1}{1 + \frac{1}{\eta_1}(\eta_2 + \eta_3 + \eta_4)}. \tag{3.14}$$

Recall η_1 is the market-wide information share, η_2 is the share of firm-specific information that is revealed through trading, η_3 is the share of public firm-specific information, and η_4 is the noise share. Equation (3.14) shows that R^2 is affected by both firm-specific information and noise. If any of the firm-specific information components of variance (η_2, η_3) are high then R^2 will be low. Similarly, if noise (η_4)

[29] See Appendix 3.B Proof 2 for the derivation.

is high, R^2 will be low. Therefore, R^2 is not unambiguously associated with more or less informationally efficient prices. We formalize the model's predictions regarding the relation between R^2 and each component of stock return variation as follows:

Hypothesis 1: R^2 is negatively related to the private and public firm-specific information components of variance, ceteris paribus.

Hypothesis 2: R^2 is negatively related to the amount of noise in prices, ceteris paribus. Because noise and liquidity are inversely related, R^2 is positively related to liquidity, ceteris paribus.

Hypothesis 3: R^2 is positively related to the market-wide information component of variance, ceteris paribus.

Following the literature on stock price synchronicity, we estimate R^2 for each stock in each year, using daily returns:

$$r_{i,t} = \alpha + \beta r_{m,t} + \varepsilon_{i,t}. \tag{3.15}$$

Each year, we compute the average R^2 from the cross-section of stocks to obtain an annual time-series, R_t^2. We do the same for each of the variance components obtained from our variance decomposition. We perform a logit transformation on both the R_t^2 and the variance shares so that they are not bounded to the [0,1] interval. To compare the relative importance of each variance share as a driver of the trends in R^2 we standardize the logit-transformed time-series to have a mean of zero and a standard deviation of one.

We estimate the following time-series regression for the pre-1997 and post-1997 periods using the standardized logit-transformed variables:

$$\begin{aligned} R_t^2 = \ & \alpha + \beta_1 PrivateInfoShare_t + \beta_2 PublicInfoShare_t \\ & + \ \beta_3 NoiseShare_t + \varepsilon_t. \end{aligned} \tag{3.16}$$

Table 3.2 reports the results. Column (1) reports the pre-1997 results, column (2) reports the post-1997 results, and column (3) reports the full period with interaction terms to capture the change in coefficients after 1997. Consistent with our theoretical model and hypotheses, noise and the two types of firm-specific information are negatively related to R_t^2. An increase in either noise or firm-specific information tends to decrease R^2. In the pre-1997 analysis, a one standard deviation increase in private firm-specific information will decrease R^2 by 0.43 standard deviations, whereas a one standard deviation increase in public firm-specific information will decrease R^2 by 0.71, and a one standard deviation increase in noise will decrease R^2 by 0.91 standard deviations. Based on the magnitudes of the independent variables, noise is the dominant factor and has a larger effect on the R^2 than firm-specific information.

The magnitude of the coefficients increases in the post-1997 period, but they each retain the same sign, consistent with our hypotheses. This result suggests that after 1997, the R^2 is even more sensitive to changes in firm-specific information and changes in the level of noise. During both sub-periods, noise is the strongest driver of R^2 by magnitude.

The regression results, together with the time-series of the variance components, help to explain the reversal of the trends in R^2 during the 1997-2015 sub-period: R^2 increases in the second sub-period primarily as a result of the large decrease in noise. While the level of firm-specific information in prices has also increased during this period, putting downward pressure on R^2, this effect is dominated by the upward pressure from decreasing noise. The evidence indicates that the market has become more efficient in the past two decades, contrary to what is implied by interpretations of R^2 as an inverse measure of efficiency.

Table 3.2. Drivers of the trend in R^2 through time.
This table reports the results from time-series regressions in which the dependent variable is stock return co-movement (R_t^2 from a market model) and the units of observation are years. The dependent variable is obtained as an annual time-series by estimating a market model for each stock-year using daily data then in each year taking the cross-sectional average of the R^2 estimates. The independent variables are cross-sectional averages of return variance shares attributable to trading on private firm-specific information ($PrivateInfoShare_t$), public firm-specific information ($PublicInfoShare_t$), and noise ($NoiseShare_t$). The R_t^2 and variance shares are logit transformed and standardized to have a mean of zero and standard deviation of one. D_t^{POST} is a dummy variable that takes the value of one in the post-1997 period and zero otherwise. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes stocks listed on NYSE, AMEX, and NASDAQ. The regression is estimated separately on time periods given in the column headings.

Dependent variable: Stock return co-movement, R_t^2			
Variable	Period 1960-1996	Period 1997-2015	Period 1960-2015
Intercept	-0.26***	0.50**	-0.26***
	(-3.31)	(2.27)	(-3.18)
$PrivateInfoShare_t$	-0.43***	-0.94***	-0.43***
	(-4.03)	(-4.81)	(-3.87)
$PublicInfoShare_t$	-0.71***	-0.89***	-0.71***
	(-9.93)	(-3.51)	(-9.54)
$NoiseShare_t$	-0.91***	-1.50***	-0.91***
	(-8.60)	(-5.58)	(-8.26)
D_t^{POST}			0.75***
			(3.46)
$D_t^{POST} \times PrivateInfoShare_t$			-0.51**
			(-2.41)
$D_t^{POST} \times PublicInfoShare_t$			-0.19
			(-0.75)
$D_t^{POST} \times NoiseShare_t$			-0.59**
			(-2.17)
$Adjusted\ R^2$	84.1%	64.9%	80.8%

3.6 Conclusion

This chapter highlights the necessity of accounting for noise in empirical decompositions of stock return variance, showing that neglecting noise leads to serious problems in drawing the correct inferences.

We first extend our baseline model as discussed in Chapter 2 by decomposing the information components into cash flow and discount rate sub-components to relate our model to this long-standing branch of the existing asset pricing literature. We find that after accounting for the noise in returns, cash flow information is more important than previously believed. Cash flow information plays a considerably larger role than discount rate information in driving individual stock returns. Discount rate information plays a relatively larger role in market-wide information than it does in information about individual firms. The analysis we conduct reconciles our decomposition methodology with the large return decomposition literature on the relative importance of cash flow versus discount rate news (e.g., Campbell and Ammer, 1993; Vuolteenaho, 2002; Chen et al., 2013).

We also revisit the puzzling time trends that emerge from the standard interpretation of the R^2 of a market model as an inverse measure of informational efficiency. More specifically, we employ the results of our baseline decomposition model to explain the increase in R^2 during the post-1997 period. We show that the recent decrease in noise is largely responsible for the increasing R^2 of a market model in the past two decades and corresponding decrease in idiosyncratic volatility. This evidence indicates that the market has become more efficient in the past two decades. An important implication is that a lower R^2 is not necessarily associated with more informationally efficient prices, in contrast to the interpretation of R^2 in prior studies.

Appendix 3.A. Variable definitions

The table below provides descriptions and notation for the variables that are components of stock return variance. Each variable is estimated separately for each stock in each year using daily observations. When aggregating across stocks, we take variance-weighted averages (as per Morck et al., 2000, 2013).

Variable	Notation	Description
Stock return co-movement	R^2	R^2 is estimated by regressing individual daily stock returns on daily market return.
Noise share	$NoiseShare$	The share of stock return variance that is attributable to noise.
Market-wide information share	$MktInfoShare$	The share of stock return variance that is attributable to market-wide information.
Private firm-specific information share	$PrivateInfoShare$	The share of stock return variance that is attributable to trading on private firm-specific information.
Public firm-specific information share	$PublicInfoShare$	The share of stock return variance that is attributable to public firm-specific information.
Firm-specific information share	$FirmInfoShare$	The share of stock return variance that is attributable to firm-specific information (sum of $PrivateInfoShare$ and $PublicInfoShare$).
Discount rate information share	$DRShare$	The share of stock return variance that is attributable to discount rate information.
Cash flow information share	$CFShare$	The share of stock return variance that is attributable to cash flow information.

Variable	Notation	Description
Market-wide discount rate information share	$MktInfoShare(DR)$	The share of stock return variance that is attributable to market-wide discount rate information.
Market-wide cash flow information share	$MktInfoShare(CF)$	The share of stock return variance that is attributable to market-wide cash flow information.
Private firm-specific discount rate information share	$PrivateInfoShare(DR)$	The share of stock return variance that is attributable to trading on private firm-specific discount rate information.
Private firm-specific cash flow information share	$PrivateInfoShare(CF)$	The share of stock return variance that is attributable to trading on private firm-specific cash flow information.
Public firm-specific discount rate information share	$PublicInfoShare(DR)$	The share of stock return variance that is attributable to public firm-specific discount rate information.
Public firm-specific cash flow information share	$PublicInfoShare(CF)$	The share of stock return variance that is attributable to public firm-specific cash flow information.

Appendix 3.B. Proofs

Proof 1.

The total realized return on the firm's shares for outsiders is calculated as the change in the investors' valuation of the firm from one period to the next including the cash flow that is paid out:

$$\tilde{r}_{t+1} = \frac{K_{t+1} + C_{t+1}}{K_t} - 1. \tag{3.B.1}$$

Substituting Equation (3.4) for K_t into (3.B.1) and rearranging gives:

$$\tilde{r}_{t+1} = \frac{\left(\frac{1+r}{1+r-\varphi}\right) C_{t+1} - \frac{\varphi}{1+r-\varphi} C_t}{\frac{(1+r)}{r} \frac{K_0 X_0}{(1+r-\varphi)} + \frac{\varphi}{1+r-\varphi} C_t}. \tag{3.B.2}$$

For investors $I_t = \{\theta_{1,t} + \theta_{2,t} + \theta_{3,t} + \theta_{4,t}\}$ and therefore,

$$\tilde{r}_{t+1} = \frac{\left(\frac{1+r}{1+r-\varphi}\right) K_0 \sum_{j=1}^{4} \theta_{j,t+1} - \frac{\varphi}{1+r-\varphi} K_0 \sum_{j=1}^{4} \theta_{j,t}}{\frac{(1+r)}{r} \frac{K_0 X_0}{(1+r-\varphi)} + \frac{\varphi}{1+r-\varphi} K_0 \sum_{j=1}^{4} \theta_{j,t}}. \tag{3.B.3}$$

Multiplying the denominator and the numerator by $(\frac{1+r-\varphi}{K_0})$ and rearranging, we get:

$$\tilde{r}_{t+1} = r + \frac{(1+r)\left[\sum_{j=1}^{4}\theta_{j,t+1} - \varphi\sum_{j=1}^{4}\theta_{j,t} - X_0\right]}{\frac{(1+r)X_0}{r} + \varphi\sum_{j=1}^{4}\theta_{j,t}}. \tag{3.B.4}$$

Given the assumptions that all of the information and noise components follow stationary AR(1) process (we do not make this assumption in the empirical model), Equation (3.B.4) becomes:

$$\tilde{r}_{t+1} = r + b_t\left(\varepsilon_{1,t+1} + \varepsilon_{2,t+1} + \varepsilon_{3,t+1} + \varepsilon_{4,t+1}\right), \tag{3.B.5}$$

where
$$b_t = \frac{(1+r)}{\frac{(1+r)X_0}{r} + \varphi(\theta_{1,t} + \theta_{2,t} + \theta_{3,t} + \theta_{4,t})}. \tag{3.B.6}$$

Proof 2.

The realized return on stock i in Equation (3.B.5) can be rewritten as:

$$\tilde{r}_{i,t+1} = r + b_t\varepsilon_{1,t+1} + b_t\varepsilon_{f,t+1}, \tag{3.B.7}$$

where
$$b_t = \frac{(1+r)}{\frac{(1+r)X_0}{r} + \varphi(\theta_{1,t} + \theta_{2,t} + \theta_{3,t} + \theta_{4,t})},$$

and $\varepsilon_{f,t+1}$ is the shock related to firm-specific information and noise,

$$\varepsilon_{f,t+1} = \varepsilon_{2,t+1} + \varepsilon_{3,t+1} + \varepsilon_{4,t+1}. \tag{3.B.8}$$

The market return is the same as the return of a stock with no idiosyncratic risk:

$$\tilde{r}_{m,t+1} = r + d_t \varepsilon_{1,t+1}, \tag{3.B.9}$$

where
$$d_t = \frac{(1+r)}{\frac{(1+r)X_0}{r} + \varphi\theta_{1,t}}. \tag{3.B.10}$$

From (3.B.9), we have:

$$\varepsilon_{1,t+1} = \frac{r_{m,t+1} - r}{d_t}. \tag{3.B.11}$$

Substitute this expression (3.B.11) for $\varepsilon_{1,t+1}$ into Equation (3.B.7):

$$\tilde{r}_{i,t+1} = r + \frac{b_t}{d_t} r_{m,t+1} - \frac{b_t}{d_t} r + b_t \varepsilon_{f,t+1}. \tag{3.B.12}$$

Conditional on $\theta_{1,t}$, $\theta_{2,t}$, $\theta_{3,t}$ and $\theta_{4,t}$, the stock return variance is therefore a function of shocks to the investors' information about the cash flow process:

$$Var(\tilde{r}_{i,t+1}) = \left(\frac{b_t}{d_t}\right)^2 Var\left(r_{m,t+1}\right) + Var(b_t \varepsilon_{f,t+1}) \tag{3.B.13}$$

Substituting (3.B.9) into (3.B.13), and rearranging, we get:

$$\begin{aligned} Var(\tilde{r}_{i,t+1}) = b_t^2 \big[& Var\left(\varepsilon_{1,t+1}\right) + Var(\varepsilon_{2,t+1}) + Var(\varepsilon_{3,t+1}) \\ & + Var(\varepsilon_{4,t+1}) \big] \end{aligned} \tag{3.B.14}$$

From Equation (3.B.12), the proportion of variance explained by the market, R^2, can be written as:

$$R^2 = \frac{\left(\frac{b_t}{d_t}\right)^2 Var(r_{m,t+1})}{Var(\tilde{r}_{f,t+1})}$$

$$= \frac{Var(\varepsilon_{1,t+1})}{Var(\varepsilon_{1,t+1}) + Var(\varepsilon_{2,t+1}) + Var(\varepsilon_{3,t+1}) + Var(\varepsilon_{4,t+1})} \qquad (3.B.15)$$

$$= \frac{1}{1 + \frac{1}{\eta_1}(\eta_2 + \eta_3 + \eta_4)}.$$

Appendix 3.C. Effect of including the covariance between noise and information

In computing the variance shares in Equation (2.9), we ignore the covariance between information (innovations in the efficient price) and noise (changes in the pricing error). Here we show that accounting for this covariance has little effect on our estimates of the variance shares.

One way to account for the covariance term is to distribute it between the information components of variance and the noise component of variance in the same proportions as the variances of these components and then recompute the variance shares from the covariance-adjusted components using the total return variance as the normalizing variable. In this approach, we allocate a fraction α of $2cov(w_t, \Delta s_t)$ to the information variance and a fraction $(1-\alpha)$ to the noise variance, where $\alpha = \frac{\sigma_w^2}{\sigma_w^2+\sigma_s^2}$. Consequently the information and noise shares of variance become:

$$InfoShare = \left(\sigma_w^2 + \frac{\sigma_w^2}{\sigma_w^2+\sigma_s^2} 2cov(w_t, \Delta s_t)\right) / \sigma_r^2$$

$$= \sigma_w^2 \left(1 + \frac{2cov(w_t, \Delta s_t)}{\sigma_w^2+\sigma_s^2}\right) / \sigma_r^2$$

$$= \sigma_w^2 \left(\frac{\sigma_w^2 + 2cov(w_t, \Delta s_t) + \sigma_s^2}{\sigma_w^2+\sigma_s^2}\right) / \sigma_r^2$$

$$= \sigma_w^2 \left(\frac{\sigma_r^2}{\sigma_w^2+\sigma_s^2}\right) / \sigma_r^2 = \sigma_w^2 / (\sigma_w^2 + \sigma_s^2) \tag{3.C.1}$$

$$NoiseShare = \left(\sigma_s^2 + \frac{\sigma_s^2}{\sigma_w^2+\sigma_s^2} 2cov(w_t, \Delta s_t)\right) / \sigma_r^2$$

$$= \sigma_s^2 \left(1 + \frac{2cov(w_t, \Delta s_t)}{\sigma_w^2+\sigma_s^2}\right) / \sigma_r^2 = \sigma_s^2 \left(\frac{\sigma_r^2}{\sigma_w^2+\sigma_s^2}\right) / \sigma_r^2$$

$$= \sigma_s^2 / (\sigma_w^2 + \sigma_s^2)$$

Equation (3.C.1) shows that after the distribution of the covariance term into information and noise components, we have exactly the same variance shares as in the baseline (Equation (2.9)).

An alternative way of distributing the covariance term is to add it entirely to either the information component of variance or to the noise component of variance thereby producing upper and lower bounds on the variance shares. Applying this approach we find that the upper and lower bounds are extremely narrow (e.g., the noise share has a lower bound of 25.73% and an upper bound of 27.86%, whereas the information share has a lower bound of 72.14% and an upper bound of 74.27%). Therefore, ignoring the covariance term in the baseline variance shares has little effect on the results.

Chapter 4: Noisy stock prices and capital allocation efficiency

4.1 Introduction

Financial markets affect the allocation of capital in several ways. For example, informative stock prices facilitate corporate investment, as investors are more willing to provide capital to firms in primary markets when there is low information asymmetry between managers and outsiders (Stiglitz and Weiss, 1981). Stock prices also affect corporate investment decisions by conveying new information to corporate managers. Recent studies show that firm-specific information incorporated into stock prices allows managers to improve their investments by learning from stock prices (Chen et al., 2007; Foucault and Frésard, 2012, 2014; Edman et al., 2017).

In this chapter, we examine whether, for a given amount of information in prices, the amount of noise in prices affects the efficiency of capital allocation. That is, whether the amount of noise in prices has an *incremental* effect on how well markets facilitate capital allocation beyond the effects of information in prices.

Noise refers to temporary deviations of prices from their equilibrium levels. These price deviations arise from microstructure frictions such as bid-ask spreads and discrete price increments, nonsynchronous trading and stale prices, and temporary price impacts or “price pressure” of uninformed trading. Noise has a significant effect on returns at daily and monthly frequencies, not just intraday horizons. For example, Blume and Stambaugh (1988) and Asparouhova, Bessembinder, and Kalcheva (2013) show that noise at daily frequencies causes an economically meaningful bias in returns, equal to 50% or more of the corrected estimate and is able to explain the size effect. Jegadeesh (1990), Lehmann (1990), and Avramov, Chordia, and Goyal (2006) document significant reversals in stock returns at weekly and monthly horizons, which reflect temporary departures from efficient prices due to imperfect liquidity. Hendershott et al. (2011) estimate that one-quarter of monthly return variance in NYSE stocks is due to transitory price changes explained by order imbalances and market-makers’ inventories. Similarly, Hendershott and Menkveld (2015) find that the ratio of “price pressure” (distortions of the midquote price) to the variance of efficient prices is 33% in NYSE stocks during 1994-2005.

While it is natural to expect the amount of noise in prices to be related to the amount of information reflected in prices, noise and information are not simply the inverse of one another: more information in prices does not necessarily imply less noise and vice versa. For example, consider the noise in prices that is caused by the “price pressure” (temporary distortions) of uninformed traders, also known as “noise traders”. An increase in noise trading could increase the amount of noise in prices. Yet, additional noise traders provide increased incentives for costly information production and acquisition by informed traders, which can increase the amount of information in prices (e.g., Grossman and Stiglitz, 1980; Kyle, 1989). As another example, consider arrivals of important new public information that make prices more informative, for example, company earnings announcements. A greater flow of such information, for example more frequent or more detailed company disclosures, can increase the amount of information in prices. Yet, the releases of such information can increase the noise in prices as adverse selection risks increase around the announcements, liquidity deteriorates, and thus order imbalances have larger price impacts. While the previous examples illustrate how information and noise can be positively related, they could also be negatively related. For example, more liquid stocks are likely to have more information and less noise in their prices than less liquid stocks. What these examples illustrate is that noise and information in prices are two related but distinct features of markets that do not necessarily just mirror one another. They can vary independently of one another and thus can have different (independent) effects on the allocation of capital and investment decisions.

Our analyses focus on two major hypotheses. Our first hypothesis is that at the firm level, more noise in stock price movements makes prices less useful to managers, as it is more difficult to extract the signal from the noise. Noise therefore decreases the extent to which managers rely on prices when making investment decisions. Under this hypothesis, noise is expected to decrease the sensitivity of corporate investment to company growth opportunities. To test this hypothesis, we measure the investment-to-Tobin’s-Q sensitivity, which has become widely used as measure of investment efficiency (e.g., Chen et al., 2011; Foucault and Frésard, 2012; McLean et al., 2012; Edmans et al., 2017). To measure noise and information in prices, we apply a new approach developed by Brogaard et al. (2019), which involves a return variance decomposition based on a vector auto-regression (VAR). The variance decomposition separates noise from various sources of information in stock price movements such as market-wide information, private firm-specific information incorporated into prices through trading, and public firm-specific information not associated with trading.

We test the relation between noise in stock prices and investment efficiency in two different settings. The first is a long time-series of US stock data (1963-2015) in which we can exploit the fact that market quality, and thus the amount of noise in stock prices, varies through time along with changes in market structure. The second involves exploiting cross-country variation using a global sample of 42 countries.

The US time-series evidence shows that consistent with our hypothesis, corporate investment is positively related to Tobin's Q, and that this relation is weaker in firms with noisier prices. An increase in noise by one standard deviation is associated with a reduction firm-level investment efficiency by 4.30% from its mean, holding other variables constant and controlling for the effects of information in prices. By controlling for different types of information, our tests capture the *incremental* effect of noise in stock prices. We also find that, consistent with Chen et al. (2007), US firms with more private firm-specific information in prices have higher investment-Q sensitivity, suggesting that private information of outside investors increases the amount of information in prices that is new to managers, and thus the extent to which managers rely on prices when making corporate decisions.

Our results allow a direct comparison of the effects of noise versus the effects of information on investment efficiency. A one standard deviation increase in private information increases the investment-Q sensitivity by 3.39%, whereas the effect of noise in the same model is 4.30%. Thus, the effect of noise on investment efficiency is economically meaningful: it is at least as strong as the effect of information in prices.

Our second approach to test this hypothesis uses firms in a cross-section of 42 countries during the period from 2003 to 2015. We find consistent evidence. Namely, firm-level investment efficiency is higher in countries with less noisy stock prices, even after controlling for the information in prices. Across countries, a one standard deviation increase in noise is associated with a 6.14% decrease in investment-Q sensitivity from its mean. The finding that there is a positive association between the investment efficiency and information in prices (e.g., Chen et al., 2007) also holds in the cross-section of countries.

Our second hypothesis concerns the impact of noise on the efficiency with which capital is allocated across industries. Overall, we expect capital to flow more to the industries with the best growth opportunities: increasing the capital supply to growing industries and limiting investment in declining industries is a key feature of efficient capital allocation. If noise in stock prices impedes this process by reducing the usefulness of market-based price signals of investment opportunities then we should see that industry-level capital allocation efficiency is worse in countries with noisier stock prices, all else equal.

We follow Wurgler (2000) in using the industry-level investment-to-value-added elasticity as a proxy for industry-level capital allocation efficiency. We restrain our sample to 32 countries used by Wurgler (2000) to facilitate comparisons. Consistent with Wurgler (2000), developed countries tend to have a higher level of investment-to-value-added elasticity. Norway ranks first, while Columbia and Philippines are the two countries with the lowest elasticity estimates. Similar to the results in the firm-level analyses, we find a significant and negative relation between industry-level capital allocation efficiency and the amount of noise in stock prices. A one standard deviation increase in noise is associated with a 15.06% decrease in the investment-to-value-added elasticity from its mean. Information in prices, particularly firm-specific information, helps to improve the allocation of capital across industries. A one standard deviation increase in the amount of firm-specific information in prices increases the investment-to-value-added elasticity by 53%.

We perform a number of tests to validate the robustness of our findings. For the US market, we test whether the results are robust to substituting Tobin's Q with Total Q.[30] We also control for cash flows, given their impact on investment. We find that the results still hold: the association between the investment-Q sensitivity and noise remains negative and significant in all specifications. We also modify the cross-country research design, by adding (removing) country/year fixed effects, and employing country-level regressions. The results are robust to these changes.

This chapter proceeds as follows. Section 4.2 presents related literature. Section 4.3 presents the data and key measures. This is followed by an examination of the

[30] Total Q is a new Q proxy that accounts for intangible assets, proposed by Peter and Taylor (2017).

impact of noise on firm-level investment efficiency (Section 4.4) and industry-level capital allocation efficiency (Section 4.5). Section 4.6 presents robustness tests and Section 4.7 concludes.

4.2 Related literature

This study is related to three main bodies of literature. The first is the literature on measures of informational efficiency. The most common measure of informativeness is stock price non-synchronicity (Roll, 1988), which is based on the correlation between stock return and market return (a high correlation implies low informativeness). A second popular measure is PIN (probability of informed trading), suggested by Easley, Kiefer, and O'Hara (1996, 1997 a,b). Using information from trading process, PIN directly captures the probability of informed trading in a stock. Recently, Brogaard et al. (2019) proposed a novel approach that allows the separation of noise from information in stock return variations. Our study is the first to apply these new measures to investigate the relation between informational efficiency and capital allocation efficiency.

Second, our study contributes to the literature on firm-level capital allocation efficiency, particularly the literature that managers learn from the market in making investment decisions. Chen et al. (2007) show that investment-Q sensitivity is particularly higher for firms with more information in stock prices, measured by both the price non-synchronicity and the probability of informed trading. Foucault and Frésard (2012) find that investment-Q sensitivity is higher in cross-listed firms, because a cross-listing makes stock prices more informative to managers. Foucault and Frésard (2014) show that firms learn from peer stock prices, particularly when managers were previously uninformed, and thus peer stock prices are more likely to contain new information. Luo (2005), Bakke and Whited (2010), and Edmans et al. (2012) also provide evidence of managerial learning from prices. Our contribution is to highlight that noise in prices also affects investment efficiency, beyond the effects of information in prices.

Finally, our study is related to the small number of studies on capital allocation efficiency across industries. Morck et al. (2000) show that banking systems controlled by tycoons or family are associated with poorer allocation of capital. Wurgler (2000) provides evidence that capital is allocated more efficiently in countries with a more developed financial sector. Lee et al. (2018) document that capital flows more to high Tobin-Q industries during the 1971-1996 period, but not during the period from 1997 to 2014 due to a decrease in the equity funding-Q sensitivity resulting mostly from the increased equity repurchases-Q sensitivity after 1996. We contribute to this literature by showing that industry-level capital allocation efficiency is worse in countries with noisier stock prices.

4.3 Sample, data, and key measures

4.3.1 Country sample selection

We test the relation between noise in stock prices and the investment-Q sensitivity in two different settings. The first is a long time-series of US stock data (1963-2015) in which we can exploit the fact that the amount of noise in stock prices varies through time along with market structure changes. The second involves exploiting cross-country variation using a global sample of 42 countries, including 20 emerging countries (Argentina, Brazil, Chile, China, Columbia, Greece, Indonesia, India, South Korea, Malaysia, Mexico, Peru, Philippines, Russia, Poland, Russia, Thailand, Turkey, Taiwan, and South Africa) and 22 developed countries (Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Hong Kong, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, United Kingdom, and the United States). Additional details related to the sample selection and multistage screening process are provided in Appendix 4.A.

4.3.2 Data description

We collect data from a variety of sources to construct firm-level variables in the 42 countries during the 13-year period from 2003 to 2015. Specifically, non-US market data including international individual stock total return index, price, and volume are obtained from Datastream. We collect market value of equity, total assets, capital expenditures, sales, cash flows, and market capitalization data from Worldscope.

Our US stock sample comprises all common stocks (share codes 10 and 11) listed in major US stock exchanges (e.g., NYSE, AMEX, NASDAQ) during the period 1963 to 2015. From CRSP (Center for Research in Security Prices), we obtain daily return, price, and trading volume. Corresponding accounting and financial data are sourced from Compustat.

We collect the country-level and industry-level data from multiple additional sources. We obtain ISIC Revision 3 industry-year data on gross fixed capital formation and value added from the United Nations' INDSTAT-2 database. We obtain the US capital goods producer price index[31] and the US finished goods producer price index[32] (based on year 1982 for both) from Fred Economic Data.

