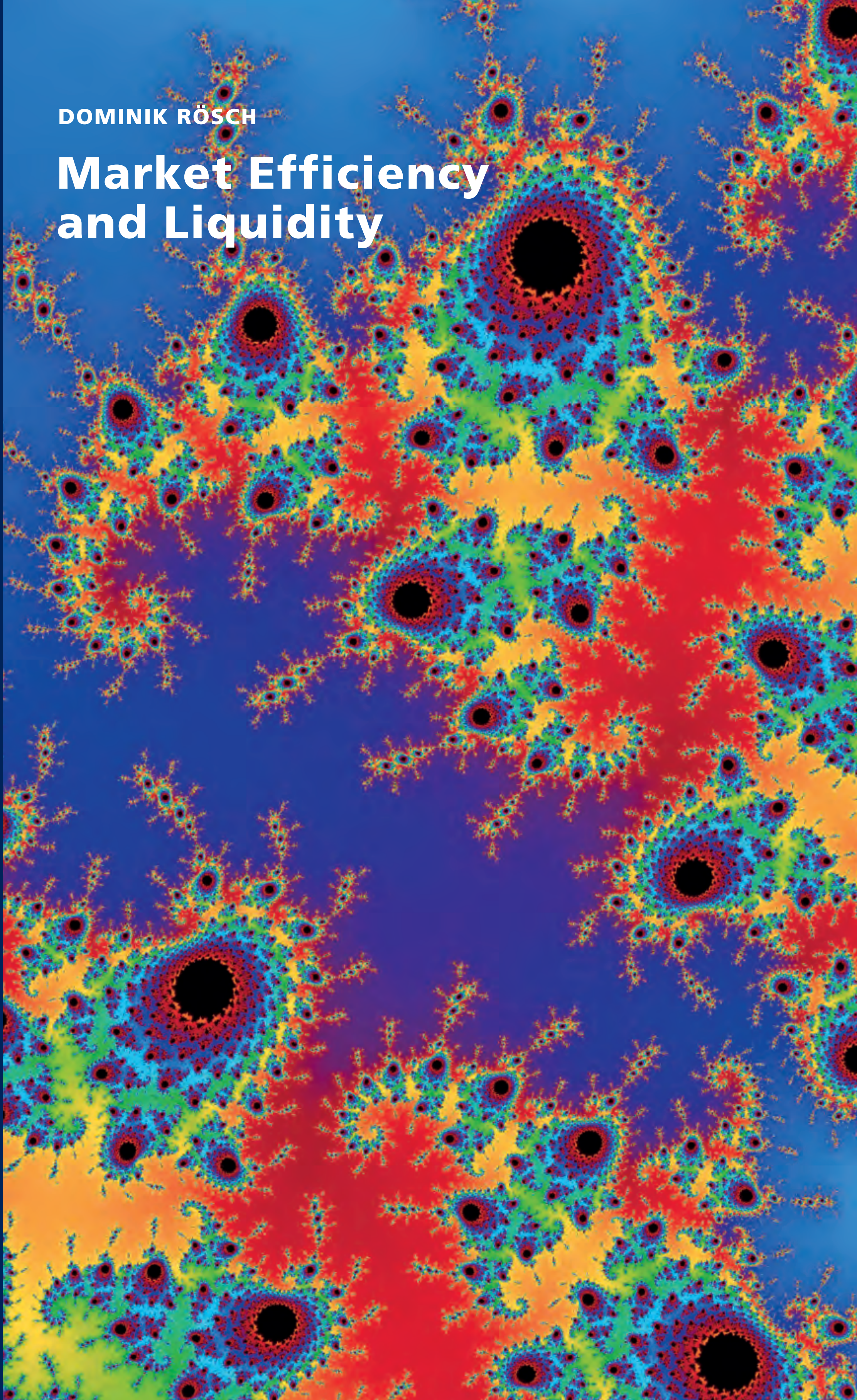
DOMINIK RÖSCH

Market Efficiency and Liquidity



Market Efficiency and Liquidity

Market Efficiency and Liquidity

Markt efficiëntie en liquiditeit

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board

The public defense shall be held on

Friday 18 December 2015 at 11:30 hrs

by

DOMINIK-MAXIMILIAN RÖSCH
born in Herdecke, Germany

Erasmus University Rotterdam

Erasmus

Doctoral Committee

Promotor: Prof.dr. M.A. van Dijk

Other members: Dr. D.G.J. Bongaerts

Prof.dr. J.J.A.G. Driessen

Prof.dr. A.J. Menkveld

Erasmus Research Institute of Management - ERIM
The joint research institute of the Rotterdam School of Management (RSM)
and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam
Internet: http://www.erim.eur.nl

ERIM Electronic Series Portal: http://repub.eur.nl/pub

ERIM PhD Series in Research in Management, 353
ERIM Reference number: EPS-2015-353-FA
ISBN 978-90-5892-425-4

Design: B&T Ontwerp en advies www.b-en-t.nl

This publication (cover and interior) is printed by haveka.nl on recycled paper, Revive© .
The ink used is produced from renewable resources and alcohol free fountain solution.
Certifications for the paper and the printing production process: Recycle, EU Flower, FSC, ISO14001.
More info: http://www.haveka.nl/greening

To all my ancestors, descendants, and everyone in between, especially to Kim, my wife.

Acknowledgments

This thesis would not have been possible without the help and support of many. In particular, I thank my supervisor, Mathijs van Dijk. Now, with the benefits of hindsight, I am aware of the importance of having the right supervisor. I am very lucky and fortunate that Mathijs taught me things I needed to learn and did not try to teach me what I already knew (except, whenever I thought I knew, but didn't). The very right supervision, indeed.

I also want to thank the other members of my committee, Dion Bongaerts, Joost Driessen, and Albert Menkveld for their feedback. Feedback is crucial and I thank everyone, who was willing to share their opinion with me.

Dominik-Maximilian Rösch
Rotterdam, The Netherlands
May, 2015

Contents

Chapter 1

Introduction

The wealth of nations is determined by the efficient usage of real assets, such as land, machinery, and natural resources. Financial assets, such as stocks and bonds, merely represent claims on these real assets.

Nevertheless, financial markets serve many important roles: they allow investors to optimize their reward to risk ratio, households to shift consumption over time, aggregate investor beliefs, and help to shift scarce resources to their optimal use (e.g., see Levine, 2005).

Through all these roles financial markets can improve the efficient usage of real assets and thereby have an effect on the real economy and improve economic growth. For example, if investors believe that a certain company has large investment opportunities, investors will buy the stock leading to an increase in the share price. This increase in the share price will make it easier for the company to raise further capital which it then can use to pursue these investments. Through this channel, financial markets play an important role in allocating scarce capital across companies.

But the efficacy of all of these roles crucially depends on prices of financial assets reflecting the true value of these assets and how well the market facilitates trading these assets. In other words, the efficacy of these roles depends on the financial market being informationally efficient and liquid. The main theme of this thesis is to investigate the interaction between market efficiency and liquidity.

Market efficiency

According to Fama (1970), an informational efficient market is "a market in which prices always 'fully reflect' available information" (p. 383). An efficient market "provide(s) accurate signals for resource allocation" (Fama and Miller, 1972, p. 335).

In his seminal classification, Fama (1970) distinguishes market efficiency based on the type of information that is reflected in prices. In particular, Fama distinguishes between semi-strong-form efficiency, in which prices reflect all public information, and strong-form efficiency, in which prices reflect all public, as well as all private information.

There is evidence that markets are not strong form efficient (e.g., see Jaffe, 1974), but a debate whether markets are semi-strong-form efficient. This debate is reflected in the work by Fama and Shiller (e.g., Fama, 1970, 1991; Shiller, 1981) who both shared the 2013 Nobel Prize in Economic Sciences with Hansen.

One of the difficulties in determining whether prices are efficient is that they need to be risk-adjusted. For example, Banz (1981) provides evidence that shares of small companies generally outperform those of big companies, but this does not necessarily mean that markets are inefficient. Investors need to properly risk-adjust the returns: holding shares of small companies, in general, may well be more risky than holding stocks of large companies, and therefore returns of small companies would be expected to be higher. This issue is often referred to as the joint-hypothesis problem (e.g., see Fama, 1991).

Trying to avoid this joint-hypothesis problem, recent empirical studies focus on deviations from the law of one price, that similar assets have similar prices, as a more direct measure for market inefficiencies (e.g., Mitchell, Pulvino, and Stafford, 2002; Lamont and Thaler, 2003; Roll, Schwartz, and Subrahmanyam, 2007; De Jong, Rosenthal, and van Dijk, 2009; Gagnon and Karolyi, 2010b). But again, observing deviations from the law of one price is not necessarily evidence of market inefficiencies. For example, Fama (1991) states that "a market is efficient, if prices reflect information to the point where the marginal benefits of acting ... do not exceed the marginal costs" (p. 1575, quoting Jensen, 1978). Following this interpretation, the market might have been inefficient when the deviation of the law of one price arose, but might well be perfectly efficient afterwards, if the costs of acting exceed the benefits. Yet, calling such a market efficient seems less than ideal, because trading costs do not necessarily lead to inefficiencies (Kyle, 1985) and because regardless of why the mispricing persists such a market can not efficiently serve its deeper roles.

Nevertheless, any study about market efficiency is also a study about trading frictions such as trading costs and market illiquidity.

Market liquidity

Illiquidity as a trading friction can not only explain why deviations from the law of one price persist, but also plays an important role in how financial markets affect the real economy. The liquidity of a market is often defined by the ease at which it allows trading, in particular liquid markets allow immediate trading of large volumes at low costs.

Wurgler (2000) provides empirical evidence that developed financial markets allow better allocation of capital. Because better developed financial markets are in general more liquid, prices are more informative about company-specific investment opportunities. More informative prices allow investors to better distinguish between good and bad investments. Wurgler (2000) concludes that the "most liquid financial markets in the world are also the ones that allocate capital most efficiently" (p. 190).

Liquid markets not only improve the efficiency of capital allocation but also directly reduce the cost of capital companies face. In their seminal study, Amihud and Mendelson (1986) find a positive relation between the illiquidity of a share and its expected return. In other words, investors discount the current share price to compensate for its illiquidity. This illiquidity premium hence increases the cost for companies to get equity funding and might result in fewer investments and hence lower economic growth. Following their seminal study, several other researchers provided evidence that liquidity is priced as a characteristic (Brennan and Subrahmanyam, 1996) as well as a source of systemic risk (Chordia, Roll, and Subrahmanyam, 2000; Pastor and Stambaugh, 2003; Acharya and Pedersen, 2005; Sadka, 2006).

Market liquidity and market efficiency

There are many finance academics that documented large time- and cross-sectional variation in trading frictions impeding trading by informed investors [such as market illiquidity in Benston and Hagerman (1974); Chordia, Roll, and Subrahmanyam (2001)], but at the same time treated market efficiency as a static concept (recent exceptions include Chordia, Roll, and Subrahmanyam, 2008; Boehmer and Kelley, 2009; Chordia, Roll, and Subrahmanyam, 2011; Mitchell and Pulvino, 2012; Hu, Pan, and Wang, 2013; Pasquariello, 2014). This seems at odds, because both efficiency and liquidity are intimately related: Arguably markets are not efficient per se, but require trading against potential inefficiencies by informed investors, whose success depends on the ease at which they can trade (market liquidity) and on their available capital (funding liquidity) (e.g., Chordia, Roll, and Subrahmanyam, 2008).

To better understand the possible interaction between market efficiency and liquidity it is important to understand why markets can neither be perfectly efficient nor liquid. Prices cannot always reflect all possible information, as in such a market informed traders would "make no (private) return from their (privately) costly activity" (Grossman and Stiglitz, 1980, p. 393) and hence would pursue other activities, leaving prices less informative. Similarly, markets cannot be perfectly liquid.

Investors that want to sell an asset first need to find an investor, who is willing to buy.

Because this matching process is not always easily achieved, intermediaries, so called liquidity providers, step in and provide immediacy by buying the asset from the first investor and holding the asset in their inventory till the second investor arrives in the market. Classic market-microstructure models provide three reasons why markets are less than perfectly liquid. First, liquidity providers face fixed costs and demand compensation for providing their service. Second, risk-averse liquidity providers face inventory risk, that the asset they buy decreases in value till they are able to sell it again (Stoll, 1978). Third, liquidity providers face adverse selection risk, the probability of trading against informed investors (Glosten and Milgrom, 1985).

In other words, while the trading of informed investors is crucial for prices to be informational efficient, at the same time, informed investors can also decrease liquidity, because liquidity providers face larger adverse selection risk.

To understand the potential interaction between market efficiency and liquidity it is crucial to understand how inefficiencies arise. If, for example, inefficiencies arise as a result of demand pressure, informed investors trade against market demand and thereby decrease inventory holding costs for liquidity providers, which improves liquidity (Gromb and Vayanos, 2010). But if inefficiencies arise as a result of differences in information then trading by informed traders is "toxic" (Foucault, Kozhan, and Tham, 2013), increases adverse selection risk, and lowers liquidity.

In the remainder of this introduction, I will provide a short introduction into the three different chapters of my thesis.

Chapter 2: An empirical analysis of co-movement in market efficiency [1]

In my second chapter, we start with exploring the idea that the degree of financial market efficiency not only varies over time, but also across stocks, and analyze co-movement in the time-varying efficiency of individual stocks. Using five stock-level measures of price efficiency, we find evidence of significant co-movement in efficiency. The degree of co-movement in efficiency is greater for more liquid stocks and varies considerably over time. In vector autoregressions, we show that shocks to funding liquidity (the TED spread), hedge fund flows, and a proxy for algorithmic trading significantly affect the degree of co-movement in efficiency. Overall, our results imply that stock price efficiency has a component that is prone to systematic improvement and deterioration, and that events and policies that impact funding

[1] This chapter is based on Rösch, Subrahmanyam, and van Dijk (2015) "An Empirical Analysis of Co-Movement in Market Efficiency" available at http://ssrn.com/abstract=2062926. For this chapter I developed several of the hypotheses and empirical tests, carried out all of the data collection and screening as well as all of the empirical analyses, and did some of the writing.

liquidity can affect the degree of co-movement in efficiency.

Co-movement in efficiency give rise to the notion of efficiency risk that is in part systematic and may be priced in the cross-section (as empirically found by Pasquariello (2014)). It also suggests that informed investors trade against inefficiencies across the whole market, rather than focus on specific segments.

Chapter 3: The impact of arbitrage on market liquidity [2]

In my third chapter, I focus on deviations from the law of one price as a measure of market inefficiency. I am especially interested to investigate whether arbitrageurs provide liquidity and thereby would not only improve the efficiency of the market but also its liquidity. Similar arguments why markets are not perfectly efficient (such as a lack of available capital), can explain why deviations of the law-of-one price persist. But much less is known why deviations from the law-of-one price arise. The reasons why these deviations arise is important, because theory predicts that the impact of trading against these deviations on liquidity depends on why these deviations arise. Theory predicts that arbitrage improves financial market liquidity when arbitrage opportunities arise as a result of temporary demand shocks and worsens liquidity when arbitrage opportunities arise as a result of differences in information. In this paper, I study the impact of arbitrage in Depositary Receipts (DRs) on market liquidity, using tick-by-tick data from the U.S. and five different home markets from 1996 to 2013. My analysis suggests that around 70% of the arbitrage opportunities in DRs arise as a result of demand shocks. Consistent with theory, I then show that an increase in arbitrage activity is associated with a reduction in market order imbalance and an improvement in liquidity. My results are robust to different proxies for arbitrage activity, different methodologies, and to instrumental variable tests. Overall, these findings indicate that arbitrageurs tend to trade against market order imbalance and thus enhance market integration and liquidity.

Chapter 4: Cross-sectional identification of informed trading [3]

In the fourth chapter, we present a new approach to measure the probability of informed trading. We propose to measure informed trading in individual securities based on a portfolio optimization model for investors facing information and liquidity shocks. These shocks induce speculative and liquidity-motivated order flow, taking into account the price impact of trading. The model allows us to back out the amount of informed trading from a security's aggregate

[2] This chapter is based on Rösch (2014) "The Impact of Arbitrage on Market Liquidity" available at http://ssrn.com/abstract=2295437.

[3] This chapter is based on Bongaerts, Rösch, and van Dijk (2014) "Cross-Sectional Identification of Informed Trading" available at http://ssrn.com/abstract=2532128. For this chapter I developed part of the theoretical analysis and most of the hypotheses and empirical tests, carried out a substantial part of the data collection and screening as well as all empirical analyses, and did some of the writing.

order flow, based on the cross-section of price impact parameters (λ) and order imbalances (OIB). Furthermore, we obtain a very simple expression for a security's aggregate private information shock: its $\lambda \times$ *OIB*, in excess of the same term for a benchmark security that is insulated from informed trading. We validate our private information measure (based on daily data for all S&P 1500 stocks over 2001-2010) by showing that it is strongly related to contemporaneous returns, and that return reversals are significantly weaker following stock-days with high private information estimates.

Taken together the results of these studies indicate that financial market efficiency (Chapter 2), arbitrage activity (Chapter 3), and informed trading (Chapter 4) varies over time. In Chapter 2 we provide evidence indicating that shocks to funding liquidity and arbitrage activity have a significant effect on the degree of co-movement in efficiency. These results indicate a time-varying component in the degree of pricing efficiency of individual stocks, partly driven by changes in funding liquidity and the intensity of arbitrage activity. On the other hand, results in Chapter 3 indicate that arbitrage activity not only improves market efficiency, but can also improve market liquidity by shifting excess trading demands across markets.

These results shed additional light on possible consequences of frictions impeding arbitrage, such as short-selling bans or transaction taxes. To curb excessive trading eleven European member states plan to introduce a transaction tax. The tax will likely have an adverse effect on arbitrage activity which might increase co-movement in individual stock efficiency and deteriorate market efficiency and liquidity.

Several open questions remain. Of particular interests are asset pricing tests, whether the co-movement in efficiency leads to a priced risk factor and whether our proxy for informed trading from Chapter 4 is priced.

Further, the effect of arbitrage activity on market liquidity seems to deserve more attention. For example, arbitrageurs seem to dampen the effect of liquidity shocks in the depositary receipt market and thereby might also decrease liquidity risk for depositary receipts. It also seems worthwhile to investigate the effect of arbitrage activity in other markets, such as in the options markets. Several recent studies (e.g., Lin, Lu, and Driessen, 2013) provide evidence that informed traders prefer the option market, and hence arbitrage opportunities might arise more often because of informational differences. In this case arbitrage might harm the liquidity of the underlying stock.

Chapter 2

An Empirical Analysis of Co-Movement in Market Efficiency*

2.1 Introduction

For most of its life, the finance profession has treated financial market efficiency as a static concept. The seminal taxonomy in Fama (1970) of weak-, semi-strong, and strong-form efficiency inspires debate on which of these best describes financial markets, but this debate has paid little heed to the possibility that the degree of market efficiency varies through time. Yet, there are sound reasons to expect such time variations. Market efficiency is enforced in part by way of arbitrage, the efficacy of which is influenced by financial frictions (such as limited capital, transaction costs, short-sales constraints, and idiosyncratic risk) whose severity varies considerably over time. Indeed, recent studies show that the efficiency of financial markets is dynamic in nature and that it varies through time with financial market liquidity and constraints on arbitrage capital (see, e.g., Boehmer and Kelley (2009); Mitchell and Pulvino (2012); Hu et al. (2013); Pasquariello (2014)).

* This chapter is based on Rösch, Subrahmanyam, and van Dijk (2015) “An Empirical Analysis of Co-Movement in Market Efficiency” available at http://ssrn.com/abstract=2062926. We thank Yakov Amihud, Tarun Chordia, Carole Comerton-Forde, Thierry Foucault, Amit Goyal, Terry Hendershott, Craig Holden, Sreeni Kamma, Andrew Karolyi, Ed Lin, Marc Lipson, Steve Mann, Christophe Pérignon, Veronika Pool, Vikas Raman, Raghu Rau, Matti Suominen, Kumar Venkataraman, Avi Wohl, Hong Yan, and participants at the 2012 Brazilian Finance Conference (São Paulo), the 2012 EFMA meetings (Barcelona), the 2012 Frontiers of Finance Conference (Warwick Business School), the 2013 Campus for Finance conference (WHU Otto Beisheim School of Management), the 2013 EFA meetings (Cambridge), and at seminars at Deakin University, Erasmus University, Goethe University Frankfurt, Indiana University, UCLA Anderson, University of Cambridge, University of Manchester, and University of South Carolina for valuable comments. This work was carried out on the National e-infrastructure with the support of SURF Foundation. We thank SURFsara, and in particular Lykle Voort, for technical support on computing and storage, and OneMarketData for the use of their OneTick software. Van Dijk gratefully acknowledges financial support from the Vereniging Trustfonds Erasmus Universiteit Rotterdam and from the Netherlands Organisation for Scientific Research through a “Vidi” grant.

Similarly, the efficiency of price formation is likely to vary across individual securities, since there is considerable cross-sectional heterogeneity in various attributes that affect the efficacy of arbitrage. For example, Benston and Hagerman (1974) and Nagel (2005) document considerable cross-sectional variation in stock-level illiquidity and short-sales constraints, respectively.

The idea that efficiency varies both over time and across stocks raises the question to what extent time-variation in market efficiency co-moves across individual stocks. And, if there is evidence of significant co-movement in efficiency, what are the fundamental forces (such as funding liquidity or other factors that affect the efficacy of arbitrage) that drive it? These questions are relevant since investors, exchange officials, and policy-makers should care about whether the efficiency of financial markets is prone to fluctuation in a systematic way, and about what factors influence the degree of such common variation.

Motivated by the above observations, in this paper, we do the following. We first compute daily market efficiency estimates for individual stocks based on five measures: intraday return predictability based on past order flow (Boehmer and Wu, 2007), variance ratios (Lo and MacKinlay, 1989; Bessembinder, 2003), the variance of intraday returns (Bessembinder, 2003), intraday Hasbrouck (1993) pricing errors, and put-call parity deviations in the corresponding options markets (Finucane, 1991; Cremers and Weinbaum, 2010) using all NYSE stocks over an extended sample period of fifteen years (based on data on 14.3 billion transactions in total).

We then construct market-wide measures of efficiency from these stock-level measures and, each month for each stock, estimate the degree of co-movement in efficiency as the R^2's from regressions of the daily stock-level measures on the market-wide measures. These analyses show that time-variation in market efficiency has a material common component across stocks, which indicates that market efficiency is prone to improvement and deterioration in a systematic way. We also find that the degree of co-movement in efficiency is considerably greater for liquid stocks than for illiquid stocks, and that it exhibits substantial time-variation.[1]

Our next goal is to analyze the economic forces that drive time-variation in the degree of co-movement in market efficiency. In particular, we are interested in whether variation in funding liquidity and other determinants of the efficacy of arbitrage affects the degree of co-movement in efficiency. We hypothesize that changes in funding liquidity and the overall intensity of arbitrage activity affect the price efficiency of many stocks at the same time,

[1] We make sure that our estimates of co-movement in stock-level efficiency are not simply a different manifestation of co-movement in stock-level liquidity by orthogonalizing stock-level efficiency with respect to stock-level liquidity before running the regressions to estimate co-movement in efficiency.

thereby elevating the degree of co-movement in stock-level efficiency.

We first create an aggregate measure of market-wide co-movement in efficiency as follows. For each of the five efficiency measures, we construct a monthly, market-wide measure of the degree of co-movement in efficiency as the equally-weighted average of the monthly R^2's from the co-movement regressions of individual stocks. We then extract the first principal component from the five resulting monthly, market-wide co-movement in efficiency measures and use it as the main variable of interest in vector autoregressions (VARs). As other endogenous variables, we include changes in the TED spread (a common indicator of funding liquidity), hedge fund flows (a proxy for changes in the amount of capital available for arbitrage activity), and the total number of quote updates divided by aggregate dollar trading volume (a proxy for algorithmic trading, inspired by Boehmer et al. (2014)). Since market efficiency is linked to market liquidity, we are careful to also include the degree of co-movement in market liquidity as an endogenous variable and to allow it to affect co-movement in efficiency in all of our VARs.

We find that shocks to funding liquidity and to variables that proxy for the intensity of arbitrage activity have a significant impact on the degree of co-movement in efficiency. In particular, shocks to the TED spread and to hedge fund flows positively affect the degree of co-movement in efficiency in the subsequent month, while a shock to algorithmic trading has a positive contemporaneous effect on efficiency co-movement. These effects are over and beyond the impact of shocks to these variables on co-movement in liquidity and are stronger for illiquid stocks than for liquid stocks. These results indicate that funding liquidity and the intensity of arbitrage activity are important forces that help us understand time-variation in the degree of co-movement in efficiency.

To the best of our knowledge, this paper is the first to study (the determinants of) co-movement in the time-varying pricing efficiency of individual stocks. We view our analysis of co-movement in market efficiency as relevant for at least three reasons. First, we show that market efficiency, rather than being a static concept, exhibits significant time-variation and co-moves across individual stocks. This result is important for academic research, since most theoretical models in finance rely on efficient financial market prices. Studying co-movement in market efficiency enhances our understanding of the extent to which the data are consistent with these models, and of whether deviations from the assumption of efficient pricing exhibit systematic variation across individual securities.

Second, we go beyond the well-known link between funding liquidity and market liquidity and demonstrate a further connection between funding liquidity and the degree of co-movement in the efficiency of price formation. The latter result suggests that policy attempts to increase funding liquidity may not only have a direct impact on trading costs, but also systematically affect the efficiency of stock market prices.

Third, our results provide a natural underpinning of Pasquariello (2014) finding that a measure of "financial market dislocations" (constructed as an average of violations of arbitrage parities in stock, foreign exchange, and money markets) is a priced factor in the cross-section of stock returns. Our analysis indicates that the degree of price efficiency of individual stocks is uncertain, and that this uncertainty cannot be fully diversified away across individual stocks, which suggests that "efficiency risk" is in part systematic and may be priced in the cross-section.

This paper is organized as follows. In Section 2.2, we discuss the estimation of the efficiency measures for individual stocks. Section 2.3 presents the sample and the estimates of the efficiency measures. In Section 2.4, we document co-movement in these measures across individual stocks and across portfolios of stocks. Section 2.5 analyzes determinants of time-series variation in the degree of co-movement in efficiency. Section 2.6 concludes.

2.2 Efficiency measures

Our analysis is based on five measures of price efficiency that we estimate each day for each stock: intraday return predictability based on past order flow, variance ratios, the variance of intraday returns, Hasbrouck (1993) pricing errors, and put-call parity deviations. In this section, we explain in detail how we estimate each of these measures.

2.2.1 Intraday return predictability

Our first efficiency measure is based on the intraday predictability of individual stock returns from past order flow. Several papers, including Hasbrouck and Ho (1987), Chan and Fong (2000), Chordia et al. (2005), and Boehmer and Wu (2007) explore and provide evidence of such return predictability, which we use as an inverse indicator of market efficiency. Chordia et al. (2005) argue that such predictability arises from a temporary disequilibrium because of dealers' inability to accommodate autocorrelated order imbalances. Their evidence suggests that trading by astute arbitrageurs removes all return predictability over intervals of five minutes or more, but some predictability remains at shorter horizons.

In line with these prior studies, we estimate the intraday return predictability of each individual stock for each day in the sample based on regressions of stock returns over short intervals within the day on order imbalance (dollar volume of buyer- minus seller-initiated trades) in the previous interval. Chordia et al. (2005) show that prices cease to be predictable from order flow in 30 minutes or less in 1996, and in around five minutes in 2002. Since our sample period lasts till 2010, we have to use intervals shorter than five minutes to still capture meaningful predictability in the later part of the sample period. In light of this consideration, we estimate predictability based on intraday returns and order imbalances measured over one-minute intervals (with a robustness check based on two-minute intervals).

We estimate the extent of short-horizon return predictability from order flow for each stock i and day d in the sample as the R^2 from the following regression, using intraday data aggregated over one-minute intervals:

$$R_{i,d,t} = a_{i,d} + b_{i,d} OIB_{i,d,t-1} + \epsilon_{i,d,t}, \tag{2.1}$$

where $R_{i,d,t}$ is the return of stock i in one-minute interval t on day d based on the mid-quote associated with the last trade to the mid-quote of the first trade in the interval (we use mid-quote returns to avoid the bid-ask bounce), and $OIB_{i,d,t-1}$ is the order imbalance for the same stock and day in the previous interval $t-1$, computed as the difference between the total dollar volume of trades initiated by buyers and sellers ($OIB\$$). A lower R^2 from the regression in Eq. (2.1) indicates greater efficiency. We refer to the efficiency measure based on this regression specification as the *OIB predictability* measure.

To assess the robustness of our results to changes in the specification of the predictability regressions, we also estimate four alternative return predictability measures, each named after the single feature that distinguishes it from the *OIB predictability* measure. The *allquotes* measure is based on returns computed using all quotes within each interval rather than only using quotes associated with trades; the *2minutes* measure is based on two-minute instead of one-minute intervals; and the *oib#* measure is based on order imbalance expressed in number of trades rather than dollars. We also present and discuss the results using the R^2 from regressions of one-minute returns on their one-minute lagged counterparts, instead of past order flows, and label this the *autocorrelation* measure. We discard stock-days with fewer than 20 observations for each of these measures. In our analyses of co-movement in market efficiency, we use a general *Predictability* measure that is constructed as the first principal component across the five alternative return predictability measures (more details are provided below).

2.2.2 Variance ratios

The second efficiency measure we consider is a daily variance ratio that examines how closely the price of individual stocks adheres to a random walk benchmark, in line with, among others, Bessembinder (2003). The stock-level *Variance ratio* measure is defined as $|1 - 30 \times Var(1\mathrm{m}in)/Var(30\mathrm{m}in)|$, where $Var(1\mathrm{m}in)$ is the return variance estimated from one-minute mid-quote returns within a day and $Var(30\mathrm{m}in)$ is the return variance estimated from 30-minute mid-quote returns within a day. Variance ratios are computed from mid-quote returns and do not utilize traded prices, mitigating the problem of non-synchronous trading. Since estimates of daily variance ratios of individual stocks can be noisy (Andersen et al., 2001), we follow Lo and MacKinlay (1989) (see their equation (5)) and Charles and Darné (2009) and estimate daily variance ratios based on overlapping intraday returns. Since expected returns over such short intervals are very close to zero, we set expected returns to zero in the computation of the variances. We discard stock-days with fewer than 20 non-zero one-minute returns. The variance ratio tends to unity as serial dependence in asset returns tends to zero as per Bessembinder (2003); therefore, it measures how closely the price adheres to a random walk.

2.2.3 Variance of intraday returns

Motivated by Bessembinder (2003), we include the variance of intraday returns as a third measure for the quality of price formation of individual stocks. Bessembinder argues that intraday return volatility is an important inverse indicator of price formation quality. We estimate the intraday return volatility of individual stocks each day as the variance of one-minute mid-quote returns and refer to this variable as the *Variance* measure. We discard stock-days with fewer than 20 non-zero one-minute returns for this measure.

2.2.4 Hasbrouck pricing errors

As a fourth efficiency measure, we estimate Hasbrouck (1993) pricing errors based on intraday trades and quotes. Hasbrouck proposes a method to decompose stock prices into random walk and stationary components. He refers to the stationary component (the difference between the efficient price and the actual price) as the pricing error, which he argues is a natural measure for price efficiency. We follow Hasbrouck and estimate vector autoregression (VAR) models to estimate these components. As in Boehmer and Kelley (2009), we estimate a five-lag VAR model based on intraday data for each stock-day with at least one hundred trades. The endogenous variables of the model are: (i) the logarithmic price return, from quote midpoints associated with trades,[2] (ii) a trade sign indicator, (iii) the signed volume (that is, the sign of

[2]Using mid-quote returns avoids the bid-ask bounce, but using returns from actual trade prices does not alter the main results.

the trade times the number of shares traded), and (iv) the sign of the trade times the square root of the number of shares traded. We sign all trades with trade prices above the prevailing quote midpoint as buyer-initiated, and seller-initiated if they are below the quote midpoint. If the trade occurred at the prevailing quote midpoint we set the sign of the trade to zero (following Hasbrouck, 1993). As in Hasbrouck (1993), we also set all lagged variables at the beginning of each day to zero. We obtain the pricing error of each trade in a stock on a given day from the vector moving average representation of the VAR system (Beveridge and Nelson, 1981) using Eq. (13) in Hasbrouck (1993). We take the maximum of the absolute pricing errors of the trades in a stock on a given day as an inverse measure of the informational efficiency for that stock on that day and label it the *Hasbrouck* measure. Since daily, stock-level estimates of the maximum intraday pricing error exhibit several large outliers, we use the logarithmic transformation of *Hasbrouck* to mitigate their influence.[3]

2.2.5 Put-call parity deviations

Our fifth proxy for the price efficiency of individual stocks is a law of one price measure derived from options markets. The use of this measure enhances our understanding of co-movement in market efficiency by extending the notion of efficiency to derivatives markets for individual stocks. This *Put-call parity* measure is estimated using the OptionMetrics database as the absolute difference between the implied volatilities of a call and a put option of the same series (i.e., pairs of options on the same underlying stock with the same strike price and the same expiration date).[4] We use end-of-day quotes from all option series with positive implied volatilities that expire in two weeks to one year and that have a strike-to-spot ratio between 0.95 and 1.05. We impose these conditions to ensure that our estimates of put-call parity deviations are based on near-the-money and relatively short maturity options, which are typically the most liquid options (following Pan (2002)). When more than one option pair satisfies these conditions for a given stock-day, we take the average of the absolute differences between the implied volatilities of the call and the put option across all option pairs.

[3] In unreported tests, we obtain similar results when we follow Boehmer and Kelley (2009) and use the daily standard deviation (instead of the daily maximum) of the intraday pricing errors as an inverse measure of price efficiency.

[4] This measure is also used in Cremers and Weinbaum (2010). These authors note that while, strictly speaking, put-call parity does not hold as an equality for the American options on individual stocks, a lower discrepancy in implied volatilities from binomial models nonetheless is indicative of more efficient options and stock markets.

2.3 Sample and efficiency estimates

To estimate the five efficiency measures, we obtain data on all trades and quotes as well as their respective sizes for individual U.S. stocks from the Thomson Reuters Tick History (TRTH) database, which contains global tick-by-tick trade and quote data across asset classes. TRTH is increasingly used in studies on high-frequency data, see, e.g., Lau et al. (2012); Marshall et al. (2012); Lai et al. (2014).[5] Our data start in March 1996, which is the earliest month available in the TRTH database. Our sample consists of all NYSE stocks that were traded at any time during our sample period from March 1996 to December 2010 and that survive our data screens. We include only NYSE stocks to prevent issues with differences in trading volume definitions across NYSE and Nasdaq, see, e.g., Gao and Ritter (2010). We use trades and (national best bid and offer) quotes on all U.S. exchanges on which these NYSE stocks are traded. We apply a variety of filters to the data that are described in Appendix A.1. We are able to use 14,253,093,209 transactions, signed by the Lee and Ready (1991) method, in our analyses.[6] Our final sample includes 2,157 NYSE stocks.

To estimate the predictability regressions in Eq. (2.1), we require at least one signed trade in both the interval over which we calculate the return as well as the previous interval. This leads us to drop a non-negligible fraction of the intraday intervals in the early years of the sample period, but since the year 2000 almost all stocks have at least one trade in almost all of the intraday intervals. We discard stock-days for which we have fewer than 20 one-minute intervals with valid data on the stock return within that interval and on the order imbalance or return in the preceding interval (in total 756,051 stock-day observations), and days for which TRTH reports a data gap that overlaps with the continuous trading session (in total 56 days). Our data filters allow us to estimate Eq. (2.1) for on average 1,711 days over the period 1996-2010 across the 2,157 stocks in our sample.

[5] To verify that our results do not depend on using TRTH instead of NYSE's Trade and Quote (TAQ) database, we compare the results based on TRTH to those based on TAQ for all 2,023 NYSE-listed common stocks that were traded at any time over the period 1996-2000 and find that they are very similar. For example, the pooled correlations between the input variables for the intraday return predictability regressions (as reported in Table 2.1: number of trades, trading volume, average one-minute mid-quote returns, average one-minute order imbalance in number of trades, and average one-minute order imbalance in US$) range from 97.9% to 99.9% for these five variables. Further details are available from the authors.

[6] The Lee/Ready algorithm classifies a trade as buyer- (seller-)initiated if it is closer to the ask (bid) of the prevailing quote. If the trade is exactly at the midpoint of the quote, the trade is classified as buyer- (seller-)initiated if the last price change prior to the trade is positive (negative). Of course, there is inevitably some assignment error, so the resulting order imbalances are imperfect estimates, see, e.g., Aitken and Frino (1996); Ellis et al. (2000); Theissen (2001) for evidence on the accuracy of the Lee/Ready algorithm for stocks traded on the Australian Stock Exchange, Nasdaq, and the Frankfurt Stock Exchange, respectively. Lee and Radhakrishna (2000); Odders-White (2000) indicate that the Lee/Ready algorithm is quite accurate for NYSE stocks, suggesting that assignment errors should have minimal impact on the results.

Table 2.1 – Summary statistics of input variables for intraday return predictability regressions

This table reports the cross-sectional (across the 2,157 NYSE stocks in the sample) mean, standard deviation ("SD"), first quartile ("25%"), median, and third quartile ("75%") of the time-series average by stock of the daily number of trades (*#trades*), daily trading volume in US$ billions (*dollar volume*), average one-minute mid-quote returns within the day in basis points (*1-min mid-quote return*), average difference between the total number of trades initated by buyers and sellers (order imbalance in number of trades) over one-minute intervals (*1-min oib#*), and the average difference between the total dollar volume of trades initiated by buyers and sellers (order imbalance in US$) over one-minute intervals (*1-min oib$*). The first column indicates the number of stocks over which the summary statistics are computed. The sample includes all 2,157 NYSE-listed common stocks from 1996 to 2010 that survive our data screens (described in the Appendix). Data to compute all variables are from TRTH.

	#Stocks	Mean	SD	25%	Median	75%
#trades	2,157	2,015	4,798	79	432	1,797
dollar volume	2,157	0.025	0.062	0.001	0.006	0.021
1-min mid-quote return	2,157	-0.007	0.255	-0.023	-0.001	0.011
1-min oib#	2,157	0.067	0.138	0.001	0.023	0.094
1-min oib$	2,157	3,052	7,077	4	520	2,616

Table 2.1 presents summary statistics of the return and order imbalance variables that serve as inputs to our predictability regressions. For these variables, the table reports cross-sectional summary statistics (the mean, standard deviation, as well as the median and the 25th and 75th percentiles) of the stock-by-stock time-series averages. The average number of trades per day is around 2,000. The average daily dollar trading volume is 0.025 or US$25m. The median one-minute mid-quote return is equal to -0.001 basis point, which corresponds to -0.4 basis points per day. There is a slight positive average order imbalance over the one-minute intervals in our sample.

Table 2.2 presents the results of the daily return predictability regressions estimated based on intraday data. As described in Section 1.1, the baseline predictability measure (*OIB predictability*) is obtained from regressions of one-minute mid-quote returns (computed using quotes associated with trades) on lagged dollar order imbalance. For robustness, we also estimate four alternative predictability measures: *allquotes*, *2minutes*, *oib#*, and *autocorrelation*.

Consistent with prior research, Table 2.2 shows that order imbalance positively predicts future returns over short intervals. The average coefficient on lagged order imbalance across the approximately 3,200,000 stock-day regressions ranges from 0.947 for the *oib#* measure to 6.169 for the *2minutes* measure. The return autocorrelation coefficient is also positive at 0.024. The first number in parentheses below the average coefficient ("t-stat avg") is the average t-statistic across all stock-day regressions. Although for all measures except perhaps one

Table 2.2 – Intraday return predictability regressions

This table reports the average results of the return predictability regressions from Eq. (2.1), estimated daily based on intraday data for each of the NYSE stocks in the sample. Each of the five columns presents the results of a different way to estimate the predictability of one-minute (or two-minute) returns from lagged order imbalance (*OIB*) or lagged returns: *OIB predictability*, *allquotes*, *2minutes*, *oib#*, and *autocorrelation*. Section 1.1 discusses all five return predictability measures in detail. The first number in each column is the average slope coefficient across all stock-day predictability regressions. The *OIB* coefficient is scaled by 10^9 for the *OIB predictability*, *allquotes*, and *2minutes* regressions and by 10^4 for the *oib#* regressions. The average *t*-statistics ("*t*-stat avg") and the average Newey-West (1994) *t*-statistics ("NW *t*-stat avg") are in parentheses below the coefficients. "% positive" is the percentage of positive coefficients, and "% + significant" is the percentage with *t*-statistics greater than 1.645 (the 5% critical level in a one-tailed test). Intercepts have been suppressed to conserve space. The last three rows report the average R^2 and adjusted R^2 across all regressions and the number of stock-day predictability regressions. The full sample includes all 2,157 NYSE-listed common stocks from 1996 to 2010 that survive our data screens (described in the Appendix). Data are from TRTH. Data to compute *Put-call parity* are from OptionMetrics.

Predictability measure:	*OIB predictability*	*allquotes*	*2minutes*	*oib#*	*autocorrelation*
OIB_{t-1}	4.380	3.792	6.169	0.947	
$Return_{t-1}$					0.024
t-stat avg	(1.254)	(0.779)	(0.863)	(1.852)	(0.370)
NW *t*-stat avg	(8.385)	(7.042)	(7.800)	(10.657)	(3.614)
% positive	81.61	72.03	74.62	88.71	58.00
% + significant	45.75	31.85	35.10	61.67	28.82
R^2	2.55	1.72	2.57	3.47	1.83
adj. R^2	1.60	0.75	1.45	2.51	0.79
# regressions	3,175,645	3,263,908	3,391,138	3,271,527	3,097,085

(*oib#*), the simple average t-statistic does not exceed critical values associated with conventional confidence levels, the t-statistics of the individual stock-day regressions can be based on as few as 20 intraday observations.

The second number in parentheses in each column ("NW t-stat avg"), is the Newey and West (1994) t-statistic computed based on the time-series of daily coefficient estimates of individual stocks, which is then averaged across stocks. These t-statistics thus exploit the power obtained from the time-series of predictability estimates obtained for each stock, while accounting for heteroskedasticity and autocorrelation using the Newey-West procedure with automatic lag selection. They are similar in spirit to the t-statistics used in the Fama and MacBeth (1973) approach, the difference being that the time-series of coefficient estimates is not obtained from cross-sectional regressions, but from intraday return predictability regressions estimated by stock-day. These average Newey-West t-statistics are highly significant for all five predictability regressions reported Table 2.2 and indicate that intraday returns exhibit significant predictability from lagged order imbalance or returns.

A potential concern about the average Newey-West t-statistics is that they could be driven by outliers, but unreported results show that median Newey-West t-statistics are actually considerably higher. We also test whether the average Newey-West t-statistic is below 1.645 (the 5% critical value of a one-tailed t-test) and reject this hypothesis with p-values below 0.001 for all predictability measures.

Table 2.2 also shows that a large fraction (around 60-90%, depending on the predictability measure) of the coefficients on lagged order imbalance and on lagged returns in the individual stock-day predictability regressions are positive, and that 30-60% of these coefficients are significant on an individual basis. The average R^2 of the regressions ranges from 1.7% for *allquotes* to 3.5% for *oib#*. Although these R^2's are modest, we note that predicting stock returns is challenging and that the results are in line with prior work on intraday return predictability (e.g., Chordia et al. (2005)).[7] The degree of predictability varies considerably over time, as well as in the cross-section. For example, the market-wide (equally-weighted) average *OIB predictability* R^2 is 6.44% in 1996 but only 1.29% in 2010, and the average *OIB predictability* R^2 in 1996 ranges from 2.4% for Sun Healthcare Group Inc. to 15.9% for Foodmaker Inc., with an interquartile range of 1.69%.

Overall, Table 2.2 provides evidence of significant intraday return predictability in our sample of all NYSE stocks over 1996-2010. The results also indicate that the degree of predictability is robust across various specifications of the predictability regressions. To compress the five return predictability measures in Table 2.2 into one measure, for each stock we take the first principal component of the daily time-series of the R^2's of the five different predictability regressions in Panel A and label it the *Predictability* measure.

