Freie Universität Berlin

Three Essays in Financial Economics

Inaugural-Dissertation zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaft
des Fachbereichs Wirtschaftswissenschaft der Freien Universität Berlin

vorgelegt von
Diplom-Volkswirt und Master of Science
Simon Willi Jurkatis
aus Berlin
Mai, 2018

Erstgutachter:	Prof. Dr. Dieter Nautz
Zweitgutachter:	Prof. Georg Weizsäcker, PhD

Tag der Disputation: 10. Juli 2018

Eigenanteil der Leistung

Diese Dissertation besteht aus drei Arbeitspapieren, von denen zwei in Zusammenarbeit mit meinem Koautor Christopher Boortz entstanden sind. Die Ergebnisse des ersten Papiers sind Teil des Diskussionspapiers SFB 649 2014-029. Dieses wurde in Kombination mit weiteren empirischen Ergebnissen in Koautorenschaft mit Dieter Nautz und Stephanie Kremer verfasst und wurden unter anderem auf der internationalen Konferenz der European Financial Managemant Association 2013 präsentiert. Die Ergebnisse des zweiten Papiers wurden unter anderem bei den internationalen Konferenzen der Deutschen Gesellschaft für Finanzwirtschaft 2016, des Vereins für Socialpolitik 2016 und des SFB "Ökonomisches Risiko" 2016, sowie bei der Deutschen Bundesbank 2016 präsentiert. Die Ergebnisse des dritten Papiers sind in Eigenleistung entstanden und wurden bereits auf den internationalen Konferenzen CFE-CMStatistik 2017 und dem Royal Economic Society PhD Meeting 2017, sowie bei der Bank of England in 2018 präsentiert.

Der Eigenanteil an Konzeption, Durchführung und Berichtsabfassung lässt sich folgendermaßen zusammenfassen:

1. Boortz, Christopher und Jurkatis, Simon:
   "The Impact of Information Risk and Market Stress on Herding in Financial Markets",
   Eigenanteil 50%.

2. Boortz, Christopher und Jurkatis, Simon:
   "How to Measure Herding in Financial Markets",
   Eigenanteil 50%.

3. Jurkatis, Simon:
   "Inferring Trade Directions in Fast Markets",
   Eigenanteil 100%

Danksagung

Ich möchte mich zunächst bei meinem Betreuer und Erstgutachter, Prof. Dieter Nautz, für seine Unterstützung und die Möglichkeit an seinem Lehrstuhl zu arbeiten bedanken. Insbesondere möchte ich mich für die Freiheit bedanken, meine wissenschaftliche Ausrichtung gemäß meiner persönlichen Interessen gestaltet haben zu können, für die Unterstützung bei Teilnahmen an internationalen Konferenzen und nicht zu Letzt für seine außerordentlich großzügige Verfügbarkeit in allen Dingen.

Ebenfalls bedanke ich mich außerordentlich bei meinem Zweitgutachter und Betreuer, Prof. Georg Weizsäcker, welcher sich immer wieder viel Zeit für Besprechungen von Forschungsideen und -umsetzungen genommen hat. Insbesondere bin ich ihm für seine Unterstützung über das letzte Dreivierteljahr, welche dazu beitrug, dass ich auch in Zukunft weiter forschend tätig sein kann, sehr verbunden.

In diesem Zusammenhang gilt meine tiefe Dankbarkeit auch Dr. Puriya Abbassi und Prof. Bidisha Chakrabarty. Puriya ermöglichte mir nicht nur einen Forschungsaufenthalt an der Deutschen Bundesbank, aus welchem der empirische Teile zum hier vorliegenden zweiten Kapitel hervor ging, und war stets zur Verfügung für anregende Diskussionen, sondern unterstützte mich auch beträchtlich in der weiteren Planung meines akademischen Arbeitens. Prof. Chakrabarty danke ich für ihre wertvollen Kommentare und Anregungen zu meinem letzten Papier und dafür, dass sie sich ohne Weiteres als Referenz für meine akademischen Bewerbungen bereit gestellt hat.

Bedanken möchte ich mich auch beim Berlin Doctoral Program in Economics and Management Science für den Einstieg in das Promotionsstudium und die finanzielle Unterstützung in den ersten drei Jahren meiner Doktorarbeit.

Ich bedanke mich ebenfalls bei meinen Kollegen und Kolleginnen am Lehrstuhl, Fachbereich oder darüber hinaus für ihre fortwährende Hilfsbereitschaft, Diskussionsbereitschaft in Forschungsfragen und offenen Ohren in allen Dingen. Hier möchte ich auch Birgit Allgermissen für ihre Unterstützung und ihren Rat in allen administrativen Belangen danken. Außerdem bedanke ich mich bei meinen

Freunden und (entfernten) Kollegen Wolfgang Strehl und Dr. Moritz Augustin für die Perspektive über den Tellerrand.

Zu guter Letzt bedanke ich mich bei meinen Eltern, Barbara und Hagen, bei meinen Geschwistern, Lena und Jan, und bei Mihaela für deren Liebe und Unterstützung während aller Phasen dieser Promotion.

Berlin, 4. Mai 2018

Overview

This thesis consists of three chapters/papers. The first two are related to the literature on herd behavior in financial markets. The third chapter is on trade classification, a method to classify trades into the orders of liquidity demanders and providers, which is a necessary first step in many studies on financial and financial economics topics, including studies on herd behavior.

Herd behavior by investors can be a significant threat to the functioning of financial markets. The distorting effects of herding range from informational inefficiency to increased stock price volatility, or even bubbles and crashes. Consequently, there exists one the one hand a large theoretical literature that shows analytically how herding arises even in rational markets, and a large empirical literature on the other that tests for the presence of herd behavior in financial markets. It has been noted, however, that these two strands of the herding literature are largely disconnected. While herd models do not provide empirical testable hypotheses, empirical works do not rigorously tie their proposed measures to the theoretical concept of herding. This thesis, particularly the first and second chapter, contributes towards closing the gap between the theoretical and empirical herding literature.

The third chapter, while contributing to the empirical herding literature as well, is a more general contribution to the empirical toolkit of financial economists by proposing a new algorithm to classify transaction data into the orders of liquidity demanders and suppliers.