4.3.3 Measures of noise and information in prices

We employ new measures of information and noise developed by Brogaard et al. (2019). Similar to Hasbrouck (1993), Brogaard et al. (2019) draw on the market microstructure toolkit that separates temporary price movements driven by frictions and uninformed trading from permanent movements driven by information. This permanent-temporary decomposition of returns involves estimating a vector autoregressive (VAR) model of stock returns, market returns, and stock trading volume for each stock in each year using daily data. The details of this procedure are described in Appendix 4.B. This approach allows a partition of the overall information reflected in prices into various sources: market-wide information, firm-specific information revealed through trading on private information, and firm-specific information revealed through public news.

4.4 The impact of noise on firm-level investment efficiency

This chapter investigates the relation between noise in stock prices and two different, but related measures of capital allocation efficiency: firm-level investment-Q sensitivity (this section) and industry-level investment-to-value-added elasticity (Section 4.5). These measures capture the sensitivity of investment to growth opportunities or to changes in the marginal product of capital. Higher sensitivity reflects a stronger tendency for capital to flow to firms or industries consistent with a more efficient allocation of capital.

[31] Producer Price Index by Commodity for Finished Goods: Capital Equipment.
[32] Producer Price Index by Commodity for Final Demand: Finished Goods.

4.4.1 Measuring firm-level investment efficiency

One popular theory is that efficient secondary market prices help investors and managers distinguish between good and bad investments because future growth opportunities are reflected in company valuations. Tobin (1969) argues that if the market value of assets is larger than its replacement value, a firm should increase its investment. Therefore, corporate investment should be sensitive to Tobin's Q (in essence, the market-to-book ratio), with higher sensitivity indicating more efficient deployment of capital to its best uses.

We estimate the investment-Q sensitivity of each country c from the following regression, similar to Chen et al. (2007) and Foucault and Frésard (2012):

$$CAPXRND_{i,c,t} = \alpha_t + \beta_{1,c} Q_{i,c,t-1} + \beta_{2,c} inv\_assets_{i,c,t-1} + \varepsilon_{i,c,t} \tag{4.1}$$

where i indexes stocks, c indexes countries, and t indexes years, respectively. The dependent variable $CAPXRND_{i,t}$ is a measure of corporate investment in year t, which is the sum of capital expenditure (CAPEX) and R&D expenses as a percentage of the beginning-of-year book value of assets ($Assets_{i,t-1}$). $Q_{i,t-1}$ is Tobin's Q: the market value of equity plus book value of assets minus the book value of equity, scaled by the book value of assets. We control for the inverse of the book value of the company's assets ($inv\_assets_{i,c,t-1}$), which is $1{,}000/Assets_{i,c,t-1}$. Higher $\beta_{1,c}$ implies higher corporate investment efficiency and therefore more efficient allocation of capital across firms.

We estimate the investment-Q sensitivity for the 42 countries during a period of 13 years from 2003 to 2015. For the US market only, due to better data availability, we are able to extend our sample back to 1963. We therefore conduct our tests separately using the long time-series of US data and then the shorter period of cross-country data.

Table 4.1. **Summary statistics for the US sample.**
This table reports the summary statistics for the US sample. The units of observations are firm-years. $CAPXRND$ is firm-level investment (annual CAPEX and R&D expenses as a percentage of assets), Q is Tobin's Q, CF is the annual cash flow from operations in million USD, and inv_assets is the inverse of the book value of assets. $Noise$, $MktInfo$, $PrivateInfo$, and $PublicInfo$ are measures of the amount of noise, market-wide information, private firm-specific information, and public firm-specific information in the stock return variance, respectively. Detailed variable definitions are in Appendix 4.C. The sample includes stocks listed on NYSE, AMEX, and NASDAQ and spans the period from 1963 to 2015.

Variable	Mean	Median	Std. Dev.	25th %	75th %
$CAPXRND$	11.542	8.524	9.879	4.438	15.327
Q	1.649	1.285	1.052	1.001	1.883
$Noise$	0.283	0.123	0.448	0.055	0.290
$MktInfo$	0.177	0.087	0.227	0.028	0.223
$PrivateInfo$	0.385	0.213	0.467	0.092	0.474
$PublicInfo$	0.504	0.259	0.670	0.118	0.572
CF	1.066	0.662	1.146	0.337	1.313
inv_assets	0.117	0.111	0.544	0.063	0.174

Tables 4.1 and 4.2 summarize the basic statistics for the key variables for each of our two settings. Table 4.1 presents the mean, median, and standard deviation of the variables in the US data during 1963 to 2015. $Noise$, $MktInfo$, $PrivateInfo$, and $PublicInfo$ are daily return variances attributable to noise, market-wide information, private information, and public information, respectively. We multiply these variables by 1,000 for readability. On average, the total amount of firm-specific information in stock return variance is 0.89 while the corresponding figures for market-wide information and noise are 0.18 and 0.28, respectively. These numbers are in units of daily return variance multiplied by 1,000 so while their absolute magnitudes are difficult to interpret, their relative magnitudes are consistent with those reported in Roll (1988), who documents that stock return variance is largely driven by firm-specific information.

Table 4.2 provides the mean of the investment-Q sensitivity, information, and noise during 2003 to 2015 by country. Panel A includes developed markets and Panel B includes emerging markets. We report the mean of variables for groups of countries and the pooled sample of 42 countries in Panel C. As expected, the investment-Q sensitivity is positive for all countries, suggesting that in general, corporate investment is sensitive to growth opportunities as reflected in company valuations. Investment-Q sensitivity tends to be higher in developed countries suggesting that such countries more efficiently allocate capital across firms.

Table 4.2. Investment-Q sensitivity, information, and noise by country

This table reports the mean of the investment-Q sensitivity, information, and noise by country in Panels A and B. The firms' investment-Q sensitivity ($\beta_{1,c}$) in each country is obtained from the following regression:

$$CAPXRND_{i,c,t} = \alpha_t + \beta_{1,c} Q_{i,c,t-1} + \beta_{2,c} inv\_assets_{i,c,t-1} + \varepsilon_{i,c,t}$$

where i indexes stocks, c indexes countries, and t indexes years. $CAPXRND$ is firm-level investment (annual CAPEX and R&D expenses as a percentage of assets), Q is Tobin's Q, and inv_assets is the inverse of the book value of assets. $Noise$, $MktInfo$, $PrivateInfo$, and $PublicInfo$ are measures of the amount of noise, market-wide information, private firm-specific information, and public firm-specific information in the stock return variance, respectively. Panel C reports the means of the variables for groups of countries and the pooled sample. Detailed variable definitions are provided in Appendix 4.C. The sample includes stocks listed on major exchanges in 42 countries and spans the period from 2003 to 2015.

Panel A: Developed markets					
Country	Investment-Q Sensitivity	$Noise$	$MktInfo$	$PrivateInfo$	$PublicInfo$
Austria	2.145	0.079	0.201	0.105	0.149
Australia	4.180	0.073	0.130	0.112	0.179
Belgium	4.937	0.056	0.114	0.073	0.093
Canada	4.590	0.071	0.118	0.118	0.173
Switzerland	1.592	0.047	0.111	0.068	0.074
Germany	2.099	0.065	0.145	0.101	0.119
Denmark	4.173	0.060	0.140	0.106	0.119
Spain	1.106	0.055	0.169	0.069	0.105
Finland	3.787	0.059	0.187	0.098	0.121
France	2.935	0.059	0.130	0.087	0.110
Hong Kong	1.683	0.098	0.230	0.150	0.187
Ireland	2.288	0.183	0.258	0.121	0.328
Italy	2.247	0.064	0.168	0.080	0.109
Japan	1.993	0.078	0.184	0.152	0.138
Netherlands	2.381	0.048	0.138	0.076	0.094
Norway	4.135	0.075	0.194	0.110	0.157
New Zealand	2.508	0.046	0.060	0.061	0.127
Portugal	5.282	0.063	0.153	0.110	0.115
Sweden	2.266	0.059	0.157	0.093	0.120
Singapore	2.893	0.097	0.178	0.110	0.136
United Kingdom	2.687	0.062	0.109	0.100	0.140
United States	3.378	0.078	0.128	0.141	0.141
Panel B: Emerging markets					
Country	Investment-Q Sensitivity	$Noise$	$MktInfo$	$PrivateInfo$	$PublicInfo$
Argentina	2.412	0.151	0.385	0.160	0.253
Brazil	0.884	0.105	0.208	0.122	0.225
Chile	3.328	0.052	0.112	0.040	0.131
China	0.847	0.164	0.410	0.208	0.263
Columbia	3.386	0.073	0.209	0.042	0.108

Table 4.2. (continued)

Country	Investment-Q Sensitivity	$Noise$	$MktInfo$	$PrivateInfo$	$PublicInfo$
Greece	3.401	0.202	0.474	0.132	0.452
Indonesia	1.562	0.166	0.265	0.216	0.287
India	1.812	0.130	0.314	0.164	0.244
South Korea	3.269	0.097	0.190	0.223	0.174
Mexico	1.853	0.050	0.131	0.073	0.127
Malaysia	2.440	0.074	0.109	0.087	0.115
Peru	1.502	0.124	0.227	0.076	0.294
Philippines	1.535	0.127	0.182	0.138	0.241
Pakistan	1.868	0.189	0.313	0.120	0.319
Poland	4.980	0.091	0.200	0.113	0.214
Russia	3.651	0.149	0.372	0.114	0.253
Thailand	2.077	0.090	0.227	0.146	0.155
Turkey	1.609	0.100	0.329	0.151	0.187
Taiwan	4.234	0.071	0.188	0.139	0.143
South Africa	2.026	0.071	0.113	0.101	0.155

Panel C: Averages

Country group	Investment-Q Sensitivity	$Noise$	$MktInfo$	$PrivateInfo$	$PublicInfo$
Pooled sample	2.796	0.091	0.198	0.117	0.174
Developed countries	2.968	0.072	0.155	0.102	0.138
Emerging countries	2.319	0.111	0.244	0.124	0.209

Figure 4.1 graphically highlights the large differences across countries in firm-level capital allocation efficiency and shows that seven out of the top ten countries with the highest $\beta_{1,c}$s are developed ones (Portugal, Belgium, Canada, Australia, Denmark, Norway, and Finland). Among the twelve countries with the lowest $\beta_{1,c}$s, nine of them are emerging countries (China, Brazil, Peru, Philippines, Indonesia, Turkey, India, Mexico, and Pakistan). The US ranks 13th of the 42 countries in investment efficiency. Therefore, while there is a tendency for more developed countries to have higher investment efficiency, the pattern in the cross-section of countries is certainly not monotonic in development suggesting factors other than simply development level affect the level of investment efficiency.

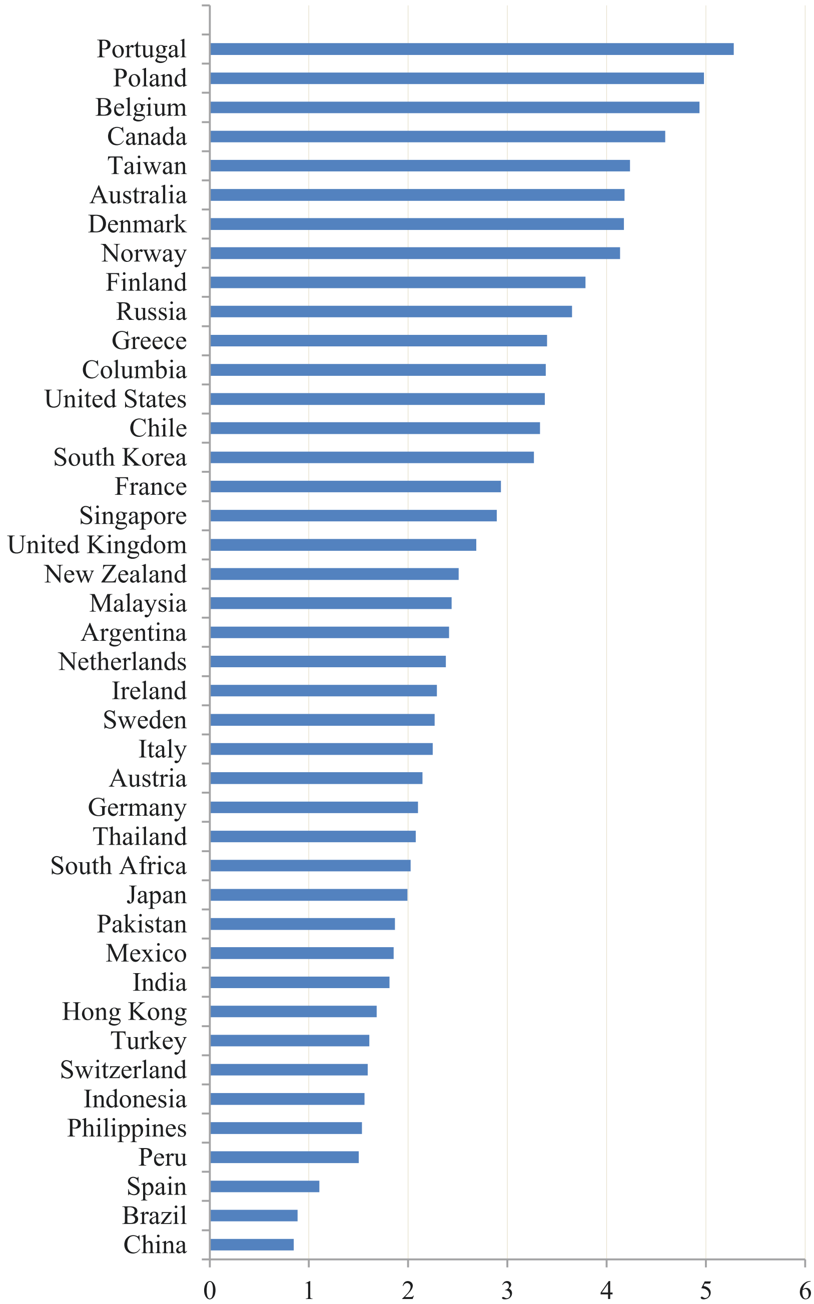


Figure 4.1. Investment-Q sensitivity by country.
This figure reports the estimates of firms' investment-Q sensitivity ($\beta_{1,c}$) in 42 countries. The estimates are obtained from the following regression, estimated for each country c:

$$CAPXRND_{i,c,t} = \alpha_t + \beta_{1,c}Q_{i,c,t-1} + \beta_{2,c}inv\_assets_{i,c,t-1} + \varepsilon_{i,c,t}$$

where i indexes stocks, c indexes countries, and t indexes years. $CAPXRND$ is firm-level investment (annual CAPEX and R&D expenses as a percentage of assets), Q is Tobin's Q, and inv_assets is the inverse of the book value of assets. Detailed variable definitions are provided in Appendix 4.C. The sample period is from 2003 to 2015.

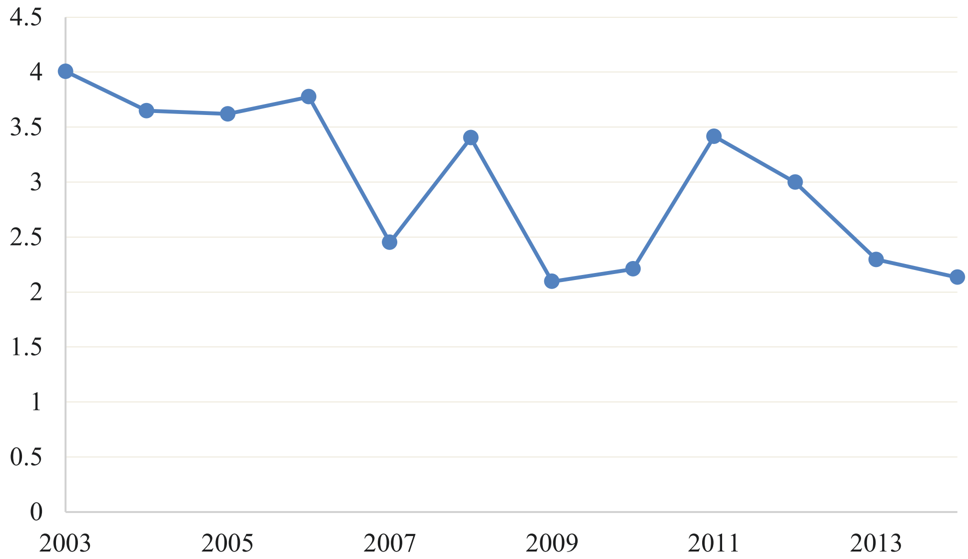


Figure 4.2. Investment-Q sensitivity by year.
This figure reports the estimates of firms' investment-Q sensitivity (β_1) for all firms in the sample. The estimates are obtained from the following regression, estimated for each year t, by using all stocks globally in year t:

$$CAPXRND_{i,c,t} = \alpha_t + \beta_1 Q_{i,c,t-1} + \beta_2 inv\_assets_{i,c,t-1} + \varepsilon_{i,c,t}$$

where i indexes stocks, c indexes countries, and t indexes years. $CAPXRND$ is firm-level investment (annual CAPEX and R&D expenses as a percentage of assets), Q is Tobin's Q, and inv_assets is the inverse of book value of assets. Detailed variable definitions are provided in Appendix 4.C. The sample period is from 2003 to 2015.

Figure 4.2 reports the yearly estimates of firms' investment-Q sensitivity ($\beta_{1,t}$) of all firms in our international sample. There is a downward trend in the investment-Q sensitivity through time. In general, for an average firm, investment is almost two times more sensitive to Tobin's Q in 2003 than in 2015. These findings are consistent with results in Lee, Shin, Stulz (2018), which indicate a decrease in the equity funding-Q sensitivity during the post-1997 period driven by an increase in the sensitivity of equity repurchases to Tobin's Q.

4.4.2 Evidence from a long time-series of US data

We first test the relation between noise and firm-level investment efficiency in the long time-series of US data back to 1963. Our approach is similar in spirit to that of Chen, Goldstein, and Jiang (2007).[33] It involves estimating the following regression:

$$CAPXRND_{i,t} = \alpha_t + \gamma_i + \beta_1 Q_{i,t-1} + \beta_2 [Noise_{i,t-1} \times Q_{i,t-1}] + \delta CONTROLS_{i,t} + \varepsilon_{i,t} \tag{4.2}$$

where subscripts i and t represent the firm and the year, respectively. $CAPXRND_{i,t}$ is a measure of corporate investment and $Q_{i,t-1}$ is Tobin's Q. $Noise_{i,t-1}$ is the contribution of noise to stock return variance. We control for different types of information impounded in prices and their interaction with Q: $PrivateInfo$, $PublicInfo$, $MktInfo$, which are the contributions to stock return variance of trading on private firm-specific information, public firm-specific information, and market-wide information. $Info$ is the sum of $MktInfo$, $PrivateInfo$, and $PublicInfo$; whereas $FirmInfo$ is the sum of the two latter types of information. We also control for the inverse of the company's book value of assets ($inv\_assets_{i,t-1}$). We include year and firm fixed effects.

We expect $\beta_1 > 0$, that is $CAPXRND_{i,t}$ is positively related with $Q_{i,t-1}$, indicating that investment is sensitive to growth opportunities as observed in the literature. However, our variable of interest is the interaction term between $Q_{i,t-1}$ and $Noise_{i,t-1}$. Our hypothesis is that more noise in stock prices makes prices less useful to managers, and thus decreases the extent to which managers rely on stock prices when making investment decisions. Under this hypothesis, the coefficient β_2, which measures the effect of noise in stock prices on the sensitivity of investment to Q, is expected to be negative and significant. We also expect the coefficients for the interaction terms between $Q_{i,t-1}$ and the information measures to be positive, that is,

[33] Chen, Goldstein, and Jiang (2007) investigate the relation between firm-specific information in prices and investment sensitivity to Q. They use firm-specific stock return variation as suggested by Roll (1988) and the PIN measure developed by Easley, Kiefer, and O'Hara (1997) as measures of private firm-specific information in stock prices, and show that the investment-Q sensitivity of firms increases with these measures.

the relation between corporate investment and Q is stronger in firms with more information in stock prices.

Table 4.3 summarizes the results from estimating Equation (4.2). Model 1 performs a univariate regression with only Tobin's Q as an independent variable. In Model 2, we add our variable of interest, which is Tobin's Q interacted with the measure of noise. Noise is also separately included in this model to control for its direct effect on investment and to make sure that this direct effect does not drive the result on β_2. We control for market-wide information, private firm-specific information, public firm-specific information, total firm-specific information, and their corresponding interaction terms with Tobin's Q in Models 3-8.

The results show that future investment and Tobin's Q are positively related in all nine models. This finding supports the observation in the literature that investment is positively related to firms' growth opportunities. From Model 1, given that the coefficient of Q is 2.04 and the standard deviation of Q is 1.05, a one standard deviation increase in Q yields a 2.14 (=2.04×1.05) increase in investment. The mean value of investment is 11.54, so a one standard deviation increase in Q is associated with a 18.54% increase in investment from its mean, or an increase of 21.62% (=2.14/9.90 ×100%) of a standard deviation of investment.

Model 2 shows that our variable of interest, the interaction between noise and Tobin's Q, has negative and significant coefficients, indicating that the investment-Q sensitivity is lower for firms whose stock prices contain more noise. The coefficient on Q is 2.15, the coefficient on $Noise \times Q$ is -0.24, and the mean of $Noise$ is 0.28, which together imply that investment-Q sensitivity evaluated at the mean level of $Noise$ is 2.08 ($=2.15-0.24\times0.28$). Holding all other variables constant, increasing noise by one standard deviation (i.e., 0.448) decreases investment-Q sensitivity by 5.17% ($= -0.24\times0.448/2.08$). This result remains significant even after controlling for the effect of various sources of information. This finding indicates that the effects of noise on investment efficiency are *incremental* to the effects of information. Consistent with Chen, Goldstein, and Jiang (2007), investment efficiency is increasing with our measure of private information in prices, consistent with the notion that the private information reflected in prices is useful to managers and thus increases the extent to which mangers rely on prices when they make investment decisions

Table 4.3. US evidence on the relation between noise and firm-level investment efficiency.
This table reports the effect of noise in stock prices on investment-Q sensitivity for the US market. The units of observations in the regression below are firm-years and the dependent variable is firm-level investment (annual CAPEX and R&D expenses as a percentage of assets). Model 1 reports the results from the baseline model, in which firm investment is regressed on Tobin's Q, controlling for the inverse of the book value of assets (inv_assets). Models 2 adds an interaction term between Q and $Noise$, which is a measure of the amount of noise in stock prices. Models 3-9 add interactions between Q and different types of information impounded into stock prices. $PrivateInfo$, $PublicInfo$, and $MktInfo$ are measures of the amount of private firm-specific information, public firm-specific information, and market-wide information in stock prices, respectively. $Info$ is the sum of $MktInfo$, $PrivateInfo$, and $PublicInfo$ whereas $FirmInfo$ is the sum of the two latter types of information. We include year and firm fixed effects in all models. Standard errors are clustered by firm. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes stocks listed on NYSE, AMEX, and NASDAQ and spans the period from 1963 to 2015.

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Q	2.04	2.15	2.18	2.10	2.18	2.15	2.16	2.14	2.17
	(43.80)***	(37.89)***	(35.12)***	(32.42)***	(34.57)***	(32.11)***	(31.78)***	(31.33)***	(31.81)***
$Noise \times Q$		-0.24	-0.17	-0.29	-0.18	-0.23	-0.21	-0.20	-0.20
		(-3.97)***	(-2.50)**	(-4.38)***	(-2.42)**	(-3.12)***	(-2.79)***	(-2.60)***	(-2.65)***
$MktInfo \times Q$			-0.26					-0.29	-0.30
			(-2.36)**					(-2.46)**	(-2.56)**
$PrivateInfo \times Q$				0.07				0.16	
				(1.38)				(2.77)***	
$PublicInfo \times Q$					-0.06			-0.06	
					(-1.49)			(-1.37)	
$FirmInfo \times Q$						-0.01			0.03
						(-0.27)			(0.81)
$Info \times Q$							-0.02		
							(-0.74)		
$Noise$		-0.41	-0.69	-0.45	-0.61	-0.58	-0.69	-0.74	-0.73
		(-3.67)***	(-5.55)***	(-3.78)***	(-4.71)***	(-4.48)***	(-5.18)***	(-5.56)***	(-5.45)***
$MktInfo$			1.74					1.73	1.73
			(7.78)***					(7.32)***	(7.34)***
$PrivateInfo$				0.20				-0.07	
				(1.81)*				(-0.57)	

$PublicInfo$					0.28 (3.25)***			0.14 (1.50)	

Table 4.3. (continued)

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$FirmInfo$						0.19 (3.18)***			0.06 (0.97)
$Info$							0.26 (4.80)***		
inv_assets	0.13 (27.60)***	0.14 (28.35)***	0.14 (28.39)***	0.14 (28.38)***	0.14 (28.34)***	0.14 (28.34)***	0.14 (28.34)***	0.14 (28.46)***	0.14 (28.39)***
R^2 (%)	14.89	15.08	15.21	15.12	15.10	15.11	15.14	15.24	15.22
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

To gauge the economic significance of the effect of noise, we compare the magnitude of its effects with the magnitude of the effect of information. In Model 8, the coefficients of the $PrivateInfo$-Q and $Noise$-Q interactions are 0.16 and -0.20, and the coefficient on Q is 2.14. Given that the mean value of $PrivateInfo$ is 0.38, investment-Q sensitivity evaluated at the mean level of $PrivateInfo$ is 2.20 (=2.14+0.16×0.38). Holding all other variables constant, increasing information by one standard deviation (i.e., 0.467) increases investment-Q sensitivity by 3.39% from 2.20 to 2.28. The corresponding effect of a one standard deviation increase in noise in this model is a 4.30% decrease in the investment-Q sensitivity from 2.08 to 1.99. The independent effect of noise is therefore at least as strong as the independent effect of private firm-specific information reflected in prices. Thus noise has an economically meaningful effect on investment efficiency.

4.4.3 Evidence from the cross-section of countries

Our second setting in which to examine how noise impacts firm-level investment efficiency is in the cross-section of 42 countries. In our baseline analysis, we use country-year level regressions after first estimating investment-Q sensitivity and noise in each country in each year. The main reason for not using firm-level regressions is that they give equal weight to each firm and therefore larger countries will drive the results.

To conduct our country-year level tests, in the first step, we estimate the investment-Q sensitivity ($\beta_{1,c,t}$) for a given country c in year t using the regression (4.1). To examine the impact of noise on the investment-Q sensitivity, we subsequently regress the estimated investment-Q sensitivity on noise and control variables, including country and year fixed effects:

$$\beta_{1,c,t} = \alpha_t + \gamma_c + \rho Noise_{c,t} + \delta CONTROLS_{i,t} + \varepsilon_{c,t}\ . \qquad (4.3)$$

We control for different types of information impounded into the stock prices in each country (e.g., total amount of information, firm-specific information, market-wide information, private information, and public information). The measures of noise and information in stock price variations for country c in year t are calculated by taking the market capitalization weighted average across stocks in that country within that year.

Our hypothesis is that ρ is negative; that is, countries with less noisy stock prices exhibit higher firm-level investment efficiency. Further, we conjecture that the investment-Q sensitivity is stronger in countries in which stock prices reflect more information.

Consistent with our hypothesis, Table 4.4 reports that investment-Q sensitivity is significantly and negatively related to noise. Taking Model 5 as an example, the coefficient of $Noise$ is -6.90 with a t-statistic of -2.55. Given the standard deviation of $Noise$ is 0.077, the standard deviation of investment-Q sensitivity is 1.769 and the mean investment-Q sensitivity is 2.796, a one standard deviation increase in $Noise$ is associated with a 19% (=$-6.90\times 0.077/2.796$) decrease in investment-Q sensitivity from its mean, or a decrease of 30% (=$-6.90\times 0.077/1.769$) of a standard deviation of the investment-Q sensitivity.

The Chen et al. (2007) result continues to hold in an international context with information being positively associated with investment-Q sensitivity. Firms in countries with a higher amount of information, in particular public information, in stock prices tend to have better investment efficiency. The coefficient of $Info$, which is the total amount of information in the stock prices, is 1.56 with a t-statistic of 1.85. Given the standard deviation of $Info$ is 0.316, a one standard deviation in $Info$ increases investment-Q sensitivity by 18% (=1.56×$0.316/2.796$) from its mean, or by 29% (=1.56×$0.316/1.769$) of its standard deviation.