On average, this first principal component explains more than 45% of the total variation in the five predictability measures for individual stocks. The loadings on the first principal component almost always have the same sign for all five predictability measures, with the exception of 91 out of the 2,058 stocks for which we could estimate the predictability regressions. Since for these 91 stocks, the first principal component across the five predictability measures cannot be unambiguously interpreted as increasing in the degree of predictability, we discard them from the sample for the remainder of our analyses. The average loading (across the remaining 1,967 stocks) of the first principal component on the underlying predictability measures is 0.57 for *OIB predictability*, 0.50 for *allquotes*, 0.37 for *2minutes*, 0.48

[7] In unreported tests, we also estimate an intraday predictability measure based on regressions that include lagged order imbalance in dollars and in trades as well as lagged returns simultaneously, and find considerably stronger return predictability based on all three variables.

for *oib#*, and 0.21 for *autocorrelation*.[8]

Table 2.3 – Summary statistics of stock-level efficiency measures

This table reports the cross-sectional mean, standard deviation ("SD"), first quartile ("25%"), median, and third quartile ("75%") of the time-series average by stock of five daily stock-level efficiency measures: *Predictability*, *Variance ratio*, *Variance*, *Hasbrouck*, and *Put-call parity*. *Predictability* is the common factor extracted via principal component analysis by stock of the daily R^2's from the five intraday return predictability measures from Panel A. *Variance ratio* is the daily, absolute difference between one and 30 times the ratio of the variance estimated from one-minute mid-quote returns to the variance estimated from 30-minute mid-quote returns. *Variance* is the daily return variance estimated from one-minute mid-quote returns. *Hasbrouck* is the daily maximum of the absolute intraday pricing errors extracted from a decomposition of observed prices into efficient prices and a stationary component (Hasbrouck, 1993). *Put-call parity* is the absolute difference between the implied volatilities of near-the-money call and put options of the same series (i.e., pairs of options on the same underlying stock with the same strike price and the same expiration date). Section 2.2 discusses all five stock-level efficiency measures in detail. The first column indicates the number of stocks over which the summary statistics are computed. The full sample includes all 2,157 NYSE-listed common stocks from 1996 to 2010 that survive our data screens (described in the Appendix). Data are from TRTH. Data to compute *Put-call parity* are from OptionMetrics.

	#Stocks	Mean	SD	25%	Median	75%
Predictability	1,967	3.58	1.97	2.09	2.72	4.84
Variance ratio	2,130	0.87	0.38	0.63	0.76	0.97
Variance	2,130	0.03	0.04	0.01	0.01	0.03
Hasbrouck	1,769	0.39	0.44	0.15	0.24	0.46
Put-call parity	1,535	2.58	2.04	1.47	2.09	3.07

Table 2.3 presents cross-sectional summary statistics of the stock-by-stock time-series averages of the five different efficiency measures. This Panel is based on the sample of stocks for which each efficiency measure could be estimated for at least 15 days over the sample period.

The mean value of the *Predictability* measure (the first principal component of the R^2's of the five predictability regressions in Panel A) across the 1,967 stocks in our sample for which the first principal component loaded on the five predictability measures with the same sign is equal to 3.58%, with an interquartile range of 2.75%.

The mean and median absolute deviations of the *Variance ratio* from unity are equal to 0.87 and 0.76, respectively. These numbers are somewhat higher than the mean of 0.53 reported by Boehmer and Kelley (2009) (see their Table 1), but that number is based 1-to-20

[8] The proportion of total variation explained by the second to fifth components (that is, their respective eigenvalues scaled by the sum of all eigenvalues) is equal to 20%, 16%, 12%, and 6%, respectively. However, for none of the remaining 1,967 stocks do the five individual predictability measures exhibit same-sign loadings on these components, so including these components in the *Predictability* measure would no longer allow us to unambiguously interpret it as increasing in the degree of predictability as picked up by the individual measures.

days variance ratios (instead of 1-to-30 minutes variance ratios as in our paper) and based on a sample of NYSE stocks that is about half the size of our sample and likely tilted towards large and liquid stocks that may be more efficiently priced.

The time-series average of the variance of one-minute returns (*Variance*) has a cross-sectional mean (median) of 0.03 (0.01). This mean corresponds to an annualized volatility of 54%. An average annualized volatility of individual stocks of 54% may seem high, but is consistent with other studies that compute volatility based on intraday returns measured over very short intervals. As a comparison, Ben-David et al. (2014) report an average return volatility of the returns of S&P500 stocks over one-second intervals of 0.022 (see their Table 2), which corresponds to 53% annually.

The mean (median) value of the *Hasbrouck* measure is 39 (24) basis points. These numbers align well with the mean pricing error of 26 basis points reported by Hasbrouck (1993) for a representative sample of 175 NYSE stocks in 1989. We would expect pricing errors to be lower in our more recent sample, but we report the maximum rather than the mean pricing error.

We are able to estimate the *Put-call parity* measure for 1,535 of the 2,157 stocks in our sample, for on average 1,448 days over our sample period 1996-2010. The mean absolute put-call parity deviation (expressed in terms of implied volatility) across stock-days in the sample is 2.58%, with an interquartile range of 1.60%. These values closely correspond to the put-call deviation estimates provided by Cremers and Weinbaum (2010) for a similarly-sized sample of U.S. stocks over 1996-2005. Panel A of their Table 1 shows an average put-call parity deviation of -0.978%, but this is an aggregation of positive and negative deviations. Taking the average of the absolute values of the percentiles of the distribution of their put-call parity deviation estimates reported in Panel B of their Table 1 yields an approximate average absolute deviation of 2.3% for their sample.

In sum, Table 2.3 illustrates that the degree of market efficiency not only varies considerably over time, but also across individual stocks. In the next section, we investigate co-movement in the time-varying efficiency of individual stocks.

2.4 Co-movement in efficiency

We now set out to accomplish one of our primary goals by examining whether there is significant co-movement in market efficiency across stocks. To estimate the degree of co-movement in efficiency across stocks, we run time-series regressions of changes in the efficiency of individual stocks on contemporaneous, lead, and lagged changes in market-wide efficiency. Specifically, we estimate the degree of co-movement in efficiency each month for each stock i in the following regression:

$$\Delta Eff_{i,d} \quad = \quad \alpha_i + \beta_i \Delta MktEff_{i,d} + \gamma_i \Delta MktEff_{i,d-1} + \delta_i \Delta MktEff_{i,d+1} + \eta_{i,d}, \tag{2.2}$$

where $\Delta Eff_{i,d}$ is the change in the efficiency of stock i on day d, and $\Delta MktEff_{i,d}$ is the change in market-wide efficiency (defined as the equally-weighted average efficiency change across all stocks in our sample excluding stock i). We require at least 15 daily observations for a given stock within the month to estimate Eq. (2.2) for that stock in that month. Inspired by Morck et al. (2000), we use the R^2 from the co-movement regressions in Eq. (2.2) as an indicator for the degree of co-movement in market efficiency across stocks.

We estimate Eq. (2.2) each month for each stock based on daily changes in our five stock-level efficiency measures: *Predictability*, *Variance ratio*, *Variance*, *Hasbrouck*, and *Put-call parity*. Our motivation for estimating Eq. (2.2) monthly based on daily stock-level efficiency estimates within the month is two-fold. First, estimating Eq. (2.2) over longer time frames (for example, one year or even the full 15-year sample period) could lead us to underestimate the degree of co-movement in efficiency, since restricting the coefficients to be constant over time while the degree of co-movement is time-varying depresses the R^2. Second, the monthly co-movement regressions yield monthly R^2 estimates of the degree of co-movement in efficiency across stocks, which we later use to examine the determinants of time-variation in co-movement in efficiency.

One issue that arises is how our stock-level efficiency measures are related to stock-level liquidity. If stock-level efficiency and liquidity are hard to distinguish empirically, our analysis of co-movement in efficiency might be perceived as a reiteration of the extensive literature on co-movement in liquidity (Chordia et al., 2000; Hasbrouck and Seppi, 2000; Huberman and Halka, 2001). However, unreported results show that our five stock-level efficiency measures are only weakly correlated with three common stock-level illiquidity proxies: the proportional quoted bid-ask spread (*PQSPR*), the proportional effective spread (*PESPR*, defined as two times the absolute difference between the transaction price and the quote midpoint, scaled by the quote midpoint), and the Amihud (2002) illiquidity proxy (*Amihud*).[9]

Nonetheless, to ensure that any co-movement in the efficiency of individual stocks we detect is not driven by underlying co-movement in their (il)liquidity, we first orthogonalize

[9] In particular, of the 15 time-series correlations (averaged across stocks) of the five efficiency measures with the three illiquidity proxies, 12 are between -0.15 and 0.15. The remaining three correlations are 0.33 between *Predictability* and *PESPR* and 0.24 and 0.49 between *Variance* and *PQSPR* and *PESPR*, respectively. Detailed results are available from the authors.

the daily changes in each of the five efficiency measures at the stock-level with respect to changes in that stock's illiquidity (*PQSPR*; we obtain similar results when we use *PESPR* or the *Amihud* measure, and slightly stronger results when we do not orthogonalize at all). We then run the co-movement regressions in Eq. (2.2) of orthogonalized changes in stock-level efficiency on contemporaneous, lead, and lagged orthogonalized changes in market efficiency (defined as the equally-weighted average changes in efficiency, orthogonalized with respect to illiquidity changes, across all stocks in our sample, excluding stock i).

Furthermore, when we subsequently analyze the determinants of time-variation in the degree of co-movement in efficiency in Section 2.5 below, we make sure to account for time-variation in co-movement in liquidity in such a way that any impact of proxies for funding liquidity and the intensity of arbitrage activity on co-movement in efficiency that we measure is over and above their effect on co-movement in liquidity.[10]

2.4.1 Monthly co-movement in efficiency across stocks

Table 2.4 presents the results of our regressions to estimate co-movement in each of the five efficiency measures across individual stocks. The table reports the average regression coefficients across all co-movement regressions estimated by stock-month for each efficiency measure. The number of stock-month regressions varies from roughly 75,000 for the *Put-call parity* measure to almost 180,000 for the *Variance ratio* and *Variance* measures.

The table reveals evidence of significant co-movement in efficiency across stocks. The average coefficient on contemporaneous changes in market-wide efficiency across the individual stock-month regressions is positive and economically substantial for all efficiency measures, ranging from 0.717 for the *Put-call parity* measure to 0.907 for the *Variance ratio* measure. The average t-statistic of this coefficient is not significant at conventional significant levels for any of the efficiency measures, but this is not to be expected in light of the fact that the individual coefficients are estimated based on at most about 20 observations (i.e., the number of trading days) per month.

As in Table 2.2, we therefore also report average Newey-West t-statistics across stocks.

[10]We also note that, as pointed out by Chordia et al. (2008), illiquidity does not necessarily imply any return predictability from order flow or past returns. In Kyle (1985), even though markets are illiquid, prices are martingales because market makers are risk-neutral. On the other hand, in inventory-based models, return predictability from order flow can arise if market makers have capital constraints or limited risk-bearing capacity that prevent them from conducting arbitrage trades that mitigate the predictability (Stoll, 1978). So, in a sense, our measure of predictability, or lack thereof, is a measure of the efficacy of such short-horizon arbitrage. Our interpretation of return predictability as a deviation from efficiency is consistent with Samuelson (1965) definition of efficiency as "properly anticipated prices fluctuate randomly."

Table 2.4 – Co-movement regressions of daily changes in stock-level efficiency on changes in market efficiency

This table reports the average results of the efficiency co-movement regressions from Eq. (2), estimated monthly based on daily data for each NYSE stock in the sample. The dependent variable $\Delta Eff_{i,d}$ is the change in the efficiency of stock i on day d, orthogonalized with respect to the change in stock i's proportional quoted spread (*PQSPR*) on day d. The independent variable $\Delta MktEff_d$ is the (orthogonalized) change in market-wide efficiency on day d, computed as the equally-weighted average change in efficiency (orthogonalized with respect to the change in *PQSPR*) of all individual stocks on day d, excluding stock i. Each co-movement regression also includes a one-day lead and lag of (orthogonalized) changes in market-wide efficiency. Each of the five columns in the table presents the results of the co-movement regressions based on a different stock-level efficiency measure: *Predictability*, *Variance ratio*, *Variance*, *Hasbrouck*, and *Put-call parity*. We refer to Table 2 for a description of all five stock-level efficiency measures. Each column presents the average slope coefficients across all stock-month co-movement regressions. The average *t*-statistics ("*t*-stat avg") and the average Newey-West (1994) *t*-statistics ("NW *t*-stat avg") are in parentheses below the coefficients. "% positive" is the percentage of positive coefficients, and "% + significant" is the percentage with *t*-statistics greater than 1.645 (the 5% critical level in a one-tailed test). Intercepts have been suppressed to conserve space. The last three rows report the average R^2 and adjusted R^2 across all regressions and the number of stock-month co-movement regressions. The full sample includes all 2,157 NYSE-listed common stocks from 1996 to 2010 that survive our data screens (described in the Appendix). Data are from TRTH. Data to compute *Put-call parity* are from OptionMetrics.

Dependent variable: $\Delta Eff_{i,d}$					
Efficiency measure:	*Predictability*	*Variance ratio*	*Variance*	*Hasbrouck*	*Put-call parity*
$\Delta MktEff_d$	0.813	0.907	0.789	0.889	0.717
t-stat avg	(0.288)	(0.539)	(1.534)	(0.430)	(0.592)
NW *t*-stat avg	(2.313)	(3.607)	(4.937)	(3.523)	(3.426)
% positive	59.95	67.26	79.94	65.49	68.60
% + significant	9.29	14.07	34.55	11.44	15.58
$\Delta MktEff_{d-1}$	0.025	0.002	0.013	0.019	0.014
t-stat avg	(0.013)	(-0.002)	(0.047)	(0.009)	(0.017)
NW *t*-stat avg	(0.043)	(-0.021)	(0.219)	(0.051)	(0.011)
% positive	50.05	49.40	50.28	50.18	49.97
% + significant	6.75	6.75	7.84	6.31	6.45
$\Delta MktEff_{d+1}$	0.024	0.000	0.027	0.037	0.025
t-stat avg	(0.012)	(-0.003)	(0.029)	(0.020)	(0.036)
NW *t*-stat avg	(0.082)	(-0.021)	(0.142)	(0.113)	(0.146)
% positive	50.30	49.62	50.45	50.28	50.50
% + significant	7.01	6.53	6.93	6.52	7.24
R^2	20.77	22.68	30.27	20.28	22.40
adj. R^2	3.86	6.01	15.19	2.99	5.18
# regressions	135,784	178,867	178,867	85,957	75,460

These t-statistics are considerably higher, at 2.3 for *Predictability*, 3.6 for *Variance ratio*, 4.9 for *Variance*, 3.5 for *Hasbrouck*, and 3.4 for *Put-call parity*, which indicates statistically significant co-movement in efficiency across individual stocks for all five efficiency measures.

As in Table 2.2, average Newey-West t-statistics could be driven by outliers, but median Newey-West t-statistics are also significant at the 5% level or better for all five co-movement regressions in Table 2.4. Furthermore, we reject the hypothesis that the average Newey-West t-statistic is below 1.645 (the 5% critical value of a one-tailed t-test) with p-values below 0.001 for all five co-movement regressions.

For each of the five efficiency measures, a clear majority (at least 59% and up to 80%) of the individual coefficients on contemporaneous changes in market-wide efficiency are positive. At least 9% (*Predictability*) and up to 34% (*Variance*) of the coefficients are positive and significant on an individual basis. The fraction of individual t-statistics that is significant is thus not overwhelming, but we note that they are based on at most around 20 observations. There is little evidence that the lead and lagged changes in market-wide efficiency are important in explaining time-variation in the efficiency of individual stocks.

The average (adjusted) R^2's of the co-movement regressions in Table 2.4 are substantial and range from 20.3% (3.0%) for *Hasbrouck* to 30.3% (15.2%) for *Variance*. The co-movement R^2's for the five efficiency measures in Table 2.4 are of the same order of magnitude as the R^2's of similar monthly regressions to estimate co-movement in liquidity as reported by Karolyi et al. (2012), who find that the R^2 for these regressions based on NYSE stocks averaged around 23% over the period 1995-2009. Co-movement in efficiency is thus roughly an equally strong phenomenon as co-movement in liquidity across individual stocks.

Overall, Table 2.4 presents evidence of economically and statistically significant co-movement in efficiency across stocks.[11]

[11] In unreported robustness tests, we estimate the co-movement regressions in Eq. (2.2) based on efficiency levels orthogonalized with respect to liquidity levels rather than based on efficiency changes orthogonalized with respect to liquidity changes, and based on contemporaneous market efficiency as the only independent variable (that is, no lead and lagged market-wide efficiency), and obtain similar results. We also obtain similar results when we compute market-wide efficiency as the value-weighted (instead of the equally-weighted) average efficiency across all stocks in our sample, excluding stock i. And although we lose a substantial number of degrees of freedom when we analyze co-movement in efficiency at the quarterly frequency instead of the monthly frequency, the main results in the paper also obtain when we estimate the degree of efficiency co-movement quarterly based on daily data within the quarter. In all of the robustness tests reported in this footnote, we orthogonalize (the changes in) each of the stock-level efficiency measures with respect to (the changes in) the stock-level *PQSPR* before estimating and analyzing the degree of co-movement in efficiency.

2.4.2 Monthly co-movement in efficiency across portfolios

In this section, we address the question whether the degree of co-movement in efficiency is different for different market segments. There are at least three reasons for why such an analysis is interesting. First, it sheds light on the question which stock-level attributes affect the degree of co-movement in efficiency. Second, the degree of co-movement in efficiency uncovered in Table 2.4 is mitigated by both estimation noise and idiosyncratic components in the efficiency of individual stocks. Looking at portfolios of stocks might alleviate estimation noise and expose a stronger image of co-movement. Third, from the perspective of portfolio management, analyzing the co-movement of the efficiency of a portfolio of stocks with the efficiency of the market is relevant, since investors that manage different portfolios of securities might be concerned about the risk that multiple portfolios are simultaneously exposed to variation in price efficiency.

Table 2.5 – Co-movement regressions of daily changes in portfolio-level efficiency on changes in market efficiency

This table reports the average results of the efficiency co-movement regressions from Eq. (2), estimated monthly based on daily data for ten "liquidity portfolios" formed yearly by sorting all NYSE stocks in the sample on the basis of their average proportional quoted spread (*PQSPR*) over the year. The dependent variable $\Delta Eff_{p,d}$ is the (orthogonalized) change in the efficiency of liquidity portfolio p on day d, which is computed as the equally-weighted average change in efficiency (orthogonalized with respect to the change in *PQSPR*) of all individual stocks in the portfolio on day d. The independent variable $\Delta MktEff_d$ is the (orthogonalized) change in market-wide efficiency on day d, computed as the equally-weighted average change in efficiency (orthogonalized with respect to the change in *PQSPR*) of all individual stocks not in the subject portfolio on day d. Each co-movement regression also includes a one-day lead and lag of (orthogonalized) changes in market-wide efficiency. The five columns in the table present the results of the portfolio-level co-movement regressions for liquidity portfolios 1 (most liquid), 2, 5, 9, and 10 (least liquid). Each column presents the results for one portfolio based on five different stock-level efficiency measures: *Predictability*, *Variance ratio*, *Variance*, *Hasbrouck*, and *Put-call parity*. We refer to Table 2 for a description of all five stock-level efficiency measures. Each column presents the average slope coefficients across all portfolio-month co-movement regressions. The average *t*-statistics ("*t*-stat avg") and the Newey-West (1994) *t*-statistics ("NW *t*-stat") are in parentheses below the coefficients. "% positive" is the percentage of positive coefficients, and "% + significant" is the percentage with *t*-statistics greater than 1.645 (the 5% critical level in a one-tailed test). Intercepts and coefficients on the lead and lagged independent variable have been suppressed to conserve space. The table also reports the average R^2 and adjusted R^2 across all regressions for each portfolio and for each efficiency measure. The full sample includes all 2,157 NYSE-listed common stocks from 1996 to 2010 that survive our data screens (described in the Appendix). Data are from TRTH. Data to compute *Put-call parity* are from OptionMetrics.

Dependent variable: $\Delta Eff_{p,d}$					
Liquidity portfolio:	liquid	2	5	9	illiquid
Efficiency measure: *Predictability*					
$\Delta MktEff_d$	0.903	0.819	0.633	0.602	0.303
t-stat avg	(2.090)	(2.327)	(1.843)	(1.164)	(0.560)
NW *t*-stat	(14.177)	(15.864)	(15.958)	(9.596)	(5.025)
% positive	93.22	94.35	87.57	81.36	70.06
% + significant	54.80	58.76	46.89	33.90	16.38
R^2	36.46	39.22	34.23	25.23	19.86
adj. R^2	24.66	27.94	21.99	11.31	4.97
Efficiency measure: *Variance ratio*					
$\Delta MktEff_d$	0.837	0.894	0.945	0.882	0.574
t-stat avg	(3.481)	(4.117)	(4.822)	(2.956)	(1.388)
NW *t*-stat	(19.994)	(23.364)	(25.082)	(16.944)	(9.896)
% positive	97.18	100.00	98.31	94.92	91.53
% + significant	81.92	80.79	87.01	74.01	36.16
R^2	53.65	57.23	60.80	46.61	27.30
adj. R^2	44.98	49.26	53.43	36.64	13.80

Table 2.5 – continued

Dependent variable: $\Delta Eff_{p,d}$					
Liquidity portfolio:	liquid	2	5	9	illiquid
Efficiency measure: *Variance*					
$\Delta MktEff_d$	0.414	0.438	0.585	1.053	1.796
t-stat avg	(4.980)	(5.382)	(5.734)	(4.359)	(1.881)
NW *t*-stat	(14.998)	(11.151)	(13.121)	(11.931)	(7.553)
% positive	97.74	98.87	98.87	93.22	85.88
% + significant	85.31	82.49	84.18	72.88	42.94
R^2	55.83	57.86	58.80	50.37	28.79
adj. R^2	47.59	49.97	51.10	41.07	15.52
Efficiency measure: *Hasbrouck*					
$\Delta MktEff_d$	1.052	1.009	0.818	0.572	0.471
t-stat avg	(2.784)	(3.050)	(2.676)	(2.017)	(1.617)
NW *t*-stat	(18.458)	(22.721)	(20.482)	(15.309)	(13.494)
% positive	98.31	98.31	97.74	88.14	88.70
% + significant	70.06	73.45	63.84	49.15	39.55
R^2	41.77	41.69	39.60	32.91	26.56
adj. R^2	30.98	30.84	28.32	20.41	12.84
Efficiency measure: *Put-call parity*					
$\Delta MktEff_d$	0.394	0.455	0.672	1.104	1.481
t-stat avg	(2.682)	(2.809)	(3.430)	(3.160)	(2.359)
NW *t*-stat	(14.112)	(21.750)	(23.604)	(17.169)	(16.398)
% positive	96.61	96.61	96.61	95.48	96.05
% + significant	61.58	72.88	77.97	70.62	62.71
R^2	37.51	39.81	46.96	43.65	34.55
adj. R^2	25.83	28.52	37.03	33.11	22.31

Table 2.5 examines the degree of co-movement in efficiency for portfolios of stocks sorted on their liquidity. As one of the most prominent limits to arbitrage, liquidity seems a natural stock-level attribute based on which to distinguish different market segments. Each year, we sort stocks into decile portfolios based on their average proportional quoted bid-ask spread (*PQSPR*) over the year. We then estimate Eq. (2.2) by running monthly regressions of daily changes in the efficiency of the ten liquidity-sorted portfolios on contemporaneous, lead, and

lagged changes in market efficiency (computed as the equally-weighted average efficiency changes across the stocks not in the subject portfolio). Just like in Table 2.4, we orthogonalize daily changes in each of the stock-level efficiency measures with respect to stock-level *PQSPR* changes before computing daily changes in portfolio-level efficiency as the equally-weighted average orthogonalized efficiency changes across all stocks in the portfolio on that day and then estimating the degree of co-movement in efficiency at the portfolio-level.

Table 2.5 shows strong co-movement in efficiency at the portfolio-level for all liquidity decile portfolios (for space considerations, Table 2.5 only reports the results for deciles 1, 2, 5, 9, and 10) based on all five efficiency measures. The coefficients on contemporaneous changes in market efficiency are positive for all decile portfolios and for all five efficiency measures and the Newey-West t-statistics are very high, indicating that the contemporaneous coefficients are all five or more standard deviations away from zero. (We note that these Newey-West t-statistics are not averages, since they are based on one time-series of coefficient estimates per portfolio for each efficiency measure.) The portfolio-level R^2's of the co-movement regressions in Eq. (2.2) are considerably greater than the individual stock level R^2's reported in Table 2.5, which suggests that estimation noise and idiosyncratic components may dampen the degree of co-movement in efficiency reported in Table 2.5.

Perhaps more interestingly, the results in Table 2.5 indicate that the degree of co-movement in efficiency is much greater for liquid stocks than for illiquid stocks. For example, for the *Predictability* measure, the adjusted R^2 is equal to 5.0% for the most illiquid decile and to 24.7% for the most liquid decile. Similarly, for the *Hasbrouck* measure, we obtain an adjusted R^2 of 12.8% for the most illiquid stocks and of 31.0% for the most liquid stocks. The finding of a considerable difference in the degree of co-movement in efficiency of liquid and illiquid stocks is remarkably consistent across the five efficiency measures (with the exception of *Put-call parity*) and is mainly driven by the relatively low degree of co-movement of illiquid stocks. This finding suggests that illiquid stocks are relatively shielded from market-wide fluctuations in the degree of pricing efficiency, and could thus be viewed as less exposed to this potential form of systematic risk.[12]

[12]In addition to these results on the degree of co-movement of the efficiency of liquidity-sorted portfolio with market-wide efficiency, in unreported analyses we also examine the degree of efficiency co-movement *within* each of the ten liquidity-sorted portfolios and *within* five industry portfolios (based on the five industries defined on the website of Ken French), and find little evidence of systematic differences in the degree of within-segment co-movement in efficiency across these different market segments.

2.5 Time-variation in co-movement in efficiency

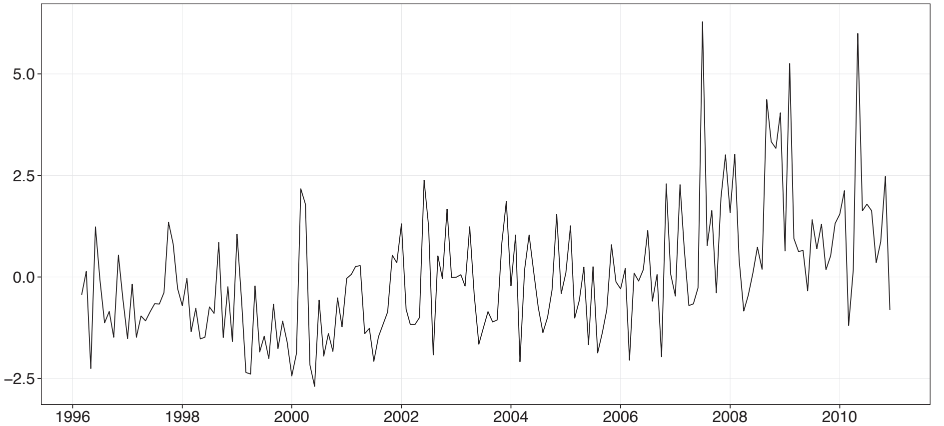
We now turn to an analysis of time-variation in the degree of co-movement in efficiency. We first aggregate the five monthly, stock-level measures of co-movement in efficiency (the R^2's from the five monthly co-movement regressions in Table 2.4 based on the five different stock-level efficiency measures: *Predictability*, *Variance ratio*, *Variance*, *Hasbrouck*, and *Put-call parity*) to the market-level by computing the equally-weighted average R^2 across individual stocks each month, separately for each efficiency measure. (We obtain similar results when taking the value-weighted average.) This procedure yields five different monthly, market-wide measures of co-movement in efficiency.

We then extract a single, comprehensive measure of monthly, market-wide co-movement in efficiency via principal component analysis (PCA) of these five different monthly, market-wide measures of co-movement in efficiency. We follow Hasbrouck and Seppi (2000) and extract the principal components based on the correlation matrix. We find that the first principal component explains 48% of the total variation in the five individual market-wide measures of co-movement in efficiency. The proportion of the total variation that each additional component or eigenvector represents (in other words, the component's eigenvalue divided by the sum of all eigenvalues) is equal to 19%, 12%, 11%, and 10% for the second to fifth components, respectively.

Importantly, the loadings of the five different co-movement in efficiency measures on the first principal component are all of the same sign, otherwise this component could not be interpreted as representing aggregate variation in co-movement in efficiency. Since the loadings on the second principal component are not of the same sign, including this component in our aggregate co-movement in efficiency measure would lead to problems in interpreting the resulting measure as being positively associated with the degree of co-movement in each of the five efficiency measures. Consequently, we use only the first principal component as representative of market-wide co-movement in efficiency. The fact that this component explains almost half of the total variation and explains almost 30% percent more variation than the next component lends credibility to the view that this component captures the dominant variation in market-wide co-movement in efficiency. The loading of the first principal component on the underlying co-movement in efficiency measures is 0.37 for co-movement in *Predictability*, 0.50 for *Variance ratio*, 0.47 for *Variance*, 0.46 for *Hasbrouck*, and 0.43 for *Put-call parity*. The first principal component is thus representative of all five co-movement in efficiency measures and is not dominated by one or more of these measures.

To get a time-series of the first principal component, we standardize each of the five co-

Figure 2.1 – Monthly variation in co-movement in efficiency, 1996 - 2010
This figure shows monthly variation in the degree of market-wide co-movement in stock-level efficiency (*Co-movement in efficiency*) from 1996 to 2010. This measure of *Co-movement in efficiency* is constructed as follows. First, each month for each NYSE stock in the sample, we estimate the degree of co-movement in that stock's efficiency with market efficiency using the co-movement regressions from Eq. (2), based on five different daily stock-level efficiency measures: *Predictability*, *Variance ratio*, *Variance*, *Hasbrouck*, and *Put-call parity*. We refer to Table 2 for a description of all five stock-level efficiency measures and to Table 3 for a description of the efficiency co-movement regressions. We then aggregate the five resulting monthly, stock-level measures of co-movement in efficiency (the R^2's from the five monthly co-movement regressions in Table 3) to the market-level by computing the equally-weighted average R^2 across individual stocks each month, separately for each efficiency measure. Subsequently, we extract a single, comprehensive measure of monthly, market-wide efficiency co-movement (*Co-movement in efficiency*) as the first principal component of these five different monthly, market-wide measures of efficiency co-movement. To get a time-series of the first principal component, we standardize each of the five co-movement in efficiency measures to have zero mean and unit standard deviation, and multiply the matrix of standardized measures by the vector of the loadings of each measure on the component. The full sample includes all 2,157 NYSE-listed common stocks from 1996 to 2010 that survive our data screens (described in the Appendix). Data are from TRTH. Data to compute *Put-call parity* are from OptionMetrics.



movement in efficiency measures to have zero mean and unit standard deviation, and multiply the matrix of standardized measures by the vector of the loadings of each measure on the component. We refer to the resulting measure as *Co-movement in efficiency* in the remainder of the paper. Figure 2.1 presents a graph of the monthly time-variation in this comprehensive measure of co-movement in efficiency. The figure shows that the degree of co-movement is considerably greater in some periods than in others. Two features of the co-movement dynamics stand out. First, the degree of co-movement tends to spike during periods of financial turmoil, such as the Asian crisis in late 1997, the LTCM / Russian debt crisis in September 1998, the burst of the internet bubble in early 2000, the quant crisis in the summer of 2007, and after the collapse of Lehman Brothers in September 2008. Second, Figure 2.1 shows a slight upward trend in the degree of co-movement in efficiency starting around 2006. A possible explanation for the latter feature is the advent of algorithmic and high-frequency trading over the last five years of our sample period (in part facilitated by the introduction of NYSE's Hybrid Market at the end of 2006), which has been linked to a generic improvement in price efficiency (Hendershott et al., 2011; Brogaard et al., 2014a).

We proceed with a formal analysis of what economic forces explain time-variation in *Co-movement in efficiency*. Since arbitrage plays a central role in enforcing efficient pricing, and since the efficacy of arbitrage, in turn, depends on the availability of arbitrage capital, our primary interest is in variables that proxy for variation in funding liquidity and the intensity of arbitrage activity.

Vector autoregressions (VARs) are a natural way to analyze the dynamics of *Co-movement in efficiency* in relation to proxies for funding liquidity and the intensity of arbitrage activity, since all of these variables are endogenous and could influence each other both contemporaneously and with a lag. We therefore estimate multivariate VARs in which *Co-movement in efficiency* is included as the last and thus most endogenous variable, which can be influenced both contemporaneously and with a lag by shocks to all of the other endogenous variables in the VARs. We also estimate separate VARs to analyze time-variation in the degree of co-movement in efficiency of liquid and illiquid stocks (defined as those in the decile portfolios of stocks with the lowest and highest proportional quoted spread or *PQSPR*, as in Table 2.5), constructed in the same way as *Co-movement in efficiency* but then based on these subsets of stocks.

Table 2.6 – Summary statistics of potential determinants of monthly co-movement in efficiency

This table reports the time-series mean, standard deviation ("SD"), first quartile ("25%"), median, and third quartile ("75%") of four potential determinants of monthly market-wide co-movement in efficiency. *TED spread* is the monthly difference between the three-month LIBOR and the three-month T-bill rate (in %), obtained from the FRED database of the Federal Reserve Bank of St. Louis (FRED ID: USD3MTD156N minus TB3MS). *Hedge fund flow* is the monthly percentage money inflow into hedge funds, obtained from Matti Suominen and LIPPER-TASS (see Jylhä, Rinne, and Suominen, 2015). *Quotes/Volume* is the total number of quote updates per month across all the NYSE stocks in our sample divided by the aggregate dollar trading volume for those stocks in the same month. This variable is scaled by 10^2. *Co-movement in liquidity* is a monthly measure of the degree of market-wide co-movement in liquidity, constructed as the equally-weighted R^2 (in %) across individual stocks each month from the equivalent co-movement regressions to Eq. (2) but then using the proportional quoted spread (*PQSPR*) as stock-level liquidity measure. Data to compute *Quotes/Volume* and *Co-movement in liquidity* are from TRTH. The full sample includes all 2,157 NYSE-listed common stocks from 1996 to 2010 that survive our data screens (described in the Appendix).

	# Obs.	Mean	SD	25%	Median	75%
TED spread	177	0.576	0.443	0.241	0.484	0.729
Hedge fund flow	177	0.618	1.817	-0.048	0.905	1.641
Quotes/Volume	177	0.02	0.02	0.00	0.01	0.03
Co-movement in liquidity	177	22.12	3.60	19.87	21.34	23.55

Since estimating multivariate VARs based on just 176 monthly observations is quite demanding, we limit the number of endogenous variables besides *Co-movement in efficiency*.[13] We focus on three key proxies for funding liquidity and the intensity of arbitrage activity. First, to the extent that fluctuations in the funding liquidity of the financial system have pervasive effects on market making and arbitrage activity (e.g., Brunnermeier (2009); Mancini-Griffoli and Ranaldo (2011)), the degree of co-movement in market efficiency can be affected by changes in funding liquidity. As a direct proxy for funding liquidity we use the *TED spread*, which is the difference between the three-month LIBOR and the three-month T-bill rate from the FRED database of the Federal Reserve Bank of St. Louis and is a widely used indicator of funding liquidity (Brunnermeier and Pedersen, 2008; Brunnermeier, 2009).[14] As per Eq. (2.2), co-movement is measured using regressions of changes in individual stock efficiency on changes in market efficiency. In other words, co-movement in efficiency arises as the result of simultaneous changes in the efficiency of many stocks. Thus, our main hypothesis is that changes in the *TED spread* induce changes in the intensity of arbitrage activity, which in turn may result in changes in the degree of price efficiency for many stocks simultaneously, thereby increasing the level of *Co-movement in efficiency*. A priori, we have no hypothesis on potential asymmetric effects of changes in funding liquidity, in the sense that an improvement of funding liquidity could have a differential impact on *Co-movement in efficiency* than a worsening of funding liquidity. Consequently, we include absolute changes in the *TED spread* (or $|\Delta TED\ spread|$) as endogenous variable in our VARs.[15] Since the *TED spread* is arguably the most exogenous of the funding liquidity measures we consider, we include it as the first variable in our VARs.

Second, we compute *Hedge fund flow* as the monthly percentage money inflow into hedge funds.[16] Greater hedge fund inflows should spur arbitrage activity. Since *Hedge fund flow* is already a flow variable that measures changes in the amount of capital available to hedge funds to engage in arbitrage activity, we simply use the absolute value of this variable ($|Hedge\ fund\ flow|$) as the second endogenous variable in our VARs.[17]

[13]We note that although our data extend over 178 months from March 1996 up to and including December 2010, we lose one month because of the one-month lag in the VAR and we cannot reliably estimate the degree of co-movement in efficiency in September 2001 as a result of the limited number of trading days in that month due to the "9/11" terrorist attacks.

[14]The notion is that the TED spread may proxy for counterparty risk, which, when elevated, can lead to funding illiquidity.

[15]In unreported tests, we separate out positive and negative changes in the *TED spread* in the VARs and find little evidence of asymmetric effects.

[16]We thank Matti Suominen and LIPPER-TASS for data on hedge fund flows. The sample includes all hedge funds that report their returns in U.S. dollars and have a minimum of 36 monthly return observations over our sample period. See Jylhä et al. (2015).

[17]In unreported tests, we find little evidence of an asymmetric effect of hedge fund flows when separating out positive and negative values of *Hedge fund flow* in the VARs.

Third, inspired by Boehmer et al. (2014), we use a proxy for the intensity of algorithmic trading defined as the total number of quote updates per month across all the stocks in our sample divided by the aggregate dollar trading volume for those stocks in the same month (*Quotes/Volume*). We include this variable to account for the marked increase in quoting activity over our sample period which has been related to the advent of algorithmic trading that could affect the efficacy of arbitrage and market making activity (Hendershott et al., 2011; Brogaard et al., 2014a). Since *Quotes/Volume* is already a flow variable that is non-negative, we do not take changes or absolute values. And since it is arguably a more direct proxy for actual arbitrage activity that could be influenced by variation in the availability of arbitrage capital as picked up by the *TED spread* and *Hedge fund flow*, we include it after these two variables as the third endogenous variable in our VARs.

An important challenge for the analysis in this section is that time-variation in *Co-movement in efficiency* could in part be driven by time-variation in the degree of co-movement in liquidity, and both can be affected by funding liquidity. Indeed, prior theoretical work (e.g., Brunnermeier (2009)) as well as empirical evidence (Coughenour and Saad, 2004; Hameed et al., 2010) establishes a link between funding liquidity and co-movement in liquidity. Hence, it may be hard to distinguish between a direct effect of funding liquidity on *Co-movement in efficiency* and an indirect effect running through co-movement in liquidity. Although this concern is mitigated by the fact that we estimate *Co-movement in efficiency* based on changes in stock-level efficiency that are orthogonalized with respect to changes in stock-level liquidity, we further tackle this challenge by also including a measure of co-movement in liquidity in the VARs. To that end, we construct a monthly measure of market-wide *Co-movement in liquidity* based on the same methodology we use to estimate *Co-movement in efficiency*, using the proportional quoted spread (*PQSPR*) as stock-level liquidity measure in equivalent co-movement regressions to Eq. (2.2).[18] We include *Co-movement in liquidity* as the fourth endogenous variable in the VARs, just before *Co-movement in efficiency*, such that shocks to the funding liquidity and arbitrage proxies can affect *Co-movement in efficiency* directly as well as indirectly through *Co-movement in liquidity*. For the VARs based on liquid and illiquid stocks, we construct *Co-movement in liquidity* based on the same decile portfolios of stocks as *Co-movement in efficiency*.

Prior to usage as endogenous variables in the VARs, we detrend all five variables with linear and quadratic trend terms (to preclude spurious results) and then standardize all detrended variables to have zero mean and a standard deviation of one (for ease of interpretation

[18]Unreported robustness tests, available from the authors, show similar results when we construct *Co-movement in liquidity* based on the proportional effective spread (*PESPR*) or the Amihud (2002) illiquidity proxy instead of based on *PQSPR*.

of the results). We do the same for the endogenous variables in the VARs based on liquid and on illiquid stocks.[19] The number of lags in the VARs is determined using the Akaike and Schwarz information criteria (AIC and SIC). For the VARs based on all stocks and based on liquid stocks, the AIC indicates four lags, while the SIC indicates one lag. For the VAR based on illiquid stocks, both AIC and SIC indicate one lag. For the sake of consistency and parsimony, we choose to report the results of one-lag VARs (as indicated by the SIC), but unreported results for four-lag VARs are similar. Table 2.6 presents summary statistics of the four potential determinants of *Co-movement in efficiency* included in the VARs.

Panels A, B, and C of Table 2.7 present the estimates of the coefficients (and their associated t-statistics) in the VARs based on, respectively, all stocks, liquid stocks, and illiquid stocks. Because we estimate one-lag VARs, these can be interpreted as the results of Granger causality tests. To save space, the table only presents the estimation results of the equations in which we are most interested, those with *Co-movement in liquidity* and *Co-movement in efficiency* as the dependent variable.

For all stocks (Panel A), we find evidence of $|\Delta TED\ spread|$ Granger causing both *Co-movement in liquidity* and *Co-movement in efficiency*, while $|Hedge\ fund\ flow|$ Granger causes *Co-movement in efficiency*. Since all variables in the VARs are standardized, the coefficients can be interpreted as the effect of a one standard deviation change in the independent variable on the dependent variable, expressed as a fraction of the standard deviation of the dependent variable. At around 0.20 standard deviations, the economic magnitude of the effect of these two proxies for funding liquidity and the intensity of arbitrage activity on the degree of co-movement in liquidity and efficiency is considerable. We also observe that our proxy for algorithmic trading negatively Granger causes *Co-movement in liquidity*, which may indicate that algorithmic trading affects the liquidity of individual stocks (consistent with Hendershott et al. (2011), but not across the board, thereby reducing the degree of co-movement.

The VAR results for liquid stocks (Panel B of Table 2.7) and illiquid stocks (Panel C) separately reveal that the effects of the funding liquidity proxies (notably, the *TED spread*) on the degree of co-movement in efficiency are concentrated in the subsample of illiquid stocks. For liquid stocks, the VAR results further show evidence of *Quotes/Volume* negatively Granger causing *Co-movement in efficiency*.

[19]In unreported results, the Augmented Dickey Fuller test rejects the null-hypothesis of a unit root for all variables included in the three VARs with p-values below 0.01.

Table 2.7 – Vector autoregressions of co-movement in efficiency: Coefficient estimates / Granger causality tests

This table reports the coefficient estimates from three multivariate vector autoregressions (VARs) with the following five endogenous variables: absolute changes in the TED spread ($|\Delta TED\ spread|$), absolute hedge fund flows ($|Fund\ flow|$), the total number of quote updates scaled by aggregate dollar trading volume (*Quotes/Volume*), co-movement in proportional quoted spreads (*Co-movement in liquidity*), and *Co-movement in efficiency*. We refer to Table 2.6 and Figure 1 for a description of these variables. In the first VAR (Panel A), the measures of co-movement in liquidity and efficiency are constructed using all NYSE stocks in the sample. In the second and third VARs (Panels B and C), the measures of co-movement in liquidity and efficiency are constructed using only the stocks in the decile portfolios of the most liquid and most illiquid stocks, respectively (these liquidity portfolios are formed based on the stocks' average proportional quoted spread, as in Table 2.5). To conserve space, intercepts have been suppressed and each panel only reports the coefficient estimates of the equations with *Co-movement in liquidity* and *Co-movement in efficiency* as the dependent variables. We estimate all three VARs with one lag, following the Schwarz information criterion (SIC), which implies that the estimation results can be interpreted as Granger causality tests. All variables in the VARs have been detrended using a linear and a quadratic time trend and then standardized to have zero mean and a standard deviation of one, which implies that the coefficients can be interpreted as the effect of a one standard deviation change in the independent variable on the dependent variable, expressed as a fraction of the standard deviation of the dependent variable. *t*-statistics are in parentheses below the coefficient estimates. Significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively. The last three rows report the R^2 and adjusted R^2 and the number of time-series observations for each equation.