Knowing the trade direction of the liquidity demanding, impatient side of a trade is key to many financial market research topics. Measures of informed trading, price efficiency and market quality all depend on the trade direction of the liquidity demander. To link this topic to the previous chapters, herding models, for example, assume that the information about an asset's value is conveyed by the impatient trader and that subsequent traders, therefore, try to learn from the action of the impatient side of the transaction.

Yet, information on the trade direction of the impatient side of a trade is generally not available and the established methods to classify transactions into the orders of liquidity demanders and suppliers face certain difficulties in today's data environments due to the increased frequency of order submissions on financial exchanges. Hence, I propose a new algorithm that overcomes these difficulties and show its superiority over the established algorithms.

Chapter 1: The Impact of Information Risk and Market Stress on Herding in Financial Markets

Theoretical work on herd behavior is not only used to motivate many empirical studies on herding, but also to inform to what extent the results obtained from proxies of herd behavior are in fact compatible with the predictions of herding theory. The theoretical predictions, however, are not rigorously derived from the model, but instead loosely inferred. In fact, herding models are highly complex and non-linear and, thus, do not allow for a straightforward derivation of the effect of parameter changes on the frequency with which herding occurs, even for a single asset, let alone for an aggregate over a heterogeneous set of assets.

The first chapter of the thesis derives empirically test-able predictions on the effect of changes in information risk and market stress on herding intensity by simulating the herding model of Park and Sabourian (2011). Information risk, the probability to encounter an informed trader as the opposite party of a trade, is a key parameter in herding models and has well-known empirical proxies. Market stress, on the other hand, is a keyword in the empirical literature that is often attached to herding and, at the same time, can be naturally translated into the herding model.

To reflect the typical empirical situation in which herding would be measured and aggregated over a sample of heterogeneous assets, the model is simulated for various parameter combinations. We find that average buy and sell herding intensity increases with information risk. For market stress we find an asymmetric effect on buy and sell herding: Interestingly, *buy* herding is more pronounced in times of high market stress than the one of sell herding.

Chapter 2: How to Measure Herding in Financial Markets

Empirical measures of herding are usually measures of some form of correlated trading and, thus, do not fully reflect the theoretical notion of herding. In particular, while the theoretical literature makes a clear distinction between the deliberate imitation of the trading decisions of others ("true" herding) and trading simply on the same type of information ("spurious" herding), the empirical literature does not. Not explicitly accounting for spurious herding is a common criticism of empirical herding measures.

In this project, we propose a new, theory-founded herding measure that separates from the observed co-ordination of traders the unintended component that is due to traders holding the same prior beliefs, i.e. spurious herding. We show by means of simulations that the new measure accurately signals herding and contrarian behavior (the counter-part of herding where traders act against the crowd) in transaction data, while the most prominent measure of investor co-ordination, the measure proposed by Lakonishok, Shleifer, and Vishny (1992), severely fails to do so.

Chapter 3: Inferring Trade Directions in Fast Markets

Established methods to classify transactions into the orders of liquidity demanders and suppliers face certain difficulties in today's data environments. These method classify trades according to the proximity of the transaction price to the best bid and offer that were in effect at the time of the trade. Due to today's high frequency of order submissions and cancellations, however, it is not clear which bid and offer was indeed in effect at the time of the trade. The wrong choice of bid and offer quote reduces the accuracy of the classification, which, in turn, impacts the analysis based on the classification leading to erroneous inference and wrong conclusions.

In this paper, I propose a new algorithm that overcomes these difficulties. The most important innovation of the proposed algorithm is the use of prices and volume changes to make an informed search for the correct correspondence between a trade and its quotes. Using a dataset of stock market transactions that contains the information on the liquidity demanding and supplying side, I test the ability of the new algorithm and the alternatives commonly applied in the literature to uncover that information. Moreover, I impose various deficiencies on the data to simulate the characteristic problems of usual data records. Testing the different methods in these environments I find that the new algorithm clearly outperforms the established methods with misclassification rates being reduced by up to half. The increase in classification accuracy also translates into considerable improvements in the estimation of statistics of informed trading and market quality, namely the order imbalance, price impact, the effective and realized spread.

Zusammenfassung

Die vorliegende Arbeit besteht aus drei Kapiteln. Die ersten beiden Kapitel stehen im starken Bezug zur Literatur über Herdenverhalten. Das dritte behandelt das Thema der sogenannten "trade classification", einer Methode um Finanzmarkttransaktionen den jeweiligen liquiditätbereitstellenden und -nehmenden Ordern zuzuordnen. Diese Zuordnung ist ein notwendiger erster Schritt in vielen Studien über finanzwissenschaftliche oder finanz-ökonomische Themen.

Herdenverhalten von Investoren kann eine signifikante Bedrohung für das Funktionieren von Finanzmärkten darstellen. Die disruptiven Effekte reichen von Preisineffizienz, im Sinne der Funktion der Informationsaggregation durch Preise, bis hin zu erhöhter Preisvolatilität und gar Preisblasen und -einstürze. Konsequenterweise existiert auf der einen Seite eine ausgiebige, theoretische Literatur, die zeigt, dass Herdenverhalten selbst in komplett rationalen Märkten entstehen kann, und eine empirische Literatur, auf der anderen Seite, die auf Herdenverhalten auf Finanzmärkten testet.

Diese beiden Stränge der Literatur stehen jedoch in einem entkoppelten Verhältnis. Während theoretische Modelle wenige, empirisch überprüfbare Hypothesen bereit hält, sind empirische Messmethoden gleichermaßen nicht streng an das theoretische Konzept von Herdenverhalten gebunden. Die vorliegende Arbeit, insbesondere die ersten beiden Kapitel, trägt zum Zusammenbringen der theoretischen und empirischen Literatur über Herdenverhalten bei.