Table 4.4. Cross-country evidence on the relation between noise and firm-level investment efficiency.
This table reports the results from panel regressions in which the units of observations are country-years and the dependent variable is the estimated investment-Q sensitivity ($\beta_{1,c,t}$). $Noise$ is a measure of the amount of noise in stock return variance. $PrivateInfo$, $PublicInfo$, and $MktInfo$ are measures of the amount of private firm-specific information, public firm-specific information, and market-wide information in stock return variance, respectively. We include year and country fixed effects. Standard errors are clustered by country and year. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes stocks listed on major exchanges in 42 countries and spans the period from 2003 to 2015.

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Noise$	-4.83	-2.29	-6.16	-5.45	-6.90	-7.39	-6.92
	(-2.05)**	(-2.47)**	(-3.46)***	(-2.88)***	(-2.55)**	(-3.44)***	(-2.78)***
$MktInfo$	1.85					1.45	1.52
	(1.24)					(0.88)	(0.92)
$PrivateInfo$		0.15				-0.66	
		(0.05)				(-0.23)	
$PublicInfo$			3.42			2.97	
			(2.60)***			(2.04)**	
$FirmInfo$				2.02			1.60
				(1.46)			(1.05)
$Info$					1.56		
					(1.85)*		
R^2 (%)	45.99	45.54	46.21	45.94	46.23	46.47	46.23
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

4.5 The impact of noise on industry-level capital allocation efficiency

In this section, we investigate whether increased informational efficiency improves the allocation of capital at the broader level across industries. We measure the efficiency of capital allocation across industries using the elasticity of industry investment to industry value added, similar to Wurgler (2000).

4.5.1 Measuring industry-level capital allocation efficiency

Here, we adopt the Wurgler (2000) framework to estimate industry-level capital allocation efficiency. Wurgler's (2000) approach is based on the notion that efficient capital allocation implies increasing investment in faster growing industries and decreasing it in declining or slower growing industries. He uses growth in industry value added and growth in industry gross fixed capital formation as proxies for industry growth and industry investment growth, respectively. We obtain estimates of investment-to-value-added elasticity for each country c from the following regression:

$$ln\left(\frac{I_{i,c,t}}{I_{i,c,t-1}}\right) = \alpha_c + \varphi_c ln\left(\frac{V_{i,c,t}}{V_{i,c,t-1}}\right) + \varepsilon_{i,c,t} \quad 4.4)$$

where $ln\left(\frac{I_{i,c,t}}{I_{i,c,t-1}}\right)$ is the log growth in industry investment; $ln\left(\frac{V_{i,c,t}}{V_{i,c,t-1}}\right)$ is the log growth in industry value added; $I_{i,c,t}$ is gross fixed capital formation; and $V_{i,c,t}$ is value added. i indexes industries, c indexes countries, and t indexes years. The elasticity of industry investment to value added (φ_c) measures the extent to which country c increases investment in its growing industries and decreases investment in its declining industries. The greater the value of φ_c, the higher the efficiency of capital allocation across industries.

We use the United Nations' General Industrial Statistics panel (INDSTAT-2) as a source of basic manufacturing statistics.[34] The data include gross fixed capital formation and value added for up to 23 two-digit SIC industries. Value added is the value of shipments of goods produced minus the cost of intermediate goods and required services, with adjustments for inventories of finished goods, work-in-progress, and raw materials. Gross fixed capital is the cost of new and used fixed assets minus the value of sales of used fixed assets, where fixed assets include land, buildings, machinery, and equipment. Both value added and gross fixed capital are in current US dollars. We compute real gross capital formation in industry i in country c in year t ($I_{i,c,t}$) by deflating the nominal series by the US capital goods producer price index (base year 1982). Real value added ($V_{i,c,t}$) is computed by deflating the nominal series by the US finished goods producer price index (base year 1982).

To enhance comparability with prior research and overcome data limitations, we use data from the same set of 32 countries examined by Wurgler (2000). Panels A and B in Table 4.5 report the summary statistics of the log industry investment growth, the log industry value added growth, and the estimates of φ_{c} in each of countries in the sample. Panel C provides the means for groups of countries and the pooled sample.

Across countries, average log industry growth is 1.2% and log value added growth is 0.9% on average. Industries in emerging countries tend to have higher investment growth and value added growth than those in developed countries over the reported period. The average investment-to-value-added elasticity is 0.497 in the pooled sample. The elasticity of industry investment to value added is greater in developed countries than that in emerging countries, consistent with Wurgler (2000).

[34] Wurgler (2000) uses INDSTAT-3 (3-digit level of ISIC Revision 2). However, the INDSTAT-3 database was discontinued and replaced by INDSTAT-2, which is a 2-digit level database of ISIC revision 3. All the data from INDSTAT-3 was converted from ISIC Revision 2 to ISIC Revision 3. Therefore, INDSTAT 2 contains a long time series starting in 1963, until the most recent year, 2016. We replicate Wurgler (2000) using INDSTAT-3 instead of INDSTAT-2 and find similar results.

Table 4.5. Industry investment growth, value added growth, and investment-to-value-added elasticity.

This table reports country averages of log industry investment growth ($ln(I_{i,c,t}/I_{i,c,t-1})$), log industry value added growth ($ln(V_{i,c,t}/V_{i,c,t-1})$), and the elasticity of industry investment to value added (φ_c). The estimates of φ_c are obtained from the following regression:

$$ln(I_{i,c,t}/I_{i,c,t-1}) = \alpha_c + \varphi_c ln(V_{i,c,t}/V_{i,c,t-1}) + \varepsilon_{i,c,t}.$$

where $I_{i,c,t}$ is gross fixed capital formation and $V_{i,c,t}$ is value added. i indexes industries, c indexes countries, and t indexes years. Panel C reports the means of the variables for groups of countries and the pooled sample

Panel A: Developed markets

Country	Time period	$ln(I_{i,c,t}/I_{i,c,t-1})$	$ln(V_{i,c,t}/V_{i,c,t-1})$	φ_c
Australia	2003-2015	-0.013	0.020	0.537
Austria	2003-2015	0.007	0.009	0.385
Belgium	2004-2015	0.013	-0.016	0.620
Denmark	2003-2015	-0.025	0.001	0.702
Finland	2003-2015	-0.004	-0.001	0.399
France	2003-2015	0.017	-0.006	0.626
Germany	2003-2015	0.020	0.002	0.453
Hong Kong	2003-2015	0.007	-0.080	0.513
Ireland	2003-2015	0.054	0.034	0.335
Italy	2003-2015	-0.019	-0.014	0.412
Japan	2003-2014	0.004	-0.030	0.413
Netherlands	2003-2015	-0.005	-0.004	0.580
New Zealand	2003-2012	0.032	0.032	0.793
Norway	2003-2015	-0.009	0.016	0.933
Portugal	2003-2015	0.005	-0.002	0.876
Singapore	2003-2015	-0.017	0.021	0.435
Spain	2003-2005	-0.017	-0.017	0.723
Sweden	2003-2015	-0.002	0.018	0.472
Switzerland	2010-2015	0.015	0.007	0.669
United Kingdom	2003-2005	-0.006	-0.015	0.676
United States	2003-2015	-0.013	-0.012	0.507

Panel B: Emerging markets

Country	Time period	$ln(I_{i,c,t}/I_{i,c,t-1})$	$ln(V_{i,c,t}/V_{i,c,t-1})$	φ_c
Columbia	2003-2005	-0.054	0.087	0.022
Chile	2003-2012	0.117	-0.038	0.200
Greece	2004-2015	-0.056	-0.047	0.566
India	2003-2014	0.100	0.088	0.335
Indonesia	2003-2013	0.037	0.072	0.193
Malaysia	2003-2015	0.043	0.028	0.294
Mexico	2003-2015	-0.004	0.024	0.420
Poland	2003-2015	0.051	0.026	0.609
Philippines	2006-2014	-0.024	0.038	-0.141
South Korea	2003-2015	0.063	0.033	0.897
Turkey	2003-2015	0.070	0.058	0.461

Table 4.5. (continued)

Panel C: Averages

Country group	Time period	$ln(I_{i,c,t}/I_{i,c,t-1})$	$ln(V_{i,c,t}/V_{i,c,t-1})$	φ_c
Pooled sample	2003-2015	0.012	0.009	0.497
Developed countries	2003-2015	0.0004	-0.001	0.559
Emerging countries	2003-2015	0.036	0.028	0.394

Figure 4.3 graphically highlights the differences across countries in their investment-to-value-added elasticity. Similar to the measures of firm-level investment efficiency, the efficiency of capital allocation across industries is not entirely monotonic in development, suggesting other factors contribute to the level of efficiency in a given country. Norway ranks first at 0.933, followed by South Korea and Portugal. The United Kingdom is the 7th highest, and the United States is ranked 16th. Columbia and the Philippines have the lowest elasticity estimates implying a low level of efficiency.

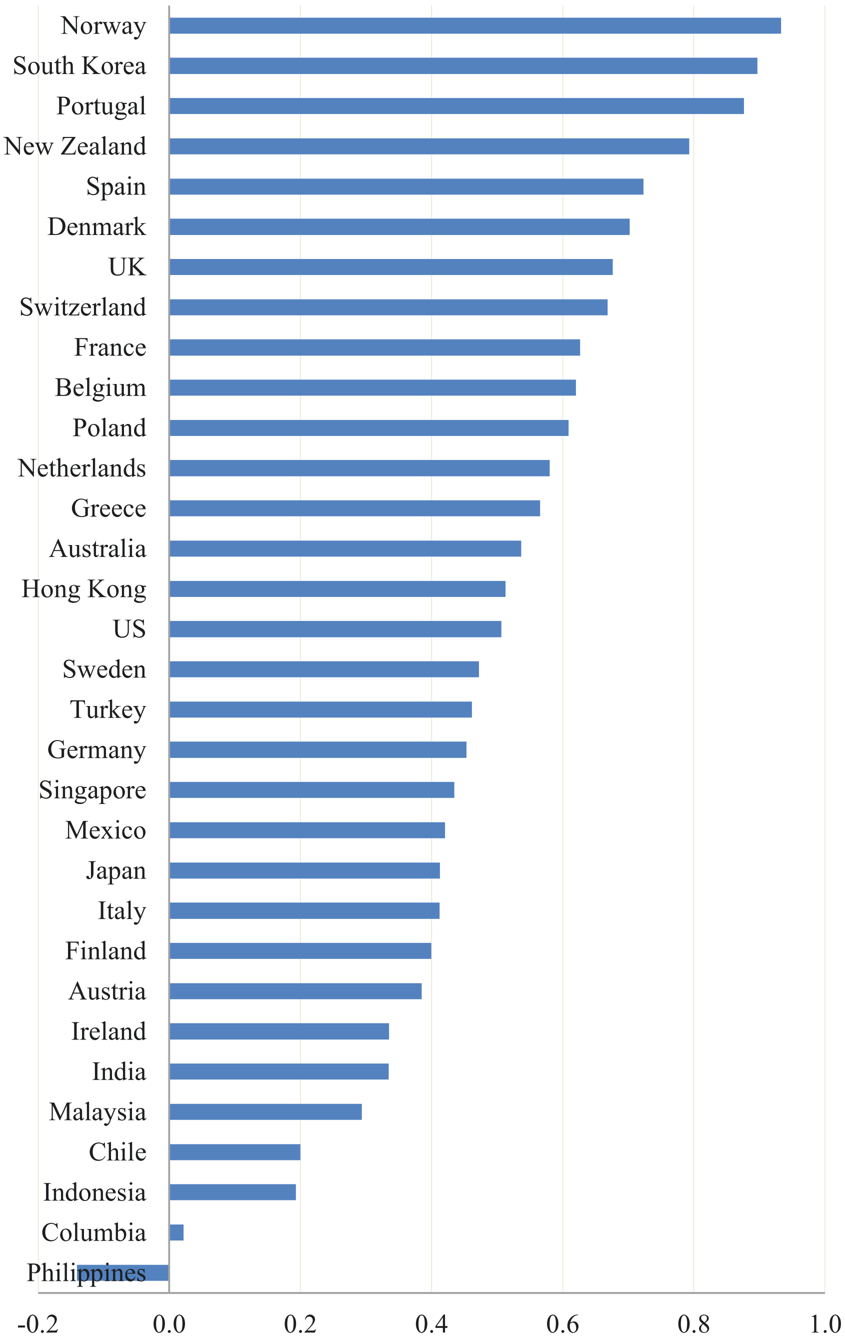


Figure 4.3. Elasticity of industry investment to value added by country.
This figure reports the estimates of the elasticity of industry investment to value added (φ_c) in 32 countries. The estimates are obtained from the following regression, estimated for each country c:

$$ln\left(\frac{I_{i,c,t}}{I_{i,c,t-1}}\right) = \alpha_c + \varphi_c ln\left(\frac{V_{i,c,t}}{V_{i,c,t-1}}\right) + \varepsilon_{i,c,t}$$

where i indexes industry, c indexes country, and t indexes year. $I_{i,c,t}$ is gross fixed capital formation and $V_{i,c,t}$ is value added. The sample period is from 2003 to 2015.

4.5.2 Evidence from the cross-section of countries

We examine the relation between noise in stock prices and industry-level capital allocation efficiency by using country-level estimates. We estimate the investment-to-added value elasticity (φ_c) for each country c using the regression (4.4). Noise and information for stocks in country c are calculated by first taking the market capitalization weighted average across stocks in that country within each year, then taking the simple average across years. We then regress φ_c on the measures of noise in prices, controlling for various information types, as follows:

$$\varphi_c = \alpha + \rho Noise_c + \delta CONTROL_c + \varepsilon_c \tag{4.5}$$

Table 4.6 shows that the country-level regression results are consistent with the notion that increased noise in stock prices reduces the efficiency with which capital is allocated across industries. Noise is negatively and significantly associated with the elasticity of investment to value added. The coefficient of $Noise$ is -1.92 with a t-statistic of -1.84 in Model 1. When considering this level of statistical significance, one must take into account that the country-level regressions use a mere 32 observations: one for each country. The standard deviation of $Noise$ is 0.039, the mean of φ_c is 0.497, and the standard deviation of φ_c is 0.237. Therefore a one standard deviation increase in $Noise$ is associated with a 15.06% (=$-1.92\times0.039/0.497$) decrease in the investment-to-value-added elasticity from its mean, or a decrease of 31.59% (=$-1.92\times0.039/0.237$) of its standard deviation. These magnitudes are similar to what we estimated earlier for the impact of noise on firm-level investment efficiency, despite considerable differences in the efficiency measures and their granularity.

The effects of noise on capital allocation efficiency in Table 4.6 are even stronger when we control for the impact of information on φ_c, indicating the effects of noise are not subsumed by the effects of information, but rather, they are incremental to the effects of information in prices. Taking Model 3 as an example, the coefficient for $Noise$ becomes -9.30 with a t-statistic of -2.69 after controlling for $MktInfo$ and $FirmInfo$.

Firm-specific information in prices is associated with higher capital allocation efficiency, φ_c. A one standard deviation increase in $FirmInfo$ increases φ_c by 53.56% (=2.61 × 0.102 / 0.497) from its mean, or an increase of 112% (=2.61×0.102/0.237) of a standard deviation of φ_c. In other words, countries with better informational efficiency—more information and less noise in prices—have a stronger tendency for investment to increase in growing industries and decrease in declining or more slowly growing industries. These results highlight the important real effects of the quality and efficiency of stock markets.

Table 4.6. Cross-country evidence on the relation between noise and industry-level capital allocation efficiency.
This table reports the results from panel regressions in which the units of observations are countries and the dependent variable is the elasticity of industry investment to value added (β_c). Model 1 reports the results from the baseline model, in which we regress β_c on a measure of the amount of noise in stock prices ($Noise$). Models 2-4 control for different types of information impounded into stock prices. $PrivateInfo$, $PublicInfo$, and $MktInfo$ are measures of the amount of private firm-specific information, public firm-specific information, and market-wide information in stock prices, respectively. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes major manufacturing industries in 32 countries from 2003 to 2015.

Variable	(1)	(2)	(3)	(4)
$Noise$	-1.92	-10.10	-9.30	-7.46
	(-1.81)*	(-2.74)**	(-2.69)**	(-2.26)**
$MktInfo$		0.32	0.39	
		(0.32)	(0.39)	
$PrivateInfo$		2.38		
		(1.85)*		
$PublicInfo$		3.18		
		(2.06)**		
$FirmInfo$			2.61	
			(2.47)**	
$Info$				1.30
				(1.90)*
R^2 (%)	10.13	28.56	28.06	21.89

4.6 Robustness tests

We test the robustness of the results in several ways. For the US evidence on the relation between noise and firm-level investment efficiency, we alter the baseline specification by accounting for the impact of cash flows on future investment. While Q is a price-based measure of a firm's investment opportunities, cash flows (CF), defined as the sum of net income before extraordinary items, depreciation and amortization expenses, and R&D expenses, scaled by the book value of assets, is a non-price-based measure. We include CF in the regressions as a separate term and as an interaction with noise and information.

Panel A in Table 4.7 presents the results of this robustness test. The impact of Q on investment remains positive and highly significant. We find that cash flows are significantly positively associated with the level of investment, confirming the findings of prior literature that investment depends positively on cash flow. The coefficients for the interaction between CF and information are negative and significant. This is consistent with the argument that if low cash flow firms that have greater information in prices can more easily raise capital, then investment should be less dependent on cash flow in these firms. Investment sensitivity to cash flow should be lower for firms facing lower external financial costs.

More importantly, the coefficients on the interaction between Q and noise remain negative and statistically significant after controlling for CF and its interaction terms. The results also suggest that investment-Q sensitivity is higher in firms with more firm-specific private information in prices.

Second, in Panel B of Table 4.7, we replace Tobin's Q with Total Q. Total Q is an improved proxy for Q, proposed by Peters and Taylor (2017). It involves adjusting the denominator of Tobin's Q (which traditionally reflects the physical assets of the firm) to include the capitalized intangible assets.[35] The results are robust to this alternative way of measuring Q with the coefficient for Total $Q \times Noise$ being negative and statistically significant at 1% level in Model 1 and Model 4, and statistically significant at 10% level in Models 2 and 3, which reinforces our main findings. Increased firm-specific information also makes investment more sensitive to Total Q.

[35] Peters and Taylor (2017) find that the intangibles-adjusted Tobin's Q explains firms' total and intangible investment rates substantially better than the traditional measure based on physical assets only.

Table 4.7. Robustness tests for the US evidence on how noise impacts firm-level investment efficiency.
This table reports variations of the tests in Table 4.3 in which the units of observations are firm-years and the dependent variable is firm-level investment (annual CAPEX and R&D expenses as a percentage of assets). Panel A uses Tobin's Q as per our baseline tests, but additionally controls for the firms' cash flows. Panel B uses Total Q instead of Tobin's Q. Total Q is a new Q proxy that accounts for intangible capital, proposed by Peters and Taylor (2017). $Noise$, $PrivateInfo$, $PublicInfo$, and $MktInfo$ are measures of the amount of noise, private firm-specific information, public firm-specific information, and market-wide information in stock return variance, respectively. $Info$ is the sum of $MktInfo$, $PrivateInfo$, and $PublicInfo$ whereas $FirmInfo$ is the sum of the two latter types of information. CF is the annual cash flow from operations in million USD and inv_assets is the inverse of the book value of assets. We include year and firm fixed effects. Standard errors are clustered by firm. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes stocks listed on NYSE, AMEX, and NASDAQ and spans the period from 1963 to 2015.

Panel A: Using Tobin's Q				*Panel B: Using Total Q*			
Variable	(1)	(2)	(3)	(1)	(2)	(3)	(4)
Q	1.78	1.70	1.73	1.09	0.73	0.76	0.84
	(19.74)***	(18.76)***	(19.28)***	(19.62)***	(9.06)***	(9.53)***	(10.40)***
$Noise \times Q$	-0.24	-0.14	-0.15	-0.20	-0.11	-0.12	-0.17
	(-3.11)***	(-1.88)*	(-1.93)*	(-4.08)***	(-1.67)*	(-1.78)*	(-2.64)***
$MktInfo \times Q$		-0.31	-0.34		-0.08	-0.09	
		(-2.54)**	(-2.71)***		(-0.69)	(-0.78)	
$PrivateInfo \times Q$		0.27			0.18		
		(4.17)***			(3.11)***		
$PublicInfo \times Q$		0.01			-0.02		
		(0.24)			(-0.50)		
$FirmInfo \times Q$			0.12			0.06	
			(3.45)***			(1.86)*	
$Info \times Q$	0.06						0.03
	(2.10)**						(1.17)
$Noise$	-0.53	-0.59	-0.57	-0.89	-1.05	-1.04	-1.14
	(-3.77)***	(-4.33)***	(-4.20)***	(-7.86)***	(-7.92)***	(-7.88)***	(-7.69)***
$MktInfo$		1.38	1.37		1.25	1.16	
		(5.94)***	(5.92)***		(5.28)***	(4.87)***	
$PrivateInfo$		-0.12			0.56		
		(-1.02)			(3.90)***		
$PublicInfo$		0.25			0.50		
		(2.70)***			(5.13)***		

Table 4.7. (continued)

Panel A: Using Tobin's Q				*Panel B: Using Total Q*			
Variable	(1)	(2)	(3)	(1)	(2)	(3)	(4)
$FirmInfo$			0.11			0.55	
			(1.83)*			(7.72)***	
$Info$	0.30						0.69
	(5.41)***						(9.97)***
CF	9.25	9.37	9.39		14.59	14.68	14.48
	(6.12)***	(6.24)***	(6.25)***		(6.42)***	(6.45)***	(6.27)***
$Noise \times CF$	-0.72	-2.23	-2.15		-4.43	-4.25	-1.85
	(-0.98)	(-5.35)***	(-5.32)***		(-6.43)***	(-6.20)***	(-1.35)
$MktInfo \times CF$		4.08	4.50		7.02	7.97	
		(4.27)***	(4.72)***		(4.77)***	(5.42)***	
$PrivateInfo \times CF$		-1.19			-1.17		
		(-1.37)			(-0.88)		
$PublicInfo \times CF$		-2.14			-3.40		
		(-5.13)***			(-5.49)***		
$FirmInfo \times CF$			-1.87			-2.76	
			(-5.87)***			(-5.79)***	
$Info \times CF$	-1.80						-2.65
	(-5.40)***						(-4.75)***
inv_assets	0.13	0.13	0.12	0.25	0.23	0.23	0.23
	(25.43)***	(25.06)***	(24.96)***	(35.76)***	(31.38)***	(31.27)***	(31.77)***
R^2 (%)	17.14	17.64	17.60	13.21	17.34	17.26	16.50
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 4.8. Robustness tests for the cross-country evidence on how noise impacts firm-level investment efficiency.
This table reports variations of the tests in Table 4.4 in which the dependent variable is the estimated investment-Q sensitivity ($\beta_{1,c,t}$). Panel A reports the results in which the units of observation are country-years as per our baseline models, but includes variations of fixed effects. Model 1 includes year fixed effects. Model 2 includes country fixed effects. Model 3 includes neither country nor year fixed effects. Panel B modifies the baseline empirical models by replacing the dependent variable with investment-Q sensitivity ($\beta_{1,c}$) estimated at the country level rather than as country-year observations. Models 1-3 control for different types of information impounded into stock prices. $Noise$, $PrivateInfo$, $PublicInfo$, and $MktInfo$ are measures of the amount of noise, private firm-specific information, public firm-specific information, and market-wide information in stock return variance, respectively. $Info$ is the sum of $MktInfo$, $PrivateInfo$, and $PublicInfo$ whereas $FirmInfo$ is the sum of the two latter types of information. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes stocks listed on major exchanges in 42 countries and spans the period from 2003 to 2015.

Panel A: Using country-year observation and different fixed effects				*Panel B: Using country-level regressions*		
Variable	(1)	(2)	(3)	(1)	(2)	(3)
$Noise$	-7.46	-6.54	-5.93	-26.32	-18.07	-13.45
	(-2.47)**	(-3.55)***	(-2.28)**	(-1.95)*	(-1.69)*	(-1.43)
$MktInfo$	-0.72	2.07	0.03	-0.90	-1.40	
	(-0.51)	(1.38)	(0.02)	(-0.28)	(-0.41)	
$PrivateInfo$	0.95	-1.94	-0.10	3.28		
	(0.31)	(-0.75)	(-0.03)	(0.65)		
$PublicInfo$	2.86	3.75	3.48	11.42		
	(1.60)	(2.33)**	(2.12)**	(1.53)		
$FirmInfo$					6.29	
					(1.24)	
$Info$						1.83
						(0.74)
R^2 (%)	9.71	40.35	1.22	10.95	8.64	4.93
Country FE	No	Yes	No	No	No	No
Year FE	Yes	No	No	No	No	No

We also perform a variety of robustness tests for the analysis of how noise is related to firm-level investment efficiency in the cross-section of countries and report the results in Table 4.8. In Panel A, we use country-year observations (as per those in the baseline models) but the regressions include different types of fixed effects. Model 1 has only year fixed effects. Model 2 excludes year fixed effects but includes country fixed effects. No fixed effects are used in Model 3. In Panel B, we first estimate the investment-Q sensitivity by country rather than by country-year using all stocks listed in each country during the 2003-2015 period. The noise and information measures for country c are calculated by first taking the market capitalization weighted average across stocks in that country within each year, then taking the simple average across years. We then estimate the regressions linking noise and firm-level investment efficiency using the 42 country-level observations rather than country-year observations.

For both of these additional robustness tests we find that our main results remain strong and qualitatively consistent with our baseline models. The variable of interest, which is $Noise$, has negative and statistically significant coefficients across the six specifications, reaffirming that more noise is associated with lower investment efficiency.

Finally, we conduct additional robustness tests for the analysis of industry-level capital allocation efficiency. The first is using industry-level observations instead of the country-level observations used in the baseline models. By doing so, we increase the number of observations and statistical power, potentially at the cost of an uneven number of observations per country.

We estimate the following specification:

$$ln\left(\frac{I_{i,c,t}}{I_{i,c,t-1}}\right) = \alpha + \varphi ln\left(\frac{V_{i,c,t}}{V_{i,c,t-1}}\right) + \theta\left[ln\left(\frac{V_{i,c,t}}{V_{i,c,t-1}}\right) \times Noise_{c,t}\right] \quad (4.6)$$

$$+\delta CONTROLS_{i,c,t} + \varepsilon_{i,c,t}$$

where our variable of interest is the interaction term $ln\left(\frac{V_{i,c,t}}{V_{i,c,t-1}}\right) \times Noise_{c,t}$. Similar to previous tests, we also control for the impact of information and its interaction with the value-added growth.

Table 4.9 reports the results of these robustness tests. Consistent with our baseline results, industry investment tends to be less sensitive to industry value added in countries with higher levels of noise. By contrast, the investment-to-value-added elasticity tends to be higher in countries with more information reflected in stock prices. The results remain robust when we add country fixed effects, year fixed effects, and both year and country fixed effects.

Table 4.9. Robustness tests for the evidence of how noise impacts industry-level capital allocation efficiency.
This table reports variations of the models in Table 4.6 testing the effect of noise on industry-level capital allocation efficiency. In all models the units of observation are industry-years and the dependent variable is industry investment growth ($ln(I_{i,c,t}/I_{i,c,t-1})$). The key independent variables are industry value added growth ($ln(V_{i,c,t}/V_{i,c,t-1})$) and its interactions with $Noise$, $PrivateInfo$, $PublicInfo$, and $MktInfo$, which are measures of the amount of noise, private firm-specific information, public firm-specific information, and market-wide information in stock return variance, respectively. $Info$ is the sum of $MktInfo$, $PrivateInfo$, and $PublicInfo$ whereas $FirmInfo$ is the sum of the two latter types of information. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes stocks listed on NYSE, AMEX, and NASDAQ and spans the period from 2003 to 2015.