	Panel A: All stocks		Panel B: Liquid stocks		Panel C: Illiquid stocks	
Dependent variable is co-movement in:	liquidity	efficiency	liquidity	efficiency	liquidity	efficiency
$\lvert\Delta TED\ spread_{m-1}\rvert$	0.250***	0.192**	0.000	0.159*	0.097	0.202**
	(2.66)	(2.07)	(0.00)	(1.83)	(1.09)	(2.28)
$\lvert Fund\ flow_{m-1}\rvert$	0.054	0.182**	-0.032	0.037	-0.089	0.109
	(0.63)	(2.16)	(-0.37)	(0.43)	(-1.05)	(1.27)
Quotes/Volume$_{m-1}$	-0.25**	-0.070	0.015	-0.18*	0.150	-0.114
	(-2.42)	(-0.70)	(0.16)	(-1.90)	(1.54)	(-1.17)
Co-movement in liquidity$_{m-1}$	-0.002	0.068	0.058	0.020	0.124	0.006
	(-0.02)	(0.67)	(0.71)	(0.24)	(1.54)	(0.08)
Co-movement in efficiency$_{m-1}$	0.058	0.025	0.053	0.154*	-0.094	0.106
	(0.57)	(0.25)	(0.65)	(1.96)	(-1.15)	(1.30)
R^2	6.01	8.85	0.89	5.52	5.45	6.62
adj. R^2	3.25	6.17	-2.03	2.74	2.67	3.87
# Obs.	176	176	176	176	176	176

Table 2.8 – Vector autoregressions of co-movement in efficiency: Residual correlations

This table reports the contemporaneous correlations between the innovations (residuals) in the following five endogenous variables in three multivariate vector autoregressions (VARs): absolute changes in the TED spread ($|\Delta TED\ spread|$), absolute hedge fund flows ($|Fund\ flow|$), the total number of quote updates scaled by aggregate dollar trading volume (*Quotes/Volume*), co-movement in proportional quoted spreads (*Co-movement in liquidity*), and *Co-movement in efficiency*. We refer to Table 2.6 and Figure 1 for a description of these variables. In the first VAR (Panel A), the measures of co-movement in liquidity and efficiency are constructed using all NYSE stocks in the sample. In the second and third VARs (Panels B and C), the measures of co-movement in liquidity and efficiency are constructed using only the stocks in the decile portfolios of the most liquid and most illiquid stocks, respectively (these liquidity portfolios are formed based on the stocks' average proportional quoted spread, as in Table 2.5). We estimate all three VARs with one lag, following the Schwarz information criterion (SIC). All variables in the VARs have been detrended using a linear and a quadratic time trend and then standardized to have zero mean and a standard deviation of one. p-values are in parentheses. Significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively.

	$\lvert\Delta TED\ spread\rvert$	$\lvert Fund\ flow\rvert$	$\frac{Quotes}{Volume}$	*Co-movement in liquidity*	*Co-movement in efficiency*
Panel A: Residual correlations of VAR based on all stocks					
$\lvert\Delta TED\ spread\rvert$	1				
$\lvert Fund\ flow\rvert$	0.139* (0.07)	1			
Quotes/Volume	0.284*** (0.00)	0.035 (0.65)	1		
Co-movement in liquidity	0.392*** (0.00)	0.151** (0.05)	0.366*** (0.00)	1	
Co-movement in efficiency	0.080 (0.29)	0.001 (0.99)	0.321*** (0.00)	0.614*** (0.00)	1
Panel B: Residual correlations of VAR based on liquid stocks					
$\lvert\Delta TED\ spread\rvert$	1				
$\lvert Fund\ flow\rvert$	0.157** (0.04)	1			
Quotes/Volume	0.310*** (0.00)	0.047 (0.54)	1		
Co-movement in liquidity	0.237*** (0.00)	0.101 (0.18)	0.154** (0.04)	1	
Co-movement in efficiency	-0.022 (0.78)	0.028 (0.71)	0.149** (0.05)	0.311*** (0.00)	1
Panel C: Residual correlations of VAR based on illiquid stocks					

Table 2.8 – continued

	$\lvert \Delta TED\ spread \rvert$	$\lvert Fund\ flow \rvert$	$\frac{Quotes}{Volume}$	*Co-movement in liquidity*	*Co-movement in efficiency*
$\lvert \Delta TED\ spread \rvert$	1				
$\lvert Fund\ flow \rvert$	0.134* (0.08)	1			
Quotes/Volume	0.296*** (0.00)	0.031 (0.68)	1		
Co-movement in liquidity	0.258*** (0.00)	0.032 (0.67)	0.201*** (0.01)	1	
Co-movement in efficiency	0.269*** (0.00)	0.037 (0.62)	0.278*** (0.00)	0.317*** (0.00)	1

Table 2.8 reports contemporaneous correlations between the innovations (residuals) in the five endogenous variables in each of the three multivariate VARs we estimate. These results show the relation between shocks to the different endogenous variables in the VARs. Panels A, B, and C show the residual correlations for the VARs based on, respectively, all stocks, liquid stocks, and illiquid stocks. Panel A shows significantly positive contemporaneous correlations between shocks to the *TED spread* and shocks to *Hedge fund flow* and *Quotes/Volume*, which indicates a positive association between shocks to funding liquidity and shocks to more direct proxies for the intensity of arbitrage activity. Shocks to *Co-movement in liquidity* are positively and significantly correlated with shocks to the *TED spread*, *Hedge fund flow*, and *Quotes/Volume*, which indicates a link between shocks to funding liquidity and liquidity co-movement, consistent with Brunnermeier (2009) and Hameed et al. (2010). There is a strong positive contemporaneous correlation between shocks to *Co-movement in liquidity* and *Co-movement in efficiency*. Shocks to our proxy for algorithmic trading (*Quotes/Volume*) also show a clear positive contemporaneous relation with shocks to *Co-movement in efficiency*. Panels B and C of Table 2.8 by and large show similar results for liquid and illiquid stocks, respectively, though the contemporaneous correlations between shocks to the funding liquidity proxies and shocks to *Co-movement in liquidity* are somewhat weaker for liquid stocks, while for illiquid stocks, we observe that shocks to the *TED spread* are positively correlated with shocks to all four other endogenous variables, including *Co-movement in efficiency*. Overall, the results in Table 2.8 indicate that shocks to several of our proxies for funding liquidity and the intensity of arbitrage activity are positively associated with shocks to the degree of co-movement in both liquidity and efficiency.

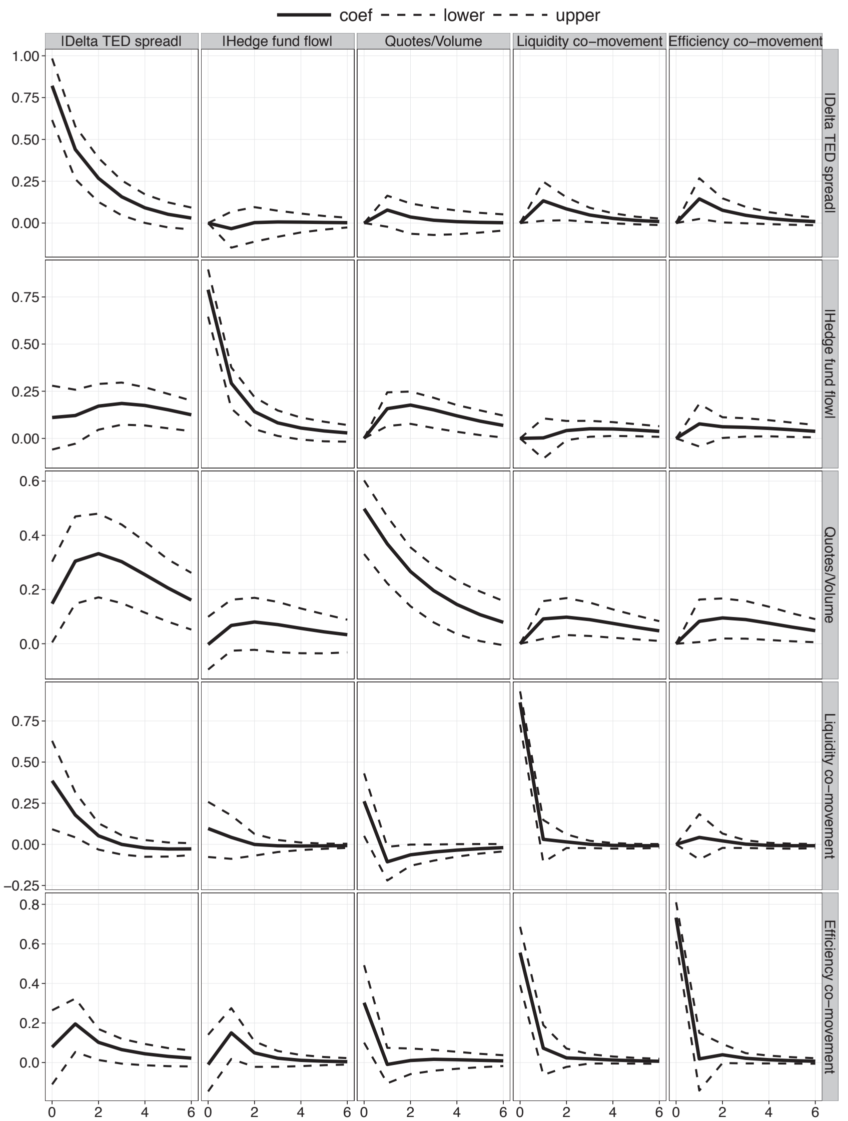
Although Tables 2.7 and 2.8 provide some initial evidence on the relations between (shocks to) the endogenous variables in the VARs, they do not account for the full dynamics of the VAR systems, and for the fact that shocks to the different endogenous variables are correlated (Table 2.8). Impulse response functions (IRFs) provide a more complete picture by tracing the impact of a one time, unit standard deviation, orthogonalized (using the inverse Cholesky decomposition) shock to one of the endogenous variables on current and future values of the other endogenous variables.

Figures 2.2, 2.3, and 2.4 present IRFs for the VARs estimated based on, respectively, all stocks, liquid stocks, and illiquid stocks. All IRF graphs show the response (measured in standard deviations) of the variable mentioned in the vertical legend to the right of the figure to a Cholesky one standard deviation shock to the variable mentioned in the horizontal legend at the top of the figure. Each IRF graph shows the response up to six months ahead (solid line labeled "coef"; month 0 on the horizontal axis of each IRF graph corresponds to the contemporaneous response), as well as the bootstrapped 95% confidence bands based on 1,000 runs (dashed lines labeled "lower" and "upper"). We note that these are not cumulative IRFs, so the responses depicted in the graphs are those pertaining to each of the individual horizons.

The main result in Figure 2.2 is that shocks to all three proxies for funding liquidity and the intensity of arbitrage activity have a significant direct effect on *Co-movement in efficiency* for at least one of the horizons under consideration. First, the response of *Co-movement in efficiency* to a shock to $|\Delta TED\ spread|$ (bottom left IRF in Figure 2.2) is positive and significant with a one- and two-month lag and is economically meaningful, at 0.1 to 0.2 standard deviations for these horizons. Second, the response of *Co-movement in efficiency* to a shock to $|Hedge\ fund\ flow|$ (second IRF on bottom row) is significantly positive with a one-month lag, and non-trivial in magnitude at around 0.15 standard deviations.[20] Third, the contemporaneous response of *Co-movement in efficiency* to a shock to *Quotes/Volume* (third IRF on bottom row) is significantly positive and economically large (at more than 0.3 standard deviations). In line with expectations, we also find a significant contemporaneous response of *Co-movement in efficiency* to a shock to *Co-movement in liquidity* (fourth IRF on bottom row). We would like to emphasize that the set-up of our VARs, with *Co-movement in efficiency* as the most endogenous variable, implies that all three measures of funding liquidity and the intensity of arbitrage activity have a significant and independent effect on *Co-movement in efficiency* that is not driven by their effect on *Co-movement in liquidity*.

[20] In some of the robustness tests reported throughout this paper, the p-value of the response of *Co-movement in efficiency* to a shock to $|Hedge\ fund\ flow|$ in the past month increases to just above 0.05.

Figure 2.2 – VAR of co-movement in efficiency: Impulse response functions (all stocks)
This figure shows impulse response functions (IRFs) for the vector autoregression (VAR) with one lag based on all NYSE stocks in the sample, with the following endogenous variables (in this order): $|\Delta TED\ spread|$, $|Hedge\ fund\ flow|$, *Quotes/Volume*, *Co-movement in liquidity*, and *Co-movement in efficiency* (see description in Tables 5 and 6 and Figure 1). Each IRF shows the response (measured in standard deviations, "coef") of the variable in the vertical legend to the right of the figure to a Cholesky one standard deviation shock to the variable in the horizontal legend at the top of the figure, and bootstrapped 95% confidence bands ("lower" and "upper").



In the IRFs that depict the response of *Co-movement in liquidity* to shocks in the other endogenous variables in the VAR (fourth row of Figure 2.2), we observe a significantly positive and large response to a shock to the *TED spread*, as measured contemporaneously as well as with a one-month lag. The effect of a shock to our proxy for algorithmic trading on *Co-movement in liquidity* is more complex. An increase in algorithmic trading is associated with an increase in *Co-movement in liquidity* in the same month, but a reduction in subsequent months.

Furthermore, in the second and third rows of Figure 2.2, we find that a shock to funding liquidity as picked up by a shock to the *TED spread* has a significantly positive effect on both $|Hedge\ fund\ flow|$ and *Quotes/Volume*, consistent with the view that funding liquidity affects the availability of arbitrage capital and the intensity of arbitrage activity. Perhaps not surprisingly, these effects are relatively long-lived, since it may take time for arbitrageurs to respond to a relaxation of funding constraints.

Figure 2.3 shows the IRFs for the VAR based on the decile portfolio of the most liquid stocks. In line with the VAR coefficient estimates in Table 2.7, the evidence of a significant response of the degree of co-movement in efficiency of the most liquid stocks in our sample to shocks to the proxies for funding liquidity and the intensity of arbitrage activity is weaker than for all stocks in Figure 2.2. The only significant effect is the significant contemporaneous response to a shock to *Quotes/Volume*. We also find a significantly positive contemporaneous response of *Co-movement in efficiency* to a shock to *Co-movement in liquidity*. *Co-movement in liquidity*, in turn, responds significantly to shocks to the *TED spread* in the same month, but not to the other proxies for funding liquidity and the intensity of arbitrage activity.

Figure 2.4 shows that, just like in Table 2.7, the main difference between the VAR results for liquid and illiquid stocks is that the degree of co-movement in efficiency of illiquid stocks does respond significantly to shocks to the *TED spread* for illiquid stocks. This positive response is prolonged and economically considerable, at 0.1-0.25 standard deviations contemporaneously and at lags of one and two months. Although the average degree of co-movement in efficiency is lower for illiquid stocks (Table 2.5), the findings in Figure 2.4 indicate that the degree of efficient co-movement of illiquid stocks is more sensitive to shocks to funding liquidity. A potential interpretation is that while there is less arbitrage activity and less pronounced common variation in price efficiency in stocks with greater frictions, a shock to the availability of arbitrage capital affects these stocks to a greater extent, possibly inducing amplified common changes in their price efficiency.[21]

[21] We note that the somewhat weaker results for the VARs based on liquid and illiquid stocks in Figures 2.3 and 2.4 may in part be driven by the more noisy estimates of *Co-movement in liquidity* and *Co-movement in efficiency* at the portfolio-level than at the market-level.

Figure 2.3 – VAR of co-movement in efficiency: Impulse response functions (liquid stocks)
This figure shows impulse response functions (IRFs) for the vector autoregression (VAR) with one lag based on the 10% most liquid stocks in the sample, with the following endogenous variables (in this order): $|\Delta TED\ spread|$, $|Hedge\ fund\ flow|$, *Quotes/Volume*, *Co-movement in liquidity*, and *Co-movement in efficiency* (see description in Tables 5 and 6 and Figure 1). Each IRF show the response (measured in standard deviations, "coef") of the variable in the vertical legend to the right of the figure to a Cholesky one standard deviation shock to the variable in the horizontal legend at the top of the figure, and bootstrapped 95% confidence bands ("lower" and "upper").

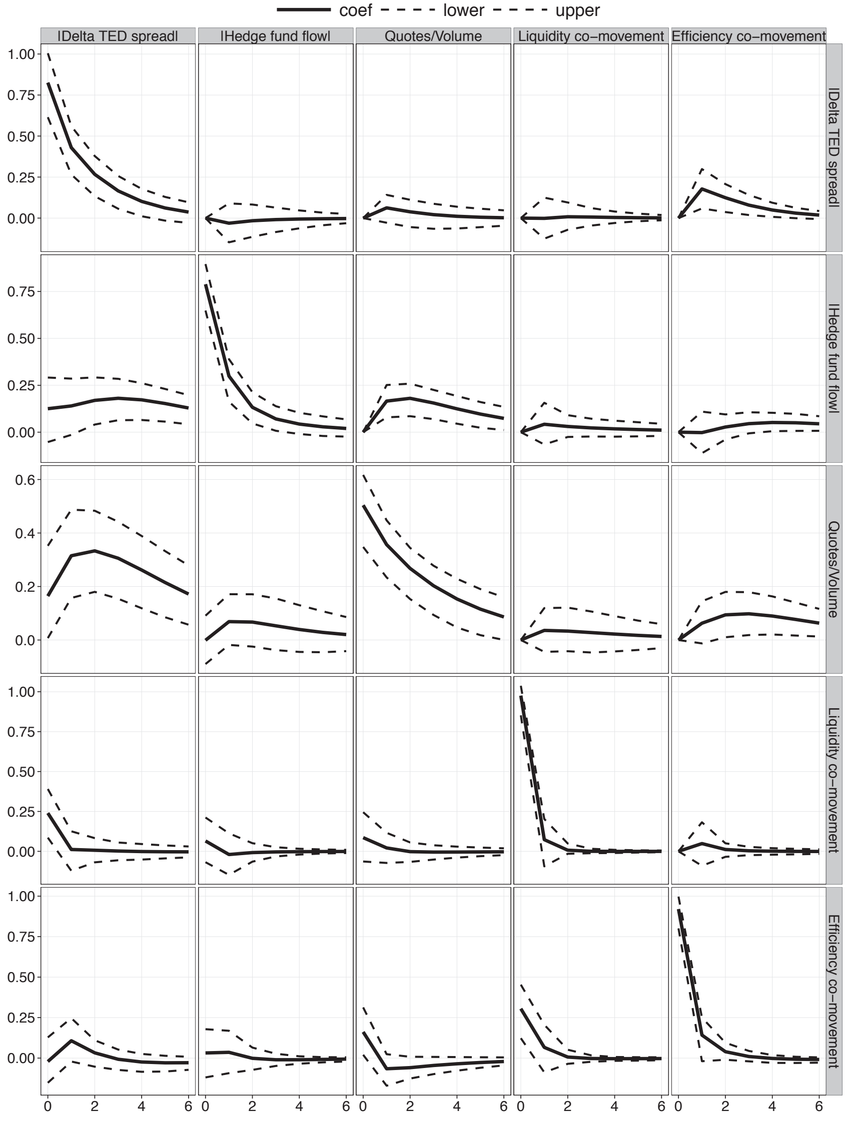
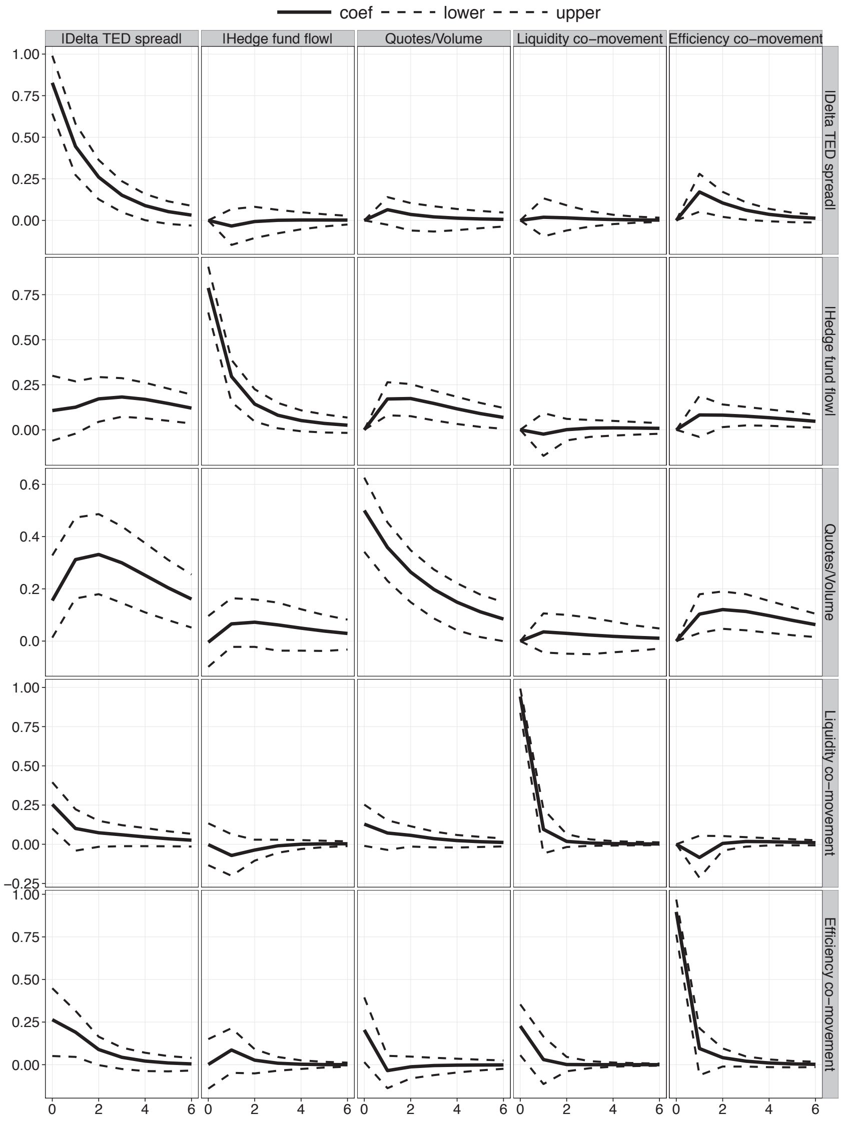


Figure 2.4 – **VAR of co-movement in efficiency: Impulse response functions (illiquid stocks)**
This figure shows impulse response functions (IRFs) for the vector autoregression (VAR) with one lag based on the 10% most illiquid stocks in the sample, with the following endogenous variables (in this order): $|\Delta TED\ spread|$, $|Hedge\ fund\ flow|$, *Quotes/Volume*, *Co-movement in liquidity*, and *Co-movement in efficiency* (see description in Tables 5 and 6 and Figure 1). Each IRF show the response (measured in standard deviations, "coef") of the variable in the vertical legend to the right of the figure to a Cholesky one standard deviation shock to the variable in the horizontal legend at the top of the figure, and bootstrapped 95% confidence bands ("lower" and "upper").



In sum, our VAR results indicate that funding liquidity and the intensity of arbitrage activity are important economic forces that help to understand time-variation in the degree of co-movement in market efficiency across individual stocks.

2.6 Conclusions

Market efficiency remains central to the study of financial markets, but most research to date has treated it as a static concept. In this paper, we consider variation in efficiency across stocks and over time, and examine the degree of co-movement in efficiency across individual stocks.

We show that five different stock-level market efficiency measures (intraday return predictability, variance ratios, the variance of intraday returns, Hasbrouck (1993) pricing errors, and put-call parity deviations) demonstrate considerable time-series and cross-sectional variation and also exhibit significant co-movement across stocks.

We then study the determinants of time-variation in the degree of co-movement in market efficiency. We first extract the first principal component across all five monthly, market-wide co-movement in efficiency measures and then include this variable as the last variable in vector autoregressions that also include proxies for funding liquidity and the intensity of arbitrage activity and a measure of co-movement in liquidity as endogenous variables. We show that shocks to funding liquidity (the TED spread) and to variables that more directly measure the intensity of arbitrage activity (hedge fund flows and a proxy for algorithmic trading) have a significant effect on the degree of co-movement in efficiency.

Overall, our results point to a significant, systematic, time-varying component in the degree of pricing efficiency of individual stocks, and to an important role of funding liquidity and the intensity of arbitrage activity in driving fluctuations in this component.

Recognizing that market efficiency is dynamic and co-moves across individual stocks opens new vistas for research. First, it would be worth exploring whether there is *global* co-movement in market efficiency across stock markets in different countries. This would allow us to ascertain the extent to which the quality of price formation in markets across the world has a systematic component, and whether fluctuations in global funding liquidity affect the degree of global co-movement in efficiency. Second, it would be worth investigating whether co-movement in market efficiency measures extends to other asset classes such as fixed income securities, foreign exchange, and derivatives. These and other issues are left for future research.

Chapter 3

The impact of arbitrage on market liquidity*

3.1 Introduction

Arbitrage enforces the law of one price and thereby improves the informational efficiency of the market, so that prices better reflect fundamentals. But how arbitrage affects other measures of market quality, in particular market liquidity, is less well understood.

This is an important question because recent changes to the trading environment (such as market fragmentation and high frequency trading) ease arbitrage, and policy choices that impede arbitrage (such as short-sell bans) might not only negatively affect the efficiency of the financial market, but also its liquidity and the cost of capital for firms.

In a recent paper by Gromb and Vayanos (2010) surveying the theoretical limits-of-

* This chapter is based on Rösch (2014) "The Impact of Arbitrage on Market Liquidity" available at http://ssrn.com/abstract=2295437. I thank Lamont Black, Ekkehart Boehmer, Dion Bongaerts, Howard Chan, Tarun Chordia, Ruben Cox, Louis Gagnon, Nicolae Gârleanu, Michael Goldstein, Amit Goyal, Allaudeen Hameed, Shing-yang Hu, Jonathan Kalodimos, Andrew Karolyi, Albert Kyle, Su Li, Albert Menkveld, Pamela Moulton, Maureen O'Hara, Louis Piccotti, Gideon Saar, Piet Sercu, René Stulz, Avanidhar Subrahmanyam, Raman Uppal, Dimitrios Vagias, Mathijs van Dijk, Manuel Vasconcelos, Kumar Venkataraman, Axel Vischer, Avi Wohl, and participants at the 2014 Financial Management Conference and Doctoral Consortium (Nashville), the 2014 Northern Finance Conference (Ottawa), the 2014 Asian Finance Conference (Bali), 2014 Eastern Finance Conference (Pittsburgh), the 2013 Erasmus Liquidity Conference (Rotterdam), the 2013 Conference on the Theories and Practices of Securities and Financial Markets (Kaohsiung), the 2013 World Finance and Banking Symposium (Beijing), and at seminars at Erasmus University and Cornell University for valuable comments. This work was carried out on the National e-infrastructure with the support of SURF Foundation. I am grateful for the hospitality of the Department of Finance at the Johnson Graduate School of Management (Cornell University), where some of the work on this paper was carried out, especially from my hosts, Andrew Karolyi and Pamela Moulton. I thank SURFsara, and in particular Lykle Voort, for technical support, and OneMarket-Data for the use of their OneTick software. I also gratefully acknowledge financial support from the Vereniging Trustfonds Erasmus Universiteit Rotterdam and from the Netherlands Organisation for Scientific Research through a "Vidi" grant.

arbitrage literature, the authors state that "arbitrageurs provide liquidity" (p. 258) because arbitrage opportunities arise from non-fundamental demand shocks, such as fire sales by mutual funds. The idea is that if arbitrage opportunities arise as a result of demand pressure, arbitrageurs trade against market demand and thereby decrease inventory holding costs for liquidity providers, which improves liquidity. For the examples given by Gromb and Vayanos (2010)—large price deviations that last for months—it is difficult to imagine otherwise.

But if arbitrage opportunities arise as a result of differences in information (for example, because local liquidity providers are slow to update their quotes) then "with arbitrage present, the adverse selection costs of domestic dealers increase, so that ... liquidity falls" (Domowitz et al., 1998), or simply: in this case arbitrage is "toxic" (Foucault et al., 2013).

In other words, whether arbitrage improves or worsens liquidity depends on the reasons why arbitrage opportunities arise. Theory predicts that if arbitrage opportunities arise as a result of demand shocks arbitrage improves liquidity (Holden, 1995; Gromb and Vayanos, 2010), but if arbitrage opportunities arise as a result of differences in information arbitrage worsens liquidity (Kumar and Seppi, 1994; Foucault et al., 2013).

Motivated by this observation, in this paper, I investigate why deviations from the law of one price arise and I estimate the impact of arbitrage on market liquidity. As price deviations can sometimes arise as a result of demand shocks and other times as a result of differences in information, so can arbitrage sometimes improve and other times worsen liquidity. In the extreme case where both effects are equally strong and cancel each other out, arbitrage will not have a visible effect on liquidity. This is my null hypothesis. Alternatively, if the effect of increased adverse selection dominates, arbitrage worsens liquidity, and if the effect of lower inventory holding costs dominates, arbitrage improves liquidity.

To study the impact of arbitrage on liquidity, I focus on the American Depositary Receipts (ADR) market. As laid out by Gagnon and Karolyi (2010b), the ADR market is particularly suitable to study arbitrage, because the ADR and the home-market share offer identical cash-flows (albeit in different currencies) and here institutions exist that facilitate arbitrage, implying that arbitrage likely occurs often in the ADR market.[1]

I examine intraday bid and ask quotes and trade prices for 72 ADRs and currency adjusted prices for the home-market share from Brazil, France, Germany, Mexico, and the U.K. over a long time frame from 1996 till 2013.

[1] Both JP Morgan and BNY Mellon (the world's largest depositaries for ADRs) confirm that arbitrage does occur in the ADR market. Employees of BNY Mellon confirmed that arbitrage is frequent in the ADR market during a meeting with the author. JP Morgan does list arbitrageurs as one investor type in the "Ownership" section of each ADR.

I first construct two intraday measures of price deviations. The first measure is the second-by-second difference between the highest bid and the lowest ask price across the ADR and the currency adjusted home-market share, which I refer to as gross *Opportunity-Profit*. The second measure is the difference in prices of "arbitrage trades", which I identify as trades on both the ADR and the home-market share within two seconds and while *Opportunity-Profit* is positive. Because recent empirical research finds that high-frequency traders significantly rely on limit orders (Menkveld, 2013; Brogaard et al., 2014b), I use unsigned trades, i.e. I consider trades regardless whether both legs are buyer or seller initiated. This way I also capture trades in which the arbitrageur trades in one market with a limit order and in the other market with a market order. This measure I refer to as gross *Traded-Profit*.

From these price deviations, I construct three (inverse) proxies for daily arbitrage activity. I assume that the market is reasonably efficient so that "prices reflect information to the point where the marginal benefits of acting ... do not exceed the marginal costs" (Fama (1991), p. 1575). In other words, the price deviations I observe reflect underlying frictions impeding arbitrage, such as risk, illiquidity, and capital constrains. Motivated by Gagnon and Karolyi (2010b)—who show that price deviations positively correlate with holding costs—and following Hu et al. (2013), I interpret large price deviations as "a symptom of a market in severe shortage of arbitrage capital" (p.2342). Consequently, I interpret the magnitude of the deviation as a proxy for the cost of capital the arbitrageur faces, which is inversely related to arbitrage activity.

The first (inverse) measure of arbitrage activity is the daily maximum *Opportunity-Profit*. Because *Opportunity-Profit* can be below, but not above the cost of capital—if profits would be above, the arbitrageur would step in and the opportunity would disappear—I interpret the daily maximum *Opportunity-Profit* as an inverse proxy for daily arbitrage activity. The second measure is the daily average *Traded-Profit*. Because the returns from each arbitrage trade should be close to the cost of capital, I interpret the daily average *Traded-Profit* as an inverse proxy for daily arbitrage activity. The third measure is the velocity at which *Opportunity-Profit* increase before an arbitrage trade occurs.

Using the intraday measures of price deviations, I start with investigating why price deviations arise. In particular, I investigate whether price deviations arise as a result of non-fundamental demand shocks or differences in information. Inspired by Schultz and Shive (2010), I identify non-fundamental demand shocks as situations in which price deviations arise as a result of temporary price movements, i.e. when one share moves to create the price deviation and later moves back to eliminate it. This identification is based on the common understanding that demand shocks are associated with price reversals and the incorporation of new information should have a permanent price effect (see, e.g. Gagnon and Karolyi (2009)). My analysis reveals that in the ADR market more than 70% of all arbitrage opportunities arise due to a non-fundamental demand shock.

In the second part of my paper I use the daily measures to investigate the impact of arbitrage on market liquidity. I find that the average daily maximum *Opportunity-Profit* is around 0.8% (as a percentage of the home-market share price) and the average daily *Traded-Profit* is around 0.5%, both similar to the cost-adjusted, absolute end-of-day price deviations reported by Gagnon and Karolyi (2010b) of 1.12%. The velocity at which *Opportunity-Profit* increase before an arbitrage trade is 4BP per minute, on average. All three measures are positively correlated.

I then estimate vector autoregressions and impulse response functions using arbitrage activity and liquidity as endogenous variables. Impulse response functions (from stock-specific and panel VARs) indicate that a positive shock to arbitrage activity predicts an increase in liquidity and a decrease in net market order imbalance. For the average home- and host-market share a one standard deviation shock to *Opportunity-Profit* predicts an increase of 0.25 and 0.13 standard deviations in quoted spreads over the next five days. Aggregating the stock-specific price deviations per exchange makes this effect become even stronger, in this case a one standard deviation shock to *Opportunity-Profit* predicts an increase of 0.35 and 0.20 standard deviations for the home- and host-market quoted spread. The average impact of arbitrage on liquidity might be lower at the stock-level than at the exchange level, because at the exchange level stock specific periods of higher adverse selection risks might get diversified.

To further highlight the impact of arbitrage on liquidity, I look at intraday differences in liquidity with and without arbitrage activity. Every day I identify periods with arbitrage as the time when both the home- and the host-market share trade, and the period when only one share trades as the period without arbitrage. I then look at impulse response functions estimated from VARs with arbitrage activity and intraday differences in liquidity as endogenous variables. Consistent with the notion of arbitrageurs as “cross-sectional market makers” (Holden, 1995), I find that a positive shock to arbitrage activity increases liquidity in the period with arbitrage relative to the period without arbitrage.

These results are robust to instrumental variables estimation, where I exploit the fact that corporate actions for the ADR and the home-market stock do not occur on the same day. For example, on days when only the ADR is cum-dividend arbitrageurs are likely less active, because the final dividend payment depends on the currency conversion rate at which the depositary bank could convert the dividends received on the home-market shares, which is in general only known weeks after the ex-date.

So far my results are consistent with theory, which predicts that if arbitrage opportunities arise as a result of demand shocks, arbitrage improves liquidity. In the last part of my paper I further investigate this prediction. I investigate whether the percentage of price deviations that arise as a result of demand shocks can explain part of the cross-sectional variation of the impact of arbitrage activity on liquidity. I find that the number of price deviations that arise in the home- and in the host-market, and the percentage of price deviations that arise as a result of demand shocks in the host-market can explain around one quarter of the variation of the impact of arbitrage on liquidity in the home-market. The percentage of price deviations that arise as a result of demand shocks in the home-market does not play a significant role in explaining the impact of arbitrage on liquidity in the home-market.

My paper relates to several different parts of the literature.

First, my paper relates to the empirical limits-of-arbitrage literature (among many other significant contributions: Mitchell et al. (2002); Lamont and Thaler (2003); De Jong et al. (2009); Gagnon and Karolyi (2010b).) The limits-of-arbitrage literature explains why arbitrage opportunities persist and how liquidity impacts arbitrage activity. I add to this literature by following previous literature to make an attempt to answer the questions (i) why arbitrage opportunities arise (Schultz and Shive, 2010) and (ii) how arbitrage impacts liquidity (Roll et al., 2007; Choi et al., 2009; Lou and Polk, 2013; Foucault et al., 2013; Ben-David et al., 2014). In a broader view, my paper is related to the literature that investigates how changes to the trading environment affect market quality (e.g. Chordia et al. (2005, 2008); Hendershott et al. (2011); Menkveld (2013); Chaboud et al. (2013); Brogaard et al. (2014b)). I add to these important contributions by empirically linking the question of why arbitrage opportunities arise to the impact arbitrage has on liquidity.

The main contribution of my study is to provide empirical evidence that arbitrage improves liquidity. I provide empirical evidence that most price deviations arise as a result of demand shocks and that an increase in arbitrage activity predicts a decrease in net market order imbalance. Both findings indicate that arbitrageurs trade against net market demand, and thereby improve international market integration and liquidity.

Second, I build upon previous work in the ADR literature. Especially, Gagnon and Karolyi (2010b) (who study price deviations in the ADR market) and Werner and Kleidon (1996); Moulton and Wei (2009) (who investigate differences in liquidity during and outside overlapping trading times). I add to this literature. In contrast to most previous studies I have access to tick-by-tick data for the home market, which allows me to combine both and study the impact of price deviations on the difference in liquidity during and outside overlapping trading times. I provide empirical evidence that an increase in arbitrage activity decreases the gap between liquidity during and outside overlapping trading times, providing further evidence that arbitrageurs improve international market integration and liquidity. These results can also provide an explanation for time-variation in liquidity differences during and outside overlapping trading times. Where Werner and Kleidon (1996) find that quoted spreads of ADRs in 1991 are higher during than outside overlapping trading times, using data from 2003 Moulton and Wei (2009) find the opposite. The increase in arbitrage activity provides one explanation for these different findings.

I consider the finding that arbitrageurs improve liquidity important for at least three reasons. First, the findings provide empirical justification for the assumption underlying the limits-of-arbitrage literature that arbitrage opportunities arise as a result of demand shocks. Second, the findings help to understand how policy changes that could hinder arbitrage activity (e.g. short-sell bans or transaction taxes) negatively impact not only the efficiency of the financial market, but also its liquidity, and ultimately the cost of capital for firms (Amihud and Mendelson, 1986). Third, the results add to the debate about how recent changes to the trading environment (such as fragmentation, and high frequency trading), seemingly helping arbitrage, affect market quality (Chordia et al., 2008; Hendershott et al., 2011; O'Hara and Ye, 2011; Menkveld, 2013).

3.2 Data and variable construction

3.2.1 Data and sample

To investigate the impact of arbitrage on market liquidity I focus on the American Depositary Receipts market (ADR), because with almost 10% of total NYSE trading value it is an important market and here institutions exist that facilitate arbitrage (Gagnon and Karolyi, 2013, 2010b). Many features are endemic to the ADR market [I refer to Karolyi (1998); Gagnon and Karolyi (2010a, 2013) for a detailed explanation and a comprehensive introduction to the ADR market], for example, the feature of convertibility—both ADR and home-market share can be converted to each other—allows to interpret price deviations between bid and ask prices at the time an arbitrageur opens the arbitrage position as (almost) risk-free profits.

If the currency adjusted bid price of the home-market share is higher than the ask price of the ADR in the host-market (and similarly, if the bid price of the host-market ADR is higher than the ask price of the home-market share) an arbitrage opportunity exists to simultaneously short sell the home-market share at the bid price, convert the proceeds from the short-sale into USD, and buy the ADR in the host-market at the ask price.[2] After that the ADR can be converted [within one business day and for less than five cents a share (Gagnon and Karolyi, 2010b)] into the home-market share either through a broker (e.g. Interactive Brokers), a crossing platform (e.g. ADR Max, or ADR Navigator), or the actual depositary bank. After the conversion the home-market share can be delivered to close down the short position, resulting in a risk-free USD profit equal to the difference between the bid of the home market and the ask of the host-market ADR *at the time the arbitrage position was opened*.

To construct my sample of ADRs and their respective home-market shares I use standard sources in the DR literature: Datastream, Bank of New York Complete Depositary Receipt Directory (www.adrbnymellon.com) and Deutsche Bank Depositary Receipts Services (adr.db.com). Details about the sample construction can be found in Appendix B.1. I focus on the NYSE as the host-market because it is the world's leading exchange in terms of listed Depositary Receipts (DR) and total trading in the DR market (Cole-Fontayn, 2011). I identify matched pairs of home/host-market shares and construct my sample based on the five home-market exchanges with the most identified pairs, and with overlapping trading times to the NYSE. This results in 72 pairs across the following five exchanges: the London Stock Exchange (the U.K., with 26 home-market shares), Sao Paolo Stock Exchange (Brazil, 17 shares), Bolsa Mexicana de Valores (Mexico, 11 shares), XETRA (Germany, 9 shares), and Euronext Paris (France, 9 shares).[3] For all matched pairs of home/host-market shares I obtain intraday data on quotes and trades (time-stamped with at least millisecond precision) as well as their respective sizes from the Thomson Reuters Tick History (TRTH) database over the sample period January, 1996, (the earliest date available in TRTH) till December, 2013. Similarly, I obtain intraday quotes on the currency pairs required to convert local prices into USD, the currency in which the ADR is quoted in, from TRTH.[4]

Quote and trade data is filtered as described in Appendix B.2. After the filtering the data contains 8,620,877,770 updates to the best bid and ask quote, of which around 50% are on ADRs, and 777,849,237 trades, of which 162,188,165 trades are on the ADR. I ignore stock-days on which the NYSE or the home-market exchange is only partly open. I also ignore stock-days in which prices of the ADR and the home-market share could not be aligned (days with price deviations above USD 10 or above 30%, as described in Appendix B.2). Further,

[2] Note that this example is for illustrative purposes only. In real markets short-selling is capital intensive, and an initial margin requirement of the initial value of the share plus 50% is required (in the US, Regulation T), which then also creates exchange rate risk.

[3] Focusing on ADRs excludes Canada and The Netherlands as potential home-markets, because, in general, stocks from Canada and The Netherlands do not list in the NYSE as ADRs, but as Canadian ordinaries and New York registered shares.

[4] The TRTH database is managed by the Securities Industry Research Center of Asia-Pacific (SIRCA) and is used in several recent studies, e.g. Marshall et al. (2011); Lau et al. (2012); Lai et al. (2014).

for the main analysis I drop stock-days in which one stock is ex- and the other stock is cum-dividend (and similar for other corporate actions). To be specific, I drop days after a corporate action occurs on one market till a corporate action occurs on the other market, with a maximum of six days.

Most of the analysis requires comparing prices across the ADR and the home-market share. To have valid, tradable quotes for both the ADR and the home-market share most of the analysis is based on overlapping trading times, i.e. when both the home- and host-market are in their continuous trading session.

Figure 3.1 – Continuous trading sessions per exchange 2008-10-15
This figure shows the hour of the day (x-axis) in which each of the five home-market exchanges (y-axis) is in their continuous trading session on one specific date, 2008-10-15 (horizonal lines). The vertical lines in the figure depict the opening (left) and closing (right) time of the continuous trading session at the host-market (NYSE). The x-axis shows the hour of the day in Coordinated Universal Time (UTC).

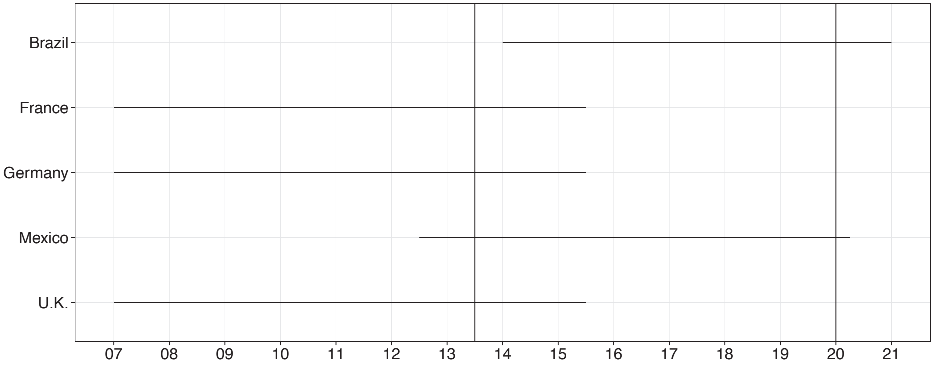


Figure 3.1 shows the continuous trading times for all five exchanges in the sample on October 15, 2008.[5] The opening and closing time at the NYSE is indicated by the left and right vertical line, respectively. The area within the vertical lines, in which the home-market is open, refers to the overlapping trading hours and is 2, 6, and 6.5 hours between Europe, Brazil, and Mexico and the NYSE, respectively.

3.2.2 Measures of price deviations

I construct two price deviation measures, the first one is based on quote prices and the second one is based on trade prices. The first measure I call *Opportunity-Profit* ($profit_{i,s}$), which I calculate for every stock i in every second s as the difference between the highest bid and

[5] Day light saving time (DST) does not follow the same rule in the USA and the other countries in the sample, which leads to variations in the overlapping trading hours within the year.

the lowest ask price across the home- and host-market relative to the mid price of the home market. If this difference is not positive, I set $profit_{i,s}$ to zero, i.e. *Opportunity-Profit* is calculated as:

$$profit_{i,s} = \max(\frac{bid.home_{i,s} - ask.adr_{i,s}}{mid.home_{i,s}}, \frac{bid.adr_{i,s} - ask.home_{i,s}}{mid.home_{i,s}}, 0) \quad (3.1)$$

where $mid.home_{i,s}$ is the last mid-quote price of stock i in second s, and $bid.home_{i,s}$ ($ask.home_{i,s}$) is the last bid (ask) of stock i in second s converted to USD using the prevailing bid (ask) of the respective currency pair, i.e. BRL for Brazil, GBP for the U.K., EUR for Germany and France (after January 1, 1999, and before DEM and FRF, respectively), and MXN for Mexico. Further $bid.adr_{i,s}$ ($ask.adr_{i,s}$) is the last bid (ask) in second s of the ADR trading at the NYSE associated to stock i, adjusted for the respective bundling ratio as described in Appendix B.2.

The second measure I call *Traded-Profit* ($trade.profit_{i,t}$), which I calculate for every stock i and for every simultaneous trade t as the absolute difference between the trade prices of the ADR and the home market stock, relative to the mid price of the home-market, i.e. $trade.profit_{i,t}$ is calculated as:

$$trade.profit_{i,t} = |\frac{trade.home_{i,t} - trade.adr_{i,t_1}}{mid.home_{i,t}}| \quad (3.2)$$

where $trade.home_{i,t}$ is the currency adjusted trade price for trade t of the home-market stock, and $trade.adr_{i,t_1}$ is the bundling adjusted trade price for trade t_1 of the ADR, such that t_1 minimizes the distance to t and both trades occur within two seconds, i.e. $|t - t_1| < 2seconds$.