Das dritte Kapitel, wenn gleich es ebenfalls zu der empirischen Literatur über Herdenverhalten beiträgt, ist ein mehr allgemeiner Beitrag zum empirischen Handwerkszeug von Ökonomen. Im dritten Kapitel schlage ich einen neuen Algorithmus zum Klassifizieren von Transaktionsdaten in die jeweiligen Order von Bereitstellern und Nehmern von Liquidität vor. Der Liquiditätsbegriff bezieht sich dabei auf die Möglichkeit beispielsweise Aktien zu großen Mengen kaufen oder verkaufen zu können, ohne einen starken Einfluss auf den Preis der Aktie auszuüben.

Die Kenntnis der Handelsrichtung, also ob Käufer oder Verkäufer, des Liquiditätsnehmers ist grundlegend für viele Studien zu finanzwissenschaftlichen und

ökonomischen Themen, wie beispielsweise informiertes Handeln auf Finanzmärkten hin zu "Insider-Trading", Preiseffizienz und Marktqualität. Um dieses Kapitel mit den vorangegangenen zu verknüpfen, in Modellen zu Herdenverhalten, beispielsweise, wird generell angenommen, dass Informationen über den fundamentalen Wert eines Assets durch die Handlung (Kaufen oder Verkaufen) des Liquiditätsnehmers transportiert wird, sodass aufeinander folgende Händler versuchen, von den Handlungen der vorangegangenen Liquiditätsnehmern etwas über den fundamentalen Wert des Assets zu lernen.

Jedoch, eine Zuordnung von Transaktionen zu der Seite des Liquiditätsnehmers und -bereitstellers sind üblicherweise nicht von vorneherein in den Daten gegeben und etablierte Methoden, um diese Information aus den Daten zu filtern, sind heutzutage durch die erhöhte Aktivität an Finanzmärkten mit gewissen Schwierigkeiten konfrontiert, welche deren Klassifizierungsgüte beeinflusst. Daher schlage ich einen neuen Algorithmus vor, der diese Schwierigkeiten überwindet und zeige seine Überlegenheit gegenüber den etablierten Methoden auf.

Kapitel 1: The Impact of Information Risk and Market Stress on Herding in Financial Markets

Theoretische Arbeiten zu Herdenverhalten werden nicht nur genutzt, um empirische Arbeiten zu motivieren, sonder auch, um diese zu informieren zu welchem Ausmaß die Ergebnisse von empirischen Proxies von Herdenverhalten tatsächlich mit den Vorhersagen der theoretischen Literatur übereinstimmen. Die theoretischen Vorhersagen, jedoch, sind dabei nicht strikt aus einem bestimmten Modell hergeleitet, sonder entstammen eher groben Interpretationen. In der Tat sind Modelle von Herdenverhalten in der Regel nicht linear und so komplex, dass sie keine analytische Herleitung von bestimmten Effekten von Parameteränderungen auf das Herdenverhalten zulassen.

Das erste Kapitel dieser Arbeit leitet empirisch überprüfbare Hypothesen über den Effekt von Änderungen in "information risk" und "market stress" auf die Intensität von Herdenverhalten mittels Simulationen des Modells von Park and Sabourian (2011) her. Information risk bezeichnet das Risiko, eine Transaktion mit einem besser informiertem Gegenüber durchzuführen. Es ist ein elementarer Parameter in Modellen zu Herdenverhalten und hat wohl-bekannte empirische Proxies. Market stress bezeichnet Phasen von negativen ökonomischen Aussichten und erhöhter Unsicherheit. Market stress ist ein Schlüsselwort in der empirischen Literatur zu Herdenverhalten und hat gleichzeitig eine natürliche Übersetzung in theoretische Modelle.

Um die typische empirische Situation zu reflektieren, bei welcher Herdenverhalten über ein heterogenes Set von Aktien oder Ähnlichem gemessen und aggregiert werden würde, simulieren wir das Modell für eine große Anzahl an Parameterkombinationen. Wir finden das Herdenverhalten auf der Käufer, sowie auf der Verkäuferseite mit erhöhtem information risk zu nimmt. Der Effekt von market stress ist asymmetrisch für das Herdenverhalten von Käufern und Verkäufern: Interessanterweise ist der positive Effekt auf das Herdenverhalten von Käufern ausgeprägter als der für Verkäufer.

Kapitel 2: How to Measure Herding in Financial Markets
Empirische Maße von Herdenverhalten sind typischer Weise ein Form von Messung korrelierten Verhaltens der Marktteilnehmer. Als solche reflektieren sie nicht in Gänze das theoretische Konzept von Herdenverhalten. Insbesondere, während die theoretische Literatur eine Unterscheidung trifft zwischen der absichtlichen Imitation des Entscheidungen anderer ("wahres" Herdenverhalten) und dem gleichgerichteten Verhalten welches lediglich aus, beispielsweise, korrelierten Informationen zwischen Markteilnehmern herrührt (sogenanntes "suprious" Herdenverhalten), tut dies die empirische Literatur nicht. Dies ist eine häufige Kritik an der empirischen Literatur.

In diesem Kaptiel schlagen wir ein neues, Theorie-fundiertes Maß für Herdenverhalten vor, welches für die Koordinierung von Investoren aufgrund dessen gleichgerichteter Informationen kontrolliert. Wir zeigen mittels Simulationen, dass unser neues Maß präzise Herdenverhalten und sogenanntes "contrarian" Verhalten, welches den Gegensatz zu Herdenverhalten darstellt, also das Handeln entgegen der Herde, anzeigt. Das prominenteste Maß für Herdenverhalten von Lakonishok, Shleifer, and Vishny (1992) hingegen verfehlt diese Aufgabe.

Kapitel 3: Inferring Trade Directions in Fast Markets
Etablierte Methoden zum Zuordnen von Transaktionen in die Order der Liquiditätsnehmer und -bereitsteller sind heutzutage mit gewissen Schwierigkeiten konfrontiert. Diese Methoden klassifizieren Transaktionen auf Basis der Nähe des Transaktionspreises zu den gegebenen Kauf- und Verkaufkursen zur Zeit der Transaktion. Aufgrund der hohen Frequenz mit der Kauf- und Verkauforder heute in den Markt gestellt werden, ist es jedoch schwer zu bestimmen welcher Verkauf- und Kaufkurs tatsächlich zur Zeit der Transaktion bestand hatte. Eine falsche Zuordnung der Transaktion zu den Kursen reduziert die Güte der Klassifizierungsmethoden und, in der Konsequenz, beeinflusst die auf ihnen basierte Analyse mit dem Risiko von fehlerhaften Schlussfolgerungen.