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$ln(V_{i,c,t}/V_{i,c,t-1})$	0.35	0.25	0.26	0.38	0.26	0.28	0.39	0.28	0.30
	(5.33)***	(3.90)***	(4.02)***	(6.83)***	(4.64)***	(5.03)***	(7.01)***	(4.97)***	(5.29)***
$Noise \times ln(V_{i,c,t}/V_{i,c,t-1})$	-2.94	-2.69	-2.63	-3.08	-3.01	-2.67	-2.71	-2.44	-2.08
	(-2.46)**	(-2.29)**	(-2.24)**	(-2.60)***	(-2.60)***	(-2.32)**	(-2.25)**	(-2.14)**	(-1.80)*
$MktInfo \times ln(V_{i,c,t}/V_{i,c,t-1})$	0.62	0.40	0.37	0.61	0.42	0.34			
	(1.54)	(1.08)	(0.98)	(1.54)	(1.12)	(0.89)			
$PrivateInfo \times ln(V_{i,c,t}/V_{i,c,t-1})$	1.47	1.37	1.35						
	(2.40)**	(2.29)**	(2.29)**						
$PublicInfo \times ln(V_{i,c,t}/V_{i,c,t-1})$	0.66	0.83	0.81						
	(1.52)	(1.86)*	(1.85)*						
$FirmInfo \times ln(V_{i,c,t}/V_{i,c,t-1})$				0.89	1.10	0.96			
				(2.23)**	(2.75)***	(2.42)**			
$Info \times ln(V_{i,c,t}/V_{i,c,t-1})$							0.70	0.70	0.58
							(2.54)**	(2.64)***	(2.16)**
$Noise$	0.59	0.39	0.43	0.60	0.32	0.46	0.55	0.23	0.42
	(2.85)***	(1.95)*	(1.98)**	(2.90)***	(1.58)	(2.13)**	(2.63)***	(1.15)	(1.91)*
$MktInfo$	0.01	0.09	0.11	0.02	0.10	0.13			
	(0.08)	(1.38)	(1.34)	(0.26)	(1.44)	(1.57)			
$PrivateInfo$	0.10	0.19	0.40						
	(0.56)	(1.79)*	(2.18)**						
$PublicInfo$	-0.46	-0.30	-0.38						
	(-5.58)***	(-3.86)***	(-4.56)***						

Table 4.9. (continued)

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$FirmInfo$				-0.34	-0.15	-0.23			
				(-4.65)***	(-2.25)**	(-3.05)***			
$Info$							-0.17	-0.03	-0.07
							(-3.41)***	(-0.58)	(-1.35)
R^2 (%)	9.44	12.24	13.15	9.27	11.98	12.87	9.05	11.81	12.64
Country FE	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes
Year FE	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

4.7 Conclusion

This chapter investigates an important real effect of stock market quality by analyzing the relation between the amount of noise in stock prices and the efficiency of capital allocation across firms and industries. The results show that stock markets, in particular those that exhibit a low level of noise, appear to provide useful signals to investors and corporate managers of investment opportunities and thereby direct capital to its most efficient uses. We find that in the long time-series of US data starting in 1963, a one standard deviation increase in noise reduces firm-level investment efficiency by 4.30%, after controlling for the information in prices and other factors. In comparison, a one standard deviation increase in the amount of information reflected in prices increases firm-level investment efficiency by 3.39%. Therefore, in an absolute sense and relative to the effects of information, our results show that noise has an economically meaningful incremental effect on capital allocation efficiency.

Our analysis of noise and investment efficiency in the cross-section of 42 countries also shows that countries with greater noise level in prices tend to have lower firm-level investment efficiency. In the cross-section of countries, a one standard deviation increase in $Noise$ is associated with a 6.14% decrease in the sensitivity of corporate investment to growth opportunities. Again, these results are robust to controlling for various forms of information in prices and therefore capture the independent or incremental effects of noise on investment efficiency.

We also analyze the efficiency of capital allocation at the broader level of across industries. Consistent with the measures of firm-level investment efficiency, we find that countries with less noise in stock prices have a stronger tendency to increase investment in their fastest growing industries and decrease it in their declining industries. A one standard deviation decrease in $Noise$ is associated with a 15.06% increase in the efficiency of capital allocation across industries. Information in prices, particularly firm-specific information, also helps to improve the allocation of capital across industries. A one standard deviation increase in firm-specific information increases industry-level capital allocation efficiency by 53%.

Our findings highlight the important real effects of secondary market quality in determining firms' investment behavior and the efficiency with which capital is allocated. Our results go some way to quantifying the benefits to the real economy of policies or structural changes in the secondary markets that make the markets more liquid and thus make prices less noisy.

Appendix 4.A. Data description

4.A.1 Sample selection

In each country, we use stocks from the largest stock exchange in that country, with the exception of several countries for which we include the two major listing exchanges: China (Shanghai and Shenzhen Stock Exchanges), Japan (Tokyo and Osaka Stock Exchanges), Germany (Deutsche Boerse AG and XETRA), and the US (NYSE and Nasdaq).

Following Griffin, Kelly, and Nardari (2010), we exclude preferred stocks, warrants, mutual funds, unit or investment trusts, duplicates, Global Depositary Receipts (GDRs), and other non-common equity. We eliminate cross-listed stocks and exclude secondary issues of shares. We restrict our analysis to stocks that trade in the company's home market and in local currency. Additionally we require that the stock has market capitalization data (in US dollars) at the end of each year. Dead stocks are included in the sample to mitigate survivorship bias.

Following Ang et al. (2009), in all non-US countries, we exclude very small firms by eliminating the 5% of firms with the lowest market capitalizations. We also omit newly listed stocks that have been traded for less than five months, as well as stocks that are about to be delisted. The initial sample includes slightly more than 38,000 stocks from around the world over the thirteen-year period from 2003 to 2015.

4.A.2 Return filters

We use the following screens in the return data to ensure tradability and avoid outliers. We exclude non-trading days, defined as days on which more than 90% of stocks listed on a given exchange have zero returns. We also exclude a stock if the number of zero-return days is more than 80% in a given month. To handle data errors in Datastream, we follow Ince and Porter (2006), and set the daily stock returns of both days d and $d-1$ to be missing if $R_{i,d}R_{i,d-1}-1 \leq 50\%$, where $R_{i,d}$ is the gross return of stock i on day d, and at least one of the two returns is 200% or greater. To avoid rounding errors, the daily return is set to be missing if the return index in Datastream for either the previous day or the current day is less than 0.01.

To construct stock liquidity measures, including Amihud's (2002) ILLIQ, which is sensitive to extreme returns, we discard stock-day observations with a daily return in the top or bottom 1% of the cross sectional distribution within a country. We exclude stock-month observations with a monthly return or stock price at the end of the previous month in the top or bottom 1% of the cross-sectional distribution within a country.

4.A.3 Firm filters

We apply the following filters when we conduct tests of the impact of noise on firm-level investment efficiency. Following Foucault and Frésard (2012), we exclude financial firms (SIC codes between 6000 and 6999) and utilities (SIC codes between 9000 and 9999) because their accounting numbers are largely dependent on statutory capital requirements. We also exclude observations with missing information on market value of equity, total assets, sales, and capital expenditure, as well as firms with total assets that are inferior to $10 million and firms with negative sales. To reduce the effects of outliers, all ratios are winsorized at 5% in each tail.

4.A.4 Industry filters

We apply the following filters when we conduct tests of the impact of noise on industry-level investment efficiency. To reduce the influence of outliers, we follow Wurgler (2000), excluding observations for which the absolute value of either log investment growth or log value added growth exceeds one. Additionally, to focus on economically important industries, we exclude those for which value added is less than 0.1% of the country's total manufacturing value added in that year.

Appendix 4.B. Measures of information and noise in stock prices

Brogaard et al. (2019) propose a variance decomposition model that separates noise from various sources of information. The Brogaard et al. (2019) approach builds on Hasbrouck (1993) with a temporary-permanent decomposition of returns, where permanent innovations are interpreted as innovations in the fundamental value (efficient price) driven by new information being impounded in prices, and temporary innovations are pricing errors (deviations from fundamentals) generically referred to as “noise”.

According to Hasbrouck (1993), the observed price can be thought of as a sum of the efficient price, which follows a random walk process, and a residual component, termed the pricing error. Brogaard et al. (2019) push the variance decomposition further to separate the information (efficient price) into market-wide information, firm-specific information revealed through trading on private information, and firm-specific information revealed through public information.

We follow the procedure described by Brogaard et al. (2019) and estimate a vector autoregressive (VAR) model of stock returns, signed dollar volume of trading, and market returns.

We start with a structural VAR with five lags to allow a full week of lagged effects:

$$\begin{aligned}
r_{m,t} &= a_0 + \textstyle\sum_{l=1}^{5} a_{1,l} r_{m,t-l} + \sum_{l=1}^{5} a_{2,l} x_{t-l} + \sum_{l=1}^{5} a_{3,l} r_{t-l} + \varepsilon_{r_m,t} \\
x_t &= b_0 + \textstyle\sum_{l=0}^{5} b_{1,l} r_{m,t-l} + \sum_{l=1}^{5} b_{2,l} x_{t-l} + \sum_{l=1}^{5} b_{3,l} r_{t-l} + \varepsilon_{x,t} \qquad (4.B.1) \\
r_t &= c_0 + \textstyle\sum_{l=0}^{5} c_{1,l} r_{m,t-l} + \sum_{l=0}^{5} c_{2,l} x_{t-l} + \sum_{l=1}^{5} c_{3,l} r_{t-l} + \varepsilon_{r,t},
\end{aligned}$$

where $r_{m,t}$ is the daily market return, x_t is the signed dollar volume of trading in the given stock (positive for net buying and negative for net selling), and r_t is the daily stock return. We estimate this VAR separately for each stock in each year using daily data.

The VAR in equation (4.B.1) is subsequently transformed into an infinite order structural vector moving average (VMA), from which we isolate temporary price movements from permanent price movements by considering the impulse response functions.

$$
\begin{aligned}
r_{m,t} &= A_0 + \textstyle\sum_{l=0}^{\infty} A_{1,l}\varepsilon_{r_m,t-l} + \sum_{l=1}^{\infty} A_{2,l}\varepsilon_{x,t-l} + \sum_{l=1}^{\infty} A_{3,l}\,\varepsilon_{r,t-l} \\
x_t &= B_0 + \textstyle\sum_{l=0}^{\infty} B_{1,l}\varepsilon_{r_m,t-l} + \sum_{l=0}^{\infty} B_{2,l}\varepsilon_{x,t-l} + \sum_{l=1}^{\infty} B_{3,l}\,\varepsilon_{r,t-l} \\
r_t &= C_0 + \textstyle\sum_{l=0}^{\infty} C_{1,l}\varepsilon_{r_m,t-l} + \sum_{l=0}^{\infty} C_{2,l}\varepsilon_{x,t-l} + \sum_{l=0}^{\infty} C_{3,l}\,\varepsilon_{r,t-l}.
\end{aligned}
\tag{4.B.2}
$$

The permanent information-driven innovation in the efficient price is given by:

$$w_t = \theta_{r_m}\varepsilon_{r_m,t} + \theta_x\varepsilon_{x,t} + \theta_r\varepsilon_{r,t} \tag{4.B.3}$$

where $\theta_{r_m} = \sum_{l=0}^{\infty} C_{1,l}$, $\theta_x = \sum_{l=0}^{\infty} C_{2,l}$, $\theta_r = \sum_{l=0}^{\infty} C_{3,l}$.

The innovation in the pricing error (the temporary component of the return) is given by:

$$\Delta s_t = r_t - \mu - w_t = r_t - A_0 - \theta_{r_m}\varepsilon_{r_m,t} - \theta_x\varepsilon_{x,t} - \theta_r\varepsilon_{r,t}. \tag{4.B.4}$$

The variance of information-driven innovations in the efficient price is therefore:

$$\sigma_w^2 = \theta_{r_m}^2 \sigma_{\varepsilon_{r_m}}^2 + \theta_x^2 \sigma_{\varepsilon_x}^2 + \theta_r^2 \sigma_{\varepsilon_r}^2 \tag{4.B.5}$$

The contribution to the efficient price variation from each of the information components is $MktInfo = \theta_{r_m}^2 \sigma_{\varepsilon_{r_m}}^2$ (market-wide information), $PrivateInfo = \theta_x^2 \sigma_{\varepsilon_x}^2$ (private firm-specific information), and $PublicInfo = \theta_r^2 \sigma_{\varepsilon_r}^2$ (public firm-specific information). The variance of the changes in the pricing errors (noise), σ_s^2, is computed from the time-series of Δs_t: $Noise = \theta_s^2$.

$FirmInfo$ is the sum of $PrivateInfo$ and $PublicInfo$. $Info$ is the sum of $FirmInfo$ and $MktInfo$. As the values of return variance are small, in the empirical tests, we multiply these measures by 1,000 for readability.

Appendix 4.C. Variable definitions

Variables	Notation	Description	Data sources
Panel A: Firm-level variables			
Investment	$CAPXRND$	Defined as the sum of annual capital expenditure and the annual R&D expenses, as a percentage of the beginning-of-year book value of assets.	Compustat, Worldscope
Tobin's Q	Q	Calculated by using the market value of equity plus book value of assets minus the book value of equity, scaled by book value of assets.	Compustat, Worldscope
Total Q	Q	An intangibles-adjusted Tobin's Q that accounts for both physical and intangible investment opportunities by additionally incorporating intangible assets into the dominator.	Peters and Taylor (2017)
Cash flow	CF	Defined as the sum of annual net income (before extraordinary items, depreciation and amortization expenses) and R&D expenses, scaled by beginning-of-year book value of assets.	Compustat
Inverse of the book value of assets	inv_assets	Defined as one thousand divided by the total book value of assets. Book value of assets is in million US dollars.	Compustat
Noise	$Noise$	A measure of the amount of noise in stock return variance. Calculated from a variance decomposition model specified in Appendix 4.B.	Datastream, CRSP
Market-wide information	$MktInfo$	A measure of the amount of market-wide information in stock return variance. Calculated from a variance decomposition model specified in Appendix 4.B.	Datastream, CRSP
Private firm-specific information	$PrivateInfo$	A measure of the amount of private firm-specific information in stock return variance. Calculated from a variance decomposition model specified in Appendix 4.B.	Datastream, CRSP

Variables	Notation	Description	Data sources
Public firm-specific information	$PublicInfo$	A measure of the amount of public firm-specific information in stock return variance. Calculated from a variance decomposition model specified in Appendix 4.B.	Datastream, CRSP
Firm-specific information	$FirmInfo$	A measure of the amount of firm-specific information in stock return variance. $FirmInfo$ is the sum of $PrivateInfo$ and $PublicInfo$.	Datastream, CRSP
Information	$Info$	A measure of the amount of information in stock return variance. $Info$ is the sum of $FirmInfo$ and $MktInfo$.	Datastream, CRSP
Panel B: Industry-level variables			
Gross fixed capital formation	I	Defined as the cost of new and used fixed assets minus the value of sales of used fixed assets, where fixed assets include land, buildings, machinery and equipment.	United Nations
Value added	V	Defined as the value of shipments of goods produced minus the cost of intermediate goods and required services, with appropriate adjustments made for inventories of finished goods, work-in-progress, and raw materials.	United Nations
Panel C:Country-level variables			
Investment-Q sensitivity	β	β is estimated by regressing firm's investment on Tobin's Q, controlling for the inverse of the book value of assets.	Estimated
Elasticity of industry investment to value added	φ	φ is estimated by regressing the log growth in industry investment on the log growth in industry value added.	Estimated

Chapter 5: How gambling in stock markets affects information and noise

5.1 Introduction

Trading stocks is a source of fun and excitement to many individuals, much like gambling in casinos or betting on sporting events (Gao and Lin, 2014). Previous studies show that some individuals trade for entertainment and sensation seeking, effectively treating the stock market as a type of casino (e.g., Goetzmann and Dhar, 2004; Grinblatt and Keloharju, 2008; Dorn and Sengmueller, 2009). Such trading impacts asset prices. For example, "lottery-like" stocks such as those with low price, high skewness, and high volatility become overpriced and subsequently have low average returns (e.g., Mitton and Vorkink, 2007; Barberis and Huang, 2008; Kumar, 2009; Bali et al., 2011).

This chapter analyzes the extent to which gambling occurs on stock markets, what determines the amount of gambling in a given country's stock markets, and how gambling impacts stock market liquidity and the informational efficiency of prices. We tackle these issues using unique global data on gambling in traditional forms (e.g., casinos, lotteries, sports betting) and novel measures of gambling in stock markets, spanning 39 countries. Our data on "traditional" gambling captures the amount of money staked in each country's casinos, lotteries, horse racing/sport betting, gaming machines, and online gambling (hereafter "casino" gambling). Together, these revenues account for 86% of estimated total global gambling revenue. We use a novel approach to measure the amount of gambling in stock markets, exploiting the turnover differences between lottery-like stocks and non-lottery stocks. Under the assumption that gambling in stock markets involves disproportionate trade in lottery-like stocks, the excess amount of trading in lottery-like stocks captures the gambling "revenue" in stock markets.

Our first finding is that the amount of gambling on stock markets is economically meaningful. In dollar terms, gambling on stock markets is at least 3.5 times the combined gambling in casinos, lotteries, horse racing/sport betting, gaming machines, and online gambling. This difference is particularly strong in countries with relatively more developed and liquid stock markets, such as the US and Hong Kong, which have the highest per capita levels of gambling on stock markets. Thus, financial markets play a significant role as one of the venues in which gambling takes place.

Our second finding is that determinants of the level of gambling on stock markets and casino gambling are remarkably similar. We find that much of the cross-country variation in both forms of gambling can be explained by a set of common drivers. Wealthier countries tend to gamble more per capita. Various cultural dimensions also explain gambling, with higher levels of gambling in societies that are less troubled by uncertainty and societies with a more individualistic (confident) orientation. Region effects also play a significant role, with higher levels of gambling in North and South America, as well as Asia, after controlling for other factors. As a result of these common drivers, the two forms of gambling are strongly positively related across countries. These findings further reinforce the notion that gambling on stock markets has many similarities with gambling in casinos.

The third main finding is that stock markets serve as a substitute or alternative to “traditional” gambling venues such as casinos, lotteries, sports betting, gaming machines, and so on. We find that when a country restricts traditional gambling, for example by banning casinos or restricting them to a state-owned monopoly, a significant volume of gambling activity spills over onto stock markets. This finding is consistent with prior evidence that stock market volume declines when weekly lottery jackpots are larger than usual (e.g., Gao and Lin, 2014; Dorn et al., 2015) and indicates that the substitution effect holds more broadly for a range of gambling forms and in a global sample.

Finally, we turn to the question of how gambling on stock markets affects various measures of liquidity and informational efficiency. Our identification strategy exploits the fact that restrictions on “traditional” forms of gambling such as casinos and sports betting vary across countries, impact the amount of gambling on stock markets due to a spillover effect, and that restrictions on gambling are unlikely to be influenced by stock market characteristics. In first-stage regressions, we isolate the variation in stock market gambling that is due to restrictions on traditional gambling such as casinos (e.g., bans or monopolies). In second-stage regressions we find that gambling in stock markets tends to make stock markets more liquid. This finding is consistent with the notion that stock market liquidity depends on the amount of stock market participation and in particular the participation of relatively uninformed traders such as gamblers (e.g., Kyle, 1985).

Perhaps more surprisingly, we find that increased gambling in stock markets does not harm the amount of information reflected in prices. Rather, gambling on stock markets tends to increase the amount of information that is reflected in prices, increase measures of informational efficiency, and reduce measures of noise in prices. This finding is consistent with theory, which predicts that higher liquidity creates more incentives for traders to gather private information and trade on it (e.g., Grossman and Stiglitz, 1980; Kyle, 1984, 1989). Therefore, even if gamblers are relatively or completely uninformed traders, they can still contribute to market efficiency by making markets more liquid and thereby encouraging informed trading.

The findings of this study have policy implications. For example, regulation or restriction of casino gambling is one mechanism to influence the amount of gambling in a country's stock market. Our results suggest that stock markets can benefit from gambling activity, which can in turn benefit the real economy through corporate investment (e.g., Dessaint et al., 2016) and a lower cost of capital (e.g., Amihud and Mendelson, 1986). Therefore stock market liquidity can be encouraged by restricting casino gambling to allow more of the country's gambling to occur within financial markets. As a market development strategy, this idea has the most potential in countries that have high gambling propensities and relatively illiquid or underdeveloped stock markets. These potential benefits, however, should be weighed up against downsides such as distortions in the pricing of lottery-like stocks and systemic risk.

5.2 Related literature

This chapter contributes to several branches of the literature. The first is asset pricing studies that find gambling and preferences for skewness lead to low average returns for lottery-like stocks including stocks with high skewness (e.g., Mitton and Vorkink, 2007; Barberis and Huang, 2008; Kumar, 2009; Bali et al., 2011). Our study shows that gambling has additional effects on stocks and stock markets, such as increasing liquidity and increasing the informational efficiency of prices.

This chapter also contributes to studies of individual investors that engage in gambling, sensation seeking, and trading for entertainment on stock markets (e.g., Goetzmann and Dhar, 2004; Grinblatt and Keloharju, 2008; Dorn and Segmeuler 2009). While such studies provide rich insights about the variety of motivations that drive individual investors to trade stocks and clearly demonstrate that gambling is among those reasons, our study contributes by quantifying the amount of trading that is associated with gambling. Our estimates, which can be considered a lower bound (as they only capture gambling in lottery-like stocks), suggest that gambling on stock markets is economically meaningful and considerably exceeds the scale of traditional gambling such as casinos and lotteries.

The third branch of literature pertains to studies of the determinants of gambling. For instance, previous gambling studies (e.g., Kallick et al., 1976; Brenner and Brenner, 1990; Clotfelter et al., 1999; Raylu and Oei, 2004; Kumar, 2009; Kumar et al., 2011) provide evidence that income, wealth, age, education, gender, religion, culture, and marital status affect lottery purchases or investments in lottery-like stocks. We contribute to this literature by showing that traditional forms of gambling and gambling on stock markets have very similar drivers, including wealth effects, culture, and geographic fixed effects.

Fourth, our study contributes to a growing body of work on the relation between lotteries and the stock market. In cross-sectional analyses, Kumar (2009) shows that state lotteries and lottery-like stocks in the US market are complements with very similar socioeconomic clienteles. In contrast, Barber et al. (2009), Gao and Lin (2014), and Dorn, Dorn, and Sengmueller (2015) employ time-series data and find a substitution effect between stock trading and lottery participation. Our study is the first to directly investigate the relation between traditional and stock market gambling globally, using more comprehensive measures of gambling (in some countries lotteries account for a negligible fraction of gambling) and country-level restrictions on traditional gambling. Our results support those of Kumar (2009) in that many of the determinants of traditional forms of gambling are also determinants of stock market gambling. Our results also show that the substitution effect between traditional gambling and stock markets holds more generally, across broader measures of gambling and in a global sample.

Finally, our study relates to the literature on how uninformed “noise traders” affect liquidity and informational efficiency. Our results support theory in which increasing noise traders makes markets more liquid (e.g., Kyle, 1985; Glosten and Milgrom, 1985). Our results also support theory that predicts the presence of more noise traders and greater liquidity incentivize more information acquisition (e.g., Grossman and Stiglitz, 1980; Kyle, 1984, 1989), which can lead to more informative prices.

5.3 Data and measures

5.3.1 Sample and data sources

We analyze the 39 countries for which gambling data are available. Among them, 19 are emerging countries (Argentina, Brazil, Chile, China, Columbia, Czech Republic, Greece, India, South Korea, Malaysia, Mexico, Peru, Philippines, Poland, Russia, Portugal, South Africa, Thailand, and Turkey) and 20 are developed countries (Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Hong Kong, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Singapore, Spain, Sweden, United Kingdom, and United States). In each country, we analyze all stocks that meet the data requirements and multistage screening process described in Appendix 5.A. Our sample period is from 2003 to 2015.

We combine data from multiple sources. We obtain data on individual stock returns, prices, volumes, and market capitalizations from Datastream for all stocks in our sample. For US stocks, we use data from CRSP (Center for Research in Security Prices). We obtain country macroeconomic data such as GDP per capita, the unemployment rate, and population from the World Bank database. We use measures of cultural dimensions proposed by Hofstede (1997, 2001) including IDV (Individualism versus Collectivism), IVR (Indulgence versus Restraint), and UAI (Uncertainty Avoidance Index) from Hofstede's website.[36] We obtain country-level data on gambling revenues and regulations in the 39 countries from a unique database collected by Global Betting and Gaming Consultants (GBGC). The data from GBGC are hand collected from various primary sources such as annual reports of gambling regulators and industry associations, financial reports from gambling license operators, local newspaper reports, ministry of finance tax reports, and experts based in the various jurisdictions covered by the data.

The next subsection details of how we use the GBGC data to construct measures of traditional gambling, following which we discuss our novel measures of

[36] https://geerthofstede.com/research-and-vsm/dimension-data-matrix/

stock market gambling. Appendix 5.C provides the definitions and sources for all variables.

5.3.2 Measures of "casino" gambling

While previous studies of the link between gambling and the stock market focus on lotteries as a proxy for gambling activities outside of the stock market (e.g., Dorn, Dorn, and Sengmueller, 2015), our unique dataset allows us to take a broader approach to measuring gambling outside of the stock market. Our measure of "traditional" gambling includes the whole spectrum of gambling activities in all land-based forms, including casinos (including Native American gaming), horse racing, sports betting, gaming machines, bingo/keno, and lotteries, in addition to online gambling.37 This broader analysis of gambling is important in drawing cross-country comparisons because the popularity of lotteries compared to other forms of gambling can vary substantially across countries; in some cases, lotteries are a tiny fraction of a country's total gambling activity. For example, our data show that in Australia and Japan lotteries represent only 7% of total gambling, while the corresponding number for gaming machines is more than 50%. As another example, in Hong Kong, where casinos and gaming machines are banned, lotteries are also only a small part of total gambling (5%), with the majority of gambling activity (82%) being attributable to horse racing and sports betting. Although our measure of gambling outside of the stock market includes many forms of gambling, for conciseness, we refer to it as "casino" gambling to differentiate it from our measures of stock market gambling.

For each country in each year, we aggregate the US dollar value of all money wagered in each of the above forms of gambling. We normalize this aggregate by the population to obtain a measure of gambling activity per capita per annum ($CasinoGambling$). The money wagered is known in the gambling industry as the "handle". It includes "recycling", which is money that is staked on more than one occasion within the same gambling session. For example, a bettor's winnings on race A that are then staked on race B or C, or winnings within a casino that are then re-bet

---

[37] Due to some challenges in the geographic allocation of online gambling revenues, in robustness tests we show that our main results are robust to only considering the land-based forms of gambling.

within the casino. Our measure therefore is a reflection of the amount of money that is actually risked by punters in the legal part of the gambling industry in each country. Our measure, $CasinoGambling$, accounts for 86% of estimated total global gambling activity.

5.3.3 Measures of stock market gambling

a. Measure development

We use novel measures of stock market gambling that exploit the turnover differences between lottery-like stocks and non-lottery stocks. The intuition for these measures is as follows. If all market participants were passive investors, merely buying and selling the market portfolio, the ratio of traded dollar volume to market capitalization for individual stocks (their turnover ratios) would be equal for all stocks (e.g., Tkac, 1999; Bhattacharya and Galpin, 2011). In contrast, with a mix of passive investors and gamblers/speculators, the stocks that are disproportionally traded more by gamblers/speculators will have higher turnover ratios than stocks that are less attractive to gamblers/speculators. Given that gamblers are drawn to stocks with particular characteristics such as positive skewness in returns (e.g., Kumar, 2009; Bali et al., 2011), the excess amount of trading in lottery-like stocks will approximate the gambling "revenue" on stock markets. More precisely, as we explain below, it is likely to provide a lower bound on the gambling in stock markets.

To more formally motivate our measure, we start with the assumption that lottery-like stocks could be traded for gambling and non-gambling purposes. Therefore the total trading volume of lottery-like stocks ($Volume_{Lot}$) in country i in month t is the sum of gambling volume and non-gambling volume for this stock category:

$$Volume_{Lot,i,t} = NGV_{Lot,i,t} + GV_{i,t} \tag{5.1}$$

where $NGV_{Lot,i,t}$ and $GV_{i,t}$ are non-gambling and gambling volumes of lottery-like stocks in country i in month t, respectively.