3.2.3 Price deviations as an inverse proxy for arbitrage activity

From the two intraday, price-deviation measures introduced in the previous section, I construct three daily proxies for arbitrage activity. Investigating the impact of arbitrage on market liquidity using daily data—following Roll et al. (2007)—seems suitable to ensure that the time is short enough to measure the effect of market order imbalance (OIB) on liquidity (as empirically found by Comerton-Forde et al. (2010) using daily data), but on the same time long enough (compared to intraday data) to capture more persistent effects.

Unfortunately, a direct measure of arbitrage activity is not available, but a possible indirect (inverse) measure is absolute price deviation.

Previous literature measured arbitrage activity by the outcome of arbitrage activity, such as absolute price deviations (Roll et al., 2007; Ben-David et al., 2014) or return correlations (Lou and Polk, 2013). Alternatively, previous literature used the excess amount of short-selling to measure arbitrage activity (Choi et al., 2009; Hanson and Sunderam, 2014), but this measure is not feasible for arbitrage positions that are open for less than three business days (as is the case in the market I look at).[6]

I note that Roll et al. (2007) interpret *end-of-day* absolute price differences (the basis) as a direct measure of arbitrage activity, so that "if the basis widens on a particular day, arbitrage forces on subsequent days ... increase." Because I consider the maximum *Opportunity-Profit within* a day, I interpret the measure as an inverse measure of arbitrage activity.

If *Opportunity-Profit* indeed indicate capital constraints (as I conjecture following Hu et al. (2013)) and hence indicate less arbitrage activity, one would expect price deviations to be correlated with other measures of capital constraints. Further, one would expect that when price deviations in the ADR market are high, arbitrage activity in other markets is low, because capital constraints would likely affect arbitrage across all markets.

This is indeed the case. Empirically, I do find a negative correlation between price deviations and bank returns and a positive correlation between price deviations and the TED spread (the difference in one month Libor and T-Bill rates). In both cases the correlations are statistically significant and indicate that if funding liquidity is low (low bank returns, or a high TED spread) price deviations are high. In both cases I also obtain statistically significant correlations when controlling for the general risk in the market (proxied by the VIX).[7]

I also find that if price deviations in the ADR market are high, arbitrage activity in the index market is low (as reported by the NYSE as part of "Program Trading"). A one standard deviation increase in *Opportunity-Profit* is associated with a decrease of 0.43 standard deviations in index arbitrage volume.[8]

[6] Equity transaction (in the US) settle "T+3", i.e. traders are required to settle the transaction within three business days, if the short-position is open less than three business days it will likely not show up in any statistic.

[7] The TED spread and bank returns are widely used as proxies for funding liquidity, see for example Brunnermeier et al. (2008); Brunnermeier (2009); Hameed et al. (2010).

[8] Because price deviations can arise from both frictions in the home market as well as the USA (the host-market), I first extract the common component ("frictions in the USA") across all five exchange specific daily price deviations using principal component analysis. I find that the first component explains around one-half of the joint variation, that all loadings have the same sign and are of similar magnitude. To derive a daily time-series I then multiply the matrix of the exchange specific daily price deviations with the vector of the loadings from the PCA. First, this time-series is positively correlated to the TED spread and daily returns are negatively correlated to bank returns (Dow Jones U.S. Financial industry index; an inverse measure of capital constraints). Both correlations are statistically significant at the 5% level over the whole sample, in both the first half and for the bank returns in the second half of the sample, and survive controlling for general risk in the market (proxied by the US Volatility Index, VIX). Second, when explaining the monthly average of this time-series by a linear trend, and the index arbitrage volume as a percentage of total NYSE volume in the given month, I find that the estimated slope coefficient for arbitrage volume is -0.28951 with a t-statistic of -2.886. A one standard deviation increase in *Opportunity-Profit* is associated with a decrease of 0.43 standard deviations in index arbitrage volume. (The NYSE reports the number of shares traded as part of an index arbitrage at a weekly frequency as reported by the NYSE as part of their "Program Trading" press releases, e.g. in 2008 http://www.nyse.com/press/2_2008.html).

Motivated by these observations I construct two daily (inverse) proxies for arbitrage activity based on intraday absolute price deviations. First, the stock-day maximum *Opportunity-Profit* within the day (from Eq. 3.1). And second, the stock-day average *Traded-Profit* across all "arbitrage trades" (defined as trades on both the home-market share and the ADR within two seconds and which occur during positive *Opportunity-Profit*, from Eq. 3.2). As motivated in the introduction, the daily maximum *Opportunity-Profit* should be a better proxy for arbitrage activity than its average because price deviations need to be sufficiently large before an arbitrageur would trade on them.

One concern with these measures might be that arbitrage activity might vary purely because of variations in the costs associated with arbitrage (beyond the bid-ask spread). Another concern might be that both measures of arbitrage activity do not take into account the time it takes for an arbitrageur to become active. Despite a large *Opportunity-Profit*, if an arbitrageur quickly trades on it, arbitrage activity might still be considered relatively high.

To address both concerns, in the next section, I construct another (inverse) proxy for arbitrage activity based on the speed at which arbitrageurs get active.

3.2.4 The velocity at which price deviations increase before "arbitrage trades" as an inverse proxy for arbitrage activity

In the spirit of an event study (where the event is a simultaneous trade on both the home-market share and the ADR) I look at *Opportunity-Profit* before and after the event. If these trades are partly driven by arbitrage motivations I would expect *Opportunity-Profit* to increase before and decrease after the trade. In this case the velocity at which *Opportunity-Profit* increase before the trade could be interpreted as an inverse proxy for arbitrage activity. For example, if the velocity is small this indicates that either a small gradual rise or a sudden bigger jump in *Opportunity-Profit* is enough for an arbitrageur to step in and trade. On the other hand if the velocity is large, it indicates that arbitrageurs wait for an extended time for the *Opportunity-Profit* to increase before they trade. The latter case indicating less arbitrage activity than the former.

For each stock-day I average *Opportunity-Profit* per second from one minute before till one minute after a simultaneous trade occurs across all simultaneous trades within the day. For each stock i and each day d I get 121 observations n with $-60 <= n <= 60$ denoted

The availability of this more direct measure of arbitrage activity makes the index market particular suitable for comparison. Because index arbitrage activity is strongly influenced by the monthly expiry date of futures contracts, I consider monthly averages. Because press releases from NYSE are not available before Jun-06, I estimate this regression using monthly data from Jun-06 to Dec-13.

$profit_{i,d}(n)$. To focus on trades potentially driven by arbitrage motivations, I only consider simultaneous trades within positive *Opportunity-Profit*, i.e. where $profit_{i,d}(0)$ is positive. I then estimate regression Eq. 3.3 for each stock-day to explain the time variation of *Opportunity-Profit* around simultaneous trades. On days with more than one simultaneous trade, I estimate Eq. 3.3 using Weighted-Least-Squares regressions, with weights equal to one over the standard deviation of the average *Opportunity-Profit* in second n.

$$profit_{i,d}(n) = \alpha_{i,d} + \beta_{1,i,d} * Before_n + \beta_{2,i,d} * Time_n + \beta_{3,i,d} * Time_n * Before_n + \epsilon_{i,d,n} \quad (3.3)$$

where, $\alpha_{i,d}$ is the intercept, $Before_n$ is a dummy variable which is set to 1 before the event, $Time_n$ is a linear time trend, and $Time_n * Before_n$ is a time trend before the event.

Panel B of Table 2.1 reports the results of stock-day regressions as in Eq. 3.3. I report the pooled average estimated slope coefficient (with both time trends scaled by 60), the pooled average Newey and West (1994) *t*-statistic (*t*-stat avg) (which are capped at -100, and +100), the percentage of coefficients that are positive (*% positive*), and the percentage of coefficients that are positive and significant at the 5% level (*% + significant*). Further I report the average R^2 and the number of regressions over which the averages are taken (*# regressions*).

The number of regressions indicates that from 1996 to 2013 for 153,157 stock-days a simultaneous trade occurs inside positive *Opportunity-Profit*. This is a big fraction of the total number of stock-days of around 200,000.

The average R^2 is above 66%, which indicates that the proposed functional form captures most of the 2-minutes time variation in *Opportunity-Profit* around simultaneous trades. Further *Opportunity-Profit* is higher before a simultaneous trade occurs (measured by $Before > 0$), and is strictly increasing ($Time + Time * Before > 0$) and decreasing ($Time < 0$) before and after the event, respectively. *Opportunity-Profit* at the time the trade occurs is on average 53BP ($0.41 + 0.06 + 1 * (0.12 - 0.06)$), which is almost 30% higher than the average opportunity profit (the estimated intercept of 41BP).

Because *Opportunity-Profit* rise before and fall after simultaneous trades during positive *Opportunity-Profit*, I interpret these trades as driven by arbitrage. Of course taking the average across all simultaneous trades with positive *Opportunity-Profit* will wrongly classify many trades as arbitrage trades and potentially miss out several arbitrage trades. However, the finding of a statistically and economically significant increase in *Opportunity-Profit* before and decline after these trades indicates that these trades are at least partly driven by arbitrage motivations.

One concern might be that I use all simultaneous trades (within positive *Opportunity-Profit*), and not only signed simultaneous trades, i.e. simultaneous trades for which the trade for one market is buyer and the other market is seller initiated. However, this would not allow arbitrageurs to use limit orders, a potentially unrealistic restriction.[9]

To establish a daily proxy for arbitrage activity I measure the *Velocity* at which *Opportunity-Profit* increase before an *arbitrage trade* as the sum of the estimated slope coefficients from both time-trends in Table 2.1.

3.2.5 Measures of market liquidity and order imbalance

As the main liquidity measure I use the proportional quoted spread (PQSPR), defined as the daily time-weighted average of the difference in the ask and the bid price, scaled by the mid-quote price. For robustness, I also consider proportional effective spread, quoted depth, and the standard deviation of the pricing error as in Hasbrouck (1993) as alternative measures of market quality. All four measures have been widely used as measures of market quality before, for example Roll et al. (2007); Boehmer and Kelley (2009); Moulton and Wei (2009); Schultz and Shive (2010).

I further construct a measure of buying or selling pressure. I sign every trade in both the home market and the ADR using the Lee and Ready (1991) algorithm.[10] Second, to derive a daily order imbalance measure for each stock I take the absolute difference between the number of buyer- and seller-initiated trades in a given day (OIB).

3.2.6 Summary statistics

Panel A of Table 3.1 presents cross-sectional summary statistics of time-series averages for the main measures of arbitrage activity and liquidity.

For the majority of all stock-days both *Traded-Profit* (78%) and *Opportunity-Profit* (92%) are nonzero. While both *Opportunity-Profit* and *Traded-Profit* cannot be negative by construction, velocity can. However, Table 3.1 indicates that for most stock-days (65%) the velocity is positive. Considering that the velocity can only be calculated on 78% of all stock-days (days with a simultaneous trade during positive *Opportunity-Profit*), this indicates that veloc-

[9] Recent empirical research finds that both high-frequency and algorithmic traders significantly rely on limit orders, for example Menkveld (2013) finds that for one particular high-frequency trader "that employs a cross-market strategy ... four out of five of its trades are passive" (also compare Brogaard et al. (2014b); Chaboud et al. (2013)).

[10] A trade is classified as buyer- (seller-) initiated if it is closer to the ask (bid) of the prevailing quote. A trade at the midpoint of the quote is classified as buyer- (seller-) initiated if the previous price change is positive (negative). Lee and Radhakrishna (2000) and Odders-White (2000) give evidence that this algorithm is quite accurate for NYSE stocks, indicating that at least for ADRs misclassifications should be minimal.

Table 3.1 – Summary statistics of time-series averages, 72 home/host stock pairs, 1996 - 2013
Panel A of this table reports the cross-sectional average, standard deviation, minimum, median, and maximum of the time-series average by stock of the daily time-weighted average proportional quoted spread for the home market (*PQSPR Home*) and the host-market (*PQSPR Host*), the difference in quoted spread for the home-market share between the overlapping trading times and from 11 UTC until the host-market opens ($\Delta PQSPR$ *Home*), the difference in quoted spread for the host-market share between the overlapping trading times and from the time the home-market closes until 17 UTC ($\Delta PQSPR$ *Host*), the absolute order imbalance (the number of buyer minus seller initiated trades) for the home-market ($|OIB\ Home|$) and the host-market ($|OIB\ Host|$), the average number of price deviations per day (*# of price deviations*), the average time in seconds it takes till the price deviation disappears (*Seconds in deviation*), the daily highest *Opportunity-Profit* (*Max. Opportunity-Profit*), the daily average *Traded-Profit* (*Avg. Traded-Profit*), and the velocity at which *Opportunity-Profit* increase before a simultaneous trade within positive *Opportunity-Profit* (*Velocity*: the sum of both time-trends from Panel B of this Table). I measure *Opportunity-Profit* as the difference between the highest bid and the lowest ask price across the home- and host-market relative to the mid price of the home market (as in Eq. 3.1). I measure *Traded-Profit* as the absolute difference in trade prices across both markets that occur within two seconds and within positive *Opportunity-Profit* (as in Eq. 3.2). The first column (*%Days+*) indicates the percentage of stock-days in which the statistics are positive.
Panel B of this table reports the average of the regressions results from Eq. 3.3 estimated per stock-day. For each stock-day d I estimate the average *Opportunity-Profit* per second (denoted nth-second *Opportunity-Profit*) from one-minute before till one-minute after any simultaneous trade (the event: a trade on both the home- and the host-market share within two seconds) across all simultaneous trades within positive *Opportunity-Profit* and within the day. The dependent variable is the nth-second *Opportunity-Profit*. The independent variables are, an intercept (*Intercept*), a dummy variable which is set to 1 before the event (*Before*), a linear time trend (*T*), and a time trend before the event (*T*Before*). Panel B reports the pooled average estimated slope coefficient (coefficients for both time trends are scaled by 60), the pooled average Newey and West (1994) *t*-stat (*t*-stat avg), the percentage of coefficients that are positive (% *positive*), and the percentage of coefficients that are positive and significant at the 5% level (% + *significant*). Further for each regression I report the average R^2 and the number of regressions over which the averages are taken (*# regressions*).

Panel A: Cross-sectional summary statistics of time-series averages

	%Days+	avg	stddev	min	median	max
PQSPR Home(%)	1.00	0.45	0.80	0.05	0.16	5.01
PQSPR Host (%)	1.00	0.40	0.56	0.07	0.25	3.82
$\Delta PQSPR$ *Home(%)*	0.24	-0.02	0.08	-0.56	-0.00	0.01
$\Delta PQSPR$ *Host (%)*	0.26	-0.04	0.06	-0.26	-0.02	0.08
$\|OIB\ Home\|$	0.98	204.73	196.97	14.19	155.49	1231.84
$\|OIB\ Host\|$	0.95	78.55	83.72	4.73	50.78	474.31
# of price deviations	0.92	81	49	6	68	216
Seconds in deviations	0.92	252	234	5	160	1,043
Max. Opportunity-Profit (%)	0.92	0.82	0.48	0.15	0.72	2.61
Avg. Traded-Profit (%)	0.78	0.50	0.38	0.06	0.44	2.43
Velocity (%)	0.65	0.04	0.04	0.00	0.03	0.26

Panel B: *Opportunity-Profit* around simultaneous trades

	$Intercept_d$	$Before_d$	T_d	$T * Before_d$	R^2_d	# regressions
Opportunity-Profit	0.41	0.06	-0.06	0.12		
t-stat avg		1.51	-1.19	1.97		
% positive		70.70	24.47	79.91		
% + significant		38.61	8.51	49.69		
					66.10	153,157

ity is negative for only around 15% of all stock days. Further the time-series average velocity across all days in the sample is positive for all stocks.

I find that cross-sectional averages for *Traded-Profit* and *Opportunity-Profit* are similar in magnitude—consistent with the interpretation that both are a measure of the cost of capital an arbitrageur faces—the average of the daily maximum *Opportunity-Profit* is 0.82%, with a maximum of 2.61% for one Brazilian stock (with RIC CPFE3.SA). The average of *Traded-Profit* is 0.5% with a maximum of 2.43%. The *Velocity* at which *Opportunity-Profit* increase before an arbitrage trade is on average 4BP per minute, which seems relatively large considering that the average *Opportunity-Profit* around arbitrage trades is just 41BP (the intercept in Panel B of Table 2.1).

While Gagnon and Karolyi (2010b) use a different sample, and use end-of-day data from 1993 to 2004, the authors find a cost-adjusted absolute average price deviations of 1.12%, which is close to the average *Opportunity-Profit* in 1996 to 2004 of 1.07% (untabulated).

3.3 Do price deviations arise as a result of demand shocks or differences in information?

Using the intraday price deviations, I first follow Schultz and Shive (2010) and investigate why price deviations arise, because theory predicts that the impact of arbitrage on liquidity depends on why arbitrage opportunities arise. If arbitrage opportunities arise as a result of non-fundamental demand shocks arbitrageurs should act as "cross-sectional market makers" (Holden, 1995) and improve liquidity. But if arbitrage opportunities arise as a result of differences in information, arbitrageurs should increase adverse selection and deteriorate liquidity.

If for one particular stock i at time $t-1$ *Opportunity-Profit* are zero, but at time t *Opportunity-Profit* are positive, at least one bid or ask quote of at least one asset changed from time $t-1$ to time t (this asset is denoted the *First mover*, either the ADR, the home-market share, or the respective currency pair). Similarly, if *Opportunity-Profit* are positive till time $\tau - 1 > t$, but zero at time τ at least one bid or ask quote of at least one asset changed (this asset is denoted the *Last mover*). In this case I say that the *First mover* creates a price deviation for stock i at time t and the *Last mover* eliminates the price deviation at time τ.[11]

[11] In case the day opens with a price deviation, I consider the asset which market opened last as the *First mover*. On the other hand if a price deviation exists and either of the markets closes I drop this price deviation from the analysis in this section, as I do not know which asset closes down the arbitrage. Both cases are infrequent and do not impact the main results in this section. Further, to reduce potential noise, I ignore any price deviation that lasts less than one second, so that $\tau > t$.

Table 3.2 – Number of daily price deviations and reasons for why they arise, 1996 - 2013

This table presents the total number of price deviations (*# Price deviations*) by the asset that moves to create the deviation (*First mover*) and by the asset that moves to eliminate it (*Last mover*). The first column ($\#Stocks$) indicates the number of home- and host-market share pairs over which the statistics are computed. The second column indicate the asset that moves to create the price deviation: either the home-market share (*Home*), the host-market share (*Host*), both the home- and the host-market share (*Both*), or the respective currency pair (*Forex*). The third column (*#Price deviations*) indicates the total number of price deviations across all stocks and days in this category. The fourth column (*%Price pressure*) indicates the percentage of all price deviations that arise because of a temporary price movement, when one share moves to create the price deviation and later moves back to eliminate it (if the $Home$-market share is the first mover *%Price pressure* is defined as $(Home + Both)/(Home + Host + Both)$, and if the $Host$-market share is the first mover *%Price pressure* is defined as $(Host + Both)/(Home + Host + Both)$). A statistically significant difference from 0.5 at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively (based on a t-test of the per-stock estimates). The rest of the columns *Home*, *Host*, *Both*, and *Forex* indicate the percentage of all price deviations that get eliminated because of a movement in the respective asset.

Panel A: By exchange				Last mover:			
#Stocks	First mover	#Price deviations	%Price pressure	*%Home*	*%Host*	*%Both*	*%Forex*
All							
72	*Home*	3,735,537	0.70***	0.46	0.26	0.17	0.10
72	*Host*	4,644,340	0.78***	0.19	0.52	0.17	0.11
72	*Both*	2,288,232		0.24	0.32	0.36	0.08
72	*Forex*	1,953,310		0.20	0.25	0.11	0.43
Brazil							
17	*Home*	1,362,010	0.77***	0.58	0.22	0.14	0.05
17	*Host*	1,377,628	0.74***	0.24	0.53	0.17	0.05
17	*Both*	613,740		0.32	0.30	0.34	0.04
17	*Forex*	235,013		0.35	0.29	0.12	0.24
France							
9	*Home*	638,708	0.69***	0.40	0.26	0.17	0.17
9	*Host*	564,650	0.75***	0.20	0.44	0.18	0.18
9	*Both*	377,439		0.23	0.28	0.37	0.12
9	*Forex*	497,131		0.20	0.21	0.10	0.49
Germany							
9	*Home*	580,559	0.69***	0.39	0.26	0.20	0.14
9	*Host*	482,000	0.75***	0.21	0.38	0.25	0.16
9	*Both*	415,813		0.20	0.29	0.42	0.10
9	*Forex*	328,090		0.23	0.23	0.14	0.39
Mexico							
11	*Home*	533,859	0.64***	0.40	0.31	0.16	0.13
11	*Host*	1,238,613	0.84***	0.15	0.63	0.14	0.09
11	*Both*	363,548		0.23	0.36	0.33	0.08
11	*Forex*	351,896		0.20	0.31	0.10	0.39
U.K.							

Table 3.2 continued

#Stocks	First mover	#Price deviations	%Price pressure	Last mover: *%Home*	*%Host*	*%Both*	*%Forex*
Panel A: By exchange							
26	*Home*	620,401	0.65***	0.38	0.32	0.20	0.10
26	*Host*	981,449	0.81***	0.16	0.50	0.18	0.16
26	*Both*	517,692		0.19	0.35	0.36	0.10
26	*Forex*	541,180		0.13	0.26	0.12	0.50
Panel B: By subperiod							
1996 to 2002							
49	*Home*	372,583	0.76***	0.66	0.23	0.04	0.07
49	*Host*	138,165	0.51*	0.46	0.44	0.04	0.06
49	*Both*	18,207		0.56	0.30	0.07	0.07
49	*Forex*	56,947		0.43	0.19	0.03	0.35
2003 to 2013							
72	*Home*	3,362,954	0.70***	0.44	0.27	0.18	0.11
72	*Host*	4,506,175	0.79***	0.19	0.53	0.18	0.11
72	*Both*	2,270,025		0.24	0.32	0.36	0.08
72	*Forex*	1,896,363		0.20	0.26	0.12	0.43

Panel A of Table 3.2 reports the total number of price deviations across all 72 home/host-market pairs and, separately, across all five exchanges. Panel B of Table 3.2 reports the number of price deviations across two different subperiods, namely 1996 to 2002 and 2003 to 2013. The subperiods are chosen as in later parts I focus on data from 2003 to 2013 to mitigate issues arising from infrequent trading and stocks entering the sample.

Following Schultz and Shive (2010) I consider a price deviation to arise as a result of price pressure (demand shocks) if the share that moves to create the price deviation later moves back to eliminate it.

Table 3.2 reports all statistics by the *First* and *Last mover*.[12] For each *First mover* I separately report the percentage of all price deviations that arise as a result of price pressure.

In total I find 12,621,419 price deviations in my sample. Price movements in the host-market ADR create 4,644,340 of these price deviations, of which 52% are later eliminated because the price of the ADR moves back, and in only 19% a price movement of the home-market share eliminate the price deviation. The percentage of all price deviations that arise

[12] In the case that the currency pair moves together with any of the other two shares, the *First mover* is considered to be the other share.

as a result of price pressure in the host-market is 78%. Similarly, the percentage of all price deviations that arise as a result of price pressure in the home-market is 70%. The percentage of price deviations that arise as a result of price pressure in the host-market is higher than in the home-market for all exchanges, except Brazil. In other words, price deviations that arise as a result of price movements in the host-market are less likely than in the home-market to arise as a result of differences in information, which is consistent with previous literature that finds that price discovery normally occurs in the home-market (Halling et al., 2007; Gagnon and Karolyi, 2009).

For all five exchanges the percentage of price deviations created by a price movement in the ADR and that arise as a result of price pressure is higher than 74%. For price deviations created by a price movement in the home-market share this percentage is somewhat lower but also higher than 64% for all exchanges. In all cases this percentage is statistically significantly higher than 50%.[13]

Table 3.2 provides evidence that the majority of all price deviations arise as a result of price pressure in either the ADR or the home-market share. Indicating, that arbitrageurs trade against net market demand and act as “cross-sectional market makers” (Holden, 1995) most of the time. Of course, the overall impact of arbitrage on liquidity might still be negative. Table 3.2 indicates that many price deviations arise because of differences in information, which potentially might lead to “toxic arbitrage” (Foucault et al., 2013) and hence that arbitrage would worsen liquidity.

To study the overall effect I now turn to proxy arbitrage activity, and investigate the joint dynamics between arbitrage activity and market liquidity in the following sections.

3.4 The impact of arbitrage on market liquidity

3.4.1 Correlations between daily arbitrage activity, liquidity and order imbalance

To understand the joint dynamics between arbitrage activity, liquidity and order imbalance a natural first step is to study pairwise correlations. However, the variables might be correlated just because of a common time trend, or because of other calendar regularities. To address these concerns I follow Roll et al. (2007) and first expunge each variable from their time trend and other calendar regularities, i.e. I replace each observation of each time-series of variable

[13] I also obtain similar percentages of price deviations that arise as a result of price pressure if I only consider the one price deviation per stock-day with the highest *Opportunity-Profit*.

y and stock i by its residual from regression Eq. 3.4.

$$y_{i,d} = \alpha + \beta_{i,1} * T + \beta_{i,2} * T^2 + \sum_{n=3}^{6} \beta_{i,n} * DOW_n + \sum_{n=7}^{17} \beta_{i,n} * MON_n + \beta_{i,18} * MEX + \beta_{i,19} * FRA + \epsilon_{i,d} \tag{3.4}$$

where the independent variables are a linear (T), and a quadratic time-trend (T^2), four day-of-the-week dummies (DOW_n), and 11 month dummies (MON_n). Further to address sudden changes in USD quoted depth I include a dummy variable for stocks and their cross-listed counterpart from France (FRA) and Mexico (MEX), which is set to 1 after 2007-02-17 (2009-09-28).[14] Results of these regressions are unreported and available upon request.

Panel A and B of Table 3.3 report pairwise Pearson and Spearman rank time-series correlations between these adjusted series of daily estimates of arbitrage activity, liquidity and order imbalance. Time-series correlations are estimated over the whole sample per stock, and then averaged across all stocks in the sample. The percentage of stocks for which the correlation is positive and significant at the 1% level is given in parenthesis.

The average stock Pearson (Spearman) correlation between *Opportunity-Profit* and *Traded-Profit* is 64% (69%) and positive and significant at the 1% level for all stocks. Both measures are also positively correlated to the velocity (albeit weaker at around 10% to 14%), and significantly so for 61% to 78% of all stocks. All three (inverse) measures of arbitrage activity are positively correlated to both PQSPR for the ADR and the home-market share with coefficients around 20%, and with around 75% of all coefficients positive and significant. Further both *Opportunity-Profit* and *Traded-Profit* are positively correlated to OIB, albeit weaker (around 5%) and only around 40% of the coefficients are positive and significant.

The rather strong positive correlations between the difference in the highest bid and the lowest ask price across the home- and the host-market share (i.e. *Opportunity-Profit*) and quoted spreads are somewhat surprising, because mechanically an increase in the spread in either the home- or the host-market would lower *Opportunity-Profit*. However, the finding supports the notion of *Opportunity-Profit* as an inverse measure of arbitrage activity, because illiquidity hampers arbitrage one would expect less arbitrage activity when illiquidity is high.

Of course, correlations do not account for the joined dynamics between arbitrage activity, liquidity and OIB, for this I estimate vector autoregressions in the next subsection.

[14] USD depth for stocks from France, and Mexico dropped by a factor of 100 on the 27-Feb-2007 and the 28-Sep-2009, respectively.

Table 3.3 – Whole sample correlations of daily arbitrage activity, quoted spread, and order imbalance
This table reports the average of the time-series correlations (estimated by stock over the whole sample) between the following daily measures: *Opportunity-Profit* (*Opportunity*), *Traded-Profit*(*Traded*), velocity at which *Opportunity-Profit* increase before an arbitrage trade (*Velocity*), home- and host-market proportional quoted spread (*PQSPR Home*, and *PQSPR Host*), home- and host-market order imbalance (*OIB Home*, and *OIB Host*). For a description of these variables I refer to Table 2.1. All variables are detrended, i.e. residuals of Eq. 3.4 are used. All measures are computed during the overlapping trading times only, i.e. when both the home- and host-market share are trading. Panel A (Panel B) reports Pearson (Spearman rank) correlation coefficients averaged across all individual stock estimates and in parenthesis the percentage of how many estimates are positive and significant at the 1% level. All data underlying the computations are from TRTH.

Panel A: Pearson correlations

	Opportunity	*Traded*	*Velocity*	*PQSPR Home*	*PQSPR Host*	*OIB Home*
Traded	64.16					
	(100.00)					
Velocity	9.91	14.04				
	(61.11)	(62.50)				
PQSPR Home	21.31	20.22	16.49			
	(80.56)	(79.17)	(68.06)			
PQSPR Host	20.54	24.19	18.55	47.26		
	(79.17)	(80.56)	(79.17)	(95.83)		
OIB Home	6.65	3.21	1.16	-0.35	3.81	
	(45.83)	(30.56)	(6.94)	(19.44)	(43.06)	
OIB Host	6.39	4.57	1.16	2.50	0.16	14.76
	(44.44)	(33.33)	(5.56)	(33.33)	(34.72)	(88.89)

Panel B: Spearman correlations

	Opportunity	*Traded*	*Velocity*	*PQSPR Home*	*PQSPR Host*	*OIB Home*
Traded	68.82					
	(100.00)					
Velocity	11.76	14.26				
	(63.89)	(63.89)				
PQSPR Home	19.88	20.36	19.72			
	(77.78)	(77.78)	(79.17)			
PQSPR Host	20.32	23.41	21.87	48.89		
	(73.61)	(69.44)	(77.78)	(94.44)		
OIB Home	6.22	2.03	0.46	-4.01	0.23	
	(44.44)	(29.17)	(9.72)	(11.11)	(26.39)	
OIB Host	6.77	4.66	0.40	0.04	-4.00	12.76
	(50.00)	(36.11)	(6.94)	(27.78)	(22.22)	(90.28)

3.4.2 Stock level: Impulse response functions of arbitrage activity and liquidity.

Vector autoregressions (VARs) regress each variable on lagged versions of itself and on lagged versions of all other variables in the system. As such using VARs allows addressing endogeneity issues from contemporaneous regressions, where all variables are likely to have a causal impact on each other. This is likely the case between liquidity and arbitrage activity, because first liquidity encourages arbitrage activity, and second arbitrageurs might trade against net order imbalance, improving liquidity.

Further the impact of arbitrage on liquidity does not need to be contemporaneous alone. O'Hara and Oldfield (1986) and Comerton-Forde et al. (2010) provide theoretical and empirical evidence that overnight inventories affect future liquidity. If arbitrageurs trade against net market demand, an increase in arbitrage activity might lead to a lower order imbalance, which could predict an increase in liquidity.

An impulse response function (IRF) estimated from a VAR tracks the response on one variable from an impulse to another variable and hence allows investigating longer term behavior from variables that are jointly determined. Using the Cholesky decomposition to calculate orthogonalized impulse responses an IRF also allows estimating contemporaneous effects. But because in the Cholesky decomposition a variable only has a contemporaneous effect on other variables, if it enters the system of equations before the other variables, theory needs to guide the ordering of the variables (Doan, 2010). By construction *Opportunity-Profit* (Eq. 3.1) is negatively correlated to quoted spread, however, Table 3.3 indicates a positive correlation across the daily measures and as such Table 3.3 indicates only a weak contemporaneous effect of quoted spread on *Opportunity-Profit*. In the following I hence fix the order to *Opportunity-Profit*, home market liquidity, and last ADR liquidity, the same order as used in Roll et al. (2007). However, to rule out that results are driven by the ordering of the variables I also estimate IRFs using all the other five possible permutations of the order of the input variables, qualitatively leaving the results unchanged.

In the following I estimate a stock specific (and later a panel) VAR with five lags (motivated below) and endogenous variables as *Opportunity-Profit* (π_t), proportional quoted spread of the home-market share ($Home\lambda_t$) and proportional quoted spread of the host-market ADR

($Host\lambda_t$) as given in Eq 3.5.[15]

$$\pi_t = \sum_{d=1}^{5} \beta_{1,1,d} * \pi_{t-d} + \sum_{d=1}^{5} \beta_{1,2,d} * Home\lambda_{t-d} + \sum_{d=1}^{5} \beta_{1,3,d} * Host\lambda_{t-d} + \epsilon_{1,t} \tag{3.5a}$$

$$Home\lambda_t = \sum_{d=1}^{5} \beta_{2,1,d} * \pi_{t-d} + \sum_{d=1}^{5} \beta_{2,2,d} * Home\lambda_{t-d} + \sum_{d=1}^{5} \beta_{2,3,d} * Host\lambda_{t-d} + \epsilon_{2,t} \tag{3.5b}$$

$$Host\lambda_t = \sum_{d=1}^{5} \beta_{3,1,d} * \pi_{t-d} + \sum_{d=1}^{5} \beta_{3,2,d} * Home\lambda_{t-d} + \sum_{d=1}^{5} \beta_{3,3,d} * Host\lambda_{t-d} + \epsilon_{3,t} \tag{3.5c}$$

As mentioned before, all variables are estimated during the overlapping periods only and are first expunged of deterministic time trends and other calendar regularities (i.e. residuals from Eq.3.4 are used). Further these series are winsorized at the 1% level, i.e. for each stock the lowest (highest) 1% are set to the 1st (99th) percentile.[16] To estimate the responses in standard deviations I further standardize each series, i.e. from each observation I subtract the time series mean and divide each observation by the series standard deviation. For comparability across stocks, the lag-length for each VAR is fixed to five. The lag-length was chosen by first using the Akaike information criteria separately for each stock, which yields a lag-length between one and ten days. A good choice seems five days, which is around the median lag-length, and one working week.

Panel A of Table 3.4 reports Granger causality tests. I find that for more than half of all stocks I cannot reject the null hypothesis that *Opportunity-Profit* Granger causes home- and host-market quoted spread. Similarly, I find that for the majority of both the home and host-market shares quoted spread Granger causes *Opportunity-Profit*. Of course, Granger causality tests are based on a single equation and do not account for the full dynamics of the VAR, something that impulse response functions (IRFs) do.

Panel B of Table 3.4 reports the contemporaneous and cumulative impulse response after five days (i.e. the sum of the day-to-day responses) to a one standard deviation shock to the causal variable. Because the VAR is estimated on standardized data the IRF measures the response in standard deviations.

For the average stock a positive shock of one standard deviation to *Opportunity-Profit*

[15] I only add order imbalance in the later part when I estimate VARs at the exchange level. Adding order imbalance to the per stock VARs only marginally changes the results.

[16] Using non-winsorized data does not affect results for developed home markets (the U.K., France, and Germany) and results for emerging markets (Brazil and Mexico) remain qualitatively unchanged.

Table 3.4 – Stock impulse response functions, 1996 - 2013
This table reports results from stock-level vector autoregressions (VARs). VARs are estimated using 5-lags and time-series of daily arbitrage activity, and home- and host-market proportional quoted spread. I use *Opportunity-Profit* as an inverse proxy for arbitrage activity. All time-series are detrended and expunged from other calendar regularities (i.e. the residuals from regression Eq. 3.4 are used). For a description of these variables I refer to Table 2.1. Panel A of this table reports Granger causality tests: the percentage of stocks for which the null hypothesis that the coefficients of the column variable are jointly equal to zero when explaining the row variable is rejected at the 5% level. Panel B of this table report impulse response functions (IRFs), the contemporaneous effect ($d = 0$) and the effect after five days ($d = 5$) of a Cholesky one standard-deviation shock to *Opportunity-Profit* on home- and host-market share proportional quoted spread: the cross-sectional average, minimum, maximum, and the 25th, 50th (median), and 75th percentile. The first three rows of Panel B report the responses in standard-deviations to a one standard deviation shock to *Opportunity-Profit*, the next three rows to a shock to home-market share quoted spread (*PQSPR Home*), and the last three rows to a shock to host-market share quoted spread (*PQSPR Host*). The second column ($\%Sig+$) gives the percentage of stocks for which the response is positive and significant at the 5% level (based on bootstrapped error bands from 1000 runs). Similarly, Panel C and Panel D report IRFs using *Traded-Profit* and *Velocity* as alternative proxies for arbitrage activity. Significance of the cross-sectional average response at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively.

Panel A: Granger causality test (% of stocks for which column variable Granger causes row variable)

	Opportunity-Profit	*PQSPR Home*	*PQSPR Host*
Opportunity-Profit		52	61
PQSPR Home	45		77
PQSPR Host	56	72	

Panel B: cumulative IRF responses

	after shock	%Sig+	avg	min	25%	Median	75%	max
effect of one standard deviation shock to *Opportunity-Profit* on:								
Opportunity-Profit	$d = 0$	100	0.77***	0.35	0.65	0.84	0.93	1.00
	$d = 5$	100	1.53***	0.78	1.41	1.54	1.71	1.90
PQSPR Home	$d = 0$	81	0.09***	-0.15	0.06	0.09	0.13	0.24
	$d = 5$	76	0.25***	-0.24	0.15	0.26	0.36	0.73
PQSPR Host	$d = 0$	48	0.03***	-0.15	-0.02	0.03	0.07	0.12
	$d = 5$	51	0.13***	-0.27	0.04	0.12	0.26	0.45
effect of one standard deviation shock to *PQSPR Home* on:								
Opportunity-Profit	$d = 0$	0	0.00	0.00	0.00	0.00	0.00	0.00
	$d = 5$	37	0.07***	-0.22	-0.01	0.05	0.14	0.41
PQSPR Home	$d = 0$	100	0.68***	0.31	0.59	0.70	0.80	0.92
	$d = 5$	100	1.58***	0.88	1.46	1.63	1.70	1.87
PQSPR Host	$d = 0$	94	0.11***	0.03	0.09	0.11	0.14	0.23
	$d = 5$	94	0.39***	0.03	0.30	0.40	0.49	0.74
effect of one standard deviation shock to *PQSPR Host* on:								
Opportunity-Profit	$d = 0$	0	0.00	0.00	0.00	0.00	0.00	0.00
	$d = 5$	45	0.08***	-0.27	0.01	0.07	0.14	0.38
PQSPR Home	$d = 0$	0	0.00	0.00	0.00	0.00	0.00	0.00
	$d = 5$	70	0.13***	-0.15	0.05	0.12	0.21	0.40
PQSPR Host	$d = 0$	100	0.56***	0.25	0.44	0.57	0.69	0.79
	$d = 5$	100	1.41***	0.72	1.23	1.46	1.62	1.91

predicts a contemporaneous increase in home- and host-market quoted spreads of 0.09 and 0.03 standard deviations which increases to 0.25 and 0.13 standard deviations after five days. Further a shock to *Opportunity-Profit* results in a cumulative impulse response in home- and host-market quoted spread that is statistically significant at the 5% level for 76% and 51% of all stocks.

In the working paper version underlying this Chapter I repeat above analysis using *Traded-Profit* and *Velocity* as alternative (inverse) proxies for arbitrage activity. In both cases I find that a positive shock to arbitrage activity (i.e., a negative shock to either of the two proxies) predicts a statistically significant increase in liquidity for most stocks. After a positive shock to arbitrage activity, the average ADR quoted spreads increases by 0.24 standard deviations, when using *Traded-Profit*, and by 0.15 standard deviations, when using *Velocity* as the proxy for arbitrage activity. The effect on home-market quoted spread is similar to *Opportunity-Profit* when using *Traded-Profit* as a proxy for arbitrage activity (0.23 standard deviations), but somewhat weaker when using *Velocity* (0.15 standard deviations).

In an unreported robustness test I control for volatility as one other important driver of liquidity in the VAR used to construct the impulse response functions. This leaves the main results from Table 3.4 unchanged.

Consistent with theory and with the findings that most price deviations (around 70%) arise as a result of demand shocks, results in this subsection indicate that an increase in arbitrage activity predicts an increase in liquidity (lower quoted spreads). However, the question remains if arbitrage has any causal impact on liquidity. Arbitrageurs might be able to predict general changes in liquidity and in anticipation of an increase in illiquidity or funding constraints step out of the market (Shleifer and Vishny, 1997). In this case liquidity would deteriorate, regardless of whether arbitrageurs step out of the market or decide to continue to be active. Alternatively, arbitrageurs might directly influence liquidity, for example through cross-sectional market making (Holden, 1995). In other words, one concern might be that an omitted variable [such as funding liquidity (Brunnermeier and Pedersen, 2008)] could drive the predictive power of arbitrage activity on market liquidity. To investigate this question, in the next subsection, I look at intraday differences in liquidity with and without arbitrage activity.

3.4.3 Stock level: Impulse response functions of arbitrage activity and liquidity differences during and outside overlapping trading times.

In the Depositary Receipt market the same stock can often be observed with and without arbitrage activity (i.e. during and outside overlapping trading hours, as depicted by Figure 3.1). For Mexico and Brazil, however, the opening hours of the home market almost exactly overlap with the opening hours of the host market (NYSE) and hence I do not consider both markets in the following.

For the home market I examine differences in proportional quoted spread during the overlapping trading time and from 11 UTC (to avoid the general effects of the opening period) until the host market opens. In a similar way I look at differences in proportional quoted spread during the overlapping time and afterwards (till 17 UTC, to avoid the general effects of the closing period) for the host market. Like in the previous section these series are first adjusted for time trends and calendar regularities, i.e. residuals from individual stock regressions Eq.3.4, are used.

By using the residuals from regressions in which the differences in illiquidity during and outside overlapping trading hours is explained by an intercept, a time trend and other calendar regularities, I especially remove any general differences in illiquidity across the same day. Hence, if arbitrageurs would predict a general decline in liquidity and in anticipation withdraw from the markets, differences in liquidity across the same day should be around zero. However, if arbitrageurs provide liquidity the difference in liquidity between overlapping and non-overlapping trading hours should decrease, on and after days arbitrageurs get less active.

Panel A of Table 3.5 reports Granger causality tests. For around 19 out of the 44 home/host-market pairs (i.e. 43%) I find that I cannot reject that *Opportunity-Profit* Granger causes the differences in quoted spread during and outside overlapping trading times in the home-market stock (and similar for 52% for the host-market stock).

Panel B of Table 3.5 reports the contemporaneous and cumulative impulse response after five days (i.e. the sum of the day-to-day responses) to a one standard deviation shock to the causal variable.

For example, for the average stock a positive shock of one standard deviation to *Opportunity-Profit* predicts a contemporaneous increase in the difference in home- and host-market quoted spreads of 0.07 and 0.02 standard deviations which increases to 0.10 and 0.06 standard deviations after five days. In other words, if arbitrageurs get less active (a positive shock to *Opportunity-Profit*) liquidity deteriorates during the overlapping trading times relative to the

Table 3.5 – Stock impulse response functions between and outside overlapping trading times
This table reports results from stock-level vector autoregressions (VARs) using all 44 home/host-market pairs from France, Germany, and the U.K. VARs are estimated using 5-lags and time-series of daily arbitrage activity, and the difference in proportional quoted spread between and outside overlapping trading times for the home- and the host-market (Δ*PQSPR Home* and Δ*PQSPR Host*). I use *Opportunity-Profit* as an inverse proxy for arbitrage activity. All time-series are detrended and expunged from other calendar regularities (i.e. the residuals from regression Eq. 3.4 are used). For a description of these variables I refer to Table 2.1. Panel A of this table reports Granger causality tests: the percentage of stocks for which the null hypothesis that the coefficients of the column variable are jointly equal to zero when explaining the row variable is rejected at the 5% level. Panel B of this table report the contemporaneous effect ($d = 0$) and the effect after five days ($d = 5$) of a Cholesky one standard-deviation shock to *Opportunity-Profit* on home- and host-market share proportional quoted spread: the cross-sectional average, minimum, maximum, and the 25th, 50th (median), and 75th percentile. The first three rows of Panel B report the responses in standard-deviations to a one standard deviation shock to *Opportunity-Profit*, the next three rows to a shock to the difference in home-market share quoted spread during and outside overlapping trading times (Δ*PQSPR Home*), and the last three rows to a shock to the difference in host-market share quoted spread during and outside overlapping trading times (Δ*PQSPR Host*). The second column ($\%Sig+$) gives the percentage of stocks for which the response is positive and significant at the 5% level (based on bootstrapped error bands from 1000 runs). Significance of the cross-sectional average response at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively.