In diesem schlage ich eine neue Methode vor, die diese Schwierigkeiten überwindet. Dabei ist die entscheidende Innovation das hinzuziehen von Preis- und Volumen-Informationen, um eine informierte Zuordnung der Transaktionen zu den jeweiligen Kursen treffen zu können. Ich nutze einen Datensatz von Aktienmarkttransaktionen, welcher die Informationen über die Liquiditätsnehmer und -bereitsteller bereits enthält, um die Güte des neuen Algorithmus sowie der etablierten Methoden zu bestimmen. Ich finde, dass der neue Algorithmus die etablierten Methoden deutlich übertrifft: Die Fehlklassifizierungsrate wird mit unter halbiert. Dieser Vorsprung in der Güte der Klassifizierung von Transaktionen übersetzt sich außerdem in verbesserte Messungen von informiertem Handeln und Marktqualität.

Contents

List of Figures

List of Tables

Paper 1

The Impact of Information Risk and Market Stress on Herding in Financial Markets*

1.1 Introduction

Herd behavior by investors can be a significant threat to the functioning of financial markets. The distorting effects of herding range from informational inefficiency to increased stock price volatility, or even bubbles and crashes as demonstrated in many theoretical works (see, e.g., Lux, 1995; Avery and Zemsky, 1998; Lee, 1998; Park and Sabourian, 2011).

Theoretical work on herd behavior, however, is not only used to motivate empirical studies on herding, but also to inform to what extent the results obtained from empirical measures are in fact compatible with the predictions of herding theory. This is done because empirical measures of herding are typically only proxies of the type of herding that is discussed in the theoretical literature (see Chapter 2 of this thesis). Hence, theoretical predictions are used to inform empirical results on which type of herding is detected (Wermers, 1999; Sias, 2004; Patterson and Sharma, 2010).[1] Yet, theoretical predictions are not rigorously derived from a particular model, but instead loosely inferred. In fact, herding models are highly complex and non-linear and, thus, do not allow for a straightforward derivation of the effect of parameter changes on the frequency with which herding occurs, even for a single asset, let alone for an aggregate over a heterogeneous set of assets as it is usually the objective in empirical applications.

Therefore, in this paper we show how theory-based predictions can be derived from a particular herding model by means of numerical simulations. Specifically,

*This paper was written in collaboration with my co-author Christopher Boortz. The results in combination with an empirical analysis are also presented in Boortz, Kremer, Jurkatis, and Nautz (2014)

[1] The type of herding may refer to spurious versus intentional (Bikhchandani and Sharma, 2001), or to the cause of herding, e.g., investigative herding versus herding due to information externalities (Froot et al., 1992; Bikhchandani et al., 1992).

we focus on the effect of information risk, i.e. the probability to trade with a counter-party who holds private information, and market stress, defined as situations where investors are both pessimistic and uncertain about a stock's value, on herding intensity over a cross-section of heterogeneous stocks. We focus on information risk as it is a key parameter in financial market herding models and, at the same time, has well-known and established empirical proxies (e.g. Easley et al., 2002). Similarly, market stress is central to the empirical literature, while it has, as we will show, a natural translation into a herding model.

Building on Glosten and Milgrom (1985) and Easley and O'Hara (1987), the literature on information risk deals with estimating the information content of trades (see, e.g., Hasbrouck, 1991; Easley et al., 1996b, 1997). The effects of information risk on herding intensity, however, are rarely considered.[2] While the probability of informed trading is a key parameter in financial market herd models (Avery and Zemsky, 1998; Park and Sabourian, 2011), to date these models have not been exploited to discover the impact of information risk on herding intensity. This is surprising, since the effects of information risk on herding intensity are far from obvious. On the one hand, an increase in information risk increases the average information content of an observed trade. As a consequence, traders update their beliefs more quickly and those investors that are susceptible to herding are more easily swayed to follow the crowd. On the other hand, increased information risk amplifies the market maker's adverse selection problem. Given the higher probability of trading at an informational disadvantage, the market maker quotes larger bid-ask spreads which tends to prevent potential herders from trading. Understanding which of these counteracting effects dominates could facilitate the detection of herds.

The impact of market stress on herd behavior has not been analyzed by the theoretical herding literature, either. Typically, herd models focus on the reverse relationship. For example, Park and Sabourian (2011) demonstrate that price paths tend to be more volatile in the presence of herd behavior. Agent based models proposed by, for example, Lux (1998) and Eguiluz and Zimmermann (2000) show that herd behavior contributes to fat tails and excess volatility in asset returns. While the models of, e.g., Avery and Zemsky (1998) and Park and Sabourian (2011) show that uncertainty of some specific form has to exist for herding to be possible, their models do not imply that *more* uncertainty actually leads to *more* herding. If such a relationship exists, it threatens to create vicious cycles of economic downturns and high volatility regimes due to herding and market stress reinforcing each other.

[2] An exception is Zhou and Lai (2009) who provide evidence that herding is positively related to information risk measured by probability of informed trading (PIN).

The prevalent unidirectional focus of the theoretical literature is particularly puzzling in light of the mixed evidence regarding the impact of market stress on herding intensity. Chiang and Zheng (2010) and Christie and Huang (1995) find that herding increases during times of market stress, whereas Kremer and Nautz (2013a,b) find that herding in the German stock market slightly decreased during the recent financial crisis, which is similar to the results of Hwang and Salmon (2004) for herding intensity during the Asian and the Russian crisis in the 1990s.

We base our theoretical analysis on the financial market herd model of Park and Sabourian (2011), which can be viewed as a generalization of the seminal work of Avery and Zemsky (1998).[3] One important extension is the broader set of different information structures that allows a differentiated discussion of how information externalities may contribute to herd behavior under various market conditions including scenarios of high and low market stress. Relating investor herding to the shape of the information structure Park and Sabourian (2011) identify more explicitly those situations in which the potential for herding is high. Consequently, the Park and Sabourian (2011) framework is more appropriate for finding and explaining high degrees of herding. In fact, experimental evidence suggests that the Avery and Zemsky (1998) framework allows for only little or no herd behavior (Cipriani and Guarino, 2009).[4] In contrast, experiments based on the Park and Sabourian (2011) model find that herding in financial markets can be substantial (Park and Sgroi, 2012, 2016).