Assuming, conservatively, that investors do not trade non-lottery stocks for gambling or speculation, then the total trading volume of non-lottery stocks ($Volume_{NonLot,i,t}$) is equal to the non-gambling volume of non-lottery stocks ($NGV_{NonLot,i,t}$):[38]

$$Volume_{NonLot,i,t} = NGV_{NonLot,i,t} \tag{5.2}$$

An implication of the ICAPM, and also an outcome of passive investing in the market capitalization weighted market portfolio, is that stocks in a given market are expected to have the same turnover (e.g., Tkac, 1999; Bhattacharya and Galpin, 2011).[39] This means that:

$$\frac{NGV_{Lot,i,t}}{Market\ Cap_{Lot,i,t}} \approx \frac{Volume_{NonLot,i,t}}{Market\ Cap_{NonLot,i,t}} \tag{5.3}$$

Thus equation (5.3) can be rewritten as:

$$\frac{Volume_{Lot,i,t} - GV_{i,t}}{Market\ Cap_{Lot,i,t}} - \frac{Volume_{NonLot,i,t}}{Market\ Cap_{NonLot,i,t}} \approx 0 \tag{5.4}$$

---

[38] To the extent that some gambling occurs in non-lottery stocks, the turnover differential between lottery and non-lottery stocks will understate the true amount of gambling and thus our measure can be viewed as a lower bound on the amount of gambling activity.

[39] Tkac (1999) shows that the implication of equal turnovers across stocks serves as a useful benchmark for isolating and characterizing abnormal trading activity. Departures from this benchmark other than those due to gambling are likely to result in our measure underestimating the amount of gambling because lottery-like stocks tend to be smaller and less liquid than most other stocks (e.g., Bali et al., 2011). For example, if active institutional investors that are not passive favour larger and more liquid stocks, they will tend to trade disproportionately in non-lottery-like stocks, adding to their turnover. Similarly, if inclusion in a major stock index increases trading volume, large and highly liquid stocks (predominantly non-lottery-like stocks) will again obtain an additional source of turnover, leading to underestimation of gambling in stock markets.

$$\therefore \frac{GV_{i,t}}{Market\ Cap_{Lot,i,t}} \approx \frac{Volume_{Lot,i,t}}{Market\ Cap_{Lot,i,t}} - \frac{Volume_{NonLot,i,t}}{Market\ Cap_{NonLot,i,t}} \tag{5.5}$$

$$\therefore GV_{i,t} \approx \left(\frac{Volume_{Lot,i,t}}{Market\ Cap_{Lot,i,t}} - \frac{Volume_{NonLot,i,t}}{Market\ Cap_{NonLot,i,t}}\right) Market\ Cap_{Lot,i,t} \tag{5.6}$$

The equation above (5.6) formalizes the intuition that, under a few assumptions set out above, the turnover differential between lottery-like and non-lottery-like stocks (the terms inside the large parentheses) reveals the amount of trading that can be attributed to gambling. Multiplying the excess turnover of lottery-like stocks by the market capitalization (as is done in (5.6)) converts the measure from units of turnover to units of dollar volume of gambling.

To render our measure of stock gambling more comparable with the gambling activity per capita in the "casino" market, we normalize the stock market gambling measure by the country's population, giving a measure of the US dollar value of stock market gambling per capita:

$$StockGambling_{i,t} \approx \left(\frac{Volume_{Lot,i,t}}{Market\ Cap_{Lot,i,t}} - \frac{Volume_{NonLot,i,t}}{Market\ Cap_{NonLot,i,t}}\right) \frac{Market\ Cap_{Lot,i,t}}{Population_{i,t}} \tag{5.7}$$

where $Population_{i,t}$ is the population in country i in month t. In case the turnover of lottery-like stocks is lower than that of non-lottery stocks, we set the value of stock gambling to zero. We annualize the measure by summing the monthly values within the year.

To operationalize our measure, we need to identify stocks with lottery-like characteristics. Drawing from the existing literature, we use five alternative proxies for lottery-like characteristics: (1) maximum daily return (MAX); (2) average of the highest five daily returns ($MAX5$); (3) idiosyncratic volatility ($IVOL$); (4) total volatility ($TVOL$); and (5) market beta ($BETA$)[40]. Hence, we have five different proxies for stock market gambling corresponding to five lottery-like features. Our results are qualitatively similar using these different measures of lottery-like stocks. In our baseline results, we therefore use a composite index of stock market gambling, which is the average of the five individual stock gambling proxies.

Our measures of stock market gambling have the advantage that they are easier to compute and require less granular data than the lottery-preference measures proposed by Kumar (2009). This simplification is what allows us to estimate our measures across a large number of countries and across many years. Our measures also differ from asset pricing measures of the price distortions created by gambling (e.g., the FMAX of Bali et al., 2017) insomuch as our measures are in units of dollar volume of gambling rather than price distortion, and measure gambling at the time it occurs, rather than measuring the ex-post effects of the asset price distortions created by *past* gambling.

b. Measure validation

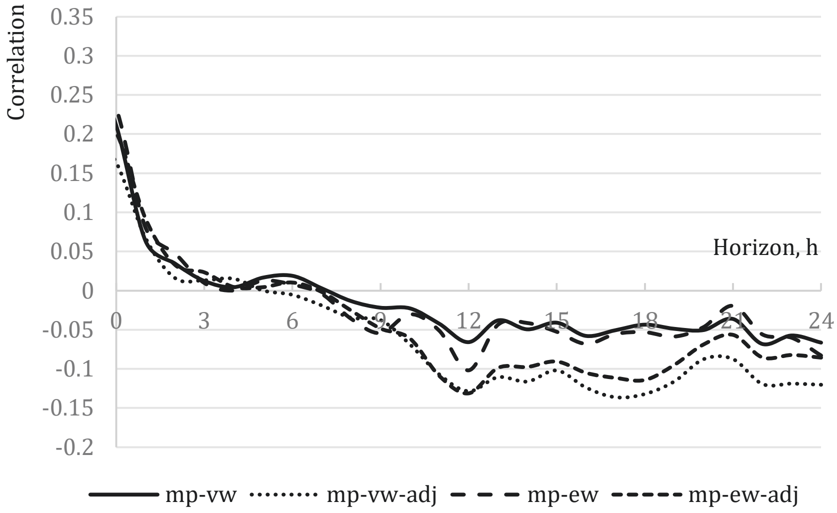
In this section, we undertake a number of validation tests of the measure of stock market gambling. For simplicity, we build the measure for the US sample in units of turnover and test whether our measure is a reasonable proxy for the gambling propensity on this stock market. In general, these tests all support the notion that the measure captures gambling in stock markets.

In the first validation test, we examine whether our measure explains the overpricing and subsequent low returns of lottery-like stocks (e.g., Bali et al., 2011). The results are shown in Figure 5.1 and Figure 5.2. Panel A of Figure 5.1 plots the correlation between the stock gambling in month t and the return of lottery-like stocks in month $t+h$. The return of lottery-like stocks is proxied by the inverse of the MAX-premium, which is obtained from the return spread of the bottom quintile and the top quintile portfolios by the MAX factor (Cheon and Lee, 2017). The correlation between gambling propensity and lottery-like stock returns is found to be positive if the horizon h is below 7 months, and negative for longer horizons. This evidence suggests

[40] These variables are defined in detail in Appendix 5.B.1.

that the US lottery-like stocks perform well when there is a high level of gambling in the US stock market, and these stocks subsequently underperform since month 7.

Panel A: Correlation between $StockGambling_t$ and $R_{t+h}^{LotteryStocks}$



Panel B: Correlation between $R_t^{LotteryStocks}$ and $StockGambling_{t+h}$

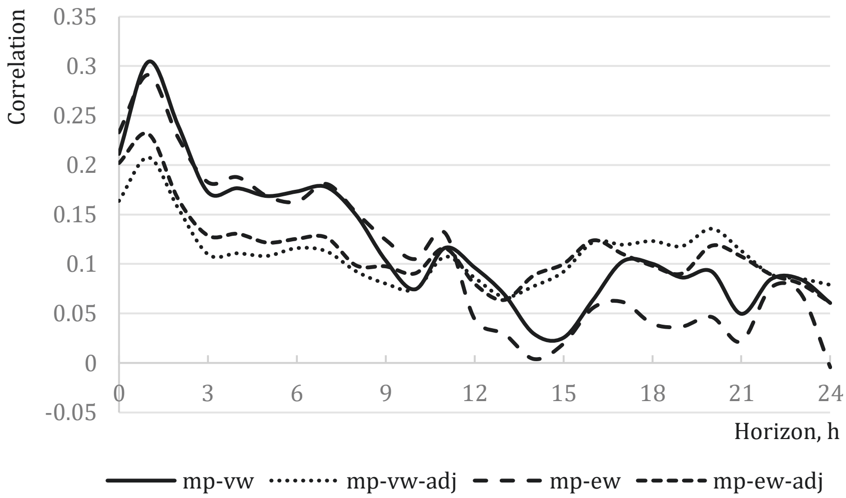


Figure 5.1. Correlation between stock gambling and return of lottery-like stocks.
This figure shows the correlation between the gambling propensity in the stock market ($StockGambling$) in month t and the return of the lottery-like stocks ($R^{LotteryStocks}$) in month $t + h$ in Panel A. Panel B reports the correlation between the return of the lottery-like stocks in month t and the gambling propensity in the stock market in month $t + h$. The return of lottery-like stocks is proxied by the inverse of the MAX-premium, which is obtained from the return spread of the bottom quintile and the top quintile portfolios by the MAX factor. Mp-vw and Mp-ew indicate that the MAX-premium is calculated using the value-weighted, and equally-weighted portfolios respectively. Mp-vw-adj and and Mp-ew-adj represent respectively the value-weighted and equally-weighted portfolios adjusted for risks. The sample includes stocks listed on NYSE, AMEX, and NASDAQ from 1963 to 2015.

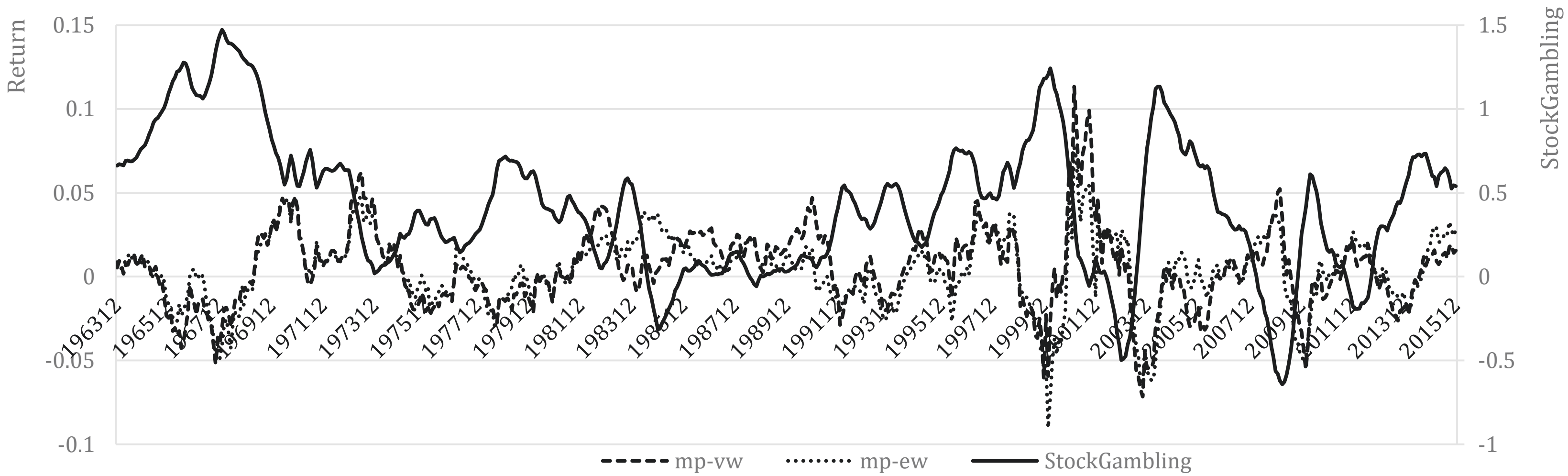


Figure 5.2. Relation between stock gambling and MAX-premium.
This figure plots the time-series of the monthly stock gambling and the monthly MAX premium. Mp-vw and Mp-ew indicate that the MAX-premium is calculated using the value-weighted, and equally-weighted portfolios respectively. The sample includes stocks listed on NYSE, AMEX, and NASDAQ from 1963 to 2015.

In Panel B, we do the reverse by reporting the correlation between the return on lottery-like stocks in month t and the stock gambling propensity in month $t + h$. The plot shows that high returns on the lottery-like stocks increase the level of gambling. By contrast, when the returns on the lottery-like stocks are low, investors gamble less as these stocks become no more attractive. This effect is weaker when the horizon increases, and ends after two years.

The time-series of stock gambling and the MAX premium are represented in Figure 5.2. We observe that the gambling measure is negatively correlated with the MAX premium. In other words, when gambling is low, the MAX premium is positive; when the gambling is high, lottery-like stocks perform well, the MAX premium is thus negative. All results are robust to using value-weighted portfolios, equally-weighted portfolios, and portfolios adjusted for risks.

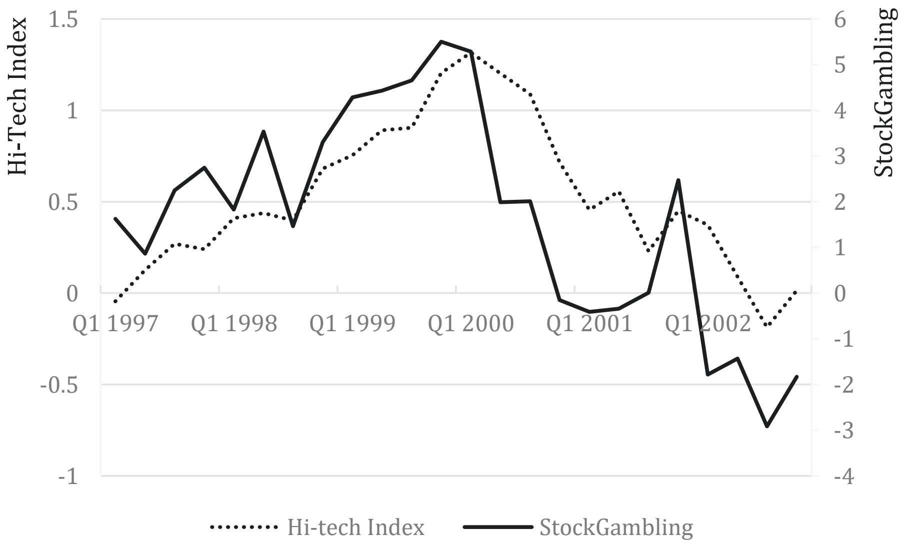


Figure 5.3. Relation between quarterly stock gambling and HiTech Index during the dot-com bubble (1997-2002).
This figure plots the time-series of the quarterly stock gambling ($StockGambling$) and Hi-Tech Index in the US market during the period from 1997 to 2002. Hi-Tech Index is calculated using the average of accumulated quarterly returns on stocks in the Hi-Tech industry. We calculate the quarterly stock gambling by summing the monthly values within the quarter. The sample includes stocks listed on NYSE, AMEX, and NASDAQ from 1997 to 2002.

As the dot-com bubble is a period of excessive speculation that occurred from 1997 to 2002 in Hi-Tech companies, we use this case study as the second validation test of our stock gambling measure. We calculate the Hi-Tech Index, which is the average of accumulated quarterly returns on stocks in the Hi-Tech industry. Figure 5.3 plots the stock gambling together with the Hi-Tech Index, showing that our gambling propensity measure rises sharply during the start of the dot com bubble, turns downward slightly ahead of the dot com bubble peak, and falls rapidly as the bubble collapses.

Finally, we observe the relation between changes in our stock gambling measure and ASVI, which is a measure of changes in people's attention toward lottery-related events, proposed by Chen, Kumar, and Zhang (2015). More specifically, $ASVI_t$ is the abnormal search volume intensity for the topic "lottery" in month t, which is calculated as the log difference between the search volume intensity for this topic from Google Trends in month t and month $t-1$. In the time-series during the period from 2004 to 2015 in Figure 5.4, the stock gambling is found to be inversely related with Google search intensity for the keyword "lottery", consistent with a time-series substitution effect between lotteries and stock market activity (Gao and Lin, 2014; Dorn et al., 2015).

$\Delta StockGambling/ASVI$

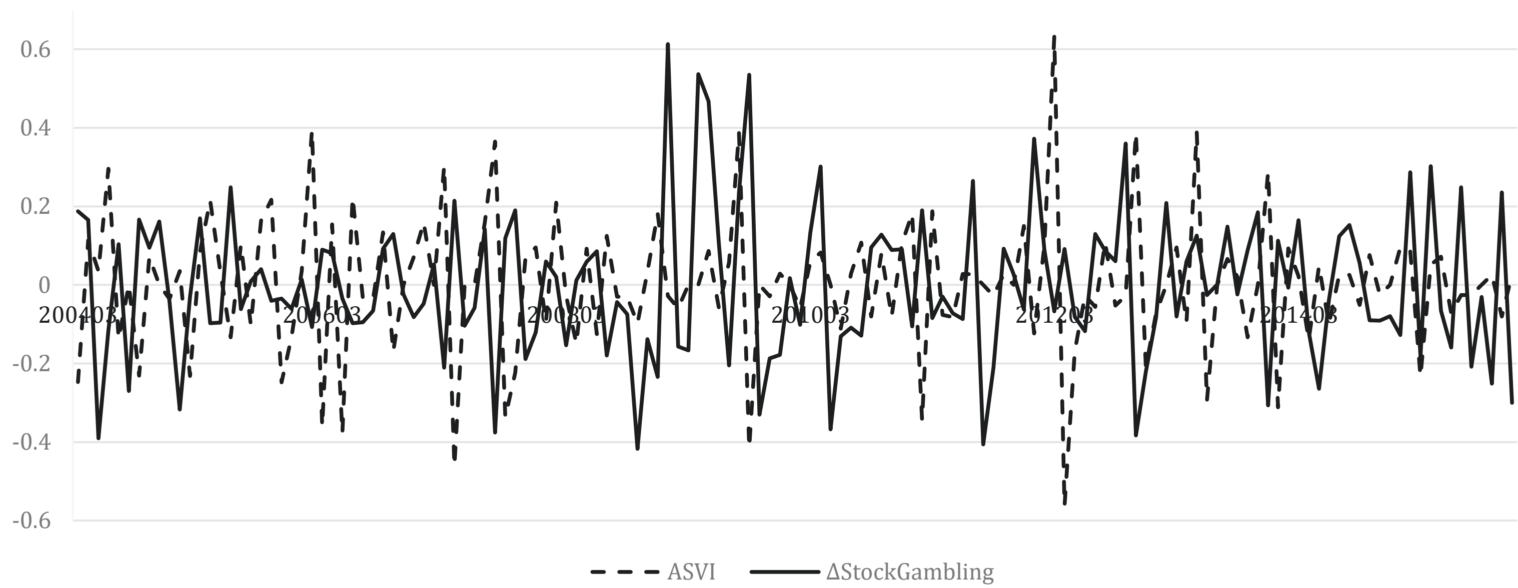


Figure 5.4. Relation between changes in stock gambling and ASVI during the period from 2004 to 2015.
This figure shows the time-series of the changes in the gambling propensity in the stock market ($\Delta StockGambling$) and the abnormal search volume intensity for the topic “lottery” (ASVI), which is calculated as the log difference between the search volume intensity for the topic “lottery” from Google Trends in the current month and the previous month. The sample includes stocks listed on NYSE, AMEX, and NASDAQ from 2004 to 2015.

5.4 How much gambling occurs in stock markets around the world?

Table 5.1 reports the estimated stock market gambling and casino gambling activity per country, as well as means for the country-level variables. Panel C reports averages for developed countries, emerging countries, and the pooled sample.

Strikingly, the estimated amount of stock market gambling is around 3.5 times that of all "traditional" forms of gambling combined. In the pooled sample, stock market gambling is estimated to amount to $2,905 per person per year, while the amount of money staked on all "traditional" forms of gambling is around $805. This overall tendency holds in both developed and emerging countries, although with larger dollar amounts wagered in developed countries ($4,952 in stock markets, $1,396 in casinos) than in emerging countries ($750 in stock markets, $183 in casinos). As discussed in the previous section, the estimates of stock market gambling are likely to provide a lower bound on the amount of gambling in stock markets. Thus, financial markets play a significant role as one of the venues in which gambling takes place.

The tendency for more money to be wagered in stock markets than in casinos and other traditional forms of gambling holds in most countries, but tends to be stronger in countries with relatively more developed and liquid stock markets, such as the US and Hong Kong, which have the highest per capita levels of gambling on stock markets. For example, in 17 of the 20 developed countries, the estimated amount of gambling in stock markets exceeds gambling in casinos and other traditional forms. The exceptions are Australia, New Zealand, and Ireland, which are particularly active in non-stock-market forms of gambling. In a further 13 of the 19 emerging countries, the estimated amount of gambling in stock markets exceeds gambling in casinos and other traditional forms.

Table 5.1. Gambling and other country-level variables by country.

This table reports country averages of the gambling activity measures, country-level variables, and gambling regulations. "Casino" gambling ($CasinoGambling$) is the amount of money wagered in casinos, lotteries, horse racing, sports betting, gaming machines, and online gambling, in USD per capita per annum. Stock market gambling ($StockGambling$) is an estimate of the dollar volume of gambling in stock markets, estimated using the approach specified in equations (1-7), in USD per capita per annum. $GDPPC$ is the GDP per capita in thousand USD; $Unemp$ is the unemployment rate in percent; IDV is the individualism dimension of culture measuring the degree of interdependence a society maintains among its members; UAI is the uncertainty avoidance dimension of culture; IVR is the indulgence versus restraint dimension of culture; and $Regulations$ specifies the gambling regulations in the country. The sample period is from 2003 to 2015.

Panel A: Developed markets

Country	$Casino$ $Gambling$	$Stock$ $Gambling$	$GDPPC$	$Unemp$	IDV	IVR	UAI	$Regulations$
Austria	1,362.53	1,465.32	45.24	5.13	55	63	70	Monopoly / Oligopoly
Australia	4,379.09	690.36	48.14	5.28	90	71	51	Competitive market
Belgium	319.39	1,831.80	42.39	7.88	75	57	94	Monopoly / Oligopoly
Canada	1,208.36	2,394.72	43.60	7.08	80	68	48	Competitive market
Germany	527.39	1,572.60	41.04	7.60	67	40	65	Monopoly / Oligopoly
Denmark	773.00	3,276.36	55.68	5.76	74	70	23	Monopoly / Oligopoly
Spain	876.22	1,666.20	29.06	16.61	51	44	86	Monopoly / Oligopoly
Finland	1,023.96	9,689.52	45.06	8.35	63	57	59	Monopoly / Oligopoly
France	599.86	2,967.96	39.31	9.24	71	48	86	Monopoly / Oligopoly
United Kingdom	1,302.15	3,739.92	42.40	6.20	89	69	35	Monopoly / Oligopoly
Hong Kong	2,730.21	16,571.88	32.46	4.52	25	17	29	Ban
Ireland	1,921.23	536.40	52.99	9.05	70	65	35	Monopoly / Oligopoly
Italy	1,056.66	2,941.32	34.55	8.91	76	30	75	Monopoly / Oligopoly
Japan	1,316.63	8,702.52	39.61	4.34	46	42	92	Monopoly / Oligopoly
Netherlands	384.78	6,322.44	47.93	4.85	80	68	53	Monopoly / Oligopoly
Norway	947.24	7,965.12	82.75	3.50	69	55	50	Monopoly / Oligopoly
New Zealand	1,573.42	535.44	33.19	4.99	79	75	49	Monopoly / Oligopoly
Sweden	784.76	4,209.12	51.01	7.34	71	78	29	Monopoly / Oligopoly
Singapore	2,789.20	6,628.44	42.47	3.85	20	46	8	Monopoly / Oligopoly

United States	2,044.53	15,335.28	48.41	6.65	91	68	46	Competitive market

Panel B: Emerging markets

Country	*Casino Gambling*	*Stock Gambling*	*GDPPC*	*Unemp*	*INV*	*IVR*	*UAI*	*Regulations*
Argentina	352.87	9.60	9.04	9.12	46	62	86	Monopoly / Oligopoly
Brazil	69.26	1,852.80	8.59	9.67	38	59	76	Monopoly / Oligopoly
Chile	73.63	159.60	11.31	7.53	23	68	86	Monopoly / Oligopoly
China	19.59	1,591.20	4.31	4.30	20	24	30	Ban
Columbia	68.22	22.44	6.51	10.61	13	83	80	Monopoly / Oligopoly
Czech	411.95	121.80	17.65	6.64	58	29	74	Competitive market
Greece	908.11	734.88	24.24	15.26	35	50	100	Monopoly / Oligopoly
India	4.36	137.16	1.33	3.68	48	26	40	Monopoly / Oligopoly
South Korea	389.38	5,461.92	21.75	3.46	18	29	85	Monopoly / Oligopoly
Mexico	60.12	108.36	9.48	4.72	30	97	82	Competitive market
Malaysia	214.11	654.60	8.18	3.28	26	57	36	Monopoly / Oligopoly
Peru	50.09	16.44	4.53	5.72	16	46	87	Monopoly / Oligopoly
Philippines	47.39	26.04	1.97	7.88	32	42	44	Monopoly / Oligopoly
Poland	82.74	149.04	11.27	11.62	60	29	93	Monopoly / Oligopoly
Portugal	302.28	646.92	20.95	10.32	27	33	99	Monopoly / Oligopoly
Russia	46.10	275.88	12.04	6.22	39	20	95	Monopoly / Oligopoly
Thailand	30.80	1,051.08	5.21	0.93	20	45	64	Ban
Turkey	63.47	924.60	9.63	9.87	37	49	85	Ban
South Africa	284.34	298.80	6.13	24.26	65	63	49	Monopoly / Oligopoly

Panel C: Averages

Countries	*Casino Gambling*	*Stock Gambling*	*GDPPC*	*Unemp*	*INV*	*IVR*	*UAI*
Developed	1,396.03	4,952.14	44.86	6.86	67	57	54
Emerging	183.10	749.64	10.22	8.16	34	48	73
All	805.11	2,904.77	27.98	7.49	51	52	63

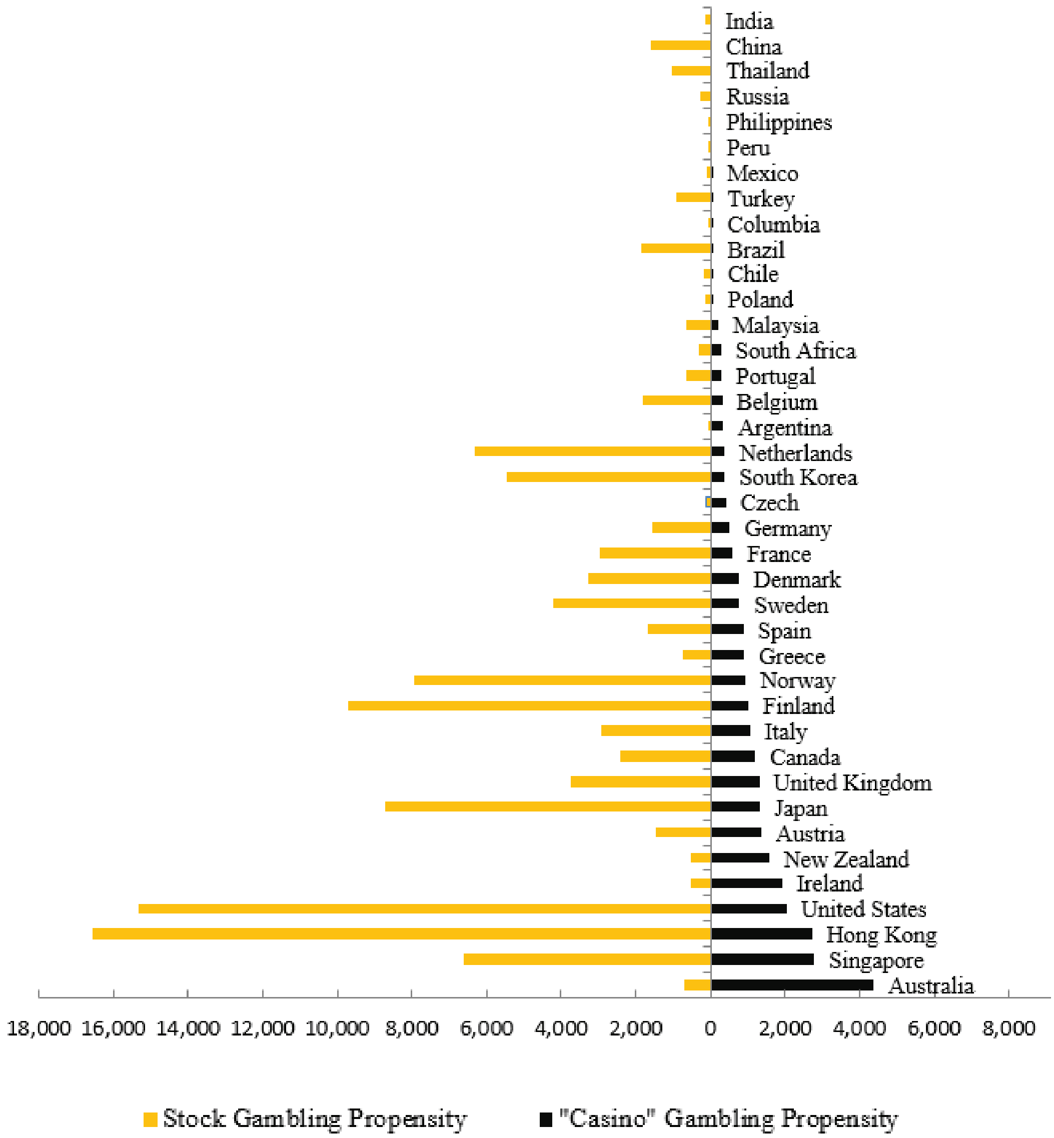


Figure 5.5. Gambling activity by country

This figure shows the country-level averages of "casino" gambling and stock market gambling. Both measures are in USD per annum, per capita of population. "Casino" gambling is the amount of money wagered in casinos, lotteries, horse racing, sports betting, gaming machines, and online gambling. Stock market gambling is an estimate of the dollar volume of gambling in stock markets, estimated from the turnover differential of lottery-like and non-lottery stocks using the approach specified in equations (1-7). The gambling activity measures are first calculated separately for each country in each year, then for each country, we take the simple average for the years 2003 to 2015.