Panel A: Granger causality test (% of stocks for which column variable Granger causes row variable)

cause:	*Opportunity-Profit*	Δ*PQSPR Home*	Δ*PQSPR Host*
Opportunity-Profit		50	47
Δ*PQSPR Home*	43		50
Δ*PQSPR Host*	52	45	

Panel B: cumulative IRF responses

	after shock	%Sig+	avg	min	25%	Median	75%	max
effect of one standard deviation shock to *Opportunity-Profit* on:								
Opportunity-Profit	$d = 0$	100	0.70***	0.35	0.54	0.71	0.88	0.98
	$d = 5$	100	1.56***	0.80	1.44	1.57	1.76	1.96
Δ*PQSPR Home*	$d = 0$	63	0.07***	0.00	0.04	0.07	0.10	0.17
	$d = 5$	47	0.10***	-0.09	0.05	0.11	0.16	0.24
Δ*PQSPR Host*	$d = 0$	31	0.02***	-0.09	-0.01	0.03	0.06	0.10
	$d = 5$	34	0.06**	-0.25	-0.02	0.07	0.14	0.36
effect of one standard deviation shock to Δ*PQSPR Home* on:								
Opportunity-Profit	$d = 0$	0	0.00	0.00	0.00	0.00	0.00	0.00
	$d = 5$	0	-0.05***	-0.19	-0.10	-0.04	0.00	0.09
Δ*PQSPR Home*	$d = 0$	100	0.97***	0.87	0.97	0.98	0.99	1.00
	$d = 5$	100	1.25***	0.94	1.11	1.24	1.34	1.64
Δ*PQSPR Host*	$d = 0$	9	0.02***	-0.03	-0.00	0.01	0.03	0.08
	$d = 5$	11	0.03**	-0.12	-0.03	0.03	0.10	0.29
effect of one standard deviation shock to Δ*PQSPR Host* on:								
Opportunity-Profit	$d = 0$	0	0.00	0.00	0.00	0.00	0.00	0.00
	$d = 5$	9	0.00	-0.21	-0.05	0.00	0.06	0.15
Δ*PQSPR Home*	$d = 0$	0	0.00	0.00	0.00	0.00	0.00	0.00
	$d = 5$	18	0.01	-0.19	-0.06	0.02	0.08	0.28
Δ*PQSPR Host*	$d = 0$	100	0.91***	0.71	0.89	0.92	0.97	1.01
	$d = 5$	100	1.54***	1.17	1.37	1.61	1.70	1.79

non-overlapping trading times, i.e the difference in quoted spreads during and outside overlapping trading times increases. While these effects are relatively small, for most of the stocks these effects are statistically significant at the 5% level. For example, for 47% of all stocks I find that the difference in home market spread statistically significantly increases after a positive shock to *Opportunity-Profit*.

In the working paper version underlying this Chapter I repeat above analysis using both alternative proxies for arbitrage activity, i.e. *Traded-Profit* and *Velocity*, respectively. In both cases impulse response functions indicate that a positive shock to either of the two proxies predicts an increase in quoted spreads, for most stocks and for the average stock. However, the effect is slightly smaller and varies from 0.02 to 0.09 standard deviations, for the impact of *Velocity* on the host- and the home-market stock, respectively. One potential reason for the somewhat weaker effects using the measures based on trade prices (Panel C and D), is that I cannot construct these measures for many stock-days, because of missing simultaneous trades. This is a problem primarily at the beginning of the sample.

Previous research investigated differences in host-market quoted spreads during and outside overlapping trading times. For example, Werner and Kleidon (1996) find that quoted spreads in 1991 are higher during than outside overlapping trading times, but Moulton and Wei (2009) find the opposite using data from 2003. Because illiquidity during the opening and closing period is often elevated, and the overlapping trading time for the home- and host-market coincides with its closing and with its opening period, in segmented markets one would expect illiquidity of both the home as well as the host-market to be higher during the overlap than outside. On the other hand, if both markets were integrated, differences in illiquidity during or outside overlapping trading hours should be minimal (illustrated in Figure 2 of Werner and Kleidon (1996)). Because an increase in arbitrage activity decreases the difference in illiquidity during and outside overlapping trading times, above results provide support that arbitrage improves market integration. Further the increase in arbitrage activity over the years provide one explanation for the opposing findings in Werner and Kleidon (1996) and Moulton and Wei (2009).

In this section I exploit the fact that every day I observe each stock with and without arbitrage activity. This allows me to control for any omitted variables that would equally affect liquidity throughout the day. However, this approach does not control for other variables that specifically would affect liquidity during the overlapping trading times.

To address this concern I estimate a fixed-effect panel regression in the next subsection.

3.4.4 Fixed-effect panel VAR and impulse response functions of arbitrage activity and liquidity.

One way to deal with omitted variables is to use a fixed-effect panel regression. For example, if time-varying funding liquidity influences stock specific arbitrage activity and liquidity this could cause an omitted variable bias. If, however, funding liquidity affects stocks equally adding time-fixed effects will control for the stock-invariant differences in time. Similarly, I can control for time-invariant heterogeneity by using individual-fixed effects.

In this subsection I estimate a panel VAR with individual and time fixed effects. Because I have many more (daily) observations over time than across stocks using Arellano-Bond estimation in the dynamic panel is not necessary. As before the lag-length of the panel VAR is five.

To reduce the impact of having an unbalanced panel I only focus on data from 2003 to 2013, this ensures that each day I observe at least 13 stocks, and on average 54 stocks per day. This also ensures that I can construct both proxies for arbitrage activity based on simultaneous trades for 85% of all stock-days. As before the endogenous variables are a proxy for arbitrage activity and home- and host-market quoted spread. All variables are winsorized at the 1% level.

In Table 3.6 I report results of these panel VARs. In contrast to stock-by-stock VARs reported in Table 3.4 where the columns reported cross-sectional summary statistics, in Table 3.6 columns report results across the three different proxies for arbitrage activity. For parsimony, in Table 3.6 I only report the results of arbitrage activity as the causal variable.

Panel A of Table 3.6 reports Granger causality tests. These tests indicate that an increase in *Opportunity-Profit* Granger causes an increase in quoted spreads both in the home- and host-market (except the null-hypothesis that *Velocity* does not Granger cause ADR PQSPR could not be rejected at any reasonable confidence level).

Results from impulse response functions support previous findings.[17] The effect of a one standard deviation shock to any of the three different (inverse) proxies for arbitrage activity on quoted spread is positive and statistically significant in all cases, but one. The contemporaneous effect on quoted spread of the home-market share range from 1.45 BP when measuring arbitrage activity by *Opportunity-Profit* to 0.65 BP when using *Traded-Profit*. The cumulative effect after five days is positive, and statistically significant, in all cases and is 5.79 BP (when

[17] Estimation is done using RATS software using generalized impulse responses, which are not sensitive to the particular order of the endogenous variables (Pesaran and Shin, 1998). For details on estimation of the error-bands and generalized impulse responses I refer to Chapter 3.1 and 5.2 in Doan (2010).

Table 3.6 – Panel vector autoregression of arbitrage activity, and quoted spread, 2003 - 2013
This table reports results from three panel vector autoregressions. Panel vector autoregressions are estimated using individual and time-fixed effects and "arbitrage activity", home- and host-market share proportional quoted spread, with a lag-length of 5-lags. The first column reports result from a VAR where arbitrage activity is proxied by *Opportunity-Profit*. The next (last) column report result from a VAR where arbitrage activity is proxied by *Traded-Profit* (*velocity*). For a description of these variables I refer to Table 2.1. Panel A of this table reports Granger causality tests: the null hypothesis that the coefficients of "arbitrage activity" are jointly equal to zero when explaining the row variable. The first row reports the test statistic (χ^2_5 *(All coeffs. = 0)*), whether all coefficients are jointly equal to zero and the next row the associate p-value. Panel B of this table report the contemporaneous effect ($d = 0$) and the effect after five days ($d = 5$) of a one standard deviation shock to "arbitrage activity" on home- and host-market share proportional quoted spread using the generalized Cholesky decomposition. Two stars in parentheses after the coefficient indicate if the estimate is more than two standard deviations away from zero (using Montecarlo simulations based on 1000 draws). The last row indicates the number of observations in each of the three different VARs. All data underlying the computations are from TRTH.

Panel A: Granger causality test of arbitrage activity on:				
		Opportunity-Profit	*Traded-Profit*	*Velocity*
PQSPR Home	χ^2_5 (All coeffs. = 0)	29.45	46.70	32.56
	p-value	0.00	0.00	0.00
PQSPR Host	χ^2_5 (All coeffs. = 0)	31.94	32.62	5.21
	p-value	0.00	0.00	0.39

Panel B: cumulative IRF responses				
	Days after shock	*Opportunity-Profit*	*Traded-Profit*	*Velocity*
effect of one standard deviation shock to arbitrage activity on:				
PQSPR Home	$d = 0$	1.45(**)	0.65(**)	0.90(**)
	$d = 5$	5.79(**)	1.92(**)	3.12(**)
PQSPR Host	$d = 0$	-0.18()	1.03(**)	0.64(**)
	$d = 5$	1.64(**)	3.36(**)	2.11(**)
No. of obs.		121,939	88,619	88,619

measuring arbitrage activity by *Opportunity-Profit*), 1.92 BP (for *Traded-Profit*), and 3.12 BP (for *Velocity*). The effects are similar, but slightly attenuated for the host-market ADR quoted spreads.

Table 3.6 indicates that if arbitrage activity increases (by one standard deviation) quoted spread contemporaneously decreases by around 1 BP, and decreases by around 2 to 6 BP after five days. Considering that, for example, for UK stocks the average quoted spread in 2013 is just 6BP with a standard deviation of 3BP, an cumulative effect of around 2 BP after five days seems large.

A different way to establish the contemporaneous effect of arbitrage activity and liquidity is to use an instrument.

3.4.5 Days between corporate actions as an instrument

Because of endogeneity issues between arbitrage activity and liquidity including contemporaneous variables in the vector regressions of the previous subsections would lead to biased estimates. There I used the Cholesky decomposition to estimate the contemporaneous effect between arbitrage activity and liquidity.

An alternative approach to the Cholesky decomposition is to find a variable that is correlated with arbitrage activity, but not directly correlated with liquidity, i.e. an instrument.

While it is challenging to motivate and statistically impossible to verify that both assumptions hold, as a suitable candidate I propose a dummy variable that is one on days after the ex-date of one asset, but before the ex-date of the other asset.

For example, the Royal Bank of Scotland (RBS) paid out a stock dividend with ex-date of May 15, 2008, and for the ADR a cash-dividend of USD 0.674089 with ex-date of May 29, 2008.[18] Accordingly, price gaps between May 15 and May 28 spiked with an average of USD 0.92 of the daily maximum difference between the bid of the ADR and the currency adjusted ask of the home-market.

While these large *Opportunity-Profit* (of almost 20%) do not reflect possible arbitrage profits (and hence I drop these days for the other analyses), these days are likely characterized by lower arbitrage activity, because of additional risk. Consider the simplest case in which holders of the home-market share receive a cash dividend. In this case the final dividend

[18] In detail, during the annual meeting on May 14, 2008 shareholders approved a distribution of rights, which was not registered under the United States Securities Act of 1933. As such these rights could not be passed on to ADR holders. Instead the Depository Bank sold of these rights in the home market and passed on the proceeds to the ADR holders as a special dividend.

payment for the ADR holder depends on the exchange rate between USD and the currency of the home-market after the payment date of the dividend in the home-market, in general weeks $after$ the ex-date. Thus, the arbitrageur introduces uncertainty when adjusting prices for the corporate action to compute their profits, making arbitrage more risky (or costly if this additional risk would be hedged away).

As such it is not surprising that during these days *Opportunity-Profit* are especially high even after adjusting quotes by corporate actions (with the exact adjustment factor only known ex-post): The average *Opportunity-Profit* on these days from 1996 to 2004 (calculated from prices adjusted by corporate actions) is 4.8%, more than four times the cost-adjusted, absolute price deviation reported by Gagnon and Karolyi (2010b). After adjusting the quotes by the dividend payment in the example before (i.e. subtracting USD 0.674089 from all bid and ask quotes of the ADR), *Opportunity-Profit* are USD 0.24, almost 5% and double and more than three standard deviations higher than the average *Opportunity-Profit* for RBS in the first quarter of 2008. It is particular interesting to note that *Opportunity-Profit* from adjusted quotes monotonically declined from around USD 0.38 on May 15 (the ex-date of RBS) to USD 0.10 on May 28 (one day before the ex-date of its ADR), indicating uncertainty of the final cash dividend for the ADR holder.

However, it is likely that liquidity on these days is also directly influenced by the corporate action, after all one reason why a stock splits is to improve liquidity (Muscarella and Vetsuypens, 1996). To address this I do not look at the liquidity by itself but rather at the difference in liquidity during and outside times in which arbitrage takes place (as in Subsection 3.4.3). As before I do not consider home-market stocks from Mexico or Brazil, because their trading times overlap with those of the host-market (the NYSE).

In this section I estimate two-stage regressions as given below:

$$AA_{i,d} = \zeta D_{i,d} + \delta Controls_{i,d} + \sum FE_i + \sum FE_d + \eta_{i,d} \tag{3.6a}$$

$$\Delta PQSPR_{i,d} = \beta \widehat{AA_{i,d}} + \gamma Controls_{i,d} + \sum FE_i + \sum FE_d + \epsilon_{i,d} \tag{3.6b}$$

where $AA_{i,d}$ is a proxy for arbitrage activity for stock i on day d, and $D_{i,d}$ is a dummy variable set to one for days after a corporate action occurred on the home-market, but before the corporate action occured on the ADR (or vice versa) (the instrument). In the second equation $\Delta PQSPR_{i,d}$ is the difference in quoted spreads during and outside overlapping trading times (for the home- or the host-market), $\widehat{AA_{i,d}}$ is the fitted value from the first equation (the first stage), $Controls_{i,d}$ are control variables as described below, FE_i and FE_d are individual- and time-fixed effects. Like in the previous section these series are winsorized at the 1% level.

Table 3.7 shows the results of panel regressions using days between corporate actions

Table 3.7 – Instrumental variable panel regression of arbitrage activity, and quoted spread, 2003 - 2013
This table reports results from five panel regressions using daily observations from 2003 to 2013, and all 44 home/host-market pairs from France, Germany, and the U.K. The first column reports the "first-stage" (for regressions reported in column two and three) panel regression of regressing Opportunity-Profit on a dummy variable, which is set to one on days in which the home-market stock is ex- and the host-market stock is cum-dividend, or vice versa (*Days between corporate actions*). The next four column report results from (second-stage) panel regressions in which *Days between corporate actions* serves as an instrument for Opportunity-Profit. In column two and four the dependent variable is the difference in quoted spread of the home-market share between and outside overlapping trading times ($\Delta PQSPR_{Home}$). In column three and five the dependent variable is the difference in quoted spread of the host-market share between and outside overlapping trading times ($\Delta PQSPR_{Host}$). The time outside the overlapping trading time for the home-market share is from 11:00 UTC till the NYSE opens and for the host-market share it is from the time the European market closes till 17:00 UTC. The independent variables are the fitted value of *Opportunity-Profit* from the first stage (*Opportunity-Profit*$*$), a dummy variable set to one if a stock is from France (*France*), and if a stock is from Germany (*Germany*), the percentage of trades in the home market scaled by the total number of trades during overlapping trading times across the home- and host-market (*TradesInEurope*), and the difference in the number of trades between and outside overlapping trading times for the home- and host-market share ($\Delta Trades_{Home}$, and $\Delta Trades_{Host}$) as a percentage of the total number of trades within the day on the home- and host-market share, respectively. The third (*time-fixed effects*) and second (*individual-fixed effects*) to last row indicate if fixed effects are applied. The last row indicates the number of observations. Significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively. Newey and West (1994) t-statistics are given in parentheses. All data underlying the computations are from TRTH.

Dependent Variable [BP]:	*Opportunity-Profit*	$\Delta PQSPR_{Home}$	$\Delta PQSPR_{Host}$	$\Delta PQSPR_{Home}$	$\Delta PQSPR_{Host}$
instrument:		*Days between corporate actions* for *Opportunity-Profit*			
DaysBetweenCorpActions	111.93(***)				
	(25.09)				
Opportunity-Profit$*$ [%]		4.44(***)	34.98(***)	0.44(***)	2.81(***)
		(4.64)	(4.87)	(2.98)	(5.72)
France				0.96(***)	4.88(***)
				(13.84)	(25.97)
Germany				0.05(*)	1.55(***)
				(1.78)	(10.74)
TradesInEurope [%]				0.16(***)	-0.37(***)
				(17.45)	(-81.98)
$\Delta Trades_{Home}[\%]$				-0.15(***)	
				(-15.51)	
$\Delta Trades_{Host}[\%]$					-0.16(***)
					(-30.10)
time-fixed effects	Yes	Yes	Yes	Yes	Yes
individual-fixed effects	Yes	Yes	Yes	No	No
No. of obs.	93,599	93,599	93,599	93,599	93,599

as an instrument for arbitrage activity, i.e. results from estimating Eq. 3.6. I report both a panel regression with stock fixed effects (column two and three) and a panel regression where instead of controlling for time-invariant heterogeneity by using stock fixed-effects I use exchange fixed-effects and directly control for two popular explanations for the difference in liquidity during and outside overlapping trading times (column four and five). In both cases I use day-fixed effects to control for stock-invariant differences over time.

Moulton and Wei (2009) examine two explanations for differences in liquidity during and outside overlapping trading times: (i) concentrated trading, and (ii) increased competition. Similar to Moulton and Wei (2009) I proxy the former by the difference between the number of trades during and outside the overlapping trading times as a percentage of all trades for both the home- ($\Delta Trades_{Home}$) and the host-market ($\Delta Trades_{Host}$). I proxy the competition from the other exchange by the percentage of trades (during the overlapping trading times) that occur on the home-market versus at the NYSE ($TradesInEurope$).

In all regressions the estimated slope coefficient of *Opportunity-Profit* is positive, and statistically significant at the 1% level. In column two and three a 1% increase in *Opportunity-Profit* is associated with an increase in the difference between quoted spreads during and outside overlapping trading times of 4.44 basis points for the home-market and 34.98 BP for the ADR. Even after controlling for alternative explanations for the difference in spreads, a 1% increase in *Opportunity-Profit* is associated with an increase in spreads of 0.44 BP for the home-market and 2.81 BP for the ADR. These effects seem economically significant, as on average the difference between quoted spreads during and outside overlapping trading times for the home-market share (ADR) is just -2 BP (-4BP) with a standard deviation of just 8 BP (6 BP) (from Table 2.1). These results indicate that if arbitrage activity increases (*Opportunity-Profit* declines) liquidity improves during the time arbitrageurs are active (during overlapping trading times the quotes spread declines) relative to when they are not active.

The estimated slope coefficients for both controls have the expected signs. For example, column five indicates that a 1% increase in trading in the home-market, an increase in competition for liquidity providers in the host-market ($TradesInEurope$), is associated with a decrease of 0.37 BP in the difference in quoted spreads between and outside overlapping trading times for the ADR, so that liquidity during the overlap improves (relative to outside the overlap).

For robustness, in unreported tests I use *Traded-Profit* as an alternative measure of arbitrage activity and get similar results.[19] To ensure that above results are not driven by stock

[19] Note that the velocity at which *Opportunity-Profit* increase before an arbitrage trade cannot be reliably estimated during these days, because *Opportunity-Profit* do not reflect possible arbitrage profits.

splits, in an unreported robustness test I exclude days after corporate actions that are not dividend payments and instead of a dummy variable I use the dividend yield (dividend payment divided by the stock price) as an alternative instrumental variable. In both cases the main results are unchanged. Because so far the main results are consistent across all three different proxies for arbitrage activity, for parsimony, in the following I focus on *Opportunity-Profit* as the main proxy for arbitrage activity. In unreported tests I used the other two measures, which does not change the main results.

In the next subsection I estimate impulse response functions at the exchange level, i.e. by first averaging stock specific estimates across all stocks from the same exchange. This sheds light on if predictability of liquidity by arbitrage activity shares a common component or is purely idiosyncratic (i.e. stock specific).

3.4.6 Exchange level: Impulse response functions of arbitrage activity and liquidity.

In this section I estimate vector autoregressions on the exchange level, i.e. by first averaging stock specific estimates across all stocks from the same exchange. Previous research provides empirical evidence that both the liquidity and the efficiency of single stocks improve and deteriorate at the same time (Chordia et al., 2000; Karolyi et al., 2012; Rösch et al., 2015). Aggregating stock specific price deviations at the exchange level should reduce noise and other stock specific variations. Especially, periods during which stock-specific arbitrage opportunities mainly arise as a result of differences in information so that arbitrageurs would lower liquidity, could be diversified at the exchange level (compare, e.g. Lai et al. (2014)). In this case predictability between arbitrage activity and liquidity should be even stronger at the market level than at the stock level.

In addition to former impulse response functions on *Opportunity-Profit* and quoted spread, at the market level, I include a proxy for market demand, namely order imbalance, the absolute difference between the number of buyer and seller initiated trades.

If arbitrageurs trade against net market demand, as would be the case if price deviations arise as a result of demand shocks, a decrease in arbitrage activity should increase net order imbalances. The increase in order imbalances could then lead to a decline in contemporaneous and future liquidity (O'Hara and Oldfield, 1986; Chordia et al., 2002; Comerton-Forde et al., 2010).

Instead of tabulating the contemporaneous and cumulative five-day responses to a shock to arbitrage activity (as before), I now report graphs to highlight the day-to-day effect.

Figure 3.2 – Responses from shocks to *Opportunity-Profit* on home and cross-listed quoted spreads and order imbalance, 2003- 2013

This figure shows impulse response functions (IRF) from vector autoregression (VAR) estimated on exchange level (i.e. equally-weighted averages across all stocks in the sample from a given exchange) daily *Opportunity-Profit*, absolute net order imbalance in the home market (*OIB Home*), absolute net order imbalance in the cross-listed market (*OIB Host*), and average proportional quoted spread in the home- and host-market (*PQSPR Home* and *PQSPR Host*). For a description of these variables I refer to Table 2.1. All timeseries are detrended and expunged from other calendar regularities (i.e. residuals of regression Eq. 3.4) The lag length of each VAR is chosen individually (for each exchange) based on the Akaike information criterion. IRF are estimated for each different exchange (in columns) separately. All IRF show responses in standard deviations measured to Cholesky one standard-deviation shocks to *Opportunity-Profit*. All variables are measured during the overlapping trading time, i.e. when both the home market and the cross-listed market are in their continuous trading session. Each figure shows bootstrapped 95% confidence bands based on 1000 runs (lower, upper). All data underlying the computations are from TRTH.

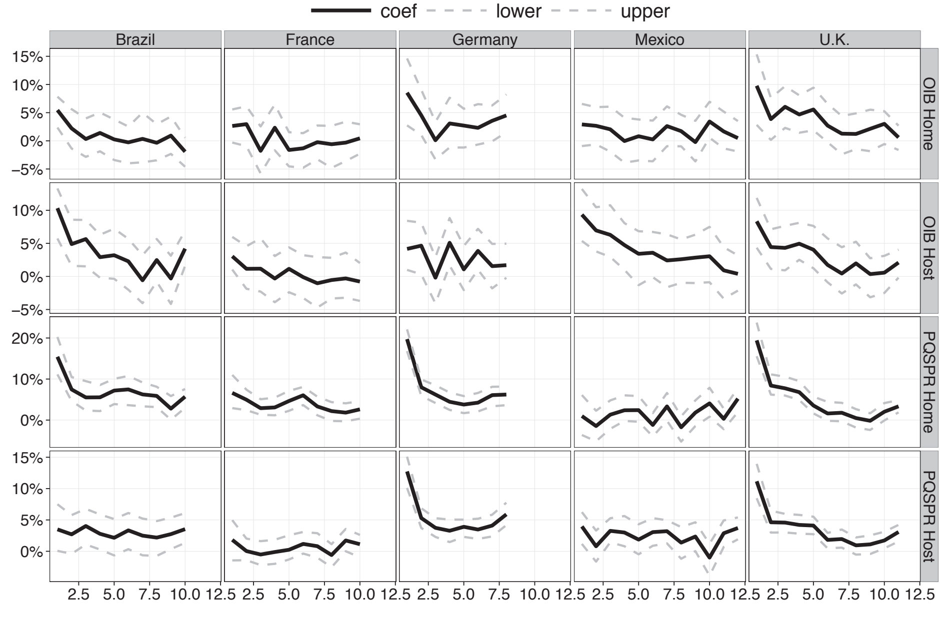


Figure 3.2 shows impulse response functions (IRFs) estimated from vector autoregressions on *Opportunity-Profit*, home and ADR order imbalance and proportional quoted spreads. These impulse response functions have been estimated by exchange (column). For parsimony Figure 3.2 only reports IRFs from shocks to *Opportunity-Profit*. The x-axis tracks the response through time starting from 1 (the contemporaneous effect) till the n-th day, the lag-length of the VAR, which was chosen by Akaike information criteria individually for each exchange and varies from 8 for Germany to 12 for Mexico.

As before, for each individual stock all five series are first detrended (i.e. residuals from Eg. 3.4 are used), and winsorized at the 1% level. I then take the equal weighted average across all stocks from a given exchange and standardize each series. These adjusted series on market *Opportunity-Profit*, order imbalance, and quoted spread are the input series for the VAR, and of this order. The order is motivated by: First, Table 3.2 indicates that most price deviations arise because of a demand shock, and hence arbitrage activity should contemporaneously affect market order imbalance. This motivates using *Opportunity-Profit* as the first variable. Second, previous literature indicates that order imbalance has a contemporaneous effect on market liquidity (Chordia et al., 2002). This motivates the order between measures of order imbalance and measures of market liquidity.

The first (second) row of Figure 3.2 shows the effect (y-axis) of an orthogonalized, one-standard deviation shock to *Opportunity-Profit* on home-market (ADR) order imbalance by day (x-axis). Similar, the third (last) row of Figure 3.2 show the effect on home-market (ADR) quoted spread.

In all but one case the IRF is positive and significant in the first few days after the shock, then decreases and becomes statistically insignificant. The negative slope in the IRFs indicates previous Dickey-Fuller tests (untabulated) that reject the existence of a unit-root in the adjusted series at the 1% level in all cases.

A one standard deviation shock to *Opportunity-Profit* leads to a contemporaneous increase in order imbalance of the home-market share (from 0.10 standard deviations in the U.K. to 0.03 in France), order imbalance of the ADR (from 0.10 in Brazil to 0.03 in France), quoted spread of the home-market (from 0.20 in Germany and the U.K. to 0.01 in Mexico), and quoted spread of the ADR (from 0.13 in Germany to 0.03 in France). One day after the shock the effect on order imbalance and quoted spread remains positive in all but one cases, and statistically significant except for order imbalance of the home-market in France and Mexico, and for quoted spread of the home-market (ADR) in Mexico (France).

This indicates that a positive shock to *Opportunity-Profit* (a decrease in arbitrage activity) predicts an increase in order imbalance and quoted spread contemporaneously and over the next few days, and provides evidence that arbitrageurs trade against market demand and thereby improve liquidity.

The effect of arbitrage activity at the exchange-level is much stronger than at the individual stock-level. For the average stock a positive shock of one standard deviation to *Opportunity-Profit* predicts an increase in home- and host-market quoted spreads of 0.25 and 0.13 standard deviations after five days (from Table 3.4). For the average exchange the impact almost doubles and increases to around 0.35 and 0.20 standard deviations for the home- and host-market quoted spread and can be as high as 0.5 standard deviations for the home-market quoted spread in Brazil, the U.K., and Germany. By aggregating estimations at the exchange-level noise and other stock-specific variation is reduced, which potentially can explain the difference in magnitudes.

I unreported robustness tests I use effective spread, quoted depth, and the standard deviation of the pricing error (Hasbrouck, 1993) as alternative measures of market quality and both *Traded-Profit* and the velocity as alternative measures of arbitrage activity. In all cases the results are similar.

The results are also robust for using a slightly different time period and a different order of the endogeneous variables (as reported in a previous version of this paper). Using data from 2001 till 2011 with the default order the shock to *Opportunity-Profit* results in a cumulative significant response (at the 5% level) after 5-days for quoted spread, effective spread, and quoted depth in 22 out of the 30 cases (three variables times five exchanges, for both the home market and the ADR). Estimating IRFs with the reverse order in which *Opportunity-Profit* is last, indicates that a positive shock to *Opportunity-Profit* predicts an cumulative increase in illiquidity (quoted spread, effective spread, and quoted depth) in 26 out of 30 cases and 15 of these are significant at the 5% level.

3.5 Does arbitrage improve market liquidity the more price deviations arise as a result of demand shocks?

Theory predicts that if arbitrage opportunities arise as a result of demand shocks arbitrageurs act as a "cross-sectional market maker" (Holden, 1995) and thereby improve market liquidity. So far the empirical results are consistent with this argumentation, I first find that most price deviations arise as a result of demand shocks (around 70% from Table 3.2), and then impulse response functions indicate that an increase in arbitrage activity improves future and contemporaneous market liquidity.

A natural question is thus, whether the percentage of price deviations that arise as a result of demand shocks can explain part of the variation of the impact of arbitrage activity on market liquidity. However, there is one other way how arbitrage might affect market liquidity. Ben-David et al. (2014) provide empirical evidence that arbitrage leads to negative spillover effects in the ETF market, such that a demand shock in the ETF leads to a liquidity shock in the underlying shares.

In other words, not only the percentage of price deviations that arise as a result of demand shocks is important, but also the number of price deviations is important, and especially how many price deviations in the opposite market arise as a result of demand shocks.

To answer these questions I look at cross-sectional regressions that explain the cumulative impulse response in quoted spreads to a shock in arbitrage activity (as reported in Table 3.4). I explain these responses by the stock-specific percentage of price deviations that arise as a result of demand shocks, and by the (logarithm of the) number of price deviations (taken from Table 3.2).

Table 3.8 reports results of these cross-sectional regressions. In column one and two I explain the impact of arbitrage on liquidity for the home-market share. In column three I use both the home- and host-market shares as dependent variables. In column one as independent variables I use price deviations that arise because of a price movement in both the home- and the host-market. In the rest of the columns I focus on the percentage of price deviations that arise as a result of demand shocks in the opposite market, i.e. in column two, the host-market.

In all specifications I find that the number of price deviations that arise in the opposite market negatively affects the impact arbitrage has on market liquidity. A stock, all else equal, but with double the number of price deviations that arise in the opposite market, is expected to have an impact of arbitrage on liquidity which is lower by 0.2 standard deviation than the other stock, i.e. for this stock a one standard deviation shock to *Opportunity-Profit* would change quoted spread by $x - 0.2$ standard deviations, where x is the impact for the other stock. Importantly, this is mitigated by how many price deviations arise as a result of demand shocks. Comparing two stocks, one for which all and the other for which no price deviations arise as a result of demand shocks, the former stock has an impact of arbitrage on market liquidity which is higher by one standard deviation. I also find that the number of price deviations that arise in the same market positively affect the impact arbitrage has on market liquidity. This is consistent with the idea of arbitrageurs as "cross-sectional market makers" (Holden, 1995), because their services would only be required if price deviations actually occur in the first

Table 3.8 – Regressions to explain the impact of arbitrage on market liquidity, 1996 - 2013
This table presents the results of cross-sectional regressions to explain the impact of arbitrage on market liquidity by the reason why price deviations arise. The dependent variable *PQSPR Home* and *PQSPR* is the five day cumulative impact of a shock to *Opportunity-Profit* on home- and host-market share quoted spread (from Table 3.4). The independent variables are the logarithm of the number of price deviations that arise because of a movement in the same (*First mover same*) or opposite market (*First mover opposite*) (*log # price deviations*), a percentage of how many of these price deviations arise as a result of price pressure (*% price pressure*)(in which the one share that moves to create the price deviation later moves back to eliminate it), and exchange dummies. The "same" or "opposite" market refers to the same or opposite market used in the dependent variable. In the first two columns regression are over the 72 home-market shares (and hence the opposite market is the host-market). The last column uses all 144 home- and host-market shares for the dependent variable. Significance at the 1%, 5%, and 10% level (using Newey and West (1994) standard errors) is indicated by ***, **, and *, respectively.

effect of shock to *Opportunity-Profit* on:	*PQSPR Home* (%)		*PQSPR* (%)
Intercept$_m$	-0.621 (-0.96)	-0.923* (-1.86)	-0.479* (-1.77)
First mover opposite			
log # price deviations	-0.232** (-2.59)	-0.190*** (-2.79)	-0.039 (-1.03)
% price pressure	0.929** (2.02)	0.888* (1.91)	0.585** (2.10)
First mover same			
log # price deviations	0.282*** (2.98)	0.244*** (3.13)	0.076* (1.84)
% price pressure	-0.442 (-1.01)		
Exchange dummies	Yes	Yes	Yes
R^2	24.89	23.54	14.17
adj. R^2	15.35	15.17	9.75
# Obs.	72	72	144

place. While the magnitude of the coefficients slightly varies across the three specifications all coefficients keep the same sign and in most cases remain statistically significant.

For robustness, in an unreported test, I repeat the above analysis using responses in home- and host-market quoted spreads estimated from a VAR with the reverse order in variables, i.e. with host-market quoted spreads first, and *Opportunity-Profit* last. The results practically stay unchanged.

3.6 Conclusion

Arbitrageurs enforce the law of one price by trading against mispricings, but if by doing so arbitrageurs improve market liquidity or not depends on the reason for the arbitrage opportunity to arise. In this paper I provide empirical evidence that is in line with the interpretation of arbitrageurs as "cross-sectional market-makers" (Holden, 1995). Arbitrageurs are improving liquidity and are indeed trading against net market demand, or as Foucault et al. (2013) put it, arbitrageurs are "leaning against the wind" (p. 336). These results confirm the limits of arbitrage literature, which in general assumes that arbitrage opportunities arise because of non-fundamental demand shocks and hence assumes that arbitrageurs are improving liquidity (Gromb and Vayanos, 2010).

These results shed additional light on possible consequences of frictions impeding arbitrage, such as short-selling bans or transaction taxes and hence might be of interest for policy makers. To curb excessive trading eleven European member states plan to introduce a transaction tax in 2016,[20] while liquidity providers might be exempted, arbitrageurs likely will not. The tax will have an adverse effect on arbitrage activity and hence on liquidity.

One way to encourage arbitrage activity is to introduce portfolio margins, where the offsetting position between the home-market stock and the associated ADR are incorporated in the margin requirements. This is already approved by the SEC for example for index options.

[20] France introduced a transaction tax on 2012-08-01. See, http://www.eurexclearing.com/clearing-en/resources/faqs?frag=187982.

Chapter 4

Cross-sectional identification of informed trading*

4.1 Introduction

The notion of private information plays an important role in many theoretical models of market microstructure, asset pricing, and corporate finance. Such models show, for example, that firms whose securities are more subject to informed trading face greater illiquidity in these securities' secondary markets, a higher cost of capital, and reduced incentives to invest.[1] However, measuring private information and informed trading empirically remains a considerable challenge.

In this paper, we propose a new way of measuring informed trading based on a portfolio optimization model for individual investors. Our approach has two main advantages. First, it allows us to identify the amount of informed trading in an individual security over a given period based on the cross-section of price impact parameters (λ) and order imbalances (OIB, or the volume of buyer- minus seller-initiated trades). Hence, our measure can be estimated for each security on each day, or even at higher frequencies. Second, our model also delivers a very simple and intuitive expression for the aggregate private information shock for a given security over a given period. In other words, in addition to estimating the prevalence of trading based on private information, we can measure the direction and magnitude of private

* This chapter is based on Bongaerts, Rösch, and van Dijk (2014) "Cross-Sectional Identification of Informed Trading" available at http://ssrn.com/abstract=2532128. Van Dijk gratefully acknowledges financial support from the Vereniging Trustfonds Erasmus Universiteit Rotterdam and from the Netherlands Organisation for Scientific Research through a "Vidi" grant.

[1] See, among many others, Glosten and Milgrom (1985); Kyle (1985); Fishman and Hagerty (1989); Manove (1989); Easley et al. (2002); Dow and Rahi (2003); Easley and O'Hara (2004); Goldstein and Guembel (2008); Edmans (2009).

information for each security on each day.

In the model, investors arrive at the market with an optimal portfolio of securities, but are then hit by liquidity shocks and private information shocks that induce them to rebalance their portfolio. Investors' order flow generates price impact that is linear in trading volume, which implies that total transaction costs are quadratic in trading volume. Individual securities differ in their price impact parameter for exogenous reasons. When hit by a liquidity shock, investors optimally spread their trading over many securities, such that the marginal transaction costs for all securities are equal. As a result, the order flow in individual securities is proportional to the inverse of their price impact parameter, which implies that most trading is done in the most liquid securities. When hit by a private information shock about a certain security, investors trade an amount in that security that is inversely related to its price impact parameter. Furthermore, investors trade other securities in the opposite direction to finance the speculative trade, where again the amount of trading in each security is inversely related to its price impact.

The aggregate order flow (across all investors) in a security thus consists of three components: (i) liquidity-motivated order flow, (ii) speculative order flow based on private information about that security, and (iii) "funding" order flow to finance the speculative trading in other securities. When we introduce a benchmark security that is insulated from informed trading to resolve underidentification, we obtain a closed-form solution to back out the amount of informed trading in any security, or component (ii), from its aggregate order flow and the aggregate order flows and price impact parameters of other securities.

We refer to our identification of informed trading as "cross-sectional" since it exploits the idea that order flow that is purely liquidity-motivated has the same sign for all securities, while trading based on private information about a certain security results in opposite-sign order flow in other securities to finance the speculative trade. Crucially, the identification of informed trading also makes use of the notion that any order flow is affected by the expected price impact of trading.

Empirically, our model allows us to measure the dollar volume of informed trading in any security over any time period based on the cross-section of price impact parameters and order imbalances for a relevant set of peer securities as well as a benchmark security. We can also compute the probability of informed trading inferred from the cross-section (or $XPIN$) as the fraction of informed trading over total trading.

Next to a measure of the volume (and probability) of informed trading, our model also provides a very simple expression for a security's aggregate private information shock (aggregated across investors): the security's order imbalance multiplied by its price impact parameter ($\lambda \times OIB$), minus the order imbalance of the benchmark security multiplied by the benchmark's price impact parameter. The intuition is that the observed order imbalance in a security is more likely to be information-driven when the price impact of trading is high, since investors only trade securities that are expensive to trade when they have valuable private information about these securities. Furthermore, any trading in the benchmark security is either liquidity-motivated or funding-motivated, so the benchmark's order imbalance (accounting for its price impact) forms a natural reference point that can be used to isolate the aggregate private information shock of an individual security.

We estimate our measures of the amount and probability of informed trading and of the aggregate private information shock for all S&P 1500 stocks each day in the period 2001-2010 based on intraday price and transaction data from the NYSE Trade and Quote (TAQ) database. We estimate daily price impact parameters based on intraday data by implementing the approach of Glosten and Harris (1988). We use each stock's moving average price impact estimate over the past 20 days as the expected price impact on the current day. We estimate the daily order imbalances of individual stocks by signing individual trades using the Lee and Ready (1991) algorithm. Our final sample consists of all 2,130 stocks (listed at NYSE, Nasdaq, or Amex) that were an S&P 1500 constituent at some point during our sample period of 2001-2010 and that survive our basic data screens. As the benchmark security, we use the SPDR S&P500 ETF (ticker "SPY"), for which we obtain consolidated trades and quotes from the Thomson Reuters Tick History (TRTH) database. We argue that the SPDR is a reasonable benchmark security since it is highly traded, since the scope for market-wide private information is arguably limited (Baker and Stein, 2004), and since the SPDR is unlikely to be used for trading on private information of individual securities.[2]

The main purpose of our empirical analyses is to assess whether cross-sectional patterns in stock returns are consistent with our private information measure picking up meaningful cross-sectional variation in aggregate private information shocks. As our key predictions are cross-sectional in nature, most of our tests are based on a further simplified version of our private information measure: a stock's order imbalance multiplied by its price impact parameter ($\lambda \times OIB$). Since the correction for the benchmark's order imbalance times its price impact is the same for all stocks on a given day, this simplification does not affect our cross-sectional tests.

We first show, in Fama-MacBeth regressions, that the cross-section of daily stock returns is positively and highly significantly related to this simplified private information measure for individual stocks estimated on the same day. This finding is consistent with the idea that

[2] This idea is similar to the rationale behind program trading facilities. These also allow better liquidity because at least 15 securities need to be traded at the same time and hence the likelihood of trading on private information on any of these securities is low.

stocks with a more positive (negative) information shock on a given day have a more positive (negative) realized stock return, but it does not rule out other interpretations of our private information measure. In particular, our measure is a positive function of a stock's order imbalance and it is well-known that stocks with a more positive (negative) order imbalance on a given day tend to have a more positive (negative) return, for reasons that may be distinct from private information (e.g., price pressure). However, we show that the positive relation between the cross-section of stock returns and our private information measure survives controlling for order imbalance and expected price impact separately. In other words, $\lambda \times$ ***OIB*** has explanatory power for the cross-section of returns that goes beyond that of λ and OIB individually. We are not aware of models that provide an alternative interpretation of $\lambda \times$ ***OIB***. Furthermore, the explanatory power of $\lambda \times$ ***OIB*** is not subsumed by other "scaled" measures of order imbalance, such as the product of OIB and the quoted bid-ask spread or OIB scaled by market capitalization.

We then follow the reasoning that if our measure picks up private information, return reversals should be weaker following stock-day observations for which our measure assumes large negative or large positive values. After all, the price impact of informed trades should be permanent, while the price impact of uninformed order flow should be temporary (e.g., Kyle (1985); Admati and Pfleiderer (1988); Glosten and Harris (1988); Sadka (2006). To test this conjecture, we run daily Fama-MacBeth regressions of the cross-section of stock returns on one-day lagged returns, interacted with the absolute value of $\lambda \times$ ***OIB***. We reproduce the common result in the literature that the one-day autocorrelation in returns is negative (e.g., Roll (1984); Cox and Peterson (1994); Nagel (2012)). The interaction effect between one-day lagged returns and the absolute value of $\lambda \times$ ***OIB*** is significantly positive, indicating that returns revert significantly less following stock-days with large negative or large positive values of the private information measure.

To get a better idea of the economic magnitude of the reduced return reversal for high private information shocks, we also take a double-sorting approach to studying the relation between return reversals and the private information measure. We first sort stocks into quintile portfolios based on their private information measure $\lambda \times$ ***OIB*** on a given day, in such a way that portfolio 1 and 5 contain stocks with, respectively, large negative and large positive values for the measure. We then sorts stocks within each quintile into winner and loser stocks based on their returns on that day. We compute the daily returns on a reversal strategy within each quintile portfolio based on a long position in that day's loser stocks and a short position in that day's winner stocks, held from the market close on that day till the market close on the next day. The results of this double sort show that the abnormal returns (alphas) on the reversal strategy of quintile portfolio 3 (consisting of stocks with values of the private information measure close to zero) are significantly greater than the abnormal returns on the reversal strategy in the two extreme private information portfolios (quintiles 1 and 5). The economic magnitude of the difference in the strength of the return reversals is substantial, at 12 basis points per day. We interpret this as further evidence consistent with the view that our measure picks up meaningful cross-sectional variation in the direction and magnitude of private information for individual stocks.

In sum, this paper proposes new measures for the amount and probability of informed trading in individual stocks based on a portfolio optimization model whose key predictions concern the cross-section of order imbalances and price impact parameters. The model also yields a simple measure of the direction and magnitude of private information for individual stocks. We provide empirical support for this measure by showing that it is positively related to contemporaneous stock returns in the cross-section, and that return reversals are significantly weaker following stock-days with high values for this measure.

We contribute to the literature on measuring informed trading by suggesting an alternative to the popular "probability of information-based trading" (PIN) measure developed by Easley et al. (1996) and Easley et al. (2002), which is based on a market microstructure model instead of a portfolio optimization model and which has a different intuition. An advantage of our approach to measuring informed trading is that it is easy to implement and that it does not require a long time-series of transaction data for individual securities (and can thus be estimated even at high frequencies), since its main data requirements are of a cross-sectional rather than a time-series nature. Our work is also related to more recent papers on the "volume-synchronized probability of informed trading" or $VPIN$, see, among others, Easley et al. (2011, 2012). A common feature of $VPIN$ and our measure of information trading is that order imbalances play a key role, but our measure is distinct in that it also takes into account the price impact of trading.

Furthermore, to the best of our knowledge, our study is the first to propose a way to measure the magnitude and direction of the aggregate private information shock in an individual security contained in its trading over a given period. Our approach complements the work of, among others, Glosten and Harris (1988); Hasbrouck (1991); Sadka (2006), who measure the information effects of a trade through its permanent price impact, but who do not attempt to extract a direct proxy for the private information shocks on which informed trades are based.

The paucity of sophisticated proxies for informed trading and private information is illustrated by the paper of Lai et al. (2014), who benchmark PIN using crude, low-frequency firm-level proxies for information asymmetry such as the number of analysts following the firm, the analyst forecast dispersion, the age of the firm, and equity index membership. We hope that our new, high-frequency measures of informed trading and private information provide useful alternatives to existing measures and offer new opportunities to test and revise existing private information models of market microstructure, asset pricing, and corporate finance.