In Park and Sabourian (2011), herding is triggered by information externalities that an investment decision by one agent imposes on subsequent agents' expectations about the asset value, similarly to the early observational learning literature (e.g. Bikhchandani et al., 1992; Banerjee, 1992).[5] Therefore, this model is a natural

[3] Similar to the bulk of the theoretical literature, both models define herd behavior as a switch in an agent's opinion toward that of the crowd, see Brunnermeier (2001). As herders make their decision irrespective of their private information, herd behavior is informationally inefficient and thus has the potential to distort prices and destabilize markets.

[4] Avery and Zemsky (1998) includes different model setups. The most basic setup extends the traditional herd model of Bikhchandani et al. (1992) by a price mechanism that prevents herd behavior. Prominent experimental tests of the Avery and Zemsky (1998) framework, Drehmann et al. (2005) and Cipriani and Guarino (2005), focus on this setup and confirm the theoretical prediction of no herding. Cipriani and Guarino (2009), on the other hand, focus on one of the more complex setups in which herd behavior is predicted, but again find only little evidence of it.

[5] Alternative drivers for herd behavior include reputational concerns as well as investigative herding. Reputational herd models modify the agents' objective functions such that their decisions are affected by positive externalities from a good reputation (see, e.g., Scharfstein and Stein, 1990; Graham, 1999; Dasgupta et al., 2011). Investigative herd models examine conditions under which investors may choose to base their decisions on the same information resulting in correlated trading behavior (see, e.g., Froot et al., 1992; Hirshleifer et al., 1994). For a survey of the early herding literature see Devenow and Welch (1996). For an in-depth discussion of how

candidate for investigating the impact of information risk on herding intensity.[6]

The history dependence of trading decisions in financial market herd models drastically impedes the derivation of analytical results on herding intensity. This may explain why these models have not yet been exploited to make empirically testable predictions on the impact of information risk and market stress. Moreover, standard empirical herding measures, including the ones proposed by Lakonishok et al. (1992) and Sias (2004), examine herding intensity on an *aggregate* level. Consequently, empirical testability of our theory-guided hypotheses requires that we analyze herding intensity aggregated over investor groups, time periods, and heterogeneous stocks. This further complicates the derivation of analytical results.

We circumvent these problems by simulating the Park and Sabourian (2011) model for more than 13,000 different parameterizations that broadly cover the theoretical parameter space, generating about 2.6 billion trades for analysis. We obtain two testable hypotheses on the model-based measure of aggregate herding intensity. First, an increase in information risk should result in a symmetric increase of buy and sell herding intensity. Second, high market stress should be found to have an asymmetric effect on herding intensity: while buy herding is predicted to surge during crisis periods, the simulation results suggest that sell herding intensity increases only moderately.

The remainder of this paper is structured as follows. In Section 1.2 we review the model of Park and Sabourian (2011). In Section 1.3 we define information risk as well as market stress and provide an initial qualitative assessment of their effect on herding intensity. Section 1.4 formalizes the concept of aggregate herding intensity. It subsequently introduces the simulation setup and derives testable hypotheses regarding the role of information risk and market stress for aggregate herding intensity. Section 1.5 summarizes the results.

1.2 A Model of Investor Herding

This section reviews the herding model of Park and Sabourian (2011) and highlights conceptual additions and modifications that are relevant to our application. Moreover, it formalizes the notion of herding intensity.

the herding literature ties into the observational learning literature see Vives (1996).

[6]Other financial market herd models such as Lee (1998), Chari and Kehoe (2004), and Cipriani and Guarino (2008), investigate how investor herding is related to transaction costs, endogenous timing of trading decisions, and informational spillovers between different assets, respectively.

1.2.1 The Model Setup

Park and Sabourian (2011) consider a sequential trading model à la Glosten and Milgrom (1985), consisting of a single asset, both informed and noise traders, and a market maker. The model assumes rational expectations and common knowledge of its structure.

The Asset: There is a single risky asset with unknown fundamental value $V \in \{V_1, V_2, V_3\}$, where $V_1 < V_2 < V_3$. Without loss of generality, let $V_1 = 0$, $V_2 = 1$ and $V_3 = 2$. The prior distribution $0 < P(V = V_j) < 1$ for $j = 1, 2, 3$ determines the degree of *public uncertainty* about the asset's true value, $\text{Var}(V)$, before trading has started. The asset is traded over T consecutive points in time.

The Traders: Traders arrive in the market one at a time in a random exogenous order and decide to buy, sell or not to trade one unit of the asset at the quoted bid and ask prices. Traders are either informed traders or noise traders. The fraction of informed traders is denoted by μ. Informed traders base their decision to buy, sell or not to trade on their expectations regarding the asset's true value.

Publicly available information consists of the history of trades $H_t := \{(a_1, p_1), ..., (a_{t-1}, p_{t-1})\}$, where a_i is the action of a trader in period i and p_i the price at which the trader's action is executed, and the risky asset's prior distribution $P(V)$.

In addition to public information, informed traders base their asset valuation on a private signal $S \in \{S_1, S_2, S_3\}$ regarding the true value of the asset. They buy (sell) one unit of the asset if their expected value of the asset $E[V \mid S, H_t]$ is strictly greater (smaller) than the ask (bid) price quoted by the market maker. Otherwise, informed traders choose not to trade. In contrast to informed traders, noise traders trade randomly, that is, they decide to buy, sell or not to trade with equal probability of $1/3$. p_t denotes the price at which the asset is traded in period t.