Figure 5.5 graphically highlights the large differences across countries in their gambling activity, both in their stock market(s) (light colored bars on the left hand side) and in casinos and other traditional forms of gambling (dark colored bars on the right hand side). The countries are sorted by the amount of casino gambling. The overall pattern observed in Figure 5.5 is a tendency for more developed countries to have higher levels of gambling on a dollars per capita basis. This is likely to be driven by a country wealth effect, which we formally test in the next section. Consequently, there appears to be a positive association between the two forms of gambling in the cross-section of countries.

A few noteworthy outliers and examples stand out in Figure 5.5. Australia, Singapore, and Hong Kong are the three countries with the highest levels of casino gambling per capita during the period from 2003 to 2015. As shown in Table 5.1, the average Australian spends about $4,400 dollars per year on traditional gambling. Indeed, as a relatively free and competitive gambling market, Australia has a flourishing gambling industry with diverse gambling products including lotto and scratch cards, electronic gaming machines, table games, horse and sports betting, casinos, keno, football pools, and so on. According to GBGC, around 65% of the Australian population considers gambling as a recreational activity.

The gambling market in Singapore is not as highly liberated as that in Australia. For example, lottery and horse racing markets in Singapore are monopolies, while gaming machines are permitted only in a number of private clubs. Casinos have been allowed since 2004 after a 40-year ban period, and introduced to gamblers from 2010. As a result, the per capita gambling spend in Singapore has increased from less than $1,200 per year during the pre-2010 period to more than $4,600 in the post-2010 period.

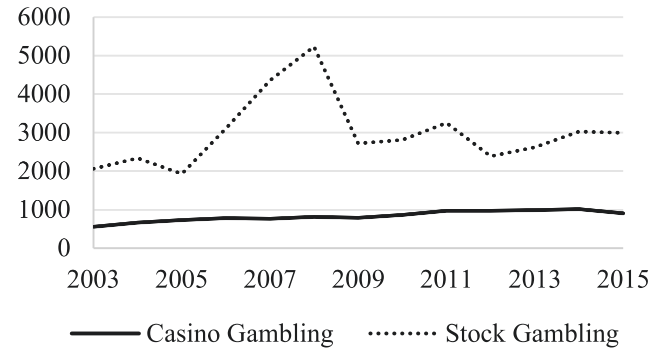
The regulations in Hong Kong are even stricter than those in Singapore, with casinos and gaming machines being prohibited. Lotteries, horse racing, and sports betting are permitted, but these activities are operated as a monopoly. Despite the unfavorable restrictions, the annual expenditure on non-stock-market gambling in Hong Kong exceeds $2,730 per capita.

Other countries with similar levels of regulation and constraints (e.g., that also have regulated monopoly gambling markets) have considerably lower levels of gambling. For example, India, Russia, Thailand, and the Philippines have the lowest annual per capita gambling activity and also regulated monopolies. These results suggest that factors other than regulations drive variation in the gambling activity across countries.

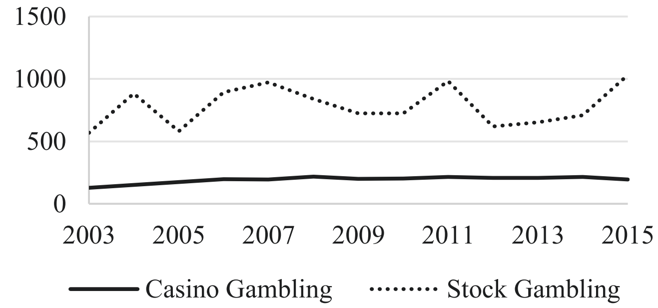
Turning to gambling in the stock market, Hong Kong and the US have the highest estimated gambling levels, with estimates above $15,000 per year per capita. The three other countries in the top-five group include Japan, Finland, and Norway, with estimates varying from $8,000 to $10,000. These high levels of stock market gambling in countries with large and liquid stock markets contrast with countries such as Argentina, Peru, Columbia, and the Philippines, where gambling on stock markets is estimated to be less than $100 per person per year.

Figure 5.6 shows the estimated levels of the two forms of gambling from 2003 to 2015. Panel A reports the pooled sample time series, while Panels B and C separate countries into emerging and developed countries, respectively. The time series show a modest tendency for both forms of gambling to increase over time. In all years stock market gambling activity is estimated to exceed that of non-stock-market gambling. Developed, but not emerging, markets show a tendency for stock market gambling to increase during the financial crisis, peaking in 2008.

Panel A: All countries



Panel B: Emerging markets



Panel C: Developed markets

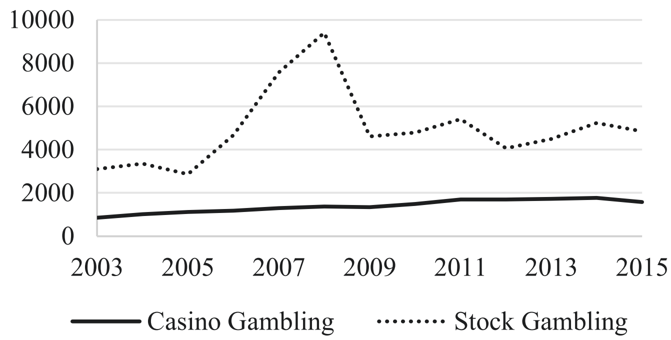


Figure 5.6. Gambling activity over time

This figure shows the annual averages of "casino" gambling and stock market gambling. Both measures are in USD per annum, per capita of population. "Casino" gambling is the amount of money wagered in casinos, lotteries, horse racing, sports betting, gaming machines, and online gambling. Stock market gambling is an estimate of the dollar volume of gambling in stock markets, estimated from the turnover differential of lottery-like and non-lottery stocks using the approach specified in equations (1-7). The gambling activity measures are first calculated separately for each country in each year, then for each year, we take the simple average for the corresponding countries (all 39 countries in Panel A, Emerging markets in Panel B, and Developed markets in Panel C).

5.5 What drives the level of gambling in stock markets and casinos?

This section examines the major drivers of gambling activity and the drivers of the cross-country variation in gambling levels. We consider the role played by economic conditions such as wealth and the unemployment rate, as well as the role of culture in explaining gambling tendencies. Section 5.5.1 presents country-level variables and hypotheses and Section 5.5.2 details our estimation of the determinants of gambling.

5.5.1 Country-level variables and hypotheses

Mikesell (1994) shows that economic activity influences lottery participation. For example, lottery sales are positively related to personal income in what is deemed a wealth effect. However, wealth effects can be complex because an increase in the unemployment rate also tends to correlate with increased lottery sales. Along similar lines, Brenner and Brenner (1990) document an increase in gambling during the Great Depression of the 1930s. Similarly, Kumar (2009) finds that individuals are likely to gamble more through lottery ticket purchases and in the stock market when the local economic conditions are not very favorable.

We therefore test the extent to which wealth effects and the unemployment rate explain the cross-country variation in gambling activity. We use the gross domestic product (GDP) per capita ($GDPPC$) as a proxy for wealth or average income of a resident in a given country. We expect both wealth and unemployment to be positively associated with gambling in both forms (casino and stock market).

We also investigate the role that culture plays in explaining the cross-country variation in gambling. We use three dimensions of national culture proposed by Hosftede (1997) including Individualism Versus Collectivism (IDV), Uncertainty Avoidance Index (UAI), and Indulgence Versus Restraint (IVR) since these three dimensions are likely to be related to gambling.

Individualism Versus Collectivism (IDV) measures the extent to which ties between individuals are loose and everyone is expected to look after him/herself and his/her immediate family. A high IDV score indicates an individualistic culture, whereas a low IDV score indicates a collectivist culture in which people are integrated into strong, cohesive in-groups that are bound by loyalty. Individualism is associated with higher overconfidence and over-optimism, which might result in a higher risk tolerance level and more gambling (e.g., Walker et al., 2007). Hence, we expect countries with high IDV to experience a higher gambling propensity than those with low IDV level.

The *Uncertainty Avoidance Index* (UAI) measures the extent to which the members of a society feel uncomfortable with uncertainty and ambiguity. In general, people in low uncertainty avoidance cultures are less skeptical about the unknown, and more accepting of unpredictable events. Therefore, they are expected to be more receptive to gambling, leading to an expected negative relation between UAI and the level of gambling.

Meanwhile, *Indulgence Versus Restraint* (IVR) measures the extent to which individuals try to control their desires and impulses based on the way they were raised. Societies with weaker controls over their desires are considered the indulgent countries, and they tend to allow free gratification of basic and natural human desires related to enjoying life and having fun. By contrast, countries with a high degree of restraint tend to suppress gratification of needs, regulated by means of strict social norms. Such countries are likely to avoid gambling. We therefore expect a positive association between gambling and IVR.

In addition, we also control for region fixed effects in case there are regional tendencies not captured by economic conditions or culture. Specifically, we include dummy variables for Asia, Europe, North America, South America, and Oceania.

5.5.2 Estimating the determinants of gambling

We test the cross-country drivers of gambling activity by estimating the following panel regression model with region and year fixed effects:

$$\begin{aligned} Y_{i,t} = \alpha_t + \beta_1 GDPPC_{i,t} + \beta_2 Unemp_{i,t} + \delta_1 IDV_i + \delta_2 UAI_i + \delta_3 IVR_i \\ + \theta_1 Asia_i + \theta_2 EU_i + \theta_3 NAmer_i + \theta_4 SAmer_i \\ + \theta_5 Oceania_i + \varepsilon_{i,t} \end{aligned} \tag{5.8}$$

where the dependent variable, $Y_{i,t}$, is either the estimated stock market gambling activity ($StockGambling_{i,t}$) or the "casino" gambling activity ($CasinoGambling_{i,t}$).

The definitions of other variables are provided in Appendix 5.C. All dependent and independent variables, except regional dummies, are standardized to have a mean of zero and a standard deviation of one so that the coefficient estimates can be directly compared within and across regression specifications.

Tables 5.2 and 5.3 report the results of the regressions testing the determinants of "casino" gambling activity (Table 5.2) and stock market gambling activity (Table 5.3). Models 1 and 2 in each table report regressions of gambling activity on the economic factors and culture variables separately, while Model 3 includes the full set of gambling determinants.

Interestingly, both forms of gambling have remarkably similar drivers. First, there is a very strong wealth effect for both forms of gambling: GDP per capita has a positive and significant (at 1% level) association with both forms of gambling, indicating that wealthier countries tend to have higher levels of gambling, after controlling for other factors. A one standard deviation increase in GDP per capital is associated with a 0.71 standard deviation increase in casino gambling (Table 5.2, Model 3) and a 0.68 standard deviation increase in stock market gambling (Table 5.3, Model 3). In dollar terms, an additional \$100 of GDP per capita is associated with an additional \$3.54 of casino gambling and an additional \$16.01 of stock market gambling.

Second, consistent with the existing literature, the unemployment rate is significantly positively related to the level of "casino" gambling activity, indicating that people are attracted more toward traditional gambling activities during bad economic times. A one standard deviation increase in the unemployment rate (an increase of 4.82 percentage points of unemployment) is associated with a 0.20 standard deviation increase in casino gambling (Table 5.2, Model 3). In dollar terms, an additional one percentage point of unemployment is associated with an additional \$42.36 per person per year wagered in "casino" gambling. The relation between unemployment and stock market gambling is weaker and although the coefficients in Table 5.3 are positive, they are not statistically significant. A possible explanation is that because stock market gambling is a more sophisticated, higher stakes form of gambling, unemployed persons are more likely to turn to lotteries, gaming machines, sports betting and so on rather than to stock markets.

Table 5.2. Determinants of "casino" gambling activity.
This table reports the results from panel regressions in which the dependent variable is "casino" gambling ($CasinoGambling$) and the units of observation are country-years. "Casino" gambling is the amount of money wagered in casinos, lotteries, horse racing, sports betting, gaming machines, and online gambling, in USD per capita per annum. $GDPPC$ is the GDP per capita in thousand USD; $Unemp$ is the unemployment rate in percent; IDV is the individualism dimension of culture measuring the degree of interdependence a society maintains among its members; UAI is the uncertainty avoidance dimension of culture; and IVR is the indulgence versus restraint dimension of culture. $Asia$, EU, $NAmer$, $SAmer$, and $Oceania$ are dummy variables that take the value of one if the country is located in Asia, Europe, North America, South America, or Oceania, respectively. All dependent and independent variables, except region dummies, are standardized to have a mean of zero and standard deviation of one. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes stocks listed on major exchanges across 39 countries. The sample period is from 2003 to 2015.

Variable	(1)	(2)	(3)
$GDPPC$	0.71		0.71
	(11.48)***		(11.78)***
$Unemp$	0.16		0.20
	(13.36)***		(16.40)***
IDV		0.16	-0.07
		(5.43)***	(-1.80)*
UAI		-0.32	-0.32
		(-5.38)***	(-6.50)***
IVR		-0.13	-0.24
		(-6.88)***	(-9.16)***
$Asia$	0.94	0.64	0.65
	(14.21)***	(14.01)***	(9.03)***
EU	-0.19	0.63	0.04
	(-1.72)	(17.28)***	(0.54)
$NAmer$	0.41	1.05	0.85
	(3.66)***	(32.27)***	(18.77)***
$Oceania$	2.02	2.61	2.41
	(20.82)***	(19.88)***	(31.86)***
$SAmer$	0.25	0.53	0.72
	(3.17)***	(25.83)***	(14.88)***
R^2 (%)	61.65	44.22	67.70
Fixed Effects	Year	Year	Year

Table 5.3. Determinants of stock market gambling activity.
This table reports the results from panel regressions in which the dependent variable is stock market gambling ($StockGambling$) and the units of observation are country-years. Stock market gambling is an estimate of the dollar volume of gambling in stock markets, estimated using the approach specified in equations (1-7), in USD per capita per annum. $GDPPC$ is the GDP per capita in thousand USD; $Unemp$ is the unemployment rate in percent; IDV is the individualism dimension of culture measuring the degree of interdependence a society maintains among its members; UAI is the uncertainty avoidance dimension of culture; and IVR is the indulgence versus restraint dimension of culture. $Asia$, EU, $NAmer$, $SAmer$, and $Oceania$ are dummy variables that take the value of one if the country is located in Asia, Europe, North America, South America, or Oceania, respectively. All dependent and independent variables, except region dummies, are standardized to have a mean of zero and standard deviation of one. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes stocks listed on major exchanges across 39 countries. The sample period is from 2003 to 2015.

Variable	(1)	(2)	(3)
$GDPPC$	0.69		0.68
	(10.11)***		(9.60)***
$Unemp$	0.03		0.04
	(0.95)		(1.64)
IDV		0.36	0.12
		(8.49)***	(2.95)**
UAI		-0.17	-0.14
		(-4.42)***	(-4.56)***
IVR		-0.16	-0.27
		(-3.07)***	(-5.18)***
$Asia$	0.73	1.25	0.55
	(5.91)***	(10.87)***	(4.50)***
EU	-0.52	0.58	-0.54
	(-3.88)***	(4.98)***	(-3.43)***
$NAmer$	0.34	1.37	0.59
	(2.27)**	(15.71)***	(5.61)***
$Oceania$	-1.07	-0.13	-0.95
	(-4.90)***	(-2.07)*	(-4.99)***
$SAmer$	-0.01	0.80	0.41
	(-0.01)	(10.1)***	(2.79)**
R^2 (%)	42.99	25.53	47.35
Fixed Effects	Year	Year	Year

National culture is also a driver of both forms of gambling. The strongest effect is for the uncertainty avoidance dimension of culture. Countries with a higher tendency to avoid uncertainty tend to gamble less in casinos and in stock markets, after controlling for other factors. This effect is economically meaningful. A one standard deviation increase in uncertainty avoidance is associated with a 0.32 standard deviation decrease in casino gambling (Table 5.2, Model 3) and a 0.14 standard deviation decrease in stock market gambling (Table 5.3, Model 3). In dollar terms, this effect translates to \$326.87 less casino gambling and \$675.63 less stock market gambling per person per year. Furthermore, more individualistic cultures have a higher propensity to engage in both forms of gambling, although the association with casino gambling disappears after controlling for economic factors (Table 5.2, Model 3). Finally, a greater cultural propensity for indulgence rather than restraint is positively associated with gambling activity in univariate tests (not tabulated) but has the opposite association after controlling for the other dimensions of culture, economic conditions, and region fixed effects, suggesting the unconditional positive link between indulgence and gambling is subsumed by other factors.

Regional fixed effects are also important determinants of both forms of gambling, even after controlling for culture and economic conditions. For example, Asia and North and South America have an increased tendency to gamble in both stock markets and non-stock-market forms of gambling, all else equal. Oceania has an increased tendency to gamble in non-stock-market forms, but a lower tendency to gamble on stock markets.

In conclusion, most country-level factors affect both forms of gambling in a similar manner. For example, wealth effects, culture effects, and region effects similarly drive cross-country variation in both forms of gambling. This commonality in the drivers of both forms of gambling supports the notion that our measures of stock market gambling measure a similar type of activity to gambling in casinos, lotteries and so forth, just in a different venue. These results suggest a close link between the two forms of gambling, prompting examination of the extent to which stock markets serve as alternatives or substitutes for casinos. This question is examined in the next section.

5.6 Are stock markets substitutes for casinos?

Given the previous section's findings of a remarkable similarity in the drivers of both forms of gambling ("casino" gambling and gambling on stock markets), we expect to find an unconditional positive relation between the two forms of gambling. If the positive relation between the two types of gambling arises due to common drivers of both forms of gambling, we expect that controlling for the common drivers, the positive relation will diminish, possibly even disappear, if we are able to capture all of the common drivers.

We test these conjectures empirically using OLS panel regressions with country-year observations as follows:

$$StockGambling_{i,t} = \alpha + \beta_1 CasinoGambling_{j,i} + \sum \gamma_j Controls_{i,t} + \varepsilon_{i,t} \quad (5.9)$$

where $StockGambling_{i,t}$ and $CasinoGambling_{i,t}$ are our measures of gambling in the stock market and in casinos plus other traditional forms for country i in year t. In $Controls_{i,t}$ we include the common drivers of gambling identified in the previous section. In these regressions, all dependent and independent variables, except regional dummies, are standardized to have a mean of zero and standard deviation of one.

Table 5.4 reports the results of the OLS panel regressions. Model 1 shows that consistent with our conjecture, there is a strong unconditional positive relation between gambling on stock markets and "casino" gambling. This result is to be expected if the two forms of gambling are similar in nature and thus share common drivers. It is consistent with Kumar's (2009) finding that greater lottery expenditure is associated with greater investments in lottery-like stocks. Unconditionally, a one standard deviation increase in casino gambling is associated with a 0.40 standard deviation increase in stock market gambling.

Models 2 to 4 in Table 5.4 add groups of common drivers as control variables. As conjectured, the positive relation between gambling on stock markets and "casino" gambling diminishes when we control for the common drivers. In Model 4, the relation between gambling on stock markets and "casino" gambling is half the strength of the unconditional relation in Model 1, indicating that much of the positive association between the two forms of gambling arises from their common drivers. It is possible that if we were able to identify further common drivers and control for them, the conditional relation between the two forms of gambling would further diminish.

The evidence thus far suggests that stock markets are indeed venues in which gambling activity, similar to that seen in casinos and lotteries, is realized. The real test of whether there are substitution effects between the two forms of gambling, however, is if we exogenously shock or shut down one of the venues for gambling, do we see a spillover into the other venue?

Table 5.4. Relation between "casino" gambling and stock market gambling (OLS regressions)

This table reports the results from panel regressions in which the dependent variable is stock market gambling ($StockGambling$) and the key independent variable is "casino" gambling ($CasinoGambling$). Stock market gambling is an estimate of the dollar volume of gambling in stock markets, estimated using the approach specified in equations (1-7), in USD per capita per annum. "Casino" gambling is the amount of money wagered in casinos, lotteries, horse racing, sports betting, gaming machines, and online gambling, in USD per capita per annum. $GDPPC$ is the GDP per capita in thousand USD; $Unemp$ is the unemployment rate in percent; IDV is the individualism dimension of culture measuring the degree of interdependence a society maintains among its members; UAI is the uncertainty avoidance dimension of culture; and IVR is the indulgence versus restraint dimension of culture. $Asia$, EU, $NAmer$, $SAmer$, and $Oceania$ are dummy variables that take the value of one if the country is located in Asia, Europe, North America, South America, or Oceania, respectively. All dependent and independent variables, except region dummies, are standardized to have a mean of zero and standard deviation of one. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes stocks listed on major exchanges across 39 countries. The sample period is from 2003 to 2015.

Variable	(1)	(2)	(3)	(4)
$CasinoGambling$	0.40	0.51	0.27	0.20
	(11.85)***	(10.85)***	(4.81)***	(3.28)***
$GDPPC$			0.50	0.54
			(6.01)***	(6.09)***
$Unemp$			-0.02	-0.01
			(-0.67)	(-0.02)
IDV		0.27		0.13
		(7.26)***		(3.79)***
UAI		-0.01		-0.08
		(-0.18)		(-2.18)*
IVR		-0.09		-0.23
		(-2.21)**		(-4.86)***
$Asia$		0.93	0.48	0.43
		(8.21)***	(4.55)***	(3.35)***
EU		0.26	-0.47	-0.55
		(2.27)**	(-3.83)***	(-3.67)***
$NAmer$		0.83	0.23	0.42
		(12.39)***	(1.73)	(4.70)***
$Oceania$		-1.46	-1.61	-1.42
		(-12.03)***	(-7.81)***	(-8.27)***
$SAmer$		0.54	-0.07	0.26
		(7.02)***	(-0.52)	(1.88)*
R^2 (%)	19.20	39.95	45.74	48.58
Fixed Effects	Year	Year	Year	Year

We perform these tests for substitution effects using gambling regulations and restrictions in a 2SLS instrumental variables (IV) framework. In each country and each year, we have data on the gambling regulations from the GBGC dataset. We use these regulations as an instrumental variable to isolate variation in casino gambling that is attributable to policy decisions within a country.

To construct variables that reflect the level of regulations/restrictions of gambling activities in a given country, we first score the level of regulation in each of the major gambling segments in the country (casinos, lotteries, sports betting/horse racing, and gaming machines). The regulation/restriction score for each gambling segment varies from 0 to 2, with 0 indicating the gambling segment is completely free and competitive, 1 indicating that the gambling services in the segment are provided by a limited number of operators (e.g., monopoly/oligopoly), and 2 indicating that gambling of the particular type is completely prohibited. We then take the average value of the four aforementioned scores for the four gambling segments within the country, and denote it as the variable $Regulations$. A higher value of $Regulations$ indicates that the gambling market is more heavily regulated/restricted in the country. In the last step, we create two dummy variables: a "few-operators" dummy variable ($FewOp$) if $0 < Regulations \leq 1$ and a "banned" dummy variable (Ban) if $Regulations > 1$.

We estimate the following 2SLS IV models with year fixed effects:

$$CasinoGambling_{i,t} = \alpha_t + \beta_1 Ban_i + \beta_2 FewOp_i + \sum \gamma_j Controls_{i,t} + \varepsilon_{i,t} \quad (5.10)$$

$$StockGambling_{i,t} = \alpha_t + \beta_1 \widehat{CasinoGambling}_{i,t} + \sum \gamma_j Controls_{i,t} + \varepsilon_{i,t} \quad (5.11)$$

The control variables include $GDPPC$ (GDP per capita), $Unemp$ (unemployment rate), the culture dimensions (IDV, UAI, and IVR), and region fixed effects. All dependent and independent variables, except regional dummies, are standardized to have a mean of zero and standard deviation of one.

Table 5.5 reports the results of the 2SLS models measuring the extent of substitution/spillover effects between the two forms of gambling. The first-stage regression results are reported in the first column of Table 5.5. The amount of gambling in casinos, lotteries, gaming machines, and sports betting is negatively associated with the extent of restrictions on gambling within the country, consistent with the expected effect of the instrumental variable. People gamble less when their ability to gamble is restricted or prohibited through country policies. The F-statistics of the two instrumental variables are between 57.50 and 591.15, suggesting that the instruments are strong.

The second column in Table 5.5 shows the results from the second stage of the 2SLS model. The results show that a decline in "casino" gambling due to restrictions and prohibitions is associated with an increase in stock market gambling: the coefficient on $\widehat{CasinoGambling}_{i,t}$ in the second column of Table 5.5 is negative and statistically significant. Thus, the data support the hypothesis of substitution effects between the two forms of gambling: when policymakers restrict the ability to gamble in casinos/lotteries/sports betting/gaming machines, gambling activity spills over into the stock market. While there is some evidence in existing studies of substitution effects consistent with our findings, the existing evidence is limited to a single country, time-series evidence, and only lotteries (e.g., Gao and Lin, 2014; Dorn et al., 2015). Our analysis shows that the substitution between stock market and non-stock-market forms of gambling is a more general phenomenon.

The magnitude of the substitution effect between the two forms of gambling is economically meaningful. For example, a one standard deviation decrease in casino gambling due to restrictions and prohibitions is associated with a 0.24 standard deviation increase in stock market gambling. In dollar terms, the spillover effects are slightly more than one-is-to-one. An exogenous $100 decrease in casino gambling is associated with a $113 increase in stock market gambling.

There are at least three potential reasons why the spillover effect is estimated to be *more* than a one-is-to-one effect. First, an additional exogenous dollar of activity on the stock market is likely to attract others to the stock market due to the positive externalities of stock market liquidity. Second, because gambling on the stock market plausibly has a lower "house take" or expected loss due to the overall positive equity premium, it is likely that gambling on the stock market involves losing wealth at a slower rate than gambling in non-stock-market venues and thus more re-betting of one's capital takes place in the stock market. Third, the minimum investment amounts in stock markets (which vary from market to market) may induce an increase in the wagered amounts for small stakes gamblers (e.g., an individual that may have spent $50 on lottery tickets, may wager a minimum of say $500 on the stock market if driven to switch from lotteries to the stock market).