4.2 Basic model assumptions and notation

In this section, we introduce the basic setup for the theoretical portfolio optimization model from which we deduce the market implied information per security to be incorporated in prices.

Our model covers one period and concerns a market for N securities. These securities are typically risky, but a riskless security can be included. The returns on the securities are collected in the vector $\mathbf{r}$ and follow a multivariate lognormal distribution with means and covariance matrix $E(\mathbf{r})$ and $\mathbf{\Sigma}$, respectively. Let us for notational convenience define $\sigma^{\mathbf{2}}$ as the array that contains the diagonal elements of $\mathbf{\Sigma}$.

There are M investors in the market, which are indexed with i. Each investor i has power utility with CRRA parameter γ_i and starting wealth W_i. We assume that investors arrive to the market with an optimal starting portfolio. Moreover, we assume that investors cannot dislocate their portfolio so much that individual securities start to dominate portfolio such that idiosyncratic risk is beyond concern . Investors are exposed to liquidity shocks as well as potential private information shocks. Liquidity shocks Z_i arrive randomly and are expressed as a fraction of initial wealth W_i such that Z_i >0 corresponds to money inflow. If no shock arrives, $Z_i = 0$. Information shocks are described in more detail below. Given the liquidity and information shocks, each investor i has to determine optimal holdings $\mathbf{x_i}$ of all securities. His starting portfolio allocation is denoted by $\mathbf{x_i^*}$.

Trading demands of investors are accommodated by a financial intermediation sector (i.e., market makers) for a fee. In particular, order flow $o_{i,j}$ of investor i in security j has price impact on security j when it is traded. This leads to a lower expected return (without affecting risk), which increases linearly with trade size. More explicitly, we express total price impact $\psi_j(o_{i,j})$ as:

$$\psi_j(o_{i,j}) = \lambda_j \delta_{i,j} o_{i,j} \;\forall\; j, \tag{4.1}$$

where λ_j is the price impact parameter for security j expressed in percentage points lower expected return over the average investor horizon per dollar traded and $\delta_{i,j}$ is a trade sign indicator for the trade by investor i in security j. Total trading costs are then given by multiplying the average shortfall or excess in price with the size of the transaction:

$$|o_{i,j}\psi_j(o_{i,j})| = \delta_{i,j}o_{i,j}\lambda_j\delta_{i,j}o_{i,j} \;\forall\; j. \tag{4.2}$$

Hence, total transaction costs (execution shortfall) are quadratic in order flow sent by an investor. We define the matrix $\mathbf{\Delta_i}$ as a diagonal matrix with $\delta_{i,j}$ as its jth diagonal element. Similarly, we define $\mathbf{\Lambda}$ as the matrix that contains the λ_js on its diagonal. We assume that for all j, $\lambda_j > 0$, also for the riskless security (if any).

4.3 Individual investor portfolio optimization

4.3.1 Liquidity shocks only

We take a somewhat unconventional approach to portfolio optimization. We assume a CAPM-like setting in which investors may be heterogeneous (due to for example background risk) and have an optimal portfolio allocation x_i^*, given information at time 0. Moreover, we assume that all securities are correctly priced; thus, $(E(r - r_f) + \frac{1}{2}\sigma^2)/\beta = \iota(E(r_m = r_f) + \frac{1}{2}\sigma_m^2) = \iota\zeta$, where ζ is the market risk premium. Under these assumptions, we can let investors optimize risk-adjusted portfolio returns.[3] When we do this, we need to impose a budget constraint to avoid that the investor loads up on risk. Combined with transaction costs, the investor would like to keep his portfolio as it is. Our motivation to use a static model with somewhat incomplete preferences is that this will give very neat and tractable solutions under relatively mild assumptions.

An investor only receiving a liquidity shock Z_i optimizes:

$$\max_{\mathbf{x_i}} \mathbf{x_i}'\iota\zeta - \frac{1}{1+Z_i}(W_i(\mathbf{x_i}(1+Z_i) - \mathbf{x_i^*})'\mathbf{\Delta_i\Lambda\Delta_i}(\mathbf{x_i}(1+Z_i) - \mathbf{x_i^*})), \tag{4.3}$$

subject to the budget constraint

$$\iota'\mathbf{x_i} = 1. \tag{4.4}$$

We note that this way of formulating the rebalancing decision problem is intuitive and parsimonious. As $\mathbf{\Lambda}$ and $\mathbf{\Delta_i}$ are diagonal matrices, their order of multiplication in (4.3) can be

[3] This approach differs from the traditional mean-variance portfolio optimization problem in that the covariance matrix is not explicitly taken into account. As such, it looks a bit like a risk-neutral setting, except for the fact that we make risk-adjustments by standardizing by β. Our motivation for doing this is to keep the model tractable and to avoid instability due to estimation error of individual elements of $\mathbf{\Sigma}$. Otherwise, in solving for optimal portfolio weights, we need to invert an investor-specific weighted sum of $\mathbf{\Sigma}$ and $\mathbf{\Lambda}$, which is highly non-linear and complex. The downside of this approach is that investors could end up with concentrated portfolios since additional diversification is not rewarded (but complete diversification is assumed). However, systematic risk is taken into account since $E(\mathbf{r})$ is scaled by β.

changed. As a result, since $\mathbf{\Delta_i}\mathbf{\Delta_i} = \mathbf{I}$, the "endogenous" parameter matrix $\mathbf{\Delta_i}$ drops out from the price impact part and we obtain a solution without any endogenous parameters.[4] Another way of seeing this is that price impact is linear in signed order flow, such that total transaction costs are quadratic in signed order flow, so that taking absolute values is irrelevant.

The problem can be optimized by standard constrained optimization techniques involving a Lagrangian multiplier.[5] The optimal portfolio weights are given by the following Lemma:

Lemma 1. *The solution to optimization problem (4.3) is given by:*

$$\mathbf{x_i} = \mathbf{Q_i}^{-1}\iota\zeta + \mathbf{Q_i}^{-1}2W_i\mathbf{\Lambda}\mathbf{x_i^*} - \mathbf{Q_i}^{-1}\iota\zeta + \mathbf{Q_i}^{-1}\iota f\frac{Z_i}{1+Z_i} \tag{4.5}$$

$$= \frac{1}{1+Z_i}\mathbf{x_i^*} + \frac{Z_i}{1+Z_i}\mathbf{\Lambda}^{-1}\iota(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}. \tag{4.6}$$

Proof. See Appendix. □

The new portfolio holdings are therefore equal to the portfolio holdings in case the liquidity shocks could be settled with a risk and friction free savings account (first term) plus a transaction cost driven adjustment (second term). This second term consists of the relative size of the shock $\left(\frac{z_i}{1+Z_i}\right)$ times the fraction of the shock that is accommodated by every security. This fraction always lies between 0 and 1 and is proportional to the inverse of the price-impact of the security, such that most trading is done in the most liquid securities.

Individual order flow is now given by:

$$\mathbf{o_i} = W_i(1+Z_i)\mathbf{x_i} - W_i\mathbf{x_i^*} \tag{4.7}$$

$$= W_iZ_i\mathbf{\Lambda}^{-1}\iota(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}. \tag{4.8}$$

One can verify that this is indeed the optimal order flow. If we pre-multiply (4.8) by $\mathbf{\Lambda}$, we see that the solution yields order flows such that the marginal transaction costs for all securities are equal, as the RHS consists solely of scalars multiplied with a unity vector. Thus, it is impossible to sell a bit more of one security and a bit less of another and thereby be better off.

[4]Bongaerts, De Jong, and Driessen (2011) use a similar setting, but their model still features these endogenous parameters since they focus on bid-ask spreads rather than on price impact.

[5]Note that incorporating other constraints, such as short sale constraints, in this framework is convenient, but comes at the cost of increased complexity. The Lagrangian multiplier μ in the proof can be interpreted as a shadow price. In this case, it is the utility loss to the investor in optimal solutions compared to the setting in which shocks can also be accommodated with a transaction cost-free risk-free account.

4.3.2 Adding information shocks

We now introduce an information shock that will create a Jensen's alpha (standardized by β) of $\mathbf{v_i}$ on the securities. In other words, in addition to the liquidity shock, each investor i receives an information shock $\mathbf{v_i}$, which is essentially a vector of the alphas gross of transaction costs that can be generated for each security. The solution to the investor optimization problem is then given by the following Lemma.

Lemma 2. *With liquidity and private information shocks, optimal portfolio weights are given by*

$$\mathbf{x_i} = \frac{1}{1+Z_i}\mathbf{x_i^*} + \frac{Z_i}{1+Z_i}\mathbf{\Lambda}^{-1}\iota(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1} + \frac{1}{2W_i(1+Z_i)}\mathbf{\Lambda}^{-1}(\mathbf{I} - (\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}\iota\iota'\mathbf{\Lambda}^{-1})\mathbf{v_i}. \tag{4.9}$$

Proof. See Appendix. □

It is worthwhile analyzing the various components of this solution. The first two components are identical to the case with only liquidity shocks. The third term consists of three parts. The first part is $\mathbf{\Lambda}^{-1}\mathbf{v_i}$. This is the solution to $\Lambda y_i = v_i$, which is a first order optimality condition as it equates for each security marginal benefits (alpha return) of an extra share to its marginal costs (price impact). The second part is most conveniently written as $((\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}\iota\iota'\mathbf{\Lambda}^{-1})(\mathbf{v_i}\mathbf{\Lambda}^{-1})$. In this form, it can be seen as a vector of shocks ($\mathbf{\Lambda}^{-1}$, resulting from the analysis above) multiplied with a matrix that tells the investor how to allocate a shock. Not surprisingly, this allocation matrix looks very similar to what we have seen before, only this time multiplied with the unity vector to account for the fact that we have a vector of shocks rather than just one funding shock. The final part is the multiplication factor $\frac{1}{2W_i(1+Z_i)}$, which follows from the fact that for wealthy investors, less is to be gained in relative terms because transaction costs quickly outweigh informational advantages.

As before, we can obtain order flow by:

$$\mathbf{o_i} = W_i(1+Z_i)\mathbf{x_i} - W_i\mathbf{x_i^*} \tag{4.10}$$
$$= W_iZ_i\mathbf{\Lambda}^{-1}\iota(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1} + \mathbf{\Lambda}^{-1}(\mathbf{I} - (\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}\iota\iota'\mathbf{\Lambda}^{-1})\mathbf{v_i}. \tag{4.11}$$

The private information induced component of the order flow can be interpreted as follows. First, the matrix $\mathbf{\Lambda}^{-1}$ dictates that the amount of trading on private information for a given security is inversely related to the price impact of trading volume, which is intuitive. Second, the matrix $(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}\iota\iota'\mathbf{\Lambda}^{-1})$ results from the budget constraint and reflects the proportions in which an information shock in one security is funded by each of the others. The rows of this matrix add up to one. Third, the setting is constructed such that each individual investor trades on information shocks in such a way that the transaction costs on a marginal dollar of trading are exactly equal to (and therefore offset by) the alpha gain. Thus, informed trading volume is independent of wealth.[6]

4.3.3 Aggregating to market level and extracting consensus information

Aggregating order flow across all investors gives:

$$\begin{aligned} \mathbf{o_m} &= \sum_i \mathbf{o_i} \\ &= \mathbf{\Lambda}^{-1}(\mathbf{I} - (\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}\iota\iota'\mathbf{\Lambda}^{-1})M\bar{\mathbf{v}} + \sum_i W_i Z_i \mathbf{\Lambda}^{-1}\iota(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}, \end{aligned} \tag{4.12}$$

where $\bar{\mathbf{v}}$ is the average (equally-weighted) information shock. In (4.12), $\mathbf{\Lambda}^{-1}M\bar{\mathbf{v}}$ refers to the aggregate speculative trading volume, $-\mathbf{\Lambda}^{-1}((\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}\iota\iota'\mathbf{\Lambda}^{-1})M\bar{\mathbf{v}}$ refers to the aggregate funding demand for the speculative trades and $\sum_i W_i Z_i \mathbf{\Lambda}^{-1}\iota(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}$ refers to the aggregate liquidity demand. $M\bar{\mathbf{v}}$ can be thought of as the aggregate amount of private information (incidence rate times size) in the market.

In our attempts to obtain a measure of informed trading, we can try to invert (4.12) to end up with an analytical expression for $M\bar{\mathbf{v}}$. However, because we allow for an information shock for each security, the matrix $\mathbf{\Lambda}^{-1}(\mathbf{I} - (\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}\iota\iota'\mathbf{\Lambda}^{-1})$ is not full rank and hence cannot be inverted. The reason for this can be seen in a two security example. Observing positive order imbalance for security 1 and negative order imbalance for security 2, could imply either (i) a positive information shock for security 1, which is associated with selling of security 2 to fund the speculative trade in security 1, or (ii) a negative information shock for security 2, leading to buying in security 1 with the funds received from selling security 2. These two are empirically indistinguishable. To solve our under-identification problem, we assume that one of our securities never suffers from informed trading. This can be a treasury bond or an information-insensitive security. In our implementation in Section 5, we use the SPDR S&P500 ETF. We refer to this security as the “benchmark security. ”

When working out $M\bar{\mathbf{v}}$, we obtain a remarkably simple expression:

Proposition 1. *The order-flow implied aggregate private information shock for security* $j \in$

[6]This assumption might be unrealistic as some of the small investors would have to go short heavily in some of their securities to fund their uninformed trading. An extra set of restrictions on non-negative holdings may resolve this issue, but leads to less tractable results that are harder to interpret.

$\{2,..,N\}$ *is given by:*

$$M\bar{v}_j = \lambda_j o_j - \lambda_1 o_1, \tag{4.13}$$

where security 1 is the benchmark security.

Proof. See appendix. □

Our model thus not only allows us to decompose a security's aggregate order flow into informed trading on the one hand and liquidity-motivated and funding-induced trading on the other hand, but also yields a very simple and intuitive expression for a security's aggregate private information shock: its $\lambda \times OIB$, in excess of the same term for a benchmark security that is insulated from informed trading. In the remainder of the paper, we set out to estimate and validate these measures of informed trading and of the aggregate private information shocks for a large sample of U.S. stocks over a prolonged time period.

4.4 Data and variable definitions

For our empirical analysis of the model introduced in Sections 2 and 3, we use a sample of S&P 1500 stocks over 2001-2010. Our motivation for using S&P 1500 stocks is that most institutional investors focus on stocks with a relatively large market capitalization, so that this sample represents a reasonable set of stocks that informed traders might consider. The choice for S&P 1500 stocks also aims to strike a balance between ensuring a sample of sufficient breadth, while at the same time excluding small and thinly traded stocks for which the estimation of order imbalance and price impact parameters based on intraday data is problematic. Our sample starts on February 1, 2001 (to prevent issues stemming from the tick size change on January 29, 2001) and runs until the end of 2010. We refer to Appendix C.3 for a detailed description of the sample selection and composition.

All of our analyses are done at the daily frequency, where the key parameters (order imbalance and price impact) are estimated each day for each stock based on intraday data. We obtain intraday price and transaction data for individual stocks from the NYSE Trade and Quote (TAQ) database. To preclude survivorship bias, we obtain data for each stock over the entire period for which we have data over 2001-2010, and not only for the period during which they were an S&P 1500 constituent. We refer to Appendix C.4 for a detailed description of the data screens and filters we apply to the TAQ data, all of which are taken from prior studies dealing with these data.

We determine the sign of each trade using the Lee and Ready (1991) algorithm, as follows. If a trade is executed at a price above (below) the quote midpoint, we classify it as a buy (sell). If a trade occurs exactly at the quote mid-point, we sign it using the previous transaction price according to the tick test. That is, we classify the trade as a buy (sell) if the sign of the last price change is positive (negative). If the price is the same as the previous trade (a zero tick), then the trade is a zero-uptick if the previous price change was positive. If the previous price change was also equal to zero, we discard the trade. We do not use a delay between a trade and its associated quote because of the decline in reporting errors (see Madhavan et al. (2002); Chordia et al. (2005)). We are able to sign the overwhelming majority of trades in this way.

For each stock on each day, we compute its order imbalance (OIB) as the dollar volume of buyer- minus seller-initiated trades based on the signed trades over that day. We express order imbalance in millions of USD.

We estimate the daily price impact parameter for each stock using the approach of Glosten and Harris (1988), based on daily regressions of the price change of a trade relative to the previous trade on the current quantity traded and the change in the sign of the trade. The coefficient on the quantity traded represents the variable costs of trading and can be interpreted as the stock's price impact parameter, in the spirit of Kyle (1985) lambda. We scale the estimate of this coefficient by the squared closing price (quote midpoint) at the end of the same trading day to make sure that, in line with the model, price impact is measured as the percentage price change per unit of dollar trading volume.

We discard stock-days with fewer than 50 trades to ensure a minimum number of observations to estimate this price impact regression. Nonetheless, individual price impact estimates are noisy and could lead to extreme estimates in our measures of informed trading and private information. Furthermore, our model assumes that investors optimize the rebalancing of their portfolio following liquidity and private information shocks based on the expected price impact of trading different securities. In other words, estimating price impact parameters over the same day as we measure the order imbalances (that within the model arise as a result of the portfolio rebalancing by individual investors) would introduce look-ahead bias into our analyses. To mitigate these concerns, we construct measures of the expected price impact of trading a given stock on a given day (λ) as the moving average of the estimated daily price impact parameters for that stock over the past 20 days, where we set negative price impact estimates to zero. To further reduce the influence of outliers, we cross-sectionally winsorize the resulting expected price impact estimates each day at the 95% level.

Our returns-based empirical analyses are based on midquote returns computed from the daily midpoint of the last quote on each day, adjusted for corporate actions using CRSP data, and cross-sectionally winsorized each day at the 99.9% level ($Return$). For some of our tests, we use a spread-based liquidity measure computed as the difference between the quoted ask and the quoted bid price scaled by the midpoint of the quotes, averaging the spread across all trades for the stock on that day ($PQSPR$). We also compute the market capitalization ($Mktcap$) of each stock based on the number of shares outstanding and prices from CRSP at the beginning of each calendar year. After estimating these variables, we drop stocks with fewer than six months of data. In addition, when the data for a stock exhibit a gap of more than two months, we only retain the longest uninterrupted period.

Our final sample consists of all 2,130 stocks (listed at NYSE, Nasdaq, or Amex) that were an S&P 1500 constituent at some point during our sample period of 2001-2010 and that survive these data screens.

We use the SPDR S&P500 ETF (ticker "SPY") as a benchmark security that is insulated from informed trading, which is needed to tackle underidentification of the model. Our motivations for choosing the SPDR as the benchmark security are that it is highly traded and that it seems unlikely that informed traders exploit their private information by trading such a passive market-wide benchmark. We obtain consolidated trades and quotes for the SPDR from the Thomson Reuters Tick History (TRTH) database. We estimate the order imbalance and the price impact parameter of the benchmark security in the same way as we do for individual stocks.

4.5 Empirical results

The main purpose of our empirical analyses is to examine whether the measures of informed trading and private information stemming from the model developed in Sections 2 and 3 can be applied to real-life data and yield results that are consistent with our theoretical interpretation of these measures.

For each stock on each day, we estimate the (signed) dollar volume of informed trading using the decomposition of the stock's aggregated order flow on that day into informed trading, liquidity trading, and funding trading, as expressed in equation (4.12). This expression is worked out in more detail in equation (C.21) in Appendix C.1. Solving for an individual stock's informed trading volume is based on our estimates of the order imbalance (OIB) and price impact parameter (λ) of the stock of interest, of all other S&P 1500 constituents in our sample on that day, and of the SPDR (our benchmark security for which we assume informed trading volume to be equal to zero) on that day. For ease of interpretation, we scale the absolute informed trading volume by total trading volume for that stock on that day. The resulting measure, which we label $XPIN$, can be interpreted as the propensity or probability of informed trading.

We also estimate the aggregate private information shock (or $M\bar{v}$) for each stock on each day based on equation (4.13). This measure of private information is based on just the estimates of the order imbalance and price impact parameter of the stock of interest and of the SPDR.

Table 4.1 – Cross-sectional summary statistics of time-series averages
This table reports the cross-sectional (across the 2,130 S&P1500 stocks in the sample) mean, standard deviation, first quartile, median, and third quartile of the time-series average by stock of the daily return from corporate action adjusted end-of-day mid-quotes winsorized at the 0.1% level ($Return$), the daily average proportional quoted spread ($PQSPR$), the price impact defined as the percentage return in prices due to a trading volume of \$1m. (λ, each day cross-sectionally winsorized at the 95% level), the daily difference between the total dollar volume of trades initiated by buyers and sellers (order imbalance in \$m.) (OIB), the ratio of absolute, daily aggregate informed trading over daily trading volume ($XPIN$), and the daily aggregate private information ($M\bar{v}$) from Eq. (4.13). The first column indicates the number of stocks over which the summary statistics are computed. The second column indicates the number of days the average stock is in the sample. The sample includes all 2,130 stocks (listed at NYSE, Nasdaq, or Amex) that were an S&P 1500 constituent at some point during our sample period of 2001-2010. Data to compute all variables in the table are from TAQ. The factor to adjust daily closing mid-quote data for corporate actions is from CRSP.

	#Stocks	Days	mean	stddev	25%	median	75%
Return [%]	2,130	1,829	0.05	0.10	0.02	0.06	0.09
PQSPR [%]	2,130	1,829	0.37	0.79	0.12	0.20	0.39
λ	2,130	1,829	0.95	1.76	0.10	0.29	0.99
OIB	2,130	1,829	1.31	3.39	-0.01	0.18	1.21
XPIN	2,130	1,829	0.16	0.07	0.12	0.15	0.18
$M\bar{v}$ [%]	2,130	1,829	0.09	0.18	-0.01	0.04	0.15

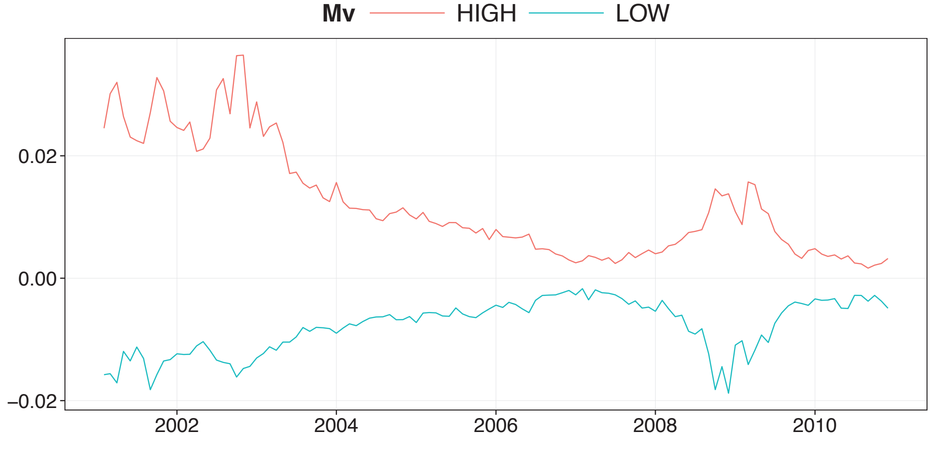
Table 4.1 presents summary statistics of the daily returns, OIB, λ, $PQSPR$, $XPIN$, and $M\bar{v}$ across all stocks in our sample over 2001-2010. The table reports cross-sectional summary statistics (mean, standard deviation, median, and 25th and 75th percentiles) of the stock-by-stock time-series averages of these variables. The table is based on all 2,130 S&P 1500 constituents in the sample, for which we have daily observations for 1,829 days on average.

The mean and median mid-quote returns are equal to, respectively, five and six basis points per day. The median OIB is slightly positive (\$0.18m.) over our sample, but, not surprisingly, exhibits substantial cross-sectional variation, with a standard deviation of \$3.39m. The median λ (scaled by 10^6) equals 0.29%, which means that the median of the average price impact across all stocks in the sample is 29 basis points for a trade of \$1m. The median $PQSPR$ is 20 basis points. The mean order imbalance and price impact estimate of the SPDR benchmark security are equal to, respectively, \$17.29m. and 0.09 basis points per \$1m trade (not tabulated), which indicates that the SPDR experienced substantial inflows over our sample period and that the average price impact of trading the SPDR is tiny, at less than one 1000th of the cross-sectional mean of the average price impact of the S&P 1500 stocks of 0.95%.

The mean and median $XPIN$ are equal to 0.15 and 0.16, respectively, which indicates that our approach identifies roughly 15% of the trading volume in individual stocks on a given day as informed. This number is comparable in magnitude to the mean and median PIN estimate of around 19% reported by Easley et al. (2002).

The mean and median $M\bar{v}$ are equal to 0.09 and 0.04, respectively, which suggests that the aggregate private information shock was slightly positive in our sample. The magnitude of $M\bar{v}$ is difficult to interpret, since it requires an assumption about the number of investors (M). However, the sign of $M\bar{v}$ does indicate whether the aggregate private information shock was positive or negative for a given stock on a given day. Furthermore, the magnitude of $M\bar{v}$ can be compared across stocks in the sense that a greater $M\bar{v}$ indicates a greater aggregate private information shock. The cross-sectional standard deviation of the average $M\bar{v}$ of individual stocks is substantial, at 0.18.

Figure 4.1 – Time-series of the average $M\bar{v}$ of the top 10% and the bottom 10% of all stocks sorted by $M\bar{v}$. This figure shows monthly time-variation in the equally-weighted, aggregate private information ($M\bar{v}$) of the 10% of all stocks with the highest and lowest private information on each given day. Aggregate private information is defined as in Eq. (4.13). Data to compute $M\bar{v}$ is from TAQ.



To get a sense of the time-series variation in private information in our sample, we plot the average $M\bar{v}$ of the top and bottom decile portfolios of stocks sorted on $M\bar{v}$ each day in Figure 4.1. Consistent with the summary statistics in Table 4.1, the aggregate private information shock tends to be somewhat larger in magnitude for stocks with positive private information shocks than for stocks with negative private information. The degree of private information is relatively high for both decile portfolios in the first few years over our sample period, then decreases slowly over time in 2003-2007 (both for positive and negative shocks), after which it shows a peak again in the period surround the start of the financial crisis in 2008-2009, to return to pre-crisis levels by 2010.

Figure 4.2 – Time-series of the average return of the top 10% and the bottom 10% of all stocks sorted by $M\bar{v}$.
This figure shows monthly time-variation of the end-of-day equally-weighted, mid-quote returns of the stocks in the top and bottom decile aggregate private information ($M\bar{v}$) portfolio. Aggregate private information is defined as in Eq. (4.13). Data to compute $M\bar{v}$ is from TAQ. The factor to adjust prices by corporate actions is from CRSP.

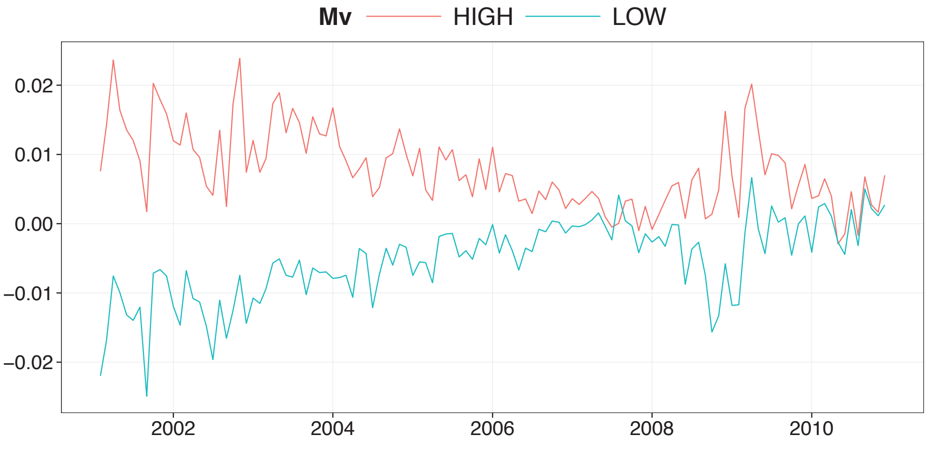


Figure 4.2 provides a first indication of the relation between $M\bar{v}$ and contemporaneous stock returns by plotting the time-series of the returns of the top and bottom decile portfolios of stocks sorted on $M\bar{v}$ each day (from Figure 4.1). The patterns in Figure 4.2 are a near mirror image of those in Figure 4.2, which suggests that the contemporaneous returns of stocks with positive (negative) private information tend to be positive (negative) and that the strength of this relation is relatively stable over time.

In our empirical tests, we focus on our measure of the aggregate private information shock ($M\bar{v}$) rather than on our measure of the probability of informed trading ($XPIN$), for two reasons. First, our private information shock measure is signed and thus contains more information. Second, the predictions about the relation with the cross-section of returns are more clear-cut for the private informed measure than for the informed trading measure. For example, we would expect $M\bar{v}$ to be linearly related to contemporaneous stock returns, but for $XPIN$ it is less clear what to expect, because $XPIN$ is unsigned but also because $XPIN$ depends on the amount of liquidity-motivated trading and not only on the underlying information signal.

Furthermore, since all of our empirical tests are cross-sectional in nature, we can use a further simplified version of our private information measure: the product of a stock's estimated order imbalance and price impact ($\lambda \times$ ***OIB***). Because the correction for the benchmark's product of order imbalance and price impact in equation (4.13) is the same for all stocks on a given day, this simplification does not affect the results.

Table 4.2 – Pooled correlations of daily private information, liquidity, order imbalance, and returns
This table reports pooled Pearson correlation coefficients between seven daily stock-specific variables: Aggregate private information ($M\bar{v}$), absolute private information ($|M\bar{v}|$), proportional quoted spread ($PQSPR$), price impact (λ), dollar order imbalance (OIB), the product of dollar order imbalance and price impact ($\lambda \times$ ***OIB***), and returns ($Return$). We refer to Table 2.1 for a description of these variables. Data to compute the variables are from TAQ and CRSP. P-values are in parentheses.

	$M\bar{v}$	$\|M\bar{v}\|$	$PQSPR$	λ	$\lambda \times$ ***OIB***	OIB	$Return$
$M\bar{v}$	1.000						
$\|M\bar{v}\|$	0.262 (0.00)	1.000					
$PQSPR$	-0.027 (0.00)	0.052 (0.00)	1.000				
λ	-0.002 (0.00)	0.003 (0.00)	0.187 (0.00)	1.000			
$\lambda \times$ ***OIB***	0.645 (0.00)	0.144 (0.00)	-0.020 (0.00)	0.013 (0.00)	1.000		
OIB	0.253 (0.00)	0.093 (0.00)	-0.032 (0.00)	-0.003 (0.00)	0.159 (0.00)	1.000	
$Return$	0.105 (0.00)	0.054 (0.00)	-0.005 (0.00)	0.003 (0.00)	0.100 (0.00)	0.050 (0.00)	1.000

Table 4.2 shows the pooled contemporaneous correlations between $M\bar{v}$, the absolute value of $M\bar{v}$, $PQSPR$, λ, the further simplified private information measure ($\lambda \times$ ***OIB***), OIB, and $Return$. As expected, a stock's quoted spread is positively correlated to the absolute magnitude of private information in that stock as well to the stock's price impact. The order imbalance is negatively correlated with both $PQSPR$ and λ. $M\bar{v}$ is highly correlated with its simplified version $\lambda \times$ ***OIB*** (at 0.645), but not perfectly, which stems from time-series variation in the product of order imbalance and price impact of the benchmark security that

will not influence our cross-sectional tests. We note that the correlations of both λ and OIB with $\lambda \times$ ***OIB*** are relatively small (at 0.013 and 0.159, respectively), which suggests that our simplified private information measure is distinct from its individual components and that any results we find for $\lambda \times$ ***OIB*** are unlikely to stem solely from λ or OIB. The correlations with returns provide some further initial evidence that our measures pick up meaningful variation in private information, since both $M\bar{v}$ and $\lambda \times$ ***OIB*** are positively and significantly related to contemporaneous stock returns. At around 0.10, these correlations are not overwhelming, but daily returns for individual stocks are noisy and we note that both correlations are more than double the magnitude of the correlation between OIB by itself and contemporaneous returns.

Table 4.3 – Daily Fama-MacBeth regressions of returns on contemporaneous private information
This table reports the time-series averages of the estimated slope coefficients from daily cross-sectional regressions to explain differences in mid-quote returns across stocks. The dependent variable is the end-of-day mid-quote price return of stock i on day d ($Return_{i,d}$). The independent variables are: the return of stock i on day $d-1$ ($Return_{i,d-1}$), the order imbalance of stock i on day d ($OIB_{i,d}$), the price impact parameter of stock i on day d calculated as the stock's average price impact estimate over the past 20 days with setting non-positive price impact estimates to zero ($\lambda_{i,d}$), the inverse of the market capitalization of stock i at the beginning of each year ($1/Mktcap_{i,y-}$), the proportional quoted spread for stock i on day $d-1$ ($PQSPR_{i,d-1}$), and various interaction terms. Fama-MacBeth t-statistics are in parentheses using Newey-West corrections. Data to compute the variables are from TAQ. Market capitalization data as well as the factor to adjust prices by corporate actions are from CRSP. Some coefficients have been scaled for ease of presentation.

	Dependent variable: $Return_{i,d}$				
$Return_{i,d-1}$	-0.07	-0.07	-0.07	-0.07	-0.07
	(-14.71)	(-14.95)	(-17.10)	(-18.35)	(-17.36)
$OIB_{i,d} \times 10^4$	0.96	0.97	0.14	0.01	-0.28
	(9.34)	(9.30)	(4.03)	(0.40)	(-4.70)
$\lambda_{i,d} \times 10^2$		0.99	0.98	0.20	1.25
		(5.44)	(6.40)	(1.34)	(6.72)
$\lambda_{i,d} \times OIB_{i,d}$			0.39	0.30	0.37
			(21.73)	(23.18)	(18.79)
$1/Mktcap_{i,y-}$				164.83	
				(8.34)	
$OIB_{i,d} \times 1/Mktcap_{i,y-}$				549.36	
				(12.59)	
$PQSPR_{i,d-1}$					-0.01
					(-0.84)
$OIB_{i,d} \times PQSPR_{i,d-1}$					0.07
					(7.76)
R^2	2.43	2.86	4.95	6.13	5.70
# regressions	2,441	2,441	2,441	2,192	2,441

In Table 4.3, we substantiate the initial evidence on the positive association between our private information measure $\lambda \times$ ***OIB*** and contemporaneous returns by running daily Fama and MacBeth (1973) regressions of the midquote returns on individual stocks on one-day lagged returns, OIB, λ, and $\lambda \times$ ***OIB***. OIB is included contemporaneously, since our approach aims to extract informed trading from the realized order imbalance on a given day. We note, however, that λ is not the contemporaneous price impact parameter for a stock on that day, but rather the expected price impact based on the moving average price impact estimates over the past 20 days (excluding the current day), since the model assumes that order flow on a given day is affected by the expected price impact of trading.

Consistent with prior studies, we find that daily stock returns exhibit a significantly negative autocorrelation. The coefficient on lagged returns is equal to -0.07 in the first model in Table 4.3, with a Fama-MacBeth t-stat of 14.7 (based on the Newey and West (1994), correction for autocorrelation in the estimated coefficients). Not surprisingly, daily stock returns are significantly higher on days with more positive OIB. However, the interpretation of this finding is ambiguous, as both liquidity-motivated and informed trading are associated with price impact. The coefficient on λ is also positive and significant in most regression models in Table 4.3. This positive effect of λ on contemporaneous returns was not clear ex ante, but may be driven by the fact that the order imbalance is positive on average in our sample.

More importantly, we find a positive and highly significant effect of our simplified private information measure $\lambda \times$ ***OIB*** on contemporaneous returns. This result suggests that returns are higher (lower) for stocks with a more positive (negative) value of $\lambda \times$ ***OIB*** on that day, which is what we would expect if $\lambda \times$ ***OIB*** measures private information. The economic magnitude of this effect is considerable. A one standard deviation increase in $\lambda \times$ ***OIB*** is associated with a 0.12 standard deviation increase in contemporaneous stock returns, which is substantial in light of the noise inherent in daily stock returns. We note that the effect of our private information measure $\lambda \times$ ***OIB*** is not driven by λ or OIB itself, and that its t-stat is considerably higher than the individual t-stats of the coefficients on λ or OIB. In other words, our new private information measure is more than the sum of its well-known parts.

In the final two regression models of Table 4.3, we examine whether the effect of $\lambda \times$ ***OIB*** disappears when we introduce other "scaled" versions of order imbalance that may be correlated with $\lambda \times$ ***OIB***. In the fourth model in Table 4.3, we include the product of OIB and the inverse of a stock's market capitalization. In the fifth model, we include the product of OIB and $PQSPR$. Although the coefficients of both $\lambda \times 1/Mktcap$ and $\lambda \times PQSPR$ are positive and significant, the effect of $\lambda \times$ ***OIB*** remains intact.

We next turn to potentially more stringent tests of our conjecture that $\lambda \times$ ***OIB*** measures private information. For this conjecture to be validated, we should observe significantly weaker return reversals following stock-days with large positive or negative values of $\lambda \times$ *OIB*, since informed trading should be associated with permanent rather than transitory price impact. We test this hypothesis in two ways.

Table 4.4 – Daily Fama-MacBeth regressions of returns on previous day private information
This table reports the time-series averages of the estimated slope coefficients from daily predictive, cross-sectional regressions to explain differences in mid-quote returns across stocks. The dependent variable is the end-of-day mid-quote price return of stock i on day d ($Return_{i,d}$). The independent variables are: the return of stock i on day $d-1$ ($Return_{i,d-1}$), the absolute order imbalance of stock i on day $d-1$ ($|OIB_{i,d-1}|$), the price impact parameter of stock i on day $d-1$ calculated as the stock's average price impact estimate over the past 20 days with setting non-positive price impact estimates to zero ($\lambda_{i,d-1}$), and various interaction terms. Fama-MacBeth t-statistics are in parentheses using Newey-West corrections. Data to compute the variables are from TAQ. The factor to adjust prices by corporate actions is from CRSP. Some coefficients have been scaled for ease of presentation.

Dependent variable: $Return_{i,d}$			
$Return_{i,d-1}$	-0.09	-0.10	-0.10
	(-11.01)	(-12.34)	(-11.56)
$\lambda_{i,d-1} \times \lvert OIB_{i,d-1}\rvert$		0.04	0.04
		(6.32)	(6.46)
$Return_{i,d-1} \times \lambda_{i,d-1} \times \lvert OIB_{i,d-1}\rvert$		2.17	2.40
		(10.35)	(11.22)
$\lvert OIB_{i,d-1}\rvert \times 10^4$			-0.01
			(-5.28)
$Return_{i,d-1} \times \lvert OIB_{i,d-1}\rvert$			0.00
			(9.91)
$\lambda_{i,d-1} \times 10^2$			0.21
			(1.15)
$Return_{i,d-1} \times \lambda_{i,d-1}$			-0.37
			(-5.51)
R^2	2.60	3.45	5.05
# regressions	2,441	2,440	2,440

Table 4.4 reports the results of daily Fama-MacBeth regressions of the midquote returns on individual stocks on one-day lagged returns, as well as one-day lagged returns interacted with one-day lagged $\lambda \times |OIB|$. If returns revert significantly less following information shocks, and if our measure is a meaningful proxy for these shocks, the coefficient on the interaction term should have the opposite sign as the coefficient on lagged returns. We note that we take the absolute value of our private information measure $\lambda \times$ *OIB* for these tests, since return reversals should be weaker following large positive or negative information shocks. However,

because λ is non-negative by construction, we only need to take the absolute value of OIB.

Consistent with Table 4.3, the first-order autoregressive coefficient is significantly negative, at -0.09 in the first model of Table 4.4. In the second model, we add lagged $\lambda \times |OIB|$ as well as lagged $\lambda \times |OIB|$ interacted with lagged returns. The coefficient on lagged $\lambda \times |OIB|$ is positive and significant, suggesting that returns tend to be higher for stocks with a more extreme private information shock on the previous day.[7]

The coefficient on the interaction term of lagged returns and lagged $\lambda \times |OIB|$ is significantly positive at 2.17, with a Fama-MacBeth Newey-West t-stat of 10.35. This finding indicates that, indeed, stock returns tend to revert significantly less following stock-days with high absolute values of our private information measure. We interpret this evidence as consistent with the view that $\lambda \times$ ***OIB*** does proxy for aggregate private information shocks. The third model of Table 4.4 shows that this result survives breaking up $\lambda \times |OIB|$ into its two separate variables and including all the relevant interactions.

To assess the economic significance of the reduced strength of return reversals following stock-days with high absolute values of $\lambda \times$ ***OIB***, we also analyze the returns on reversal strategies separately for stock-day observations with low and high private information. To that end, we first sort stocks into quintile portfolios on day $d - 1$ based on their $\lambda \times$ ***OIB***. Quintile portfolios 1 and 5 thus contain stocks with, respectively, large negative and large positive private information estimates on that day. Subsequently, we sort stocks within each private information quintile into five subportfolios based on their returns on day $d - 1$. We then compute the returns on a simple reversal strategy within each private information quintile that is long in day $d - 1$'s loser stocks (subportfolio 1) and short in day $d - 1$'s winner stocks (subportfolio 5) in that quintile. The returns of the reversal strategy are based on these stocks' next day's returns computed from the market close on day $d - 1$ till the market close on day d. The difference between the abnormal returns on the reversal strategies within the low and high private information quintiles can be interpreted as a measure for how large the reduction in the strength of return reversals is following high $\lambda \times$ ***OIB*** stock-days.

The results of this second, 5×5 double-sorts approach to analyzing the strength of return reversals following low and high private information stock-days are in Panel A of Table 4.5. The first four columns of the panel report the estimates of time-series regressions of the daily returns on the reversal strategy for private information quintile 3 (which contains stocks whose aggregate private information estimate is close to zero) on various commonly used asset pric-

[7] This effect may be driven by our finding in Figure 4.1 that over sample period positive information shocks tend to be somewhat greater than negative shocks. However, we note that the lagged effect of $\lambda \times |OIB|$ is much less significant in both statistical and economic terms compared to the contemporaneous effect of $\lambda \times$ ***OIB*** reported in Table 4.3, which is what we would expect.

Table 4.5 – The returns on reversal strategies conditional on private information
This table reports the results of time-series regressions of factor models to explain profits from two different investment strategies, based on a double-sorting approach. In Panel A, we sort all stocks in our sample into five portfolios based on $\lambda \times$ *OIB* on day $d-1$. We then sort all stocks in the median $\lambda \times$ *OIB* portfolio into five subportfolios based on their return on day $d-1$. The dependent variable in Panel A is the equally-weighted return on day d of going long the "losers" (i.e., the bottom quintile portfolio sorted by past returns) and short the "winners" (i.e., the top quintile portfolio) within the median $\lambda \times$ *OIB* portfolio. In Panel B, we sort all stocks in our sample into five portfolios based on their return on day $d-1$. We then sort all stocks in the "winner" and "loser" portfolio into five subportfolios based on $\lambda \times$ *OIB* on day $d-1$. The dependent variable in Panel B is the equally-weighted return on day d of going long the high $\lambda \times$ *OIB* stocks in the "loser" portfolio and short the low $\lambda \times$ *OIB* stocks in the "winner" portfolio. Independent variables in the regressions are: the daily market excess return ($Mkt-RF$), the daily return difference between small and large stocks (SMB), the daily return difference between high and low book-to-market stocks (HML), the daily return difference between past medium-term winner and loser stocks ($Momentum$), the daily return difference between past short-term loser and winner stocks ($Reversal$). The last columns in both Panel A and Panel B report the results of investing in the above strategies and subtracting the profits following a reversal strategy in the "opposite" $\lambda \times$ *OIB* portfolio, called a "control" strategy. In Panel A, the "control" strategy is going long the "losers" and short the "winners" in the two extreme $\lambda \times$ *OIB* portfolios. In Panel B, the "control" strategy is going long the low $\lambda \times$ *OIB* stocks in the "loser" portfolio and short the high $\lambda \times$ *OIB* stocks in the "winner" portfolio. Newey-West t-statistics are in parentheses. Data to compute the variables are from TAQ. The factor to adjust prices by corporate actions is from CRSP. Daily factor portfolio returns are from the website of Ken French.