The Private Signal: The distribution of the private signals S_1, S_2, S_3 is conditional on the true value of the asset. Denote the conditional signal matrix by $P(S = S_i \mid V = V_j) = (p^{ij})_{i,j=1,2,3}$. For each column j, the matrix is leftstochastic, i.e. $\sum_{i=1}^{3} p^{ij} = 1$. For each row i, $\sum_{j=1}^{3} p^{ij}$ is the likelihood that an informed trader receives the signal S_i. An informed trader's behavior is critically dependent on the shape of her private signal. Specifically, Park and Sabourian (2011) define a signal S_i to be

- monotonically decreasing iff $p^{i1} > p^{i2} > p^{i3}$,

- monotonically increasing iff $p^{i1} < p^{i2} < p^{i3}$,
- U-shaped iff $p^{i1} > p^{i2}$ and $p^{i2} < p^{i3}$.

Traders with monotone signals are confident about the asset's true value and rarely change their trading decision. That is, an optimistic trader with an increasing signal will only buy or hold, whereas a pessimistic trader with a decreasing signal will only sell or hold.

In contrast, traders with U-shaped signals face a high degree of uncertainty and may decide to buy, sell or hold. U-shaped traders are more easily swayed to change their initial trading decision as they observe trade histories H_t with a strong accumulation of traders on one side of the market. In fact, Park and Sabourian (2011) show that a U-shaped signal is a necessary condition for herding.

Park and Sabourian (2011) also introduce hill-shaped signals which are necessary for contrarian behavior. Since contrarian behavior is self-defeating, its destabilizing effects are limited and thus only of secondary importance for financial markets. Consequently, we exclude hill-shaped signals from our analysis.

In the following, we assume that S_1 is monotone decreasing, S_2 is U-shaped and S_3 is monotone increasing. The conditional private signal distribution $P(S \mid V)$ determines the degree of information asymmetry between market maker and informed traders. The less noisy the signal, the higher the informational advantage of the informed traders.

The Market Maker: Trading takes place in interaction with a market maker who quotes a bid and an ask price. The market maker has access only to public information and is subject to perfect competition such that he makes zero-expected profit. Accordingly, he sets the ask (bid) price equal to his expected value of the asset given a buy (sell) order and the public information. Formally, he sets $ask_t = E[V|H_t \cup \{a_t = buy\}]$ and $bid_t = E[V|H_t \cup \{a_t = sell\}]$.

1.2.2 Herding Intensity

Park and Sabourian (2011) describe herding as a "history-induced switch of opinion in the direction of the crowd" (p. 985). Thus, only informed traders can herd. More precisely, a herding trade is defined as follows:

Definition 1.1 (Herding).
Let b_t (s_t) be the number of buys (sells) observed until period t. An informed trader with signal S ***buy herds*** *in t at history H_t if the following three conditions hold:*

(BH1) $E[V|S] < E[V]$, *i.e. an informed trader with signal S does not buy initially and is more pessimistic regarding the asset's true value than is the market maker.*

(BH2) $E[V|S,H_t] > ask_t$, *i.e. an informed trader with signal S buys in t.*

(BH3) $b_t > s_t$, *i.e. the history of trades contains more buys than sells: the crowd buys.*

Analogously, an informed trader with signal S ***sell herds*** *in period t at history H_t if and only if (SH1) $E[V|S] > E[V]$, (SH2) $E[V|S,H_t] < bid_t$, and (SH3) $b_t < s_t$ hold simultaneously.*

Note that (BH1) and (SH1) imply that either buy or sell herding is possible for a given model parameterization. Our definition of herding is less restrictive than the one used in Park and Sabourian (2011), who, for example, define buy herding as an extreme switch from selling initially to buying. In our definition, buy herding also includes switches from *holding* to buying, provided that the trader leans toward selling initially (see (BH1) and (BH2) in Definition 1.1).[7] From an empirical perspective, including switches from holding to selling or buying is important as these actions may drive amplified stock price movements.

(BH3) and (SH3) also differ slightly from Park and Sabourian (2011) where, for example, buy herding requires $E[V|H_t] > E[V]$. This condition is based on the idea that prices rise when there are more buys than sells. However, this only holds if the prior distribution of the risky asset $P(V)$ is symmetric around the middle state V_2, i.e. $P(V_1) = P(V_3)$.[8] For asymmetric $P(V)$ it is possible that even though a history H_t contains more buys than sells, the price of the asset goes down (i.e. $E[V|H_t] < E[V]$). From an empirical perspective, asymmetric prior distributions $P(V)$ should not be ruled out. Therefore, we modify the herding definition to ensure that a herder always *follows the crowd*.

The above definition enables us to decide whether or not a particular trade by a *single* investor at a specific point in time is a herd trade. In contrast, empirical herding measures are based on a number of trades by different investors observed over a certain time interval, see, e.g., Lakonishok et al. (1992) and Sias (2004). Since we aim to derive theory-based predictions on herd behavior that can be tested empirically, we need to aggregate herding in the model over time as well

[7] According to Park and Sabourian (2011), such an extension of the herding definition is theoretically legitimate. They focus on the stricter version to be consistent with earlier theoretical work on herding.

[8] Note that Park and Sabourian (2011) assume symmetry of the risky asset's prior distribution throughout their paper (see Park and Sabourian, 2011, p. 980).

as over investors. We aggregate over time by considering all relevant trades from $t = 1, \ldots, T$. We aggregate over investors by calculating herding intensity for the whole group of informed traders. Therefore, we define *herding intensity (HI)* as the share of herding trades in the total number of informed trades.

Definition 1.2 (Herding Intensity).
Let b_T^{in} and s_T^{in} be the number of buys and sells of informed traders observed until period T, i.e. during the entire time interval under consideration. Let b_T^h and s_T^h denote the corresponding number of buy and sell herding trades. Then,

$$\text{Buy herding intensity (BHI)} =: \frac{b_T^h}{b_T^{in} + s_T^{in}}$$
$$\text{Sell herding intensity (SHI)} =: \frac{s_T^h}{b_T^{in} + s_T^{in}}$$

Standard empirical herding measures including those of Lakonishok et al. (1992) and Sias (2004) are calculated using only buys and sells. To be consistent with empirical herding measures, we exclude holds when calculating the number of informed trades in the definition of theoretical herding intensity.