Table 5.5. Substitution between "casino" gambling and stock market gambling.
This table reports the results from 2SLS IV estimation of the relation between "casino" gambling and stock market gambling when casino gambling is constrained through regulation or prohibition. Model (1) reports the result of the first-stage regression in which regulation dummy variables are used as instrumental variables. Model (2) reports the results for the second-stage regression that estimates the impact on stock market gambling from restricting casino gambling. $\widehat{CasinoGambling}$ is the fitted value (from the first stage) of "casino" gambling ($CasinoGambling$), which is the amount of money wagered in casinos, lotteries, horse racing, sports betting, gaming machines, and online gambling, in USD per capita per annum. Stock market gambling ($StockGambling$) is an estimate of the dollar volume of gambling in stock markets, estimated using the approach specified in equations (1-7), in USD per capita per annum. $FewOp$ and Ban are regulation dummy variables that equal one if the gambling market is a monopoly/oligopoly, and if gambling activities are prohibited, respectively. $GDPPC$ is the GDP per capita in thousand USD; $Unemp$ is the unemployment rate in percent; IDV is the individualism dimension of culture measuring the degree of interdependence a society maintains among its members; UAI is the uncertainty avoidance dimension of culture; and IVR is the indulgence versus restraint dimension of culture. $Asia$, EU, $NAmer$, $SAmer$, and $Oceania$ are dummy variables that take the value of one if the country is located in Asia, Europe, North America, South America, or Oceania, respectively. All dependent and independent variables, except region dummies, are standardized to have a mean of zero and standard deviation of one. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes stocks listed on major exchanges across 39 countries. The sample period is from 2003 to 2015.

Variable	First Stage	Second Stage
	(1) $CasinoGambling$	(2) $StockGambling$
$\widehat{CasinoGambling}$		-0.24 (-3.75)***
Ban	-0.80 (-7.58)***	
$FewOp$	-0.94 (-19.78)***	
$GDPPC$	0.73 (13.01)***	0.85 (7.98)***
$Unemp$	0.24 (20.10)***	0.09 (5.07)***
IDV	-0.06 (-1.52)	0.10 (2.56)**
UAI	-0.30 (-5.48)***	-0.22 (-5.28)***
IVR	-0.16 (-4.95)***	-0.33 (-6.28)***
R^2 (%)	70.87	47.53
Fixed Effects	Year	Year

In sum, we find strong evidence that stock markets are an economically meaningful alternative to casinos as a venue in which gambling activity is realized. The two forms of gambling have an unconditional positive relation largely due to common drivers. An exogenous shock to casino gambling due to country-level restrictions and prohibitions leads to significant spillovers of gambling activity into the country's stock markets.

5.7 How does gambling impact stock markets?

Given the above evidence of economically meaningful volumes of gambling on stock markets, it is natural to explore how this gambling affects stock markets, in particular how it impacts liquidity and the informational efficiency of prices. Given that gamblers are likely to be relatively or completely uninformed traders, we hypothesize that gambling on stock markets increases liquidity. This hypothesis follows from market microstructure theory, which demonstrates that an increase in uninformed trading makes stock markets more liquid, reducing the price impacts of order flow and bid-ask spreads (e.g., Kyle, 1985; Glosten and Milgrom, 1985).

The expected effect of gambling on informational efficiency is less straightforward. On the one hand, uninformed or "noise" traders could distort prices, in particular if their trading decisions are highly correlated (e.g., De Long et al., 1990). Yet on the other hand, models of endogenous information acquisition show that an increase in uninformed "noise" traders could make prices more efficient by creating greater incentives (profit opportunities) for agents to engage in costly information acquisition and become informed (e.g., Grossman and Stiglitz, 1980; Kyle, 1989). It is thus an empirical question as to how a spillover of gamblers into the stock market will affect informational efficiency.

Testing the effects of gambling on stock markets is challenging due to a potential endogeneity problem. The amount of gambling on stock markets might be influenced by the liquidity or efficiency of the stock market. To overcome this challenge we draw on the findings of the previous section and exploit the regulation/restrictions of "casino" gambling as an instrument for gambling on stock markets. Table 5.5 shows that regulation and restrictions on casino gambling lead to a spillover of gambling into stock markets. Yet, it is unlikely that restrictions of casino gambling are affected by the liquidity or efficiency of the country's stock markets. It is also unlikely that regulation of casinos affects stock market quality through any channels other than through its effects on gambling activity.

We therefore estimate 2SLS IV models to analyze the effects of gambling on stock market characteristics, using our gambling restrictions variables as instruments:

$$StockGambling_{i,t} = \alpha_t + \beta_1 Ban_i + \beta_2 FewOp_i + \sum \gamma_j Controls_{i,t} + \varepsilon_{i,t} \tag{5.12}$$

$$Y_{i,t} = \alpha_t + \beta_1 \widehat{StockGambling}_{i,t} + \sum \gamma_j Controls_{i,t} + \varepsilon_{i,t} \tag{5.13}$$

where the dependent variable, $Y_{i,t}$, is a measure of stock market liquidity or efficiency, $\widehat{StockGambling}$ is the fitted value of stock gambling obtained from the first stage (5.12), $Controls_{i,t}$ includes the determinants of gambling as described previously (GDP per capita, unemployment, culture, and region fixed effects), and α_t are year fixed effects. As per the previous regressions, all dependent and independent variables, except region dummies, are standardized to have a mean of zero and standard deviation of one.

We measure liquidity using Amihud's (2002) liquidity measure, but transformed so that it measures liquidity rather than illiquidity (LIQ, see Appendix 5.C for details). We measure informational efficiency using four different measures as follows. The first, the Hou and Moskowitz (2005) delay metric ($Delay$), measures the delay with which individual stock prices reflect market-wide information and thus higher values indicate lower efficiency. The second, first order absolute autocorrelations of daily returns ($Autocorrel$), measures the extent to which prices depart from a random walk and are predictable using past returns, so again higher values indicate lower efficiency (e.g., Boehmer and Wu, 2013). The third, an estimate of the noise variance in stock returns ($Noise$) from a variance decomposition in the spirit of Hasbrouck (1993) also takes higher values when prices are noisier and thus less efficient. Finally, the fourth measure is the share of information in return variance ($InfoShare$), which is also estimated from a variance decomposition in the spirit of Hasbrouck (1993). Unlike the other three efficiency measures, higher values of

$InfoShare$ indicate more efficient and informative prices. All variables are more formally defined in Appendix 5.C, and the variance decomposition approach is described in Appendix 5.B.

Table 5.6 reports the results of the 2SLS models testing the impact of gambling on stock markets (equations (5.12) and (5.13)). The first stage regression in Model 1 shows that restrictions on “casino” gambling have the expected positive and significant effects on the amount of stock market gambling as expected. The monotonic pattern for regulation dummy variables’ coefficients suggests that stock gambling increases with more stringent restrictions or prohibitions on “casino” gambling.

The second-stage regression estimates in Model 2 show that gambling in the stock market tends to improve the liquidity of the stock market, reducing the price impacts of order flow. This positive effect of gambling is consistent with microstructure theory, which demonstrates that an increase in uninformed trading makes stock markets more liquid (e.g., Kyle, 1985; Glosten and Milgrom, 1985). It is also consistent with the basic notion that market liquidity increases with the number of market participants. The magnitude of the effect is economically meaningful: a one standard deviation increase in stock market gambling is associated with a 1.38 standard deviation increase in stock market liquidity.

Table 5.6. Impact of stock market gambling on stock market liquidity and efficiency.
This table reports the results from 2SLS IV estimation of the relation between stock market gambling and stock market liquidity and informational efficiency. Model (1) reports the result of the first-stage regression in which regulation dummy variables are used as instrumental variables. Model (2) reports the results for the second-stage regression that estimates the impact of stock market gambling. Stock market gambling ($StockGambling$) is an estimate of the dollar volume of gambling in stock markets, estimated using the approach specified in equations (1-7), in USD per capita per annum. $FewOp$ and Ban are regulation dummy variables that equal one if the gambling market is a monopoly/oligopoly, and if gambling activities are prohibited, respectively. $GDPPC$ is the GDP per capita in thousand USD; $Unemp$ is the unemployment rate in percent; IDV is the individualism dimension of culture measuring the degree of interdependence a society maintains among its members; UAI is the uncertainty avoidance dimension of culture; and IVR is the indulgence versus restraint dimension of culture. $Asia$, EU, $NAmer$, $SAmer$, and $Oceania$ are dummy variables that take the value of one if the country is located in Asia, Europe, North America, South America, or Oceania, respectively. LIQ is Amihud's (2002) liquidity measure. $Delay$ is the Hou and Moskowitz (2005) delay metric, which measures the delay with which individual stock prices reflect market-wide information. $Autocorrel$ is the absolute value of the first-order autocorrelation of daily returns. $Noise$ is the stock return variance that is attributable to noise. $InfoShare$ is the share of the total amount of stock return variance that is attributable to information. All dependent and independent variables, except region dummies, are standardized to have a mean of zero and standard deviation of one. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes stocks listed on major exchanges across 39 countries. The sample period is from 2003 to 2015.

Variable	First stage	Second stage				
	(1)	(2)	(3)	(4)	(5)	(6)
	$StockGambling$	LIQ	$Delay$	$Autocorrel$	$Noise$	$InfoShare$
$\widehat{StockGambling}$		1.38	-1.20	-1.73	-1.24	1.63
		(8.31)***	(-6.77)***	(-9.09)***	(-6.75)***	(13.24)***
Ban	1.43					
	(6.18)***					
$FewOp$	0.66					
	(5.79)***					
$GDPPC$	0.79	-0.79	1.00	1.33	1.00	-1.38
	(8.34)***	(-6.55)***	(7.53)***	(7.87)***	(7.06)***	(-11.43)***
$Unemp$	0.09	-0.57	0.08	0.08	0.41	-0.18
	(5.09)***	(-7.07)***	(1.82)*	(2.09)*	(10.17)***	(-6.45)***
IDV	0.12	0.09	0.20	-0.20	-0.11	0.28
	(3.81)***	(1.62)	(3.16)***	(-3.38)***	(-2.85)**	(5.69)***
UAI	0.05	0.17	-0.14	-0.39	-0.45	0.22
	(2.11)*	(7.57)***	(-4.12)***	(-6.43)***	(-6.46)***	(5.16)***
IVR	-0.23	0.41	-0.23	-0.49	-0.69	0.34
	(-5.03)***	(4.53)***	(-3.22)***	(-10.34)***	(-6.53)***	(8.13)***
Region controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2 (%)	49.53	35.01	49.35	29.22	52.91	35.23
Fixed Effects	Year	Year	Year	Year	Year	Year

The informational efficiency measures in Models 3-6 of Table 5.6 all consistently indicate that gambling in stock markets improves informational efficiency. $Delay$ decreases, suggesting stock prices become faster at reflecting market-wide information. Absolute autocorrelations ($Autocorrel$) decrease, suggesting prices more closely resemble a random walk. The noise in stock prices ($Noise$) declines and the information content of stock returns ($InfoShare$) increases. The magnitudes of the effects are economically meaningful: a one standard deviation increase in stock market gambling is associated with improvements in the market efficiency measures by between 1.20 and 1.73 standard deviations.The evidence that market efficiency increases as the amount of gambling in stock markets increases may seem counterintuitive in light of the behavioral finance literature that shows in various circumstances noise traders can distort prices, but is in fact supported by a large number of studies of price formation and the link between informational efficiency and liquidity. For example, Chordia et al. (2008), among many others, show that more liquid markets tend to be more informationally efficient. Given that gambling in stock markets makes them more liquid, it is perhaps not that surprising that an increase in informational efficiency also follows.

The results are also consistent with a large body of literature on price formation in markets that emphasizes the important role played by incentives to gather and trade on private information. For example, the Grossman and Stiglitz paradox implies that markets cannot be highly efficient unless informed traders are able to substantially profit from gathering private information (Grossman and Stiglitz, 1980). Microstructure theory with endogenous information acquisition reaches a similar conclusion (e.g., Kyle, 1984, 1989). The intuition is simple: sharks need prey to survive. Similarly, hedge funds and other sophisticated investors need fools or uninformed traders in the market to exploit to cover their costs of analyzing stocks and becoming informed. Our results are consistent with the notion that the gamblers that spill over into stock markets when “casino” gambling is restricted are the relatively uninformed “prey” that facilitates information acquisition by the “sharks” (hedge funds and other sophisticated investors). This notion is consistent with the evidence that individual investors, and in particular more overconfident investors, tend to underperform the market (e.g., Barber and Odean, 2000; Barber et al., 2008).

Table 5.7. Robustness tests for the impact of stock market gambling.
This table reports variations on the 2SLS IV estimation of the relation between stock market gambling and stock market liquidity and informational efficiency. The baseline models are presented in Table 5.6. The variations are as follows. Panel A uses Ban as the only instrumental variable. Panels B and C split the regulations variables into regulation of casinos and non-casino gambling activities, respectively. Panel D adds an interaction term that captures differences in how gambling affects the stock markets of developed and emerging countries. The variables are defined in Appendix 5.C. All dependent and independent variables, except region dummies, are standardized to have a mean of zero and standard deviation of one. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. The sample includes stocks listed on major exchanges across 39 countries. The sample period is from 2003 to 2015.

Panel A					
Variable	LIQ	$Delay$	$Autocorrel$	$Noise$	$InfoShare$
$Stock\widehat{Gambling}$	1.35	-1.26	-1.85	-0.75	1.78
	(8.79)***	(-6.87)***	(-8.09)***	(-3.93)***	(11.62)***
Controls	Yes	Yes	Yes	Yes	Yes
R^2 (%)	33.61	49.12	29.04	48.55	35.53
Fixed Effects	Year	Year	Year	Year	Year
Panel B					
Variable	LIQ	$Delay$	$Autocorrel$	$Noise$	$InfoShare$
$Stock\widehat{Gambling}$	0.88	-1.26	-1.16	-0.23	0.98
	(10.06)***	(-7.43)***	(-6.96)***	(-1.92)*	(8.81)***
Controls	Yes	Yes	Yes	Yes	Yes
R^2 (%)	29.94	49.13	21.71	46.73	27.65
Fixed Effects	Year	Year	Year	Year	Year
Panel C					
Variable	LIQ	$Delay$	$Autocorrel$	$Noise$	$InfoShare$
$Stock\widehat{Gambling}$	0.52	-1.26	-1.79	-0.82	1.62
	(3.09)***	(-6.52)***	(-7.33)***	(-4.35)***	(9.20)***
Controls	Yes	Yes	Yes	Yes	Yes
R^2 (%)	28.06	48.80	27.70	48.84	33.08
Fixed Effects	Year	Year	Year	Year	Year
Panel D					
Variable	LIQ	$Delay$	$Autocorrel$	$Noise$	$InfoShare$
$Stock\widehat{Gambling}$	2.02	-1.38	-1.94	-1.57	2.09
	(8.06)***	(-7.40)***	(-8.14)***	(-6.97)***	(14.77)***
$Stock\widehat{Gambling} \times DEV$	-1.24	0.36	0.43	0.64	-0.88
	(-5.56)***	(2.66)**	(2.28)**	(4.40)***	(-4.74)***
Controls	Yes	Yes	Yes	Yes	Yes
R^2 (%)	41.86	49.93	30.04	54.73	38.71
Fixed Effects	Year	Year	Year	Year	Year

We conduct a series of robustness tests, including alternative instrumental variables and changes to the regression models. Table 5.7 summarizes the results, focusing only on the stock market gambling results for brevity. In Table 5.7 Panel A, we re-estimate regression (5.12) with Ban dummy as the sole instrumental variable. This test overcomes the potential concern that only full prohibitions of gambling are unambiguously the result of policymaker decisions. The results are robust, with stock gambling being positively associated with liquidity and informational efficiency.

Another potential concern is that gambling in casinos, in some countries, is partly attributable to tourists, not the local population, thereby distorting the impact of regulation of casinos. To address this issue, we split the regulations variables into those that restrict casino activities and those that restrict non-casino activities (sports betting, horse racing, lotteries, and gaming machines). The results in Table 5.7 Panels B and C show that our main results hold for both versions of the gambling restrictions variables.

Finally, we test whether our results on the impact of gambling are different in developed versus emerging countries. In Table 5.7 Panel D, we add an interaction term between stock gambling and a developed market dummy variable. The results show that the effect of gambling is similar in both developed and emerging markets, increasing liquidity and informational efficiency, but the magnitudes of the effects are stronger in emerging markets.

5.8 Conclusion

This paper uses global data on gambling in casinos, lotteries, gaming machines, and sports betting, as well as novel measures of gambling in stock markets to assess the extent to which stock markets are used as an alternative to casinos, and how gambling in stock markets affects their liquidity and efficiency. Our analysis leads to four main findings.

First, the amount of gambling that occurs on stock markets is economically meaningful. In dollar terms, it is at least 3.5 times the combined amount of gambling in casinos, lotteries, horse racing/sports betting, gaming machines, and online gambling. Particularly high volumes of gambling on stock markets occur in countries with relatively more developed and liquid stock markets, such as the US and Hong

Kong. Thus, financial markets play a significant role as one of the venues in which gambling takes place.

Second, the determinants of the amount of gambling on stock markets and casino gambling are remarkably similar. We find that much of the cross-country variation in both forms of gambling can be explained by a country's wealth, unemployment rate, national culture, and regional fixed effects. These findings further reinforce the notion that gambling on stock markets has many similarities with gambling in casinos.

Third, we find strong evidence that stock markets are an economically meaningful alternative to casinos as a venue in which gambling activity is realized. The two forms of gambling have an unconditional positive relation largely due to common drivers. An exogenous shock to casino gambling due to country-level restrictions and prohibitions leads to significant spillovers of gambling activity into the country's stock markets.

Finally, using restrictions on "traditional" forms of gambling such as casinos and sports betting for identification, we find that gambling in stock markets tends to make them more liquid and informationally efficient. These findings are consistent with theory that predicts an increase in uninformed traders increases liquidity and by providing increased incentives for information acquisition by sophisticated market participants, can make prices more informative and efficient.

Our findings imply that regulation of a country's traditional gambling sector, such as banning or liberalizing the markets for casinos, lotteries, sports betting, and gaming machines, is likely to affect a country's capital markets through the level of gambling that occurs in these markets. The liquidity and efficiency of a country's capital markets can be promoted through policies that encourage broader participation in these markets instead of participation in the traditional gambling sector. As a market development strategy, this idea has the most potential in countries that have high gambling propensities and relatively illiquid or underdeveloped stock markets. These potential benefits, however, should be weighed up against potential downsides such as distortions in the pricing of lottery-like stocks and systemic risk.

Appendix 5.A. Data description

5.A.1 Sample selection

In each country, we use stocks from the largest stock exchange in that country, with the exception of several countries for which we include the two major listing exchanges: China (Shanghai and Shenzhen Stock Exchanges), Japan (Tokyo and Osaka Stock Exchanges), Germany (Deutsche Boerse AG and XETRA), and the US (NYSE and Nasdaq).

Following Griffin, Kelly, and Nardari (2010), we exclude preferred stocks, warrants, mutual funds, unit or investment trusts, duplicates, Global Depositary Receipts (GDRs), and other non-common equity. We eliminate cross-listed stocks and exclude secondary issues of shares. We restrict our analysis to stocks that trade in the company's home market and in local currency. Additionally we require that the stock has market capitalization data (in US dollars) at the end of each year. Dead stocks are included in the sample to mitigate survivorship bias.

Following Ang et al. (2009), in all non-US countries, we exclude very small firms by eliminating the 5% of firms with the lowest market capitalizations. We also omit newly listed stocks that have been traded for less than five months, as well as stocks that are about to be delisted. The initial sample includes slightly more than 38,000 stocks from around the world over the thirteen-year period from 2003 to 2015.

5.A.2 Return filters

We use the following screens in the return data to ensure tradability and avoid outliers. We exclude non-trading days, defined as days on which more than 90% of stocks listed on a given exchange have zero returns. We also exclude a stock if the number of zero-return days is more than 80% in a given month. To handle data errors in Datastream, we follow Ince and Porter (2006), and set the daily stock returns of both days d and $d-1$ to be missing if $R_{i,d}R_{i,d-1}-1 \leq 50\%$, where $R_{i,d}$ is the gross return of stock i on day d, and at least one of the two returns is 200% or greater. To avoid rounding errors, the daily return is set to be missing if the return index in Datastream for either the previous day or the current day is less than 0.01.

To construct stock liquidity measures, including Amihud's (2002) ILLIQ, which is sensitive to extreme returns, we discard stock-day observations with a daily return in the top or bottom 1% of the cross sectional distribution within a country. We exclude stock-month observations with a monthly return or stock price at the end of the previous month in the top or bottom 1% of the cross-sectional distribution within a country.

Appendix 5.B. Variable construction

5.B.1 Lottery-like characteristics

The existing literature identifies a number of characteristics that make a stock "lottery-like." Following the literature, we measure five different characteristics and define a stock as "lottery-like" if its score on the particular characteristic exceeds the median value of that characteristic in the given country in the given year.

(1) MAX is the maximum daily return within a month.

$$MAX_{i,t} = \max(r_{i,d}), \quad d = 1, \dots, D_t \tag{5.B.1}$$

where $r_{i,d}$ is the return on stock i on day d, and D_t is the number of trading days in month t.

(2) $MAX5$ is the average of the five highest daily returns within a month.

(3) $IVOL$ is the standard deviation of the daily residuals within a month. We follow Bali et al. (2011) to calculate the monthly idiosyncratic volatility (IVOL) in month t from a single-factor model:

$$r_{i,d} = \alpha_i + \beta_i r_{m,d} + \varepsilon_{i,d} \tag{5.B.2}$$

where $r_{m,d}$ is the market return and $\varepsilon_{i,d}$ is the idiosyncratic return on stock i on day d in month t.

$$IVOL_{i,t} = \sqrt{\mathrm{var}(\varepsilon_{i,d})} \tag{5.B.3}$$

(4) $TVOL$ is the total volatility within a month. The total volatility (TVOL) of stock i in month t is defined as the standard deviation of daily returns within month t:

$$TVOL_{i,t} = \sqrt{\mathrm{var}(r_{i,d})} \tag{5.B.4}$$

(5) $BETA$ is the market beta within a month. To take into account nonsynchronous trading, we follow Scholes and Williams (1977) and Dimson (1979) when estimating the market beta:

$$r_{i,d} = \alpha_i + \beta_{1i} r_{m,d-1} + \beta_{2i} r_{m,d} + \beta_{3i} r_{m,d+1} + \varepsilon_{i,d} \tag{5.B.5}$$

We estimate the equation (5.B.5) for each stock using daily returns within a month. The market beta of stock i in month t is defined as $\beta_i = \beta_{1i} + \beta_{2i} + \beta_{3i}$. Each month needs at least 15 observations for the estimation.

5.B.2 Stock liquidity

We measure liquidity using Amihud's (2002) illiquidity measure with daily returns and daily dollar volumes. For a given stock, $ILLIQ$ is the average of its daily absolute returns divided by its daily dollar volumes. Following Karolyi et al. (2012), we add a constant to the Amihud measure and take logs, to reduce the impact of outliers. We multiply the measure by -1 to arrive at a variable that is increasing with the liquidity of individual stocks:

$$LIQ_{i,t} = -log\left(1 + ILLIQ_{i,t}\right) \tag{5.B.6}$$

5.B.3 Delay

We construct a modified version of the Hou and Moskowitz (2005) delay metric for stock i in year t as:

$$Delay_{i,t} = 1 - \frac{R^2_{constrained}}{R^2_{unconstrained}} , \tag{5.B.7}$$

where $R^2_{unconstrained}$ is the R^2 from the unconstrained regression of daily returns for stock i on daily market returns, r_m, and ten lags:

$$r_{i,d} = \alpha_i + \beta_i r_{m.d} + \sum_k^{10} \delta_{i,k} r_{m,d-k} + \varepsilon_{i,d} , \tag{5.B.8}$$

and $R^2_{constrained}$ is the R^2 from the restricted regression constraining all $\delta_{i,k}$ to zero.

If market-wide information is perfectly and instantly reflected in the stock's prices, the two R^2 are equal and $Delay_{i,t} = 0$, but if stock prices are sluggish in reflecting market-wide information, then $R^2_{unconstrained} > R^2_{constrained}$ and $Delay_{i,t} > 0$.

5.B.4 Return autocorrelation

$Autocorrel$ is the absolute value of the first-order daily return autocorrelations for each stock-year:

$$Autocorrel_{i,t} = \left|corr(r_{i,d}, r_{i,d-1})\right| , \tag{5.B.9}$$

5.B.5 Stock return variance components

We follow Hasbrouck (1993) to decompose stock return variance into two components: information and noise. The separation of information and noise is based on a temporary-permanent decomposition of prices, where permanent innovations are interpreted as innovations in the fundamental value (efficient price) driven by new information being impounded in prices, and temporary innovations are pricing errors (deviations from fundamentals), generically referred to as "noise". Following Hasbrouck (1993), the observed price is the sum of the efficient price, which follows a random walk, and a pricing error. To estimate the variance of the pricing error, we employ a vector autoregressive (VAR) model of returns and two other variables, including the signed dollar volume of trading and market return.

We start with a structural VAR with five lags to allow a full week of lagged effects:

$$
\begin{aligned}
r_{m,t} &= a_0 + \textstyle\sum_{l=1}^{5} a_{1,l} r_{m,t-l} + \sum_{l=1}^{5} a_{2,l} x_{t-l} + \sum_{l=1}^{5} a_{3,l} r_{t-l} + \varepsilon_{r_m,t} \\
x_t &= b_0 + \textstyle\sum_{l=0}^{5} b_{1,l} r_{m,t-l} + \sum_{l=1}^{5} b_{2,l} x_{t-l} + \sum_{l=1}^{5} b_{3,l} r_{t-l} + \varepsilon_{x,t} \\
r_t &= c_0 + \textstyle\sum_{l=0}^{5} c_{1,l} r_{m,t-l} + \sum_{l=0}^{5} c_{2,l} x_{t-l} + \sum_{l=1}^{5} c_{3,l} r_{t-l} + \varepsilon_{r,t},
\end{aligned}
\tag{5.B.10}
$$

where $r_{m,t}$ is the market return, x_t is the signed dollar volume of trading in the given stock (positive for net buying and negative for net selling), and r_t is the stock return. We estimate this VAR separately in every stock-year using daily returns and daily volumes.

The VAR in equation (5.B.10) is subsequently transformed into an infinite order structural vector moving average (VMA), from which we isolate temporary price movements from permanent price movements.

$$
\begin{aligned}
r_{m,t} &= A_0 + \textstyle\sum_{l=0}^{\infty} A_{1,l} \varepsilon_{r_m,t-l} + \sum_{l=1}^{\infty} A_{2,l} \varepsilon_{x,t-l} + \sum_{l=1}^{\infty} A_{3,l}\, \varepsilon_{r,t-l} \\
x_t &= B_0 + \textstyle\sum_{l=0}^{\infty} B_{1,l} \varepsilon_{r_m,t-l} + \sum_{l=0}^{\infty} B_{2,l} \varepsilon_{x,t-l} + \sum_{l=1}^{\infty} B_{3,l}\, \varepsilon_{r,t-l} \\
r_t &= C_0 + \textstyle\sum_{l=0}^{\infty} C_{1,l} \varepsilon_{r_m,t-l} + \sum_{l=0}^{\infty} C_{2,l} \varepsilon_{x,t-l} + \sum_{l=0}^{\infty} C_{3,l}\, \varepsilon_{r,t-l}.
\end{aligned}
\tag{5.B.11}
$$

The information-driven innovation in the efficient price is given by:

$$w_t = \theta_{r_m}\varepsilon_{r_m,t} + \theta_x\varepsilon_{x,t} + \theta_r\varepsilon_{r,t} \quad (5.B.12)$$

where $\theta_{r_m} = \sum_{l=0}^{\infty} C_{1,l}$, $\theta_x = \sum_{l=0}^{\infty} C_{2,l}$, $\theta_r = \sum_{l=0}^{\infty} C_{3,l}$.

The innovation in the pricing error is given by:

$$\Delta s_t = r_t - \mu - w_t = r_t - A_0 - \theta_{r_m}\varepsilon_{r_m,t} - \theta_x\varepsilon_{x,t} - \theta_r\varepsilon_{r,t}. \quad (5.B.13)$$

The variance of information-driven innovations in the efficient price is therefore:

$$\sigma_w^2 = \theta_{r_m}^2\sigma_{\varepsilon_{r_m}}^2 + \theta_x^2\sigma_{\varepsilon_x}^2 + \theta_r^2\sigma_{\varepsilon_r}^2 \quad (5.B.14)$$

The variance of noise, σ_s^2, is computed from the time-series of Δs_t.

We construct two measures of informational efficiency from this variance decomposition. The first is $Noise = \sigma_s^2$, being the stock return variance that is attributable to noise or pricing errors. The second is $InfoShare = \frac{\sigma_w^2}{\sigma_w^2+\sigma_s^2}$, being the share of stock return variance that is attributable to information.