Panel A: Return reversal in median information portfolio

					REV - Control
Intercept	0.46	0.47	0.47	0.46	0.12
	(5.56)	(6.87)	(7.61)	(7.67)	(4.81)
Mkt - RF	0.08	0.08	0.11	0.08	-0.07
	(3.98)	(4.09)	(6.10)	(4.16)	(-2.29)
SMB		-0.12	-0.14	-0.11	-0.08
		(-2.42)	(-2.77)	(-2.49)	(-1.85)
HML		-0.01	0.00	0.03	0.09
		(-0.34)	(0.05)	(0.60)	(2.11)
Momentum			0.08	0.08	-0.02
			(3.12)	(2.76)	(-0.57)
Reversal				0.13	-0.06
				(3.98)	(-1.51)
# Obs.	2,453	2,453	2,453	2,453	2,453
R^2	1.02	1.50	2.03	3.32	2.12

Table 4.5 – continued

Panel B: Return reversal in extreme return portfolios					REV - Control
Intercept	0.36	0.37	0.36	0.35	0.11
	(10.52)	(11.93)	(9.99)	(11.34)	(4.28)
Mkt - RF	0.15	0.16	0.20	0.15	0.02
	(4.85)	(4.99)	(6.25)	(5.66)	(0.76)
SMB		-0.21	-0.22	-0.19	-0.23
		(-4.24)	(-4.49)	(-4.06)	(-4.65)
HML		-0.06	-0.04	-0.00	0.05
		(-0.98)	(-0.71)	(-0.08)	(1.01)
Momentum			0.10	0.10	0.01
			(2.85)	(2.56)	(0.37)
Reversal				0.17	-0.05
				(3.33)	(-1.49)
# Obs.	2,453	2,453	2,453	2,453	2,453
R^2	3.23	4.42	5.09	6.96	2.00

ing factors. The columns correspond to, respectively, the CAPM, the Fama and French (1993) three-factor model, the Carhart (1997) four-factor model, and the Carhart model supplemented with a fifth factor based on short-term reversals (Jegadeesh, 1990). We obtain daily returns on these factors from the website of Ken French. All four models indicate economically large and statistically highly significant abnormal returns (alphas) of 46-47 basis points per day, which indicate strong daily return reversals for stocks with low private information estimates.[8]

The final column of Panel A shows the five-factor alpha of the difference between the reversal strategy for low private information stocks (private information quintile 3, as in columns 1-4) and the reversal strategy for high private information stocks (private information quintiles 1 and 5 combined). This alpha is significantly positive at 0.12 (Newey-West t-stat 4.81), which implies that the strength of return reversals is 12 basis points per day less following stock-days with high private information when compared to stock days following low private information, an effect that is significant from an economic perspective.[9]

[8] These abnormal return estimates on reversal strategies are somewhat higher than the mean reversal returns reported by Nagel (2012) of 18 basis points per day based on midquote returns and 30 basis points per day based on trade returns. This difference in magnitudes can likely be explained by differences in the sample, by the fact that Nagel's reversal strategy returns are based on all stocks rather than only the extreme winner and loser stocks, and by the fact that the first four models in Panel A of Table 4.5 use only stocks for which we estimate the amount of private information to be low. We note that neither one of these reversal strategy return estimates is realistic in the sense that they do not take into account transaction costs and short-sales constraints. We also note that we obtain qualitatively similar results when using trade returns instead of midquote returns.

[9] There are still significant return reversals following stock-days with high private information, but we note

In Panel B of Table 4.5, we reverse the 5×5 double sorting procedure by first sorting stocks into quintile portfolios based on their return on day $d-1$ and then sorting winner and loser stocks into five subportfolios based on their private information on day $d-1$. We then create an alternative reversal strategy that is long loser stocks with very positive values of $\lambda \times$ *OIB* and short winner stocks with very negative values of $\lambda \times$ *OIB*. The idea is that stock-days with very positive private information but very negative returns or with very negative private but very positive returns are likely characterized by a large amount of liquidity-motivated trading in the opposite direction of the private information signal, and should thus exhibit strong reversals on the next day. The first four columns of Panel B show that the one-, three-, four-, and five-factor alphas of this strategy are economically and statistically large, at 35-37 basis points per day, with t-stats close to 10. The final column of Panel B compares the return on this strategy to the return on a reversal strategy that is long loser stocks with very negative values of $\lambda \times$ *OIB* and short winner stocks with very positive values of $\lambda \times$ *OIB*, since the reversals should be weaker on these categories of stocks if $\lambda \times$ *OIB* is a meaningful proxy for private information. The significant difference in the abnormal returns on these two strategies of 12 basis points per day indicates that return reversals are considerably weaker when the returns on loser and winner stocks are more likely to be driven by private information.

Overall, our tests show that, consistent with our private information measure picking up meaningful cross-sectional variation in aggregate information shocks, stocks with a more positive value of $\lambda \times$ *OIB* tend to have significantly more positive contemporaneous returns, and stocks with very negative or very positive private information estimates subsequently exhibit significantly weaker return reversals. Both of these results support the theoretical interpretation of our new private information measure.

4.6 Conclusion

This paper proposes new measures of both the amount of informed trading in individual securities and the direction and magnitude of the aggregate private information shock for these securities. Both measures are derived from a portfolio optimization model for individual investors who are exposed to information and liquidity shocks. Our identification of informed trading is cross-sectional in the sense that it is based on the cross-section of price impact parameters and order imbalances for a given day (or intraday period).

We validate our private information measure by estimating it for all S&P 1500 stocks each

that our model does not rule out non-trivial liquidity-motivated trading on those stock-days, which could explain those return reversals.

day over 2001-2010. In particular, we show that it is strongly related to contemporaneous returns, and that return reversals are significantly weaker following stock-days with high private information estimates. Both pieces of evidence are consistent with the conjecture that our private information measure is indeed associated with the aggregate private information shock of individual securities.

An appealing feature of our private information measure is that it is intuitive and easy to estimate, even at high frequencies. In cross-sectional applications, it simplifies to a security's order imbalance multiplied by its price impact parameter ($\lambda \times OIB$). Furthermore, in contrast to other measures that proxy for private information, our measure also conveys the direction of the private information signal. We hope that our measure will be useful in a host of applications in market microstructure, asset pricing, and corporate finance. In future work, we plan to investigate the asset pricing applications of our private information measure.

Appendices

Appendix A

An Empirical Analysis of Co-Movement in Market Efficiency

A.1 Data Filters

This appendix describes the data filters applied to the high-frequency data. Each day in our sample period from March 1996 to December 2010, we include all NYSE-listed common stocks (i.e., CRSP PERMNOs with sharecode 10 or 11) with a previous day closing price above $5 in our sample. We collect data on all trades and quotes for these stocks from the Thomson Reuters Tick History (TRTH) database (using consolidated data across all U.S. exchanges). We discard trades that fall outside the continuous trading session (9:30 am till 4:00 pm U.S. EST/EDT) on the NYSE (in total 67,561,089 trades). We also discard trades with a negative price (51,905 trades) or a price that is more than 10% different from the trade price of the ten surrounding trades (130,930 trades). We further drop trades of more than 100,000 shares (1,750,630 trades) since large trades are often negotiated before they get reported (Glosten and Harris, 1988). We discard quotes outside the continuous trading session (252,266,477 quotes), quotes with a non-positive bid or ask price (651,568 quotes), quotes of which the bid price exceeds the ask price (128,701,417 quotes). We also discard a number of quotes we regard as outliers, defined as those for which (i) the bid (ask) price is more than 10% different from the average bid (ask) price of the ten surrounding quotes, (ii) the ask price is more than $5 higher than the bid price, or (iii) the proportional quoted spread is greater than 25%. A total of 6,550,437 quotes are discarded because of these criteria. We note that while the absolute numbers of trades and quotes excluded because of these data screens are large, they are small relative to the total number of trades and quotes in the sample. Our data screens lead us to discard less than 0.05% of all trades and less than 1% of all quotes. Our final sample consists of 2,157 stocks and 14,253,093,209 trades, of which 99.6% could be signed using the Lee and Ready (1991) algorithm. Because of a decrease in reporting errors since 1998 (Madhavan, Richardson, and Roomans, 2002), we do not use a delay between a trade and its associated quote.

Appendix B

The impact of arbitrage on market liquidity

B.1 Sample construction

This appendix describes details of the sample construction. I first retrieve all dead and alive American and global Depositary Receipts (DRs) from Datastream which returns (in Dec-13) 7700 different DRs of which around 10% (732) are traded at the New York Stock Exchange (NYSE), the focus in this study.

The home market share, associated to any of the ADRs, can be identified using data from adrbnymellon.com or adr.db.com. Both websites offer a list of DRs and an ISIN code for the home market share.

As the analysis requires intraday data for which I use the Thomson Reuters Tick History (TRTH) database, I filter out any DR for which I could not establish the RIC (the primary identifier in TRTH) for either the DR or the home-market stock. Upon request Datastream provides a RIC field, however this field is empty for around 50% of all DRs. In the case of a missing RIC field for the ADR or for the home market shares I use the TRTH API to search for a RIC code by ISIN.

For every ISIN the RIC from the major exchange of the home market country is chosen. This way 199 out of the 732 stocks remain. A possible reason for this significant drop in identified home-market/ADR pairs is that either the ADR got delisted from the NYSE, or that the home-market share got delisted from the home-market exchange before 1996, the beginning of the TRTH database.

A similar setup (i.e. using intraday data from TRTH for ADRs, albeit for an event study) is considered by Berkman and Nguyen (2010), who are able to identify 277 ADR-home market pairs, but of which only 44 trade at NYSE. Further Gagnon and Karolyi (2010b) identifies 506 ADR-home market pairs using Datastream, but the ADR can be listed on either NYSE, Amex, or Nasdaq. The above matching results in 199 pairs where the ADR is traded at the NYSE.

I now proceed to use the top five home market exchanges, in terms of having the most identified cross-listed ADRs trading in NYSE and having an overlapping trading time with the NYSE (to avoid non-synchronous prices). These exchanges are the London Stock Exchange (the U.K., with 29 stocks), Sao Paolo Stock Exchange (Brazil, 20 stocks), Bolsa Mexicana de Valores (Mexico, 14 stocks), XETRA (Germany, 9 stocks), and Euronext Paris (France, 9 stocks). Of these 81 stocks I filter out 6, because I could not find intraday data for either the home market or the cross-listed ADR for at least one year. Further I exclude three stocks from Brazil and Mexico from my sample because I could not align prices of the home market with prices of the ADR, as described in more detail on page 115.

B.2 Data filters

This appendix describes the quote and trade data filters. I discard non-positive bid and ask quotes (in total 5030 quotes), quotes where the ask is lower or equal to the bid quote (2,486,756 quotes), and quotes outside the continuous trading session (68,866,555 quotes). Further, I remove outliers (25,764 quotes). An outlier is defined as a bid (ask) quote that differs by more than 10% of the average of the ten surrounding bid (ask) quotes. Despite that I discard many quotes from the sample, as a fraction of the total 8.6 billion quotes these numbers are marginal. In a similar way trade prices are filtered. Of the in total one billion trades, I discard trades that fall outside the continuous trading session (as depicted in Figure 3.1) on the NYSE (in total 752,434 trades) and on the respective home-market (17,008,820). I also discard trades with a non-positive price (65 trades) or a price that is more than 10% different from the trade price of the ten surrounding trades in the NYSE (93 trades) and in the home-market (2,193). Further, I discard trades of more than 100,000 shares in the NYSE (20,024 trades) and in the home-market (3,057,514), because large trades are often negotiated before they get reported.

To make prices comparable between the home market stock and the ADR, bid and ask quotes of the ADR are converted according to the bundling ratio which I got from either adrbnymellon.com or adr.db.com. Unfortunately, bundling ratios can be time-varying and both websites only report the latest bundling ratio. To adjust the bundling ratio over time, I get all corporate actions from TRTH for both the ADR's and the home-market share. Changes in the bundling ratio can occur, for example because of solo stock-splits. To verify the accuracy of the bundling ratio I plot the daily average, currency adjusted mid-quote ratios for each stock in the sample (unreported). If the ratio does not vary around one and does not resemble a step function the stock is dropped from the sample. As such three stocks from Brazil and Mexico dropped from the sample because prices of the home market could not be aligned to prices of the ADR.

For 20 ADRs the bundling ratio changed over the sample, with a maximum of three changes for one ADR with RIC ICA.N referring to a stock in Mexico (Empresas ICA) and for 16 ADRs the bundling ratio changed once over the sample.

To further ensure that stocks are mapped properly and prices are adjusted correctly I drop any stock-day if *Opportunity-Profit* is higher than USD 10 or higher than 30% (of the mid-quote of the home-market) in any second within the day. Such high discrepancies in the law of one price frequently occur just after corporate actions on the home market stock, for example when the home market stock is *ex dividend*, but the ADR is *cum dividend*. Ignoring days after corporate actions such high price deviations occur relatively seldom, with less than five stock-days in the U.K., Germany, and Mexico, but 47 and 68 stock-days for Brazil and France, respectively. For 54 out of the 72 stocks the highest profit on every day is below USD 10 and below 30%.

Appendix C

Cross-sectional identification of informed trading

C.1 Proofs

Proof of Lemma 1

We solve the problem by a standard Lagrangian multiplier technique . We define

$$L(\mathbf{x_i}, \mu) = \mathbf{x_i}'\iota\zeta - \frac{1}{1+Z_i}(W_i(\mathbf{x_i}(1+Z_i) - \mathbf{x_i^*})'\mathbf{\Lambda}(\mathbf{x_i}(1+Z_i) - \mathbf{x_i^*})) - \mu(\iota'\mathbf{x_i} - 1). \tag{C.1}$$

The necessary FOCs for optimality are given by

$$\frac{\partial L(\mathbf{x_i}, \mu)}{\partial \mathbf{x_i}} = 0, \qquad \frac{\partial L(\mathbf{x_i}, \mu)}{\partial \mu} = 0. \tag{C.2}$$

As $L(\mathbf{x_i}, \mu)$ contains only polynomial terms of at most second order, we can write the FOCs as a system of linear equations and solve it as is shown below. In matrix form, the FOCs are given by

$$\begin{bmatrix} -\mathbf{a_i} \\ 1 \end{bmatrix} = \begin{bmatrix} -\mathbf{Q_i} & \iota \\ \iota' & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x_i} \\ \mu \end{bmatrix}, \tag{C.3}$$

where

$$\mathbf{Q_i} = 2W_i(1+Z_i)\mathbf{\Lambda} \tag{C.4}$$

$$\mathbf{a_i} = \iota\zeta + 2W_i\mathbf{\Lambda}\mathbf{x_i^*}. \tag{C.5}$$

Using the partitioned inverse (see Greene (2000), p. 34), we obtain our solution:

$$\mu = -(\iota'\mathbf{Q_i}^{-1}\iota)^{-1}\iota'\mathbf{Q_i}^{-1})\mathbf{a_i} + (\iota'\mathbf{Q_i}^{-1}\iota)^{-1} \tag{C.6}$$

$$\mathbf{x_i} = \mathbf{Q_i}^{-1}(\mathbf{I} - \iota(\iota'\mathbf{Q_i}^{-1}\iota)^{-1}\iota'\mathbf{Q_i}^{-1})\mathbf{a_i} + \mathbf{Q_i}^{-1}\iota(\iota\mathbf{Q_i}^{-1}\iota')^{-1}. \tag{C.7}$$

$$= \mathbf{Q_i}^{-1}\mathbf{a} + \mathbf{Q_i}^{-1}\iota\mu. \tag{C.8}$$

If we define $f = (\iota'\mathbf{Q_i}^{-1}\iota)^{-1}$, we can work out μ:

$$\mu = -f\iota'\mathbf{Q_i}^{-1}(\iota\zeta + 2W_i\mathbf{\Lambda}\mathbf{x_i^*}) + f \tag{C.9}$$

$$= -f(\iota'\mathbf{Q_i}^{-1}\iota)\zeta - f\iota'\mathbf{Q_i}^{-1}2W_i\mathbf{\Lambda}\mathbf{x_i^*} + f. \tag{C.10}$$

Substituting back f gives

$$\mu = -\zeta - f\iota'\mathbf{Q_i}^{-1}2W_i\mathbf{\Lambda}\mathbf{x_i^*} + f. \tag{C.11}$$

Substituting $\mathbf{Q_i}$ back gives

$$\mu = -\zeta - f\iota'\frac{1}{1+Z_i}\mathbf{x_i^*} + f. \tag{C.12}$$

Realizing that $\iota'\mathbf{x_i^*} = 1$ and multiplying f with $\frac{1+Z_i}{1+Z_i}$ gives

$$\mu = -\zeta + f\frac{Z_i}{1+Z_i}. \tag{C.13}$$

Now working out (C.8) gives

$$\mathbf{x_i} = \mathbf{Q_i}^{-1}\iota\zeta + \mathbf{Q_i}^{-1}2W_i\mathbf{\Lambda}\mathbf{x_i^*} - \mathbf{Q_i}^{-1}\iota\zeta + \mathbf{Q_i}^{-1}\iota f\frac{Z_i}{1+Z_i} \tag{C.14}$$

$$= \frac{1}{1+Z_i}\mathbf{x_i^*} + \frac{Z_i}{1+Z_i}\mathbf{\Lambda}^{-1}\iota(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1}. \tag{C.15}$$

Proof of Lemma 2

With information shocks, (C.1) changes to

$$\begin{bmatrix} -\mathbf{a_i^v} \\ 1 \end{bmatrix} = \begin{bmatrix} -\mathbf{Q_i} & \iota \\ \iota' & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x_i} \\ \mu \end{bmatrix}, \tag{C.16}$$

where

$$\mathbf{Q_i} = 2W_i(1+Z_i)\mathbf{\Lambda} \tag{C.17}$$

$$\mathbf{a_i^v} = \iota\zeta + 2W_i\mathbf{\Lambda}\mathbf{x_i^*} + \mathbf{v_i}. \tag{C.18}$$

Working through, we get the solution

$$\mathbf{x_i} = \frac{1}{1+Z_i}\mathbf{x_i^*} + \frac{Z_i}{1+Z_i}\mathbf{\Lambda}^{-1}\iota(\iota'\mathbf{\Lambda}^{-1}\iota)^{-1} + \frac{1}{2W_i(1+Z_i)}\mathbf{\Lambda}^{-1}(\mathbf{I} - (i'\mathbf{\Lambda}^{-1}i)^{-1}ii'\mathbf{\Lambda}^{-1})\mathbf{v_i}. \tag{C.19}$$

Proof of Proposition 1

Equation (4.12) writes like (using H, the harmonic average lambda):

$$\mathbf{o_m} = \mathbf{\Lambda}^{-1} \times \left[1 - \begin{pmatrix} \frac{H}{N\lambda_1} & \frac{H}{N\lambda_2} & \cdots & \frac{H}{N\lambda_N} \\ \frac{H}{N\lambda_1} & \frac{H}{N\lambda_2} & & \frac{H}{N\lambda_N} \\ \vdots & & \ddots & \vdots \\ \frac{H}{N\lambda_1} & \frac{H}{N\lambda_2} & \cdots & \frac{H}{N\lambda_N} \end{pmatrix} \right] \times M\bar{\mathbf{v}} + \begin{pmatrix} \frac{H\sum_j o_j}{N\lambda_1} \\ \frac{H\sum_j o_j}{N\lambda_2} \\ \vdots \\ \frac{H\sum_j o_j}{N\lambda_N} \end{pmatrix}, \tag{C.20}$$

where $\bar{v}_1 = 0$. We hence have:

$$o_j = \frac{H\sum_j o_j}{N\lambda_j} + \lambda_j^{-1} M v_j \left(1 - \frac{H}{N\lambda_j}\right) - \lambda_j^{-1} \times HM \sum_{-j} \frac{v_k}{N\lambda_k}, \tag{C.21}$$

or:

$$\frac{N\lambda_j}{H} o_j - \sum_k o_k = \frac{NMv_j}{H} - M \sum_k \frac{v_k}{\lambda_k} \tag{C.22}$$

or (by subtracting this equation for $j = 1$ from the equation for any other j):

$$Mv_j = \lambda_j o_j - \lambda_1 o_1. \tag{C.23}$$

C.2 Overview of notation used

Parameters		
Symbol	*Support*	*Description*
ζ	$\mathbb{R}$	Market price of risk
λ_j	$\mathbb{R}^+$	price impact parameter of security j
$\mathbf{\Lambda}$	$\mathbb{R}^{N+} \times \mathbb{R}^{N+}$	Diagonal matrix containing all λ_js
W_i	$\mathbb{R}^+$	Starting wealth of investor i
Z_i	$\mathbb{R}$	Liquidity shock of investor i
$\mathbf{x_i^*}$	$\mathbb{R}^N$	Starting portfolio of investor i
Indices		
i	$\{1, .., M\}$	Investors
j	$\{1, .., N\}$	Securities
Decision variables		
$\mathbf{x_i}$	$\mathbb{R}$	Portfolio allocation of investor i
$\delta_{i,j}$	$\{-1, 1\}$	Trading direction of investor i in security j
$\mathbf{\Delta_i}$	$\mathbb{Z}^N \times \mathbb{Z}^N$	Diagonal matrix containing all $\delta_{i,j}$s of investor i
$\mathbf{o_i}$	$\mathbb{R}^N$	Order flow of investor i

C.3 Sample selection and composition

This appendix describes the selection and composition of our sample of S&P 1500 stocks. Our starting point is a list of all 2,553 stocks that were a constituent of the S&P 1500 index at some point in the period from January 2001 till December 2010 (including tickers, CUSIPs, and begin and end dates of the S&P 1500 index membership) – downloaded on February 3, 2011 from Compustat Monthly Updates North America Index Constituents. There are 2,392 unique tickers in this list.[1] As TAQ is organized by ticker (or symbol in TAQ terminology),

[1] In most cases, multiple identical tickers occur on the list when the same stock (same name and same 8-digit CUSIPs) is listed as an S&P 1500 index constituent multiple times. In several cases, these different entries refer to distinct periods of S&P 1500 membership (such as Ace Ltd., which has entries for the period from January 30, 2002 till July 17, 2008 and for the period from July 15, 2010 till the end of our sample period). However, in a substantial number of cases, the different entries refer to consecutive periods of S&P 1500 membership for the same stock, with at most one trading day—quite often this day is Friday, August 1, 2003—in between the periods (such as U.S. Steel, which has entries for the period from February 1, 2001 till July 31, 2003, from

we use the TAQNAMES file downloaded on January 1, 2010 (and for later years the monthly TAQ Master files for December 2009 and December 2010 downloaded on 27 July 2011, as TAQNAMES is no longer available) to check whether the Compustat tickers are available in TAQ. Of the 2,392 unique tickers, 346 cannot be found in TAQ. For the stocks with these tickers, we check whether an adjusted ticker that refers to the same stock is available in TAQ (based on a comparison of the 8-digit CUSIP and/or stock name on Compustat and TAQ). We make adjustments to 331 of the tickers. We note that most of these adjustments are trivial, such as removing "." or ".1" at the end of the ticker. We discard 15 stocks for which we could not find a corresponding ticker in TAQ. As we want to analyze only stocks listed on NYSE, AMEX, or Nasdaq, we obtain a list of all Compustat stocks and their stock exchange (data item EXCHG – which is the most recent exchange the stock was listed on) – downloaded on May 26, 2011 from Compustat. We also need the exchange of each stock because we follow prior studies and only download quotes for each stock from their own exchange. If the exchange in this list does not equal NYSE, AMEX, or Nasdaq, we manually check (primarily using internet searches) whether the stock was listed on one of these exchanges in an earlier period. Most stocks on our S&P 1500 constituents list for which Compustat indicates a different exchange than these three are stocks that went into bankruptcy or went private but used to be listed. For stocks that change from one of these three exchanges to another one of these three exchanges during our sample period, we only use the data for the most recent exchange the stock was listed on. After this procedure, there are 2,342 unique adjusted tickers, for which we download and process intraday TAQ data over the period 2001-2010 to construct daily measures of order imbalance and price impact. As the same ticker can be used on TAQ by multiple stocks in different periods, it is important to check whether the downloaded TAQ data for each ticker actually corresponds to the same stock in our list of S&P 1500 constituents. To that end, we look up each ticker in our list of S&P 1500 stocks in the TAQNAMES and/or TAQ Master files and verify that it is the same stock based the stock name, the 8-digt CUSIP, and the begin and end dates of the presence of the stock on TAQ. This verification has to be carried out manually, because TAQNAMES often contains different rows for the same ticker and even the same stock. If a stock's ticker is not in our TAQNAMES file (which covers the period till the end of 2008), we check whether it is in the TAQ Master files of December 2009 and/or December 2010. If that is the case, we use the start and end of those years as the begin and end dates on TAQ, realizing that TAQ data may not be available over those full years. If the period during which a stock appears on TAQ does not overlap with the period during which

August 4, 2003 till August 28, 2005, and from August 29, 2005 till the end of our sample period). We treat these consecutive periods with at most one trading day in between as one continuous index membership period. When the different S&P 1500 membership periods for a particular stock are non-consecutive, we download the entire data history available in TAQ for those stocks, though we later retain only the longest uninterrupted period for stocks for which there is a gap in the data of more than two months.

it is an S&P 1500 constituent, we discard the stock.[2] In line with the recommendation of WRDS, we use the 8-digit CUSIP to match the TAQ data with CRSP based on the historical CUSIP (data item "NCUSIP") in CRSP and obtain the CRSP "PERMNO" identifier for each stock in our list. We manually check whether the names in CRSP match those of our list of stocks, and whether different names refer to the same stock using the PERMNO and/or internet searches. We discard one stock for which we cannot find a match on CRSP. The resulting dataset consists of 2,302 different stocks (with 2,282 unique adjusted tickers), of which 1,408 are NYSE listings, 12 are AMEX listings, and 882 are Nasdaq listings. We note that we discard some more stocks based on further data screens discussed in Section 4 and Appendix D.

C.4 Data screens and filters applied to the TAQ data

This appendix describes the data screens and filters we apply to our sample of S&P 1500 stocks. We follow Hasbrouck (2007) and set the price of the first trade on a day to missing to cope with issues surrounding overnight price changes and special features of the opening. We discard bid and ask quotes that are less than or equal to 0, bid and ask sizes that are less than or equal to 0, and quote conditions (mode) that are not in 4, 7, 9, 11, 13, 14, 15, 19, 20, 27, 28, following WRDS recommendations. We only retain quotes from the primary listing exchange of each stock, but we use trades from all trading venues, not just the primary listing exchange, following Hasbrouck (2007). We discard trades that are out of sequence (as indicated by a sale condition that is in O, Z, B, T, L, G, W, J, K, following WRDS recommendations), recorded before the market open or after the market close (following Chordia et al. (2001), with special settlement conditions (as indicated by a correction indicator that is not in 0,1,2), or with a price less than or equal to 0 or a trade size less than or equal to 0, again following WRDS recommendations. We also discard trades with (i) a quoted spread less than $0 or greater than $5, (ii) a ratio of effective spread to quoted spread greater than 4, or (iii) a ratio of proportional effective spread to proportional quoted spread greater than 4 (following Chordia et al. (2001).

[2] In a small number of cases, the TAQ CUSIP is different from the Compustat CUSIP (usually only the seventh digit, which identifies the exact issue – where the first six digits identify the issuer), but the stock name and period correspond and there are no other stocks with the same symbol in TAQNAMES. In these cases, we retain the TAQ CUSIP, as this is the historical CUSIP that corresponds to the data we downloaded from TAQ for that stock. In some cases, TAQNAMES shows multiple lines for the same ticker with the same name and the same 8-digit CUSIP. If the begin and end dates of those different lines are consecutive, we treat them as representing a single stock. If not, and if TAQ only covers the period listed on one of the lines, we use that period. If one of the lines lists a longer period on TAQ that encompasses the shorter period listed on the other line, we use the longer period.

Nederlandse samenvatting (Summary in Dutch)

De welvaart van landen wordt bepaald door de mate waarin efficiënt gebruik wordt gemaakt van reële activa zoals land, machines en kennis. Financiële activa zijn slechts claims op deze reële activa.

Desalniettemin hebben financiële markten verschillende belangrijke taken: ze maken het mogelijk om de verhouding tussen rendement en risico te optimaliseren, om consumptie te verplaatsen naar de toekomst, om belangrijke informatie van beleggers te aggregeren en om schaarse middelen optimaal aan te wenden. Door het vervullen van deze taken kunnen financiële markten het efficiënt gebruik van reële activa stimuleren en daarmee een reëel effect op de economie hebben en economische groei bevorderen. Bijvoorbeeld, als investeerders geloven dat een bepaald bedrijf goede investeringsmogelijkheden heeft, dan zullen zij aandelen in het bedrijf gaan kopen wat de aandelenprijs omhoog stuwt. De hogere aandelenprijs maakt het voor het bedrijf makkelijker om extra kapitaal op te halen; het opgehaalde kapitaal kan dan door bedrijf ingezet worden om de investeringsmogelijkheden te benutten. Op deze wijze spelen financiële markten een belangrijke rol in het toewijzen van schaars kapitaal aan bedrijven.

De efficiëntie waarmee deze rollen worden vervuld hangt af van de mate waarin de prijzen van financiële activa de echte waarde van deze activa weerspiegelen en de mate waarin markten het handelen in deze activa faciliteren. Met andere woorden, de voortvarendheid waarmee deze rollen worden vervuld hangt af van de mate waarin financiële markten efficiënt en liquide zijn.

Academici hebben grote variatie in liquiditeit over tijd en tussen markten gedocumenteerd, maar hebben tegelijkertijd marktefficiëntie als een statisch concept behandeld. Dit lijkt vreemd, aangezien liquiditeit en efficiëntie nauw verwant zijn aan elkaar. Er valt wat voor te zeggen dat markten niet efficiënt zijn per se, maar dat efficiëntie handel van geïnformeerde investeerders tegen mogelijke inefficiënties vereist, waarbij het bereiken van efficiëntie af-hangt van het gemak waarmee de investeerders kunnen handelen (marktliquiditeit) en van de hoeveelheid kapitaal die zij voor het handelen voor handen hebben (financieringsliquiditeit). For example, Grossman and Stiglitz (1980) argue that markets cannot always be perfectly efficient. Prices cannot always reflect all possible information, as in such a market informed traders would "make no (private) return from their (privately) costly activity" (p. 393) and hence would pursue other activities, leaving prices less informative.

Het hoofdthema van deze thesis is het onderzoeken van de interactie tussen marktefficiëntie en marktliquiditeit. Specifiek het documenteren van variatie in marktefficiëntie over tijd en tussen markten en of de efficiëntie van individuele aandelen meebeweegt met marktefficiëntie; het onderzoeken waarom inefficiënties ontstaan en hoe het handelen tegen deze inefficiënties marktliquiditeit beïnvloedt; en het aandragen van een nieuwe maatstaf voor de waarschijnlijkheid van geïnformeerd handelen.

References

Acharya, V. V. and L. H. Pedersen (2005). Asset pricing with liquidity risk. *Journal of Financial Economics 77*, 375–410.

Admati, A. R. and P. Pfleiderer (1988). A Theory of Intraday Patterns: Volume and Price Variability. *Review of Financial Studies 1*, 3–40.

Aitken, M. and A. Frino (1996). The accuracy of the tick test: Evidence from the Australian stock exchange. *Journal of Banking and Finance 20*, 1715–1729.

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

Amihud, Y. and H. Mendelson (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics 17*, 223–249.

Andersen, T., T. Bollerslev, and A. Das (2001). Variance-ratio statistics and high-frequency data: Testing for changes in intraday volatility patterns. *Journal of Finance 56*, 305–327.

Baker, M. and J. C. Stein (2004). Market liquidity as a sentiment indicator. *Journal of Financial Markets 7*, 271–299.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics 9*, 3 – 18.

Ben-David, I., F. Franzoni, and R. Moussawi (2014). Do ETFs Increase Volatility? *working paper, Ohio State University*.

Benston, G. and R. Hagerman (1974). Determinants of bid-asked spreads in the over-the-counter market. *Journal of Financial Economics 1*, 353–364.

Berkman, H. and N. H. Nguyen (2010). Domestic liquidity and cross-listing in the United States. *Journal of Banking & Finance 34*, 1139–1151.

Bessembinder, H. (2003). Trade execution costs and market quality after decimalization. *Journal of Financial and Quantitative Analysis 38*, 747–777.

Beveridge, S. and C. Nelson (1981). A new approach to the decomposition of economic time series into permanent and transitory components with particular attention to the measurement of the business cycle. *Journal of Monetary Economics 7*, 151–174.

Boehmer, E., K. Fong, and J. Wu (2014). International evidence on algorithmic trading. *working paper, Singapore Management University*.

Boehmer, E. and E. K. Kelley (2009). Institutional investors and the informational efficiency of prices. *Review of Financial Studies 22*, 3563–3594.

Boehmer, E. and J. Wu (2007). Order flow and prices. *working paper, University of Georgia*.

Bongaerts, D., D. Rösch, and M. van Dijk (2014). Cross-Sectional Identification of Informed Trading. *working paper, Erasmus University, Rotterdam School of Management*.

Brennan, M. J. and A. Subrahmanyam (1996). Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of Financial Economics 41*, 441 – 464.

Brogaard, J., T. Hendershott, and R. Riordan (2014a). High frequency trading and price discovery. *Review of Financial Studies 27*, 2267–2306.

Brogaard, J., T. Hendershott, and R. Riordan (2014b). High frequency trading and price discovery. *Review of Financial Studies 27*, 2267–2306.

Brunnermeier, M. K. (2009). Deciphering the liquidity and credit crunch 2007-2008. *Journal of Economic Perspectives 23*, 77–100.

Brunnermeier, M. K., S. Nagel, and L. H. Pedersen (2008). Carry trades and currency crashes. *working paper, Princeton University*.

Brunnermeier, M. K. and L. H. Pedersen (2008). Market Liquidity and Funding Liquidity. *Review of Financial Studies 22*, 2201–2238.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance 52*, 57–82.

Chaboud, A., B. Chiquoine, E. Hjalmarsson, and C. Vega (2013). Rise of the machines: Algorithmic trading in the foreign exchange market. *Journal of Finance forthcoming*.

Chan, K. and W. Fong (2000). Trade size, order imbalance, and the volatility-volume relation. *Journal of Financial Economics 57*, 247–273.

Charles, A. and O. Darné (2009). Variance-ratio tests of random walk: An overview. *Journal of Economic Surveys 23*, 503–527.

Choi, D., M. Getmansky, and H. Tookes (2009). Convertible bond arbitrage, liquidity externalities, and stock prices. *Journal of Financial Economics 91*, 227–251.

Chordia, T., R. Roll, and A. Subrahmanyam (2000). Commonality in liquidity. *Journal of Financial Economics 56*, 3–28.

Chordia, T., R. Roll, and A. Subrahmanyam (2001). Market liquidity and trading activity. *Journal of Finance 56*, 501–530.

Chordia, T., R. Roll, and A. Subrahmanyam (2002). Order imbalance, liquidity, and market returns. *Journal of Financial Economics 65*, 111–130.

Chordia, T., R. Roll, and A. Subrahmanyam (2005). Evidence on the speed of convergence to market efficiency. *Journal of Financial Economics 76*, 271–292.

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

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

Cole-Fontayn, M. (2011). The Depositary Receipt Market 2010 Yearbook. *BNY Mellon*.

Comerton-Forde, C., T. Hendershott, C. M. Jones, P. C. Moulton, and M. S. Seasholes (2010). Time Variation in Liquidity: The Role of Market-Maker Inventories and Revenues. *Journal of Finance 65*, 295–331.

Coughenour, J. and M. Saad (2004). Common market makers and commonality in liquidity. *Journal of Financial and Quantitative Analysis 45*, 335–367.

Cox, D. R. and D. R. Peterson (1994). Stock Returns Following Large One-Day Declines: Evidence on Short-Term Reversals and Longer-Term Performance. *Journal of Finance 49*, 255–67.

Cremers, M. and D. Weinbaum (2010). Deviations from Put-Call Parity and Stock Return Predictability. *Journal of Financial and Quantitative Analysis 45*, 335–367.

De Jong, A., L. Rosenthal, and M. van Dijk (2009). The Risk and Return of Arbitrage in Dual-Listed Companies. *Review of Finance 13*, 495–520.

Doan, T. A. (2010). RATS Handbook for Vector Autoregressions. *working paper, Estima*.

Domowitz, I., J. Glen, and A. Madhavan (1998). International Cross-Listing and Order Flow Migration : Evidence from an Emerging Market. *Journal of Finance 53*, 2001–2027.

Dow, J. and R. Rahi (2003). Informed Trading, Investment, and Welfare. *Journal of Business 76*, 439–454.

Easley, D., S. Hvidkjaer, and M. O'Hara (2002). Is Information Risk a Determinant of Asset Returns? *Journal of Finance 57*, 2185–2221.

Easley, D., N. M. Kiefer, M. O'Hara, and J. B. Paperman (1996). Liquidity, information, and infrequently traded stocks. *Journal of Finance 51*, 1405–1436.

Easley, D., M. M. López de Prado, and M. O'Hara (2011). The Microstructure of the 'Flash Crash'. *Journal of Portfolio Management 37*, 118–128.

Easley, D., M. M. López de Prado, and M. O'Hara (2012). Flow Toxicity and Liquidity in a High-frequency World. *Review of Financial Studies 25*, 1457–1493.

Easley, D. and M. O'Hara (2004). Information and the cost of capital. *Journal of Finance 59*, 1553–1583.

Edmans, A. (2009). Blockholder Trading, Market Efficiency, and Managerial Myopia. *Journal of Finance 64*, 2481–2513.

Ellis, K., R. Michaely, and M. O'Hara (2000). The accuracy of trade classification rules: Evidence from Nasdaq. *Journal of Financial and Quantitative Analysis 35*, 529–551.

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance 25*, 383–417.

Fama, E. and J. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*, 607–636.

Fama, E. and M. Miller (1972). The Theory of Finance. *Holt, Rinehart Winston, NewYork*.

Fama, E. F. (1991). Efficient capital markets: II. *Journal of Finance 46*, 1575–1617.

Fama, E. F. and K. R. French (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics 33*, 3–56.

Finucane, T. (1991). Put-call parity and expected returns. *Journal of Financial and Quantitative Analysis 26*, 445–457.

Fishman, M. J. and K. M. Hagerty (1989). Disclosure decisions by firms and the competition for price efficiency. *Journal of Finance 44*, 633–646.

Foucault, T., R. Kozhan, and W. W. Tham (2013). Toxic Arbitrage. *working paper, HEC School of Management*.

Foucault, T., M. Pagano, and A. Röell (2013). *Market Liquidity Theory, Evidence, and Policy*. New York: Oxford University Press.

Gagnon, L. and A. G. Karolyi (2010a). Do International Cross-listings Still Matter? *working paper, Cornell University*.

Gagnon, L. and A. G. Karolyi (2010b). Multi-market trading and arbitrage. *Journal of Financial Economics 97*, 53–80.

Gagnon, L. and A. G. Karolyi (2013). *Evidence on Financial Globalization and Crises*. London: Elsevier.

Gagnon, L. and G. A. Karolyi (2009). Information, Trading Volume, and International Stock Return Comovements: Evidence from Cross-Listed Stocks. *Journal of Financial and Quantitative Analysis 44*, 953–986.

Gao, X. and J. Ritter (2010). The marketing of seasoned equity offerings. *Journal of Financial Economics 97*, 33–52.

Glosten, L. and L. Harris (1988). Estimating the components of the bid/ask spread. *Journal of Financial Economics 21*, 123–142.

Glosten, L. and P. Milgrom (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics 14*, 71–100.

Goldstein, I. and A. Guembel (2008). Manipulation and the Allocational Role of Prices. *Review of Economic Studies 75*, 133–164.

Gromb, D. and D. Vayanos (2010). Limits of arbitrage: The state of the theory. *Annual Review of Financial Economics 2*, 251–275.

Grossman, S. J. and J. E. Stiglitz (1980). On the Impossibility of Informationally Efficient Markets. *American Economic Review 70*, 393–408.

Halling, M., M. Pagano, O. Randl, and J. Zechner (2007). Where Is the Market? Evidence from Cross-Listings in the United States. *Review of Financial Studies 21*, 725–761.

Hameed, A., W. Kang, and S. Viswanathan (2010). Stock market declines and liquidity. *Journal of Finance 65*, 257–293.

Hanson, S. G. and A. Sunderam (2014). The growth and limits of arbitrage: Evidence from short interest. *Review of Financial Studies 27*, 1238–1286.

Hasbrouck, J. (1991). Measuring the Information Content of Stock Trades. *Journal of Finance 46*, 179–207.

Hasbrouck, J. (1993). Assessing the quality of a security market: A new approach to transaction-cost measurement. *Review of Financial Studies 6*, 191–212.

Hasbrouck, J. (2007). *Empirical Market Microstructure*. Oxford University Press.

Hasbrouck, J. and T. Ho (1987). Order arrival, quote behavior, and the return-generating process. *Journal of Finance 42*, 1540–6261.

Hasbrouck, J. and D. Seppi (2000). Common factors in prices, order flows and liquidity. *Journal of Financial Economics 59*, 383–411.

Hendershott, T., C. M. Jones, and A. J. Menkveld (2011). Does algorithmic trading improve liquidity? *Journal of Finance 66*, 1–33.

Holden, C. W. (1995). Index arbitrage as cross-sectional market making. *Journal of Futures Markets 15*, 423–455.

Hu, G. X., J. Pan, and J. Wang (2013). Noise as Information for Illiquidity. *Journal of Finance 68*, 2341–2382.

Huberman, G. and D. Halka (2001). Systematic liquidity. *Journal of Financial Research 24*, 161–178.

Jaffe, J. F. (1974). Special information and insider trading. *The Journal of Business 47*, 410–428.

Jegadeesh, N. (1990). Evidence of Predictable Behavior of Security Returns. *Journal of Finance 45*, 881–98.

Jensen, M. C. (1978). Some anomalous evidence regarding market efficiency. *Journal of Financial Economics 6*, 95 – 101.

Jylhä, P., K. Rinne, and M. Suominen (2015). Do hedge funds supply or demand immediacy? *Review of Finance forthcoming*.

Karolyi, G. A. (1998). Why do companies list shares abroad?: A survey of the evidence and its managerial implications. *Financial Markets, Institutions & Instruments*, 1–60.

Karolyi, G. A., K.-H. Lee, and M. van Dijk (2012). Understanding commonality in liquidity around the world. *Journal of Financial Economics 105*, 82–112.

Kumar, P. and D. J. Seppi (1994). Information and index arbitrage. *Journal of Business 67*, 481–509.

Kyle, A. (1985). Continuous auctions and insider trading. *Econometrica 53*, 1315–1335.

Lai, S., L. Ng, and B. Zhang (2014). Does PIN Affect Equity Prices Around the World? *Forthcoming Journal of Financial Economics*.

Lamont, O. a. and R. H. Thaler (2003). Can the Market Add and Subtract? Mispricing in Tech Stock Carve-outs. *Journal of Political Economy 111*, 227–268.

Lau, S., L. Ng, and B. Zhang (2012). Information environment and equity risk premium volatility around the world. *Management Science 58*, 1322–1340.

Lee, C. M. and B. Radhakrishna (2000). Inferring investor behavior: Evidence from TORQ data. *Journal of Financial Markets 3*, 83–111.

Lee, C. M. and M. J. Ready (1991). Inferring trade direction from intraday data. *Journal of Finance 46*, 733–746.

Levine, R. (2005). Finance and Growth: Theory and Evidence. *Handbook of Economic Growth 1*, 865–934.

Lin, T.-C., X. Lu, and J. Driessen (2013). Why Do Options Prices Predict Stock Returns? *Working paper*.

Lo, A. and C. MacKinlay (1989). The size and power of the variance ratio test in finite samples: A Monte Carlo investigation. *Journal of Econometrics 40*, 203–238.

Lou, D. and C. Polk (2013). Comomentum: Inferring arbitrage activity from return correlations. *working paper, London School of Economics*.

Madhavan, A., M. Richardson, and M. Roomans (2002). Why do security prices change? A transaction-level analysis of NYSE stocks. *Review of Financial Studies 10*, 1035–1064.

Mancini-Griffoli, T. and A. Ranaldo (2011). Limits to arbitrage during the crisis: Funding liquidity constraints and covered interest parity. *working paper, Swiss National Bank*.

Manove, M. (1989). The Harm from Insider Trading and Informed Speculation. *Quarterly Journal of Economics 104*, 823–45.

Marshall, B., N. Nguyen, and N. Visaltanachoti (2012). Commodity liquidity measurement and transaction costs. *Review of Financial Studies 25*, 599–638.

Marshall, B. R., N. H. Nguyen, and N. Visaltanachoti (2011). Commodity Liquidity Measurement and Transaction Costs. *Review of Financial Studies 25*, 599–638.

Menkveld, A. (2013). High frequency trading and the new-market makers. *Journal of Financial Markets 16*, 712–740.

Mitchell, M. and T. Pulvino (2012). Arbitrage crashes and the speed of capital. *Journal of Financial Economics 104*, 469–490.

Mitchell, M., T. Pulvino, and E. Stafford (2002). Limited Arbitrage in Equity Markets. *Journal of Finance 57*, 551–584.

Morck, R., B. Yeung, and W. Yu (2000). The information content of stock markets: Why do emerging markets have synchronous stock price movements? *Journal of Financial Economics 58*, 215–260.

Moulton, P. C. and L. Wei (2009). A tale of two time zones: The impact of substitutes on cross-listed stock liquidity. *Journal of Financial Markets 12*, 570–591.

Muscarella, C. and M. Vetsuypens (1996). Stock splits: Signaling or liquidity? The case of ADR 'solo-splits'. *Journal of Financial Economics 42*, 3–26.