1.3 Translation of Information Risk and Market Stress into the Model

This section shows how the concepts of information risk and market stress are translated into the Park and Sabourian (2011) model. It also provides a qualitative assessment how each concept impacts herding intensity.

1.3.1 Information Risk

In Easley et al. (1996a), information risk is the probability that a trade is executed by an informed trader. Hence, information risk coincides with the parameter μ, the fraction of informed traders, in the Park and Sabourian (2011) model.

From a theoretical perspective, the effect of changes in μ on herding intensity is ambiguous. On the one hand, herding may increase with information risk because a higher μ implies that there are more potential herders (U-shaped traders) in the market. Due to the self-enforcing nature of herd behavior a higher μ contributes to longer-lasting herds and, hence, stronger herding intensity. Moreover, a higher fraction of informed traders implies that the average information content of a single trade increases. As a consequence, informed traders update their beliefs more quickly and those traders that are susceptible to herd behavior are more easily swayed to change from buying to selling and vice versa.

On the other hand, a rise in μ may also reduce herding intensity. Since the average information content per trade increases in μ, herds tend to break up more quickly as traders stop herding after observing few trades on the opposite side of the market. Higher information risk further amplifies the market maker's adverse selection problem. Given the higher probability of trading at an informational disadvantage, the market maker quotes larger bid-ask spreads in order to avoid losses. The larger spread, in turn, requires potential herders to observe much stronger accumulation of traders on one side of the market before they alter their trading decision.

1.3.2 Market Stress

Times of high market stress and crisis periods are typically understood as situations where investors are confronted with a deteriorating economic outlook and increased uncertainty about stock values, compare e.g. Schwert (2011).

A negative economic outlook in the Park and Sabourian (2011) model is captured by low expectations regarding the asset's true value $E[V]$. A low $E[V]$ not only describes a deteriorated outlook by the public but also a high degree of pessimism among informed traders. First, lower public expectations $E[V]$ result in lower private expectations $E[V|S]$ for all informed traders. Second, there tend to be more decreasing signals (pessimists) among informed traders as well as fewer increasing signals (optimists) for low $E[V]$ than for high $E[V]$.

Uncertainty in the Park and Sabourian (2011) can be sorted into two types: public uncertainty and informed trader uncertainty. *Public uncertainty* is given by the variance of the risky asset $\text{Var}(V)$. *Informed trader uncertainty* (IU) is measured by the probabilities that informed traders receive a U-shaped signal conditional on V_j, $j = 1, 2, 3$: $\text{IU} := \sum_{j=1}^{3} p^{2j}$. The higher IU, the more traders there are in the market with U-shaped signals and, hence, the higher the uncertainty among informed traders.[9] In light of the recent financial crisis, we are particularly interested in comparing herding intensity in times of high market stress with the herding intensity predicted for more optimistic periods.

The overall effect of market stress on herding intensity is not obvious and

[9] Note that an increase in $\text{Var}(V)$ may reduce the number of U-shaped traders in the market. This effect is not necessarily offset by an increase in IU. One could circumvent this issue by additionally imposing that the total probability that an informed trader receives a U-shaped signal $P(S_2) = \sum_{j=1}^{3} p^{2j} P(V = V_j)$ must also be high in times of market stress. Since this does not affect the results of our simulation, we choose not to complicate the model by adding this characteristic to the uncertainty definition.

crucially depends on model parameterization. Particularly, buy and sell herding intensity may react differently to changes in market stress. Consider, for example, an increase in market stress due to a decrease in $E[V]$. More specifically, assume a shift of probability mass from V_3 to lower values.

First, if, for a given model parameterization, buy herding is possible (and hence sell herding is impossible), a marginal reduction in $P(V_3)$ would result in a *decrease* in buy herding intensity, whereas sell herding intensity would remain constant at 0. Similarly, if sell herding is possible for a given model parameterization (and buy herding impossible), a marginal reduction in $P(V_3)$ would result in an *increase* in sell herding intensity while buy herding intensity would remain unaffected. This converse effect on buy and sell herding intensity is due to the fact that a reduction in $P(V_3)$ diminishes the probability of buy-dominated trade histories and increases the probability of sell-dominated histories. Hence, potential sell (buy) herders are more (less) likely to be confronted with a trade history that sways them into herding.

Second, if the U-shaped signal is positively biased, i.e. $P(S_2|V_1) < P(S_2|V_3)$, a reduction of $P(V_3)$ diminishes the number of U-shaped traders in the market and, hence, tends to decrease buy as well as sell herding intensity. Finally, for a whole range of model parameterizations, a lower $E[V]$ may even contribute to an increase in buy herding intensity and a decrease in sell herding intensity. Since a lower $E[V]$ implies that more informed traders are initially inclined to sell, the number of potential sell herders declines. Correspondingly, buy herding becomes more likely.

These complex and partly counteracting effects in conjunction with the history-dependent updating of beliefs lead to a low analytical tractability of herding intensity in the Park and Sabourian (2011) model.[10] This particularly applies to the empirically relevant case where herding intensity is considered as an average over a set of stocks with heterogeneous characteristics. In the following, therefore, empirically testable predictions about the effects of information risk and market stress on average herding intensity are derived by simulating the model over a broad set of model parameterizations.

[10]The Appendix to this paper makes this point very explicit.

1.4 Simulation of the Herd Model for a Heterogeneous Stock Index

1.4.1 Average Herding Intensity

Empirical studies on herd behavior typically derive results for herding intensity as an average for a large set of stocks and over certain time intervals. The stocks under consideration are likely to differ in their characteristics implying that each stock is described by a distinct parameterization for the fraction of informed traders, the prior distribution of the asset, and the distribution of the private signals. In accordance with the empirical literature, we are particularly interested in herding intensity defined as an *average* over a broad range of model parameterizations that reflects the heterogeneity in stock market indices. Specifically, we define average herding intensity as follows:

Definition 1.3 (**Average Herding Intensity**).
For a given set of model parameterizations $\mathcal{I}$ and length T of the trading period, average buy herding intensity is defined as

$$\overline{BHI} = \frac{\sum_{i \in \mathcal{I}} w_i BHI_i}{\sum_{i \in \mathcal{I}} w_i},$$

where BHI_i stands for the buy herding intensity obtained for model parameterization i and the weights $w_i = b^{in}_{T,i} + s^{in}_{T,i}$ correspond to the number of informed trades observed for that parameterization.
The definition for average sell herding intensity $\overline{SHI}$ follows analogously.