Appendix 5.C. Summary of variable definitions (further details are provided in Appendix 5.B)

Variable	Notation	Description	Data Sources
Panel A: Gambling measures			
"Casino" gambling activity	$CasinoGambling$	Dollar volume of gambling per capita in casinos, lotteries, horse racing, sports betting, gaming machines, and online gambling.	GBGC
Stock gambling activity	$StockGambling$	A proxy for the per capita dollar volume of gambling in lottery-like stocks.	Datastream, CRSP
Few-operators dummy	$FewOp$	Set to one if the gambling activities are provided by a limited number of operators.	GBGC
Ban dummy	Ban	Set to one if gambling activities are prohibited.	GBGC
Panel B: Country-level gambling determinants			
GDP per capita	$GDPPC$	Gross domestic product (GDP) per capita in thousand USD.	World Bank
Unemployment rate	$Unemp$	Annual unemployment rate in percent.	World Bank
Uncertainty Avoidance Index	UAI	Culture dimension measuring the extent to which cultures tolerate ambiguity and uncertainty.	Hosftede's website
Indulgence Versus Restraint	IVR	Culture dimension measuring the extent to which individuals try to control their desires.	Hosftede's website
Individualism Versus collectivism	IDV	Culture dimension measuring the degree of interdependence a society maintains among its members.	Hosftede's website
Panel C: Stock market quality variables			
Amihud's (2002) liquidity	LIQ	Defined as $-\log(1 + ILLIQ)$, where $ILLIQ$ is the daily Amihud's (2002) illiquidity measure.	Datastream, CRSP
Delay	$Delay$	Defined as $\lvert 1 - (R^2_{constrained})/(R^2_{unconstrained}) \rvert$, where $R^2_{unconstrained}$ is the R^2 from the unconstrained regression of daily returns for a given stock on daily market returns, and ten lags. $R^2_{constrained}$ is the R^2 from the constrained regression of the stock's daily returns on daily market returns.	Datastream, CRSP
Return autocorrelation	$Autocorrel$	Defined as the absolute value of the first-order daily return autocorrelation.	Datastream, CRSP
Noise	$Noise$	The amount of stock return variance that is attributable to noise.	Datastream, CRSP
Information share	$InfoShare$	The share of stock return variance that is attributable to information.	Datastream, CRSP

Chapter 6: Conclusions

This chapter summarizes the conclusions that can be drawn from the four related studies that comprise this thesis. The conclusions are arranged under the following seven areas:

(i) What do the new measures of noise and information in stock prices tell us about what drives stock price movements?

(ii) How have the levels of noise and information that are reflected in stock prices changed through time as the structure of stock markets has evolved?

(iii) How does noise in stock prices affect inference in existing empirical models including cash flow/discount rate news decompositions and measures of idiosyncratic volatility?

(iv) What are the real effects of informational efficiency: how does noise in stock prices affect the efficiency of corporate investment and capital allocation across firms and industries?

(v) How does gambling in stock markets affect their liquidity, informational efficiency, and noise?

(vi) How much gambling occurs in stock markets and what drives the amount of gambling?

This chapter ends with suggestions for future research.

6.1 What do the new measures of noise and information in stock prices tell us about what drives stock price movements?

Existing measures of market efficiency and decompositions of stock return variance have a number of limitations, including: (i) neglecting the role of noise in stock returns as distinct from information; (ii) an absence of methods to partition information into more refined categories; and (iii) the inability to estimate some of the key models at sufficiently high frequencies to study time-series variance in the drivers of stock return variance.

Recognizing these limitations, Chapter 2 develops a novel return variance decomposition model that isolates the role of noise in stock returns, and at the same time partitions information into various sources. At the stock level, the stock return variance is split into four components: market-wide information, private firm-specific information revealed through trading, firm-specific information revealed through public sources, and noise.

Overall, we find that noise plays a meaningful role in markets, contributing 31% of the total stock return variance in the US. Further, firm-specific information accounts for the majority (61%) of stock return variance, with market-wide information accounting for the remaining 8% of variance in the full sample. Public firm-specific information plays a larger role than private firm-specific information that is impounded into prices through trading (37% and 24%, respectively).

6.2 How have the levels of noise and information that are reflected in stock prices changed through time as the structure of stock markets has evolved?

Our new approach to separating noise and different types of information (Chapter 2) allows the decomposition of stock return variance at annual frequencies, and therefore enables analysis of how different drivers of stock return variance evolve through time and in the cross-section of stocks.

We observe that in US stock markets, the role of noise increases from the 1970s to the mid-1990s, in particular around a period of collusion by dealers to effectively widen bid-ask spreads, and has declined since then. This decline in noise is attributable in part to narrower tick sizes, which reduces bid-ask bounce, and a general improvement in liquidity and increase in turnover. Separating the sample by firm size shows that the noise component decreases monotonically with firm size: larger firms have less noisy prices. When separating the sample by industry we observe only minor differences across industries in the composition of variance, suggesting that the findings are not specific to a particular industry, nor are they driven by a certain segment of the economy.

The role of firm-specific information has increased through time, driven largely by increases in the amount of public firm-specific information that is reflected in prices, consistent with increasing informational efficiency through time. The increasing importance of public firm-specific information in stock prices is also consistent with a variety of regulatory reforms such as the Sarbanes Oxley Act (2002) and Regulation Fair Disclosure (2000) aimed at improving both the quality and quantity of corporate disclosure.

The results also show that market-wide information has become a less important driver of stock returns over time. In general, market-wide information has decreased from around 15% of variance to around 5-10% in recent years, except dramatic spikes during financial crises.

6.3 How does noise in stock prices affect inference in existing empirical models?

The second study in this thesis (Chapter 3) shifts the focus to the effects of noise on the inferences that can be made in existing empirical models of variance decomposition. In particular, we examine two widely used models in the literature.

The first model is standard cash flow/discount rate news decomposition developed by Campbell and Shiller (1988a, 1988b) and Campbell (1991). We find that after accounting for noise in the return generating process, cash flow information plays a considerably larger role in driving individual stock returns than previously believed, and discount rate information plays a smaller role. In fact, the ratio of cash flow news to discount rate news for individual stocks increases from around five to twenty five times after we account for the impact of noise. The difference in the importance of discount rate news compared to previous studies stems largely from the fact that noise impacts return predictability. Without accounting for noise, the noise-induced variation in predicted returns is attributed to variation in the discount rate thereby overstating the role of discount rate news.

In addition, the ratio of cash flow news to discount rate news is higher in firm-specific information than in market-wide information. Therefore, despite the larger share of cash flow news as compared to the share of discount rate news in stock return variation, cash flow news tends to be more idiosyncratic and discount rate news more systematic. This result helps reconcile existing studies in which cash flow news is found to be more important in stock-level analyses (e.g., Vuolteenaho, 2002), while discount rate news plays a larger role in portfolio-level analyses in which much of the idiosyncratic variation is removed (e.g., Campbell, 1991).

The second empirical model for which we explore the effects of noise is idiosyncratic variance, which is often used in the literature as a measure of the amount of firm-specific information in prices and thus a proxy for informational efficiency. A puzzle in this literature, which we are able to explain, is why idiosyncratic volatility has declined in the US since 1997, despite other evidence that US markets have become more efficient in recent decades.

We find that a lower level of idiosyncratic volatility (a higher level of market level R^2) is *not* necessarily associated with less informationally efficient prices, in contrast to the interpretation of R^2 in prior studies. In fact, the idiosyncratic volatility decreases in the post-1997 period as a result of the decrease in noise during this period. The evidence indicates that the market has become more efficient in the past two decades, contrary to what is implied by interpretations of R^2 as an inverse measure of efficiency.

6.4 What are the real effects of informational efficiency: how does noise in stock prices affect the efficiency of corporate investment and capital allocation across firms and industries?

Chapter 4 examines the real effects of stock market efficiency, in particular how noise in stock prices affects the efficiency of corporate investment and capital allocation across firms and industries. We draw on evidence from a long time-series of US data from 1963 and a cross-section of 42 countries. Consistent with the notion that noise decreases investment efficiency, we find strong evidence that the sensitivity of corporate investment to growth opportunities (as proxied by Tobin's Q) is higher for firms whose stock prices contain less noise. We also find that corporate investment is more sensitive to Tobin's Q for firms with greater firm-specific information. Importantly, the effects of noise are *incremental* to the effects of information.

Moreover, we find that noise in stock prices has a negative impact on capital allocation efficiency at the industry level. Our findings highlight the important real effects of secondary market quality in determining firms' investment behaviour and the efficiency with which capital is allocated.

6.5 How does gambling in stock markets affect their liquidity, informational efficiency, and noise?

Chapter 5 provides evidence on how gambling in stock markets affects various measures of liquidity and informational efficiency. Our identification strategy exploits the fact that restrictions on "traditional" forms of gambling (such as casinos and sports betting) vary across countries, impact the amount of gambling on stock markets due to a spillover effect, and that restrictions on gambling are unlikely to be influenced by stock market characteristics.

We find that gambling in stock markets tends to make stock markets more liquid. This finding is consistent with the notion that stock market liquidity depends on the amount of stock market participation and in particular the participation of relatively uninformed traders such as gamblers (e.g., Kyle, 1985).

Perhaps more surprisingly, we find that increased gambling in stock markets does not harm the amount of information reflected in prices. Rather, gambling on stock markets tends to increase the amount of information that is reflected in prices, increase measures of informational efficiency, and reduce measures of noise in prices. This finding is consistent with theory, which predicts that higher liquidity creates more incentives for traders to gather private information and trade on it (e.g., Grossman and Stiglitz, 1980; Kyle, 1984, 1989). Therefore, even if gamblers are relatively or completely uninformed traders, they can still contribute to market efficiency by making markets more liquid and thereby encouraging informed trading. These findings have implications regarding the use of gambling regulations as a policy instrument to affect financial market quality.

6.6 How much gambling occurs in stock markets and what drives the amount of gambling?

In Chapter 5, we propose novel measures of gambling in stock markets and compare the estimates of these measures with gambling in traditional forms such as casinos and lotteries. We find that the amount of gambling in stock markets is economically meaningful. In dollar terms, gambling in stock markets is around 3.5 times the combined gambling in casinos, lotteries, horse racing/sports betting, gaming machines, and online gambling. This difference is particularly strong in countries with relatively more developed and liquid stock markets, such as the US and Hong Kong, which have the highest per capita levels of gambling on stock markets. Thus, financial markets play a significant role as one of the venues in which gambling takes place.

We observe that determinants of the level of gambling in stock markets and casino gambling are remarkably similar. We find that much of the cross-country variation in both forms of gambling can be explained by a set of common drivers. Wealthier countries tend to gamble more per capita. Various cultural dimensions also explain gambling, with higher levels of gambling in societies that are less troubled by uncertainty and societies with a more individualistic orientation. Region effects also play a significant role, with higher levels of gambling in North and South America, as well as Asia, after controlling for other factors. As a result of these common drivers, the two forms of gambling are strongly positively related across countries. These findings further reinforce the notion that gambling on stock markets has many similarities with gambling in casinos.

Finally, we find that stock markets serve as a substitute or alternative to "traditional" gambling venues such as casinos, lotteries, sports betting, gaming machines, and so on. We find that when a country restricts traditional gambling, for example by banning casinos or restricting them to a state-owned monopoly, a significant volume of gambling activity spills over onto stock markets. This finding is consistent with prior evidence that stock market volume declines when weekly lottery jackpots are larger than usual (e.g., Gao and Lin, 2014; Dorn et al., 2015) and indicates that the substitution effect holds more broadly for a range of gambling forms and in a global sample.

6.7 Future research

As suggested in Chapter 2, the new measures of noise and information utilized in this thesis could be used to shed light on recent issues concerning the information content of prices. For example, the concern that the growth in indexing and passive investing in recent years and corresponding decline in active funds management might harm the amount of firm-specific information in prices is not intuitively supported by the observed time trends, but further investigation of the causal effects would be valuable. Another issue that could be addressed using the new approach is the question of the effects of high frequency trading (HFT) on the information and noise in stock prices.

References

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31–56.

Ang, A., Hodrick, R.J., Xing, Y., Zhang, X., 2009. High idiosyncratic volatility and low returns: International and further US evidence. Journal of Financial Economics 91, 1–23.

Asparouhova, E., Bessembinder, H., Kalcheva, I., 2010. Liquidity biases in asset pricing tests. Journal of Financial Economics 96, 215–237.

Asparouhova, E., Bessembinder, H., Kalcheva, I., 2013. Noisy prices and inference regarding returns. Journal of Finance 68, 665–714.

Avramov, D., Chordia, T., Goyal, A., 2006. Liquidity and autocorrelations in individual stock returns. Journal of Finance 61, 2365–2394.

Bai, J., Philippon, T., Savov A., 2016. Have financial markets become more informative? Journal of Financial Economics 122, 625–654.

Bakke, T.E., Whited, T.M., 2010. Which firms follow the market? An analysis of corporate investment decisions. Review of Financial Studies 23, 1941–1980.

Baldauf, M., Mollner, J., 2018. High-frequency trading and market performance. Working paper.

Ball, R., Brown, P., 1968. An empirical evaluation of accounting income numbers. Journal of accounting research 6, 159–178.

Bali, T.G., Brown, S.J., Murray, S., Tang, Y., 2017. A lottery-demand-based explanation of the beta anomaly. Journal of Financial and Quantitative Analysis 52, 2369–2397.

Bali, T.G., Cakici, N., Whitelaw, R.F., 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. Journal of Financial Economics 99, 427–446.

Barber, B.M., Lee, Y.T., Liu, Y.J., Odean, T., 2008. Just how much do individual investors lose by trading?. Review of Financial Studies 22, 609–632.

Barber, B.M. and Odean, T., 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. Journal of Finance 55, 773–806.

Barberis, N., Huang, M., 2008. Stocks as lotteries: The implications of probability weighting for security prices. American Economic Review 98, 2066–2100.

Basu, S., 1977. Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. Journal of Finance 32, 663–682.

Bhattacharya, U. and Galpin, N., 2011. The global rise of the value-weighted portfolio. Journal of Financial and Quantitative Analysis 46, 737–756.

Black, F., 1986. Noise. Journal of Finance 41, 528–543.

Blume, M.E., Stambaugh, R.F., 1983. Biases in computed returns: An application to the size effect. Journal of Financial Economics 12, 387–404.

Boehmer, E. and J. Wu, 2013. Short selling and the price discovery process. Review of Financial Studies 26, 287–322.

Bramante, R., Zappa, D. and Petrella, G., 2013. On the interpretation and estimation of the market model R-square. Electronic Journal of Applied Statistical Analysis 6, 57–66.

Brealey, R.A., Myers, S.C., Allen, F., 2005. Corporate Finance: 8th Edition. New York: McGraw–Hill Irwin.

Brenner, R., Brenner, G.A., 1990. Gambling and speculation: A theory, a history, and a future of some human decisions. Cambridge University Press.

Brogaard, J., Hendersott, T., Riordan, R., 2014. High-frequency trading and price discovery. Review of Financial Studies 27, 2267-2306.

Brogaard, J., Nguyen, T.H., Putniņš, T.J., Wu, E., 2019. What Moves Stock Prices? The Role of News, Noise, and Information. Working paper.

Brogaard, J., Shi, W., Wei, K., You, H., 2018. Do analysts improve investment efficiency? Working paper.

Campbell, J.Y., 1991. A variance decomposition for stock returns. Economic Journal 101, 157–179.

Campbell, J.Y., Ammer, J., 1993. What moves the stock and bond markets? A variance decomposition for long-term asset returns. Journal of Finance 48, 3–37.

Campbell, J.Y., Lettau, M., Malkiel, B.G., Xu, Y., 2001. Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. Journal of Finance 56, 1–43.

Campbell, J.Y., Shiller, R.J., 1988a. The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1, 195–228.

Campbell, J.Y., Shiller, R.J., 1988b. Stock prices, earnings, and expected dividends. Journal of Finance 43, 661–676.

Carlin, W. and Mayer, C., 2003. Finance, investment, and growth. Journal of Financial Economics 69, 191–226.

Chen, L., Da, Z., Zhao, X., 2013. What drives stock price movements?. Review of Financial Studies 26, 841–876.

Chen, Q., Goldstein, I., and Jiang, W., 2007. Price informativeness and investment sensitivity to stock price. Review of Financial Studies 20, 619–650.

Chen, F., Hope, O.K., Li, Q. and Wang, X., 2011. Financial reporting quality and investment efficiency of private firms in emerging markets. Accounting Review 86, 1255–1288.

Christie, W., Schultz, P., 1994. Why do NASDAQ market makers avoid odd-eighth quotes? Journal of Finance 49, 1813–1840.

Chordia, T., Roll, R., Subrahmanyam, A., 2008. Liquidity and market efficiency. Journal of Financial Economics 87, 249–268.

Chordia, T., Roll, R., Subrahmanyam, A., 2011. Recent trends in trading activity and market quality. Journal of Financial Economics 101, 243–263.

Clotfelter, C.T., Cook, P.J., 1999. State lotteries at the turn of the century: Report to the national gambling impact study commission. Working paper.

Comerton-Forde, C., Putnins, T., 2015. Dark trading and price discovery. Journal of Financial Economics 118, 70–92.

Cong, L.W., Xu, D.X., 2017. Rise of factor investing: Asset prices, informational efficiency, and security design, Working paper.

Corwin, S.A., Schultz, P., 2012. A simple way to estimate bid-ask spreads from daily high and low prices. Journal of Finance 67, 719–759.

Dasgupta, S., Gan, J., Gao, N., 2010. Transparency, price informativeness, and stock return synchronicity: Theory and evidence. Journal of Financial and Quantitative Analysis 45, 1189–1220.

De Bondt, W.F., Thaler, R., 1985. Does the stock market overreact?. Journal of Finance 40, 793–805.

De Long, J.B., Shleifer, A., Summers, L.H., and Waldmann, R.J., 1990. Noise trader risk in financial markets. Journal of Political Economy 98, 703–738.

DeLisle, J., French, D.W., Schutte, M.G., 2015. Co-movement, passive investing and price informativeness. Working paper.

Dessaint, O., Foucault, T., Frésard, L., Matray, A., 2018. Noisy stock prices and corporate investment. Review of Financial Studies 32, 2625–2672.

Dorn, A.J., Dorn, D., Sengmueller, P., 2015. Trading as gambling. Management Science 61, 2376–2393.

Dorn, D., Sengmueller, P., 2009. Trading as entertainment?. Management Science 55, 591–603.

Durnev, A., Li, K., Morck, R., Yeung, B., 2004. Capital markets and capital allocation: Implications for economies in transition. Economics of Transition 12, 593–634.

Easley, D., Kiefer, N.M., O'hara, M., Paperman, J.B., 1996. Liquidity, information, and infrequently traded stocks. Journal of Finance 51, 1405–1436.

Easley, D., Kiefer, N.M., O'Hara, M., 1997a. One day in the life of a very common stock. Review of Financial Studies 10, 805–835.

Easley, D., Kiefer, N.M., O'Hara, M., 1997b. The information content of the trading process. Journal of Empirical Finance 4, 159–186.

Edmans, A., Goldstein, I., Jiang, W., 2012. The real effects of financial markets: The impact of prices on takeovers. Journal of Finance 67, 933–971.

Edmans, A., Jayaraman, S., Schneemeier, J., 2017. The source of information in prices and investment-price sensitivity. Journal of Financial Economics 126, 74–96.

Fama, E.F.. 1965. Random walks in stock market prices. Financial Analyst Journal 21, 55–59.

Fama, E., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. Journal of Finance 25, 383–417.

Fama, E.F., Fisher, L., Jensen, M.C., Roll, R., 1969. The adjustment of stock prices to new information. International Economic Review 10, 1–21.

Fama, E.F., French, K.R., 1988. Dividend yields and expected stock returns. Journal of Financial Economics 22, 3–25.

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

Foucault, T., Frésard, L., 2012. Cross-listing, investment sensitivity to stock price, and the learning hypothesis. Review of Financial Studies 25, 3305–3350.

Foucault, T., Fresard, L., 2014. Learning from peers' stock prices and corporate investment. Journal of Financial Economics 111, 554–577.

French, K., Roll, R., 1986. Stock return variances: The arrival of information and the reaction of traders. Journal of Financial Economics 17, 5–26.

Gao, X., Lin, T.C., 2014. Do individual investors treat trading as a fun and exciting gambling activity? Evidence from repeated natural experiments. Review of Financial Studies 28, 2128–2166.

Gassen, J., Skaife, H.A., Veenman, D., 2015. Illiquidity and R2. Working paper.

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, 71–100.

Griffin, J., Kelly, P., Nardari, F., 2010. Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets. Review of Financial Studies 23, 3225–3277.

Grinblatt, M., Keloharju, M., 2009. Sensation seeking, overconfidence, and trading activity. Journal of Finance 64, 549–578.

Grossman, S.J. and Stiglitz, J.E., 1980. On the impossibility of informationally efficient markets. American Economic Review 70, 393–408.

Hasbrouck, J., 1988. Trades, quotes, inventories and information, Journal of Financial Economics 22, 229–252.

Hasbrouck, J., 1991a. Measuring the information content of stock trades. Journal of Finance 46, 179–207.

Hasbrouck, J., 1991b. The summary informativeness of stock trades: An econometric analysis. Review of Financial Studies 4, 571–595.

Hasbrouck, J., 1993. Assessing the quality of a security market: A new approach to transaction-cost measurement. Review of Financial Studies 6, 191–212.

Haugen, R.A., Lakonishok, J., 1988. The incredible January effect: The stock market's unsolved mystery. Business One Irwin.

Hendershott, T., Li, S., Menkveld, A.J., Seasholes, M.S., 2011. Risk sharing, costly participation, and monthly returns. Working paper.

Hendershott, T., Menkveld, A., 2014. Price pressures. Journal of Financial Economics 114, 405–423.

Henrikson, R.D., 1984. Market timing and mutual fund performance: An empirical investigation. Journal of business 57, 73–96.

Ho, T., Stoll, H.R., 1981. Optimal dealer pricing under transaction and return uncertainty. Journal of Financial Economics 9, 47–73.

Hong, H., Kacperczyk. M., 2010. Competition and bias. Quarterly Journal of Economics 125, 1683–1725.

Hofstede, G. 1997. Cultures and organizations, New York: McGraw-Hill.

Hofstede, G., 2001. Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations, Sage Publications.

Hou, K., Moskowitz, T.J., 2005. Market frictions, price delay, and the cross-section of expected returns, Review of Financial Studies 18, 981–1020.

Ince, O.S., Porter, R.B., 2006. Individual equity return data from Thomson Datastream: Handle with care!. Journal of Financial Research 29, 463–479.

Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. Journal of Finance 45, 881–898.

Jensen, M.C., 1968. The performance of mutual funds in the period 1945–1964. Journal of Finance 23, 389–416.

Jensen, M.C., 1978. Some anomalous evidence regarding market efficiency. Journal of Financial Economics 6, 95–101.

Jin, L., Myers, S.C., 2006. R^2 around the world: New theory and new tests. Journal of Financial Economics 79, 257–292.

Kallick, M., Suits, D., Dielman, T.E., Hybels, J.H., 1976. A survey of American gambling attitudes and behavior. Survey Research Center Research Report. University of Michigan.

Kelly, P.J., 2014. Information efficiency and firm-specific return variation. Quarterly Journal of Finance 4, 1450018.

Kelly, B., Ljungqvist, A., 2012. Testing asymmetric-information asset pricing models. Review of Financial Studies 25, 1366–1413.

Keown, A.J., Pinkerton, J.M., 1981. Merger announcements and insider trading activity: An empirical investigation. Journal of Finance 36, 855–869.

Kumar, A., 2009. Who gambles in the stock market?. Journal of Finance 64, 1889–1933.

Kumar, A., Page, J.K., Spalt, O.G., 2011. Religious beliefs, gambling attitudes, and financial market outcomes. Journal of Financial Economics 102, 671–708.

Kyle, A.S., 1984. Market structure, information, futures markets, and price formation. In: Storey, G.G., Schmitz, A., Sarris, A.H. (Eds.), International Agricultural Trade: Advanced Readings in Price Formation, Market Structure, and Price Instability. Westview Press, Boulder and London, pp. 45–64.

Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica: Journal of the Econometric Society, 1315–1335.

Kyle, A.S., 1989. Informed speculation with imperfect competition. Review of Economic Studies 56, 317–356.

Lakonishok, J., Shleifer, A., Vishny, R.W., 1994. Contrarian investment, extrapolation, and risk. Journal of Finance 49, 1541–1578.

Lee, D.W., Shin, H.H., Stulz, R.M., 2018. Does Capital Flow More to High Tobin's Q Industries?. Working paper.

Lehmann, B., 1990. Fads, martingales, and market efficiency. Quarterly Journal of Economics 105, 1–28.

Lo, A.W., MacKinlay, A.C., 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. Review of Financial Studies 1, 41–66.

Luo, Y., 2005. Do insiders learn from outsiders? Evidence from mergers and acquisitions. Journal of Finance 60, 1951–1982.

Malkiel, B.G., 2005. Reflections on the efficient market hypothesis: 30 years later. Financial Review 40, 1–9.

Mandelbrot, B., 1966. Forecasts of future prices, unbiased markets, and" martingale" models. Journal of Business 39, 242–255.

McLean, R.D., Zhang, T. and Zhao, M., 2012. Why does the law matter? Investor protection and its effects on investment, finance, and growth. Journal of Finance 67, 313–350.

Mikesell, J.L., 1994. State lottery sales and economic activity. National Tax Journal 47, 165–171.

Mitton, T., Vorkink, K., 2007. Equilibrium underdiversification and the preference for skewness. Review of Financial Studies 20, 1255–1288.

Morck, R., Yeung, B., Yu, W., 2000. The information content of stock markets: Why do emerging markets have synchronous stock price movements?. Journal of Financial Economics 58, 215–260.

Morck, R., Yeung, B., Yu., W., 2013. R2 and the economy. Annual Review of Financial Economics 5, 143–166.

Nagel, S., 2012. Evaporating liquidity. Review of Financial Studies 25, 2005–2039.

Pastor, L., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685.

Peters, R.H. and Taylor, L.A., 2017. Intangible capital and the investment-Q relation. Journal of Financial Economics 123, 251–272.

Raylu, N., Oei, T.P., 2004. Role of culture in gambling and problem gambling. Clinical Psychology Review 23, 1087–1114.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. Journal of Finance 39, 1127–1139.

Roll, R., 1988. R2. Journal of Finance 43, 541–566.

Rosch, D., Subrahmanyam, A., van Dijk , M., 2017. The dynamics of market efficiency. Review of Financial Studies 30, 1151–1187.

Rozeff, M.S., 1984. Dividend yields are equity risk premiums. Journal of Portfolio Management, 68–75.

Rozeff, M.S., Kinney Jr, W.R., 1976. Capital market seasonality: The case of stock returns. Journal of Financial Economics 3, 379–402.

Scholes, M., Williams, J., 1977. Estimating betas from nonsynchronous data, Journal of Financial Economics 5, 309–327.

Shiller, R., 1981. Do stock prices move too much to be justified by subsequent changes in dividends? American Economic Review 71, 421–436.

Stiglitz, J.E., Weiss, A., 1981. Credit rationing in markets with imperfect information. American Economic Review 71, 393–410.

Stoll, H.R., 1978. The supply of dealer services in securities markets. Journal of Finance 33, 1133–1151.

Subrahmanyam, A., 1991. Risk aversion, market liquidity, and price efficiency. Review of Financial Studies 4, 417–441.

Teoh, S.H., Yang, Y., Zhang, Y., 2007. R-square: Noise or firm-specific information. Working paper.

Tkac, P.A., 1999. A trading volume benchmark: Theory and evidence. Journal of Financial and Quantitative Analysis 34, 89–114.

Tobin, J., 1969. A general equilibrium approach to monetary theory. Journal of Money, Credit and Banking 1, 15–29.

Vuolteenaho, T., 2002. What drives firm-level returns? Journal of Finance 57, 233–264.

Walker, M., Schellink, T., and Anjoul, F., 2007. Explain why people gamble. In M. Zangeneh, A. Blaszczynski, & N.E. Turner (Eds.), In the pursuit of winning: Problem gambling theory, research and treatment, 11–20. New York: Springer.

Wurgler, J., 2000. Financial markets and the allocation of capital. Journal of Financial Economics 58, 187–214.

Zhang, Y.C., 1999. Toward a theory of marginally efficient markets. Physica A: Statistical Mechanics and its Applications 269, 30–44.