Nagel, S. (2005). Short-sales, institutional investors, and the cross-section of stock returns. *Journal of Financial Economics 78*, 277–309.

Nagel, S. (2012). Evaporating liquidity. *Review of Financial Studies 25*, 2005–2039.

Newey, W. and K. West (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies 61*, 631–653.

Odders-White, E. R. (2000). On the occurrence and consequences of inaccurate trade classification. *Journal of Financial Markets 3*, 259–286.

O'Hara, M. and G. S. Oldfield (1986). The microeconomics of market making. *Journal of Financial and Quantitative Analysis 21*, 361–376.

O'Hara, M. and M. Ye (2011). Is market fragmentation harming market quality? *Journal of Financial Economics 100*, 459–474.

Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics 63*, 3–50.

Pasquariello, P. (2014). Financial market dislocations. *Review of Financial Studies 27*, 1868–1914.

Pastor, L. and R. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*, 642–685.

Pesaran, H. and Y. Shin (1998). Generalized impulse response analysis in linear multivariate models. *Economics Letters 58*, 17–29.

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. W., E. S. Schwartz, and A. Subrahmanyam (2007). Liquidity and The Law of One Price: The Case of the Cash/Futures Basis. *Journal of Finance 62*, 2201–2234.

Rösch, D. (2014). The Impact of Arbitrage on Market Liquidity. *working paper, Erasmus University, Rotterdam School of Management*.

Rösch, D., A. Subrahmanyam, and M. van Dijk (2015). An Empirical Analysis of Co-Movement in Market Efficiency. *working paper, Erasmus University, Rotterdam School of Management*.

Sadka, R. (2006). Momentum and Post-Earnings-Announcement Drift Anomalies: The Role of Liquidity Risk. *Journal of Financial Economics 80*, 309–349.

Samuelson, P. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review 6*, 41–49.

Schultz, P. and S. Shive (2010). Mispricing of dual-class shares: Profit opportunities, arbitrage, and trading. *Journal of Financial Economics 98*, 524–549.

Shiller, R. J. (1981). Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? *American Economic Review 71*, 421–36.

Shleifer, A. and R. W. Vishny (1997). The Limits of Arbitrage. *Journal of Finance 52*, 35–55.

Stoll, H. (1978). The supply of dealer services in securities markets. *The Journal of Finance 33*, 1133–1151.

Theissen, E. (2001). A test of the accuracy of the Lee/Ready trade classification algorithm. *Journal of International Financial Markets, Institutions and Money 11*, 147–165.

Werner, I. M. and A. W. Kleidon (1996). UK and US trading of British cross-listed stocks: An intraday analysis of market integration. *Review of Financial Studies 9*, 619–664.

Wurgler, J. (2000). Financial markets and the allocation of capital. *Journal of Financial Economics 58*, 187–214.

Biography

Dominik Rösch is currently an Assistant Professor of Finance at the State University of New York at Buffalo. Before starting his PhD at the Rotterdam School of Management he received a Master's degree (or its German equivalent) in Mathematics from the University of Bonn and in Finance from the University of London. He worked several years as a programmer and as a quantitative analyst, of which four years he worked as the Head of Quants for the Regentmarkets Group. His research interests are in empirical asset pricing and market microstructure. In particular, he is interested in market efficiency, liquidity, and arbitrage. He also has a passion for programming and database management: During his PhD he developed a framework that allows to conveniently analyze terabytes of worldwide tick-by-tick data. Dominik presented his research in several international conferences, such as the annual meetings of the European Finance Association, and received an outstanding paper award at the 50th Eastern Finance Association meeting. He also spend a semester at Cornell and served as a referee for the Review of Financial Studies.

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT (ERIM)

ERIM PH.D. SERIES RESEARCH IN MANAGEMENT

The ERIM PhD Series contains PhD dissertations in the field of Research in Management defended at Erasmus University Rotterdam and supervised by senior researchers affiliated to the Erasmus Research Institute of Management (ERIM). All dissertations in the ERIM PhD Series are available in full text through the ERIM Electronic Series Portal: http://repub.eur.nl/pub. ERIM is the joint research institute of the Rotterdam School of Management (RSM) and the Erasmus School of Economics at the Erasmus University Rotterdam (EUR).

DISSERTATIONS LAST FIVE YEARS

Abbink, E.J., *Crew Management in Passenger Rail Transport*, Promotor(s): Prof.dr. L.G. Kroon & Prof.dr. A.P.M. Wagelmans, EPS-2014-325-LIS, http://repub.eur.nl/pub/76927

Acar, O.A., *Crowdsourcing for Innovation: Unpacking Motivational, Knowledge and Relational Mechanisms of Innovative Behavior in Crowdsourcing Platforms*, Promotor(s): Prof.dr.ir. J.C.M. van den Ende, EPS-2014-321-LIS, http://repub.eur.nl/pub/76076

Acciaro, M., *Bundling Strategies in Global Supply Chains*, Promotor(s): Prof.dr. H.E. Haralambides, EPS-2010-197-LIS, http://repub.eur.nl/pub/19742

Akin Ates, M., *Purchasing and Supply Management at the Purchase Category Level: strategy, structure and performance*, Promotor(s): Prof.dr. J.Y.F. Wynstra & Dr. E.M. van Raaij, EPS-2014-300-LIS, http://repub.eur.nl/pub/50283

Akpinar, E., *Consumer Information Sharing*, Promotor(s): Prof.dr.ir. A. Smidts, EPS-2013-297-MKT, http://repub.eur.nl/pub/50140

Alexander, L., *People, Politics, and Innovation: A Process Perspective*, Promotor(s): Prof.dr. H.G. Barkema & Prof.dr. D.L. van Knippenberg, EPS-2014-331-S&E, http://repub.eur.nl/pub/77209

Alexiev, A.S., *Exploratory Innovation: The Role of Organizational and Top Management Team Social Capital*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2010-208-STR, http://repub.eur.nl/pub/20632

Almeida e Santos Nogueira, R.J. de, *Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty*, Promotor(s): Prof.dr.ir. U. Kaymak & Prof.dr. J.M.C. Sousa, EPS-2014-310-LIS, http://repub.eur.nl/pub/51560

Bannouh, K., *Measuring and Forecasting Financial Market Volatility using High-frequency Data*, Promotor(s): Prof.dr. D.J.C. van Dijk, EPS-2013-273-F&A, http://repub.eur.nl/pub/38240

Ben-Menahem, S.M., *Strategic Timing and Proactiveness of Organizations*, Promotor(s): Prof.dr. H.W. Volberda & Prof.dr.ing. F.A.J. van den Bosch, EPS-2013-278-S&E, http://repub.eur.nl/pub/39128

Benning, T.M., *A Consumer Perspective on Flexibility in Health Care: Priority Access Pricing and Customized Care*, Promotor(s): Prof.dr.ir. B.G.C. Dellaert, EPS-2011-241-MKT, http://repub.eur.nl/pub/23670

Berg, W.E. van den, *Understanding Salesforce Behavior using Genetic Association Studies*, Promotor(s): Prof.dr. W.J.M.I. Verbeke, EPS-2014-311-MKT, http://repub.eur.nl/pub/51440

Betancourt, N.E., *Typical Atypicality: Formal and Informal Institutional Conformity, Deviance, and Dynamics*, Promotor(s): Prof.dr. B. Krug, EPS-2012-262-ORG, http://repub.eur.nl/pub/32345

Bezemer, P.J., *Diffusion of Corporate Governance Beliefs: Board independence and the emergence of a shareholder value orientation in the Netherlands*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2010-192-STR, http://repub.eur.nl/pub/18458

Binken, J.L.G., *System markets: Indirect network effects in action, or inaction?*, Promotor(s): Prof.dr. S. Stremersch, EPS-2010-213-MKT, http://repub.eur.nl/pub/21186

Bliek, R. de, *Empirical Studies on the Economic Impact of Trust*, Promotor(s): Prof.dr. J. Veenman & Prof.dr. Ph.H.B.F. Franses, EPS-2015-324-ORG, http://repub.eur.nl/pub/78159

Blitz, D.C., *Benchmarking Benchmarks*, Promotor(s): Prof.dr. A.G.Z. Kemna & Prof.dr. W.F.C. Verschoor, EPS-2011-225-F&A, http://repub.eur.nl/pub/22624

Boons, M., *Working Together Alone in the Online Crowd: The Effects of Social Motivationsand Individual Knowledge Backgrounds on the Participation and Performance of Members of Online Crowdsourcing Platforms*, Promotor(s): Prof.dr. H.G. Barkema & Dr. D.A. Stam, EPS-2014-306-S&E, http://repub.eur.nl/pub/50711

Borst, W.A.M., *Understanding Crowdsourcing: Effects of motivation and rewards on participation and performance in voluntary online activities*, Promotor(s): Prof.dr.ir. J.C.M. van den Ende & Prof.dr.ir. H.W.G.M. van Heck, EPS-2010-221-LIS, http://repub.eur.nl/pub/21914

Brazys, J., *Aggregated Marcoeconomic News and Price Discovery*, Promotor(s): Prof.dr. W.F.C. Verschoor, EPS-2015-351-F&A, http://repub.eur.nl/pub/78243

Budiono, D.P., *The Analysis of Mutual Fund Performance: Evidence from U.S. Equity Mutual Funds*, Promotor(s): Prof.dr. M.J.C.M. Verbeek & Dr.ir. M.P.E. Martens, EPS-2010-185-F&A, http://repub.eur.nl/pub/18126

Burger, M.J., *Structure and Cooptition in Urban Networks*, Promotor(s): Prof.dr. G.A. van der Knaap & Prof.dr. H.R. Commandeur, EPS-2011-243-ORG, http://repub.eur.nl/pub/26178

Byington, E., *Exploring Coworker Relationships: Antecedents and Dimensions of Interpersonal Fit,Coworker Satisfaction, and Relational Models*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2013-292-ORG, http://repub.eur.nl/pub/41508

Camacho, N.M., *Health and Marketing: Essays on Physician and Patient Decision-Making*, Promotor(s): Prof.dr. S. Stremersch, EPS-2011-237-MKT, http://repub.eur.nl/pub/23604

Cancurtaran, P., *Essays on Accelerated Product Development*, Promotor(s): Prof.dr. F. Langerak & Prof.dr.ir. G.H. van Bruggen, EPS-2014-317-MKT, http://repub.eur.nl/pub/76074

Caron, E.A.M., *Explanation of Exceptional Values in Multi-dimensional Business Databases*, Promotor(s): Prof.dr.ir. H.A.M. Daniels & Prof.dr. G.W.J. Hendrikse, EPS-2013-296-LIS, http://repub.eur.nl/pub/50005

Carvalho, L. de, *Knowledge Locations in Cities: Emergence and Development Dynamics*, Promotor(s): Prof.dr. L. Berg, EPS-2013-274-S&E, http://repub.eur.nl/pub/38449

Carvalho de Mesquita Ferreira, L., *Attention Mosaics: Studies of Organizational Attention*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens & Prof.dr. J. van Oosterhout, EPS-2010-205-ORG, http://repub.eur.nl/pub/19882

Cox, R.H.G.M., *To Own, To Finance, and To Insure - Residential Real Estate Revealed*, Promotor(s): Prof.dr. D. Brounen, EPS-2013-290-F&A, http://repub.eur.nl/pub/40964

Defilippi Angeldonis, E.F., *Access Regulation for Naturally Monopolistic Port Terminals: Lessons from Regulated Network Industries*, Promotor(s): Prof.dr. H.E. Haralambides, EPS-2010-204-LIS, http://repub.eur.nl/pub/19881

Deichmann, D., *Idea Management: Perspectives from Leadership, Learning, and Network Theory*, Promotor(s): Prof.dr.ir. J.C.M. van den Ende, EPS-2012-255-ORG, http://repub.eur.nl/pub/31174

Deng, W., *Social Capital and Diversification of Cooperatives*, Promotor(s): Prof.dr. G.W.J. Hendrikse, EPS-2015-341-ORG, http://repub.eur.nl/pub/77449

Desmet, P.T.M., *In Money we Trust? Trust Repair and the Psychology of Financial*

Compensations, Promotor(s): Prof.dr. D. de Cremer, EPS-2011-232-ORG, http://repub.eur.nl/pub/23268

Dietvorst, R.C., *Neural Mechanisms Underlying Social Intelligence and Their Relationship with the Performance of Sales Managers*, Promotor(s): Prof.dr. W.J.M.I. Verbeke, EPS-2010-215-MKT, http://repub.eur.nl/pub/21188

Dollevoet, T.A.B., *Delay Management and Dispatching in Railways*, Promotor(s): Prof.dr. A.P.M. Wagelmans, EPS-2013-272-LIS, http://repub.eur.nl/pub/38241

Doorn, S. van, *Managing Entrepreneurial Orientation*, Promotor(s): Prof.dr. J.J.P. Jansen, Prof.dr.ing. F.A.J. van den Bosch, & Prof.dr. H.W. Volberda, EPS-2012-258-STR, http://repub.eur.nl/pub/32166

Douwens-Zonneveld, M.G., *Animal Spirits and Extreme Confidence: No Guts, No Glory?*, Promotor(s): Prof.dr. W.F.C. Verschoor, EPS-2012-257-F&A, http://repub.eur.nl/pub/31914

Duca, E., *The Impact of Investor Demand on Security Offerings*, Promotor(s): Prof.dr. A. de Jong, EPS-2011-240-F&A, http://repub.eur.nl/pub/26041

Duursema, H., *Strategic Leadership: Moving Beyond the Leader-Follower Dyad*, Promotor(s): Prof.dr. R.J.M. van Tulder, EPS-2013-279-ORG, http://repub.eur.nl/pub/39129

Eck, N.J. van, *Methodological Advances in Bibliometric Mapping of Science*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2011-247-LIS, http://repub.eur.nl/pub/26509

Ellen, S. ter, *Measurement, Dynamics, and Implications of Heterogeneous Beliefs in Financial Markets*, Promotor(s): Prof.dr. W.F.C. Verschoor, EPS-2015-343-F&A, http://repub.eur.nl/pub/78191

Eskenazi, P.I., *The Accountable Animal*, Promotor(s): Prof.dr. F.G.H. Hartmann, EPS-2015-355-F&A, http://repub.eur.nl/pub/78300

Essen, M. van, *An Institution-Based View of Ownership*, Promotor(s): Prof.dr. J. van Oosterhout & Prof.dr. G.M.H. Mertens, EPS-2011-226-ORG, http://repub.eur.nl/pub/22643

Evangelidis, I., *Preference Construction under Prominence*, Promotor(s): Prof.dr. S.M.J. van Osselaer, EPS-2015-340-MKT, http://repub.eur.nl/pub/78202

Feng, L., *Motivation, Coordination and Cognition in Cooperatives*, Promotor(s): Prof.dr. G.W.J. Hendrikse, EPS-2010-220-ORG, http://repub.eur.nl/pub/21680

Fourne, S.P., *Managing Organizational Tensions: A Multi-Level Perspective on Exploration, Exploitation and Ambidexterity*, Promotor(s): Prof.dr. J.J.P. Jansen & Prof.dr.

S.J. Magala, EPS-2014-318-S&E, http://repub.eur.nl/pub/76075

Gharehgozli, A.H., *Developing New Methods for Efficient Container Stacking Operations*, Promotor(s): Prof.dr.ir. M.B.M. de Koster, EPS-2012-269-LIS, http://repub.eur.nl/pub/37779

Gils, S. van, *Morality in Interactions: On the Display of Moral Behavior by Leaders and Employees*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2012-270-ORG, http://repub.eur.nl/pub/38027

Ginkel-Bieshaar, M.N.G. van, *The Impact of Abstract versus Concrete Product Communications on Consumer Decision-making Processes*, Promotor(s): Prof.dr.ir. B.G.C. Dellaert, EPS-2012-256-MKT, http://repub.eur.nl/pub/31913

Gkougkousi, X., *Empirical Studies in Financial Accounting*, Promotor(s): Prof.dr. G.M.H. Mertens & Prof.dr. E. Peek, EPS-2012-264-F&A, http://repub.eur.nl/pub/37170

Glorie, K.M., *Clearing Barter Exchange Markets: Kidney Exchange and Beyond*, Promotor(s): Prof.dr. A.P.M. Wagelmans & Prof.dr. J.J. van de Klundert, EPS-2014-329-LIS, http://repub.eur.nl/pub/77183

Hakimi, N.A., *Leader Empowering Behaviour: The Leader's Perspective*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2010-184-ORG, http://repub.eur.nl/pub/17701

Hensmans, M., *A Republican Settlement Theory of the Firm: Applied to Retail Banks in England and the Netherlands (1830-2007)*, Promotor(s): Prof.dr. A. Jolink & Prof.dr. S.J. Magala, EPS-2010-193-ORG, http://repub.eur.nl/pub/19494

Hernandez-Mireles, C., *Marketing Modeling for New Products*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS-2010-202-MKT, http://repub.eur.nl/pub/19878

Heyde Fernandes, D. von der, *The Functions and Dysfunctions of Reminders*, Promotor(s): Prof.dr. S.M.J. van Osselaer, EPS-2013-295-MKT, http://repub.eur.nl/pub/41514

Heyden, M.L.M., *Essays on Upper Echelons & Strategic Renewal: A Multilevel Contingency Approach*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2012-259-STR, http://repub.eur.nl/pub/32167

Hoever, I.J., *Diversity and Creativity*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2012-267-ORG, http://repub.eur.nl/pub/37392

Hogenboom, F.P., *Automated Detection of Financial Events in News Text*, Promotor(s): Prof.dr.ir. U. Kaymak & Prof.dr. F.M.G. de Jong, EPS-2014-326-LIS, http://repub.eur.nl/pub/77237

Hoogendoorn, B., *Social Entrepreneurship in the Modern Economy: Warm Glow, Cold*

Feet, Promotor(s): Prof.dr. H.P.G. Pennings & Prof.dr. A.R. Thurik, EPS-2011-246-STR, http://repub.eur.nl/pub/26447

Hoogervorst, N., *On The Psychology of Displaying Ethical Leadership: A Behavioral Ethics Approach*, Promotor(s): Prof.dr. D. de Cremer & Dr. M. van Dijke, EPS-2011-244-ORG, http://repub.eur.nl/pub/26228

Hout, D.H. van, *Measuring Meaningful Differences: Sensory Testing Based Decision Making in an Industrial Context; Applications of Signal Detection Theory and Thurstonian Modelling*, Promotor(s): Prof.dr. P.J.F. Groenen & Prof.dr. G.B. Dijksterhuis, EPS-2014-304-MKT, http://repub.eur.nl/pub/50387

Houwelingen, G.G. van, *Something To Rely On*, Promotor(s): Prof.dr. D. de Cremer & Prof.dr. M.H. van Dijke, EPS-2014-335-ORG, http://repub.eur.nl/pub/77320

Huang, X., *An Analysis of Occupational Pension Provision: From Evaluation to Redesign*, Promotor(s): Prof.dr. M.J.C.M. Verbeek, EPS-2010-196-F&A, http://repub.eur.nl/pub/19674

Hurk, E. van der, *Passengers, Information, and Disruptions*, Promotor(s): Prof.dr. L.G. Kroon & Prof.mr.dr. P.H.M. Vervest, EPS-2015-345-LIS, http://repub.eur.nl/pub/78275

Hytonen, K.A., *Context Effects in Valuation, Judgment and Choice: A Neuroscientific Approach*, Promotor(s): Prof.dr.ir. A. Smidts, EPS-2011-252-MKT, http://repub.eur.nl/pub/30668

Iseger, P. den, *Fourier and Laplace Transform Inversion with Applications in Finance*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2014-322-LIS, http://repub.eur.nl/pub/76954

Jaarsveld, W.L. van, *Maintenance Centered Service Parts Inventory Control*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2013-288-LIS, http://repub.eur.nl/pub/39933

Jalil, M.N., *Customer Information Driven After Sales Service Management: Lessons from Spare Parts Logistics*, Promotor(s): Prof.dr. L.G. Kroon, EPS-2011-222-LIS, http://repub.eur.nl/pub/22156

Kagie, M., *Advances in Online Shopping Interfaces: Product Catalog Maps and Recommender Systems*, Promotor(s): Prof.dr. P.J.F. Groenen, EPS-2010-195-MKT, http://repub.eur.nl/pub/19532

Kappe, E.R., *The Effectiveness of Pharmaceutical Marketing*, Promotor(s): Prof.dr. S. Stremersch, EPS-2011-239-MKT, http://repub.eur.nl/pub/23610

Karreman, B., *Financial Services and Emerging Markets*, Promotor(s): Prof.dr. G.A. van der Knaap & Prof.dr. H.P.G. Pennings, EPS-2011-223-ORG, http://repub.eur.nl/pub/22280

Khanagha, S., *Dynamic Capabilities for Managing Emerging Technologies*, Promotor(s): Prof.dr. H.W. Volberda, EPS-2014-339-S&E, http://repub.eur.nl/pub/77319

Kil, J., *Acquisitions Through a Behavioral and Real Options Lens*, Promotor(s): Prof.dr. H.T.J. Smit, EPS-2013-298-F&A, http://repub.eur.nl/pub/50142

Klooster, E. van 't, *Travel to Learn: the Influence of Cultural Distance on Competence Development in Educational Travel*, Promotor(s): Prof.dr. F.M. Go & Prof.dr. P.J. van Baalen, EPS-2014-312-MKT, http://repub.eur.nl/pub/51462

Koendjbiharie, S.R., *The Information-Based View on Business Network Performance: Revealing the Performance of Interorganizational Networks*, Promotor(s): Prof.dr.ir. H.W.G.M. van Heck & Prof.mr.dr. P.H.M. Vervest, EPS-2014-315-LIS, http://repub.eur.nl/pub/51751

Koning, M., *The Financial Reporting Environment: The Role of the Media, Regulators and Auditors*, Promotor(s): Prof.dr. G.M.H. Mertens & Prof.dr. P.G.J. Roosenboom, EPS-2014-330-F&A, http://repub.eur.nl/pub/77154

Konter, D.J., *Crossing Borders with HRM: An Inquiry of the Influence of Contextual Differences in the Adoption and Effectiveness of HRM*, Promotor(s): Prof.dr. J. Paauwe & Dr. L.H. Hoeksema, EPS-2014-305-ORG, http://repub.eur.nl/pub/50388

Korkmaz, E., *Bridging Models and Business: Understanding Heterogeneity in Hidden Drivers of Customer Purchase Behavior*, Promotor(s): Prof.dr. S.L. van de Velde & Prof.dr. D. Fok, EPS-2014-316-LIS, http://repub.eur.nl/pub/76008

Kroezen, J.J., *The Renewal of Mature Industries: An Examination of the Revival of the Dutch Beer Brewing Industry*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens, EPS-2014-333-S&E, http://repub.eur.nl/pub/77042

Kysucky, V., *Access to Finance in a Cros-Country Context*, Promotor(s): Prof.dr. L. Norden, EPS-2015-350-F&A, http://repub.eur.nl/pub/78225

Lam, K.Y., *Reliability and Rankings*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS-2011-230-MKT, http://repub.eur.nl/pub/22977

Lander, M.W., *Profits or Professionalism? On Designing Professional Service Firms*, Promotor(s): Prof.dr. J. van Oosterhout & Prof.dr. P.P.M.A.R. Heugens, EPS-2012-253-ORG, http://repub.eur.nl/pub/30682

Langhe, B. de, *Contingencies: Learning Numerical and Emotional Associations in an Uncertain World*, Promotor(s): Prof.dr.ir. B. Wierenga & Prof.dr. S.M.J. van Osselaer, EPS-2011-236-MKT, http://repub.eur.nl/pub/23504

Larco Martinelli, J.A., *Incorporating Worker-Specific Factors in Operations Management Models*, Promotor(s): Prof.dr.ir. J. Dul & Prof.dr.ir. M.B.M. de Koster, EPS-2010-217-

LIS, http://repub.eur.nl/pub/21527

Leunissen, J.M., *All Apologies: On the Willingness of Perpetrators to Apologize*, Promotor(s): Prof.dr. D. de Cremer & Dr. M. van Dijke, EPS-2014-301-ORG, http://repub.eur.nl/pub/50318

Li, D., *Supply Chain Contracting for After-sales Service and Product Support*, Promotor(s): Prof.dr.ir. M.B.M. de Koster, EPS-2015-347-LIS, http://repub.eur.nl/pub/78526

Li, Z., *Irrationality: What, Why and How*, Promotor(s): Prof.dr. H. Bleichrodt, Prof.dr. P.P. Wakker, & Prof.dr. K.I.M. Rohde, EPS-2014-338-MKT, http://repub.eur.nl/pub/77205

Liang, Q.X., *Governance, CEO Identity, and Quality Provision of Farmer Cooperatives*, Promotor(s): Prof.dr. G.W.J. Hendrikse, EPS-2013-281-ORG, http://repub.eur.nl/pub/39253

Liket, K., *Why 'Doing Good' is not Good Enough: Essays on Social Impact Measurement*, Promotor(s): Prof.dr. H.R. Commandeur & Dr. K.E.H. Maas, EPS-2014-307-STR, http://repub.eur.nl/pub/51130

Loos, M.J.H.M. van der, *Molecular Genetics and Hormones: New Frontiers in Entrepreneurship Research*, Promotor(s): Prof.dr. A.R. Thurik, Prof.dr. P.J.F. Groenen, & Prof.dr. A. Hofman, EPS-2013-287-S&E, http://repub.eur.nl/pub/40081

Lovric, M., *Behavioral Finance and Agent-Based Artificial Markets*, Promotor(s): Prof.dr. J. Spronk & Prof.dr.ir. U. Kaymak, EPS-2011-229-F&A, http://repub.eur.nl/pub/22814

Lu, Y., *Data-Driven Decision Making in Auction Markets*, Promotor(s): Prof.dr.ir. H.W.G.M. van Heck & Prof.dr. W. Ketter, EPS-2014-314-LIS, http://repub.eur.nl/pub/51543

Manders, B., *Implementation and Impact of ISO 9001*, Promotor(s): Prof.dr. K. Blind, EPS-2014-337-LIS, http://repub.eur.nl/pub/77412

Markwat, T.D., *Extreme Dependence in Asset Markets Around the Globe*, Promotor(s): Prof.dr. D.J.C. van Dijk, EPS-2011-227-F&A, http://repub.eur.nl/pub/22744

Mees, H., *Changing Fortunes: How China's Boom Caused the Financial Crisis*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS-2012-266-MKT, http://repub.eur.nl/pub/34930

Meuer, J., *Configurations of Inter-firm Relations in Management Innovation: A Study in China's Biopharmaceutical Industry*, Promotor(s): Prof.dr. B. Krug, EPS-2011-228-ORG, http://repub.eur.nl/pub/22745

Micheli, M.R., *Business Model Innovation: A Journey across Managers' Attention and Inter-Organizational Networks*, Promotor(s): Prof.dr. J.J.P. Jansen, EPS-2015-344-S&E, http://repub.eur.nl/pub/78241

Mihalache, O.R., *Stimulating Firm Innovativeness: Probing the Interrelations between Managerial and Organizational Determinants*, Promotor(s): Prof.dr. J.J.P. Jansen, Prof.dr.ing. F.A.J. van den Bosch, & Prof.dr. H.W. Volberda, EPS-2012-260-S&E, http://repub.eur.nl/pub/32343

Milea, V., *News Analytics for Financial Decision Support*, Promotor(s): Prof.dr.ir. U. Kaymak, EPS-2013-275-LIS, http://repub.eur.nl/pub/38673

Naumovska, I., *Socially Situated Financial Markets: A Neo-Behavioral Perspective on Firms, Investors and Practices*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens & Prof.dr. A. de Jong, EPS-2014-319-S&E, http://repub.eur.nl/pub/76084

Nielsen, L.K., *Rolling Stock Rescheduling in Passenger Railways: Applications in short term planning and in disruption management*, Promotor(s): Prof.dr. L.G. Kroon, EPS-2011-224-LIS, http://repub.eur.nl/pub/22444

Nijdam, M.H., *Leader Firms: The value of companies for the competitiveness of the Rotterdam seaport cluster*, Promotor(s): Prof.dr. R.J.M. van Tulder, EPS-2010-216-ORG, http://repub.eur.nl/pub/21405

Noordegraaf-Eelens, L.H.J., *Contested Communication; A Critical Analysis of Central Bank Speech*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, Prof.dr. J. de Mul, & Prof.dr. D.J.C. van Dijk, EPS-2010-209-MKT, http://repub.eur.nl/pub/21061

Nuijten, A.L.P., *Deaf Effect for Risk Warnings: A Causal Examination applied to Information Systems Projects*, Promotor(s): Prof.dr. G.J. van der Pijl, Prof.dr. H.R. Commandeur & Prof.dr. M. Keil, EPS-2012-263-S&E, http://repub.eur.nl/pub/34928

Oosterhout, M. van, *Business Agility and Information Technology in Service Organizations*, Promotor(s): Prof.dr.ir. H.W.G.M. van Heck, EPS-2010-198-LIS, http://repub.eur.nl/pub/19805

Osadchiy, S.E., *The Dynamics of Formal Organization: Essays on bureaucracy and formal rules*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens, EPS-2011-231-ORG, http://repub.eur.nl/pub/23250

Otgaar, A.H.J., *Industrial Tourism: Where the Public Meets the Private*, Promotor(s): Prof.dr. L. Berg, EPS-2010-219-ORG, http://repub.eur.nl/pub/21585

Ozdemir, M.N., *Project-level Governance, Monetary Incentives, and Performance in Strategic R&D Alliances*, Promotor(s): Prof.dr.ir. J.C.M. van den Ende, EPS-2011-235-LIS, http://repub.eur.nl/pub/23550

Peers, Y., *Econometric Advances in Diffusion Models*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS-2011-251-MKT, http://repub.eur.nl/pub/30586

Peters, M., *Machine Learning Algorithms for Smart Electricity Markets*, Promotor(s): Prof.dr. W. Ketter, EPS-2014-332-LIS, http://repub.eur.nl/pub/77413

Pince, C., *Advances in Inventory Management: Dynamic Models*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2010-199-LIS, http://repub.eur.nl/pub/19867

Porck, J., *No Team is an Island: An Integrative View of Strategic Consensus between Groups*, Promotor(s): Prof.dr. P.J.F. Groenen & Prof.dr. D.L. van Knippenberg, EPS-2013-299-ORG, http://repub.eur.nl/pub/50141

Porras Prado, M., *The Long and Short Side of Real Estate, Real Estate Stocks, and Equity*, Promotor(s): Prof.dr. M.J.C.M. Verbeek, EPS-2012-254-F&A, http://repub.eur.nl/pub/30848

Poruthiyil, P.V., *Steering Through: How organizations negotiate permanent uncertainty and unresolvable choices*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens & Prof.dr. S.J. Magala, EPS-2011-245-ORG, http://repub.eur.nl/pub/26392

Potthoff, D., *Railway Crew Rescheduling: Novel approaches and extensions*, Promotor(s): Prof.dr. A.P.M. Wagelmans & Prof.dr. L.G. Kroon, EPS-2010-210-LIS, http://repub.eur.nl/pub/21084

Pourakbar, M., *End-of-Life Inventory Decisions of Service Parts*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2011-249-LIS, http://repub.eur.nl/pub/30584

Pronker, E.S., *Innovation Paradox in Vaccine Target Selection*, Promotor(s): Prof.dr. H.J.H.M. Claassen & Prof.dr. H.R. Commandeur, EPS-2013-282-S&E, http://repub.eur.nl/pub/39654

Pruijssers, J.K., *An Organizational Perspective on Auditor Conduct*, Promotor(s): Prof.dr. J. van Oosterhout & Prof.dr. P.P.M.A.R. Heugens, EPS-2015-342-S&E, http://repub.eur.nl/pub/78192

Retel Helmrich, M.J., *Green Lot-Sizing*, Promotor(s): Prof.dr. A.P.M. Wagelmans, EPS-2013-291-LIS, http://repub.eur.nl/pub/41330

Rietveld, N., *Essays on the Intersection of Economics and Biology*, Promotor(s): Prof.dr. A.R. Thurik, Prof.dr. Ph.D. Koellinger, Prof.dr. P.J.F. Groenen, & Prof.dr. A. Hofman, EPS-2014-320-S&E, http://repub.eur.nl/pub/76907

Rijsenbilt, J.A., *CEO Narcissism: Measurement and Impact*, Promotor(s): Prof.dr. A.G.Z. Kemna & Prof.dr. H.R. Commandeur, EPS-2011-238-STR, http://repub.eur.nl/pub/23554

Roelofsen, E.M., *The Role of Analyst Conference Calls in Capital Markets*, Promotor(s): Prof.dr. G.M.H. Mertens & Prof.dr. L.G. van der Tas, EPS-2010-190-F&A, http://repub.eur.nl/pub/18013

Roza-van Vuren, M.W., *The Relationship between Offshoring Strategies and Firm Performance: Impact of innovation, absorptive capacity and firm size*, Promotor(s): Prof.dr. H.W. Volberda & Prof.dr.ing. F.A.J. van den Bosch, EPS-2011-214-STR, http://repub.eur.nl/pub/22155

Rubbaniy, G., *Investment Behaviour of Institutional Investors*, Promotor(s): Prof.dr. W.F.C. Verschoor, EPS-2013-284-F&A, http://repub.eur.nl/pub/40068

Schellekens, G.A.C., *Language Abstraction in Word of Mouth*, Promotor(s): Prof.dr.ir. A. Smidts, EPS-2010-218-MKT, http://repub.eur.nl/pub/21580

Shahzad, K., *Credit Rating Agencies, Financial Regulations and the Capital Markets*, Promotor(s): Prof.dr. G.M.H. Mertens, EPS-2013-283-F&A, http://repub.eur.nl/pub/39655

Sotgiu, F., *Not All Promotions are Made Equal: From the Effects of a Price War to Crosschain Cannibalization*, Promotor(s): Prof.dr. M.G. Dekimpe & Prof.dr.ir. B. Wierenga, EPS-2010-203-MKT, http://repub.eur.nl/pub/19714

Sousa, M.J.C. de, *Servant Leadership to the Test: New Perspectives and Insights*, Promotor(s): Prof.dr. D.L. van Knippenberg & Dr. D. van Dierendonck, EPS-2014-313-ORG, http://repub.eur.nl/pub/51537

Spliet, R., *Vehicle Routing with Uncertain Demand*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2013-293-LIS, http://repub.eur.nl/pub/41513

Srour, F.J., *Dissecting Drayage: An Examination of Structure, Information, and Control in Drayage Operations*, Promotor(s): Prof.dr. S.L. van de Velde, EPS-2010-186-LIS, http://repub.eur.nl/pub/18231

Staadt, J.L., *Leading Public Housing Organisation in a Problematic Situation: A Critical Soft Systems Methodology Approach*, Promotor(s): Prof.dr. S.J. Magala, EPS-2014-308-ORG, http://repub.eur.nl/pub/50712

Stallen, M., *Social Context Effects on Decision-Making: A Neurobiological Approach*, Promotor(s): Prof.dr.ir. A. Smidts, EPS-2013-285-MKT, http://repub.eur.nl/pub/39931

Tarakci, M., *Behavioral Strategy: Strategic Consensus, Power and Networks*, Promotor(s): Prof.dr. D.L. van Knippenberg & Prof.dr. P.J.F. Groenen, EPS-2013-280-ORG, http://repub.eur.nl/pub/39130

Teixeira de Vasconcelos, M., *Agency Costs, Firm Value, and Corporate Investment*, Promotor(s): Prof.dr. P.G.J. Roosenboom, EPS-2012-265-F&A, http://repub.eur.nl/pub/37265

Tempelaar, M.P., *Organizing for Ambidexterity: Studies on the pursuit of exploration and exploitation through differentiation, integration, contextual and individual attributes*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2010-191-STR, http://repub.eur.nl/pub/18457

Tiwari, V., *Transition Process and Performance in IT Outsourcing: Evidence from a Field Study and Laboratory Experiments*, Promotor(s): Prof.dr.ir. H.W.G.M. van Heck & Prof.mr.dr. P.H.M. Vervest, EPS-2010-201-LIS, http://repub.eur.nl/pub/19868

Troster, C., *Nationality Heterogeneity and Interpersonal Relationships at Work*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2011-233-ORG, http://repub.eur.nl/pub/23298

Tsekouras, D., *No Pain No Gain: The Beneficial Role of Consumer Effort in Decision-Making*, Promotor(s): Prof.dr.ir. B.G.C. Dellaert, EPS-2012-268-MKT, http://repub.eur.nl/pub/37542

Tuijl, E. van, *Upgrading across Organisational and Geographical Configurations*, Promotor(s): Prof.dr. L. van den Berg, EPS-2015-349-S&E, http://repub.eur.nl/pub/78224

Tuncdogan, A., *Decision Making and Behavioral Strategy: The Role of Regulatory Focus in Corporate Innovation Processes*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch, Prof.dr. H.W. Volberda, & Prof.dr. T.J.M. Mom, EPS-2014-334-S&E, http://repub.eur.nl/pub/76978

Tzioti, S., *Let Me Give You a Piece of Advice: Empirical Papers about Advice Taking in Marketing*, Promotor(s): Prof.dr. S.M.J. van Osselaer & Prof.dr.ir. B. Wierenga, EPS-2010-211-MKT, http://repub.eur.nl/pub/21149

Uijl, S. den, *The Emergence of De-facto Standards*, Promotor(s): Prof.dr. K. Blind, EPS-2014-328-LIS, http://repub.eur.nl/pub/77382

Vaccaro, I.G., *Management Innovation: Studies on the Role of Internal Change Agents*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch, Prof.dr. H.W. Volberda, & Prof.dr. J.J.P. Jansen, EPS-2010-212-STR, http://repub.eur.nl/pub/21150

Vagias, D., *Liquidity, Investors and International Capital Markets*, Promotor(s): Prof.dr. M.A. van Dijk, EPS-2013-294-F&A, http://repub.eur.nl/pub/41511

Veelenturf, L.P., *Disruption Management in Passenger Railways: Models for Timetable, Rolling Stock and Crew Rescheduling*, Promotor(s): Prof.dr. L.G. Kroon, EPS-2014-327-LIS, http://repub.eur.nl/pub/77155

Venus, M., *Demystifying Visionary Leadership: In search of the essence of effective vision communication*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2013-289-ORG, http://repub.eur.nl/pub/40079

Verheijen, H.J.J., *Vendor-Buyer Coordination in Supply Chains*, Promotor(s): Prof.dr.ir. J.A.E.E. van Nunen, EPS-2010-194-LIS, http://repub.eur.nl/pub/19594

Visser, V.A., *Leader Affect and Leadership Effectiveness:How leader affective displays influence follower outcomes*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2013-286-ORG, http://repub.eur.nl/pub/40076

Vlam, A.J., *Customer First? The Relationship between Advisors and Consumers of Financial Products*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS-2011-250-MKT, http://repub.eur.nl/pub/30585

Waard, E.J. de, *Engaging Environmental Turbulence: Organizational Determinants for Repetitive Quick and Adequate Responses*, Promotor(s): Prof.dr. H.W. Volberda & Prof.dr. J. Soeters, EPS-2010-189-STR, http://repub.eur.nl/pub/18012

Waltman, L., *Computational and Game-Theoretic Approaches for Modeling Bounded Rationality*, Promotor(s): Prof.dr.ir. R. Dekker & Prof.dr.ir. U. Kaymak, EPS-2011-248-LIS, http://repub.eur.nl/pub/26564

Wang, T., *Essays in Banking and Corporate Finance*, Promotor(s): Prof.dr. L. Norden & Prof.dr. P.G.J. Roosenboom, EPS-2015-352-F&A, http://repub.eur.nl/pub/78301

Wang, Y., *Information Content of Mutual Fund Portfolio Disclosure*, Promotor(s): Prof.dr. M.J.C.M. Verbeek, EPS-2011-242-F&A, http://repub.eur.nl/pub/26066

Wang, Y., *Corporate Reputation Management: Reaching Out to Financial Stakeholders*, Promotor(s): Prof.dr. C.B.M. van Riel, EPS-2013-271-ORG, http://repub.eur.nl/pub/38675

Weenen, T.C., *On the Origin and Development of the Medical Nutrition Industry*, Promotor(s): Prof.dr. H.R. Commandeur & Prof.dr. H.J.H.M. Claassen, EPS-2014-309-S&E, http://repub.eur.nl/pub/51134

Wolfswinkel, M., *Corporate Governance, Firm Risk and Shareholder Value*, Promotor(s): Prof.dr. A. de Jong, EPS-2013-277-F&A, http://repub.eur.nl/pub/39127

Wubben, M.J.J., *Social Functions of Emotions in Social Dilemmas*, Promotor(s): Prof.dr. D. de Cremer & Prof.dr. E. van Dijk, EPS-2010-187-ORG, http://repub.eur.nl/pub/18228

Xu, Y., *Empirical Essays on the Stock Returns, Risk Management, and Liquidity Creation of Banks*, Promotor(s): Prof.dr. M.J.C.M. Verbeek, EPS-2010-188-F&A, http://repub.eur.nl/pub/18125

Yang, S., *Information Aggregation Efficiency of Prediction Markets*, Promotor(s): Prof.dr.ir. H.W.G.M. van Heck, EPS-2014-323-LIS, http://repub.eur.nl/pub/77184

Zaerpour, N., *Efficient Management of Compact Storage Systems*, Promotor(s):

Prof.dr.ir. M.B.M. de Koster, EPS-2013-276-LIS, http://repub.eur.nl/pub/38766

Zhang, D., *Essays in Executive Compensation*, Promotor(s): Prof.dr. I. Dittmann, EPS-2012-261-F&A, http://repub.eur.nl/pub/32344

Zhang, X., *Scheduling with Time Lags*, Promotor(s): Prof.dr. S.L. van de Velde, EPS-2010-206-LIS, http://repub.eur.nl/pub/19928

Zhou, H., *Knowledge, Entrepreneurship and Performance: Evidence from country-level and firm-level studies*, Promotor(s): Prof.dr. A.R. Thurik & Prof.dr. L.M. Uhlaner, EPS-2010-207-ORG, http://repub.eur.nl/pub/20634

Zwan, P.W. van der, *The Entrepreneurial Process: An International Analysis of Entry and Exit*, Promotor(s): Prof.dr. A.R. Thurik & Prof.dr. P.J.F. Groenen, EPS-2011-234-ORG, http://repub.eur.nl/pub/23422

MARKET EFFICIENCY AND LIQUIDITY

The wealth of nations is determined by the efficient usage of real assets, such as its land, machinery, and knowledge. Financial assets merely represent claims on these real assets.

Nevertheless, financial markets serve many important roles: they allow to optimize the reward to risk ratio, to shift consumption over time, can contain important information of aggregate investor beliefs, and can help to shift scarce resources to its optimal usage.

But the efficacy of all of these roles depends on prices of financial assets reflecting the true value of these assets and how well the market facilitates trading these assets. In other words, the efficacy of these roles depend on the financial market being efficient and liquid.

Finance academics documented large time- and cross-sectional variation in market liquidity, but at the same time, in general, treated market efficiency as a static concept. This seems at odds, because both efficiency and liquidity are intimately related. Arguably markets are not efficient per se, but require trading against potential inefficiencies by informed investors, who's success depends on the ease at which they can trade (market liquidity) and on their available capital (funding liquidity).

The main theme of this thesis is to investigate the interaction between market efficiency and liquidity. In particular to document time- and cross-sectional variation in market efficiency, and whether individual stock efficiency co-moves with aggregate market efficiency; to investigate why inefficiencies arise and how trading against these inefficiencies affects market liquidity; and to provide a new measure for the probability of informed trading.

ERiM

The Erasmus Research Institute of Management (ERIM) is the Research School (Onderzoekschool) in the field of management of the Erasmus University Rotterdam. The founding participants of ERIM are the Rotterdam School of Management (RSM), and the Erasmus School of Economics (ESE). ERIM was founded in 1999 and is officially accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW). The research undertaken by ERIM is focused on the management of the firm in its environment, its intra- and interfirm relations, and its business processes in their interdependent connections.

The objective of ERIM is to carry out first rate research in management, and to offer an advanced doctoral programme in Research in Management. Within ERIM, over three hundred senior researchers and PhD candidates are active in the different research programmes. From a variety of academic backgrounds and expertises, the ERIM community is united in striving for excellence and working at the forefront of creating new business knowledge.

Erasmus Research Institute of Management - ERiM

ERIM PhD Series
Research in Management

Erasmus Research Institute of Management - ERiM
Rotterdam School of Management (RSM)
Erasmus School of Economics (ESE)
Erasmus University Rotterdam (EUR)
P.O. Box 1738, 3000 DR Rotterdam,
The Netherlands

Tel. +31 10 408 11 82
Fax +31 10 408 96 40
E-mail info@erim.eur.nl
Internet www.erim.eur.nl

Design & layout: B&T Ontwerp en advies (www.b-en-t.nl) Print: Haveka (www.haveka.nl)