Weights w_i ensure that average herding intensity is not biased upward by simulation outcomes with a low number of informed trades.[11]

1.4.2 The Simulation Setup

We choose μ, the fraction of informed traders, from

$$\mathcal{M} = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}.$$

[11] Consider, for example, a situation where we observe a herding intensity of 0.5 as 2 out of 4 informed trades are herd trades. Now assume that for another simulation the herding intensity is 0, as 0 out of 16 informed trades are herd trades. In this case, the *unweighted* average of simulated herding intensities would be 0.25, which overestimates herding intensity as only 2 out of overall 20 trades were herd trades.

That is, we simulate the model for $|\mathcal{M}| = 9$ different levels of information risk.[12]

The prior distribution of the risky asset $P(V)$ is chosen from

$$\mathcal{P} = \{P(V) \in \{0.1, 0.2, \dots, 0.9\}^3 : \sum_{i=1}^{3} P(V_i) = 1\}.$$

Since we impose that V takes each value $V_1 = 0, V_2 = 1, V_3 = 2$ with positive probability, $P(V_i)$ cannot be 0.9, which gives us $|\mathcal{P}| = 36$ different prior distributions.

The conditional signal distribution $P(S|V) = (p^{ij})_{i,j=1,2,3}$ has to be chosen from the space of leftstochastic 3-by-3 matrices. As before, we discretize this space by imposing a grid ranging from 0.1 to 0.9. All elements of $P(S|V)$ are positive, that is, all signals are noisy in the sense that an informed trader cannot with certainty rule out any of the three possible states for V. Following Park and Sabourian (2011), there are always optimists ($p^{31} < p^{32} < p^{33}$), pessimists ($p^{11} > p^{12} > p^{13}$), and U-shaped traders ($p^{21} > p^{22}, p^{22} < p^{23}$) in the market. Finally, informed traders tend to be well-informed, that is, if the bad state $V = V_1$ comes true, most of the informed traders are pessimistic and only few are optimistic ($p^{11} > p^{21} > p^{31}$) and vice versa for $V = V_3$ ($p^{13} < p^{23} < p^{33}$). This implies that the set of simulated signal structures ($\mathcal{C}$) can be summarized as follows:

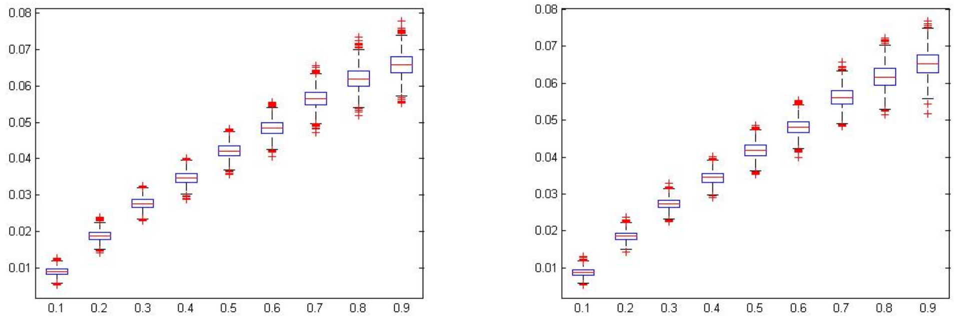
$$\begin{aligned}\mathcal{C} = \{&P(S|V) = (p^{ij})_{i,j=1,2,3} \text{ leftstochastic} : \; p^{ij} \in \{0.1, 0.2, \dots, 0.9\},\\ &p^{11} > p^{21} > p^{31},\; p^{13} < p^{23} < p^{33},\\ &p^{11} > p^{12} > p^{13},\; p^{31} < p^{32} < p^{33},\; p^{21} > p^{22},\; p^{22} < p^{23}\},\end{aligned}$$

which leads to $|\mathcal{C}| = 41$ different signal structures used in the simulation.

Considering all combinations, one obtains the simulation set $\Omega := \mathcal{M} \times \mathcal{P} \times \mathcal{C}$, where $|\Omega| = 9 \cdot 36 \cdot 41 = 13{,}284$. Each element $\omega = (\mu, P(V), P(S|V)) \in \Omega$ describes the characteristics of a specific stock.Park and Sabourian (2011) derive upper bounds for μ that have to hold in order for herding to be possible. One can check that these upper bounds are never binding for $\omega \in \Omega$, i.e. in each of the following simulations, either sell or buy herding is possible (see Park and Sabourian (2011), pp. 991-992, 1011-1012). Each stock is traded over $T = 100$ points of time. For each stock, the simulation is repeated $2{,}000$ times, which produces more than 2.6 billion simulated trades for analysis.

[12] In the German stock market, for example, the share of institutional trading (which might be considered as a proxy for informed trading) for the sample period studied in Kremer and Nautz (2013a) ranges from 0.2 to 0.7.

Figure 1.1: Information risk and herding intensity



(a) Sell Herding **(b)** Buy Herding

Notes: $\overline{SHI}$ and $\overline{BHI}$ are plotted against information risk. On the ordinate we plot average herding intensity. Information risk μ is plotted along the horizontal. Average herding intensity is calculated as the weighted cross-sectional average for the simulated $\overline{SHI}$ and $\overline{BHI}$ of stocks contained in $\{\mu\} \times \mathcal{P} \times \mathcal{C}$. The weights correspond to the observed number of informed trades. The boxplots show the variation across 2,000 simulations of average herding intensity for a fixed level of information risk μ.

1.4.3 Simulation Results: Information Risk and Average Herding Intensity

To discover the impact of information risk on average herding intensity, we fix $\mu \in \mathcal{M}$ and calculate average herding intensity as the cross-sectional average over all parameterizations in $\{\mu\} \times \mathcal{P} \times \mathcal{C}$, where $|\{\mu\} \times \mathcal{P} \times \mathcal{C}| = 1 \cdot 36 \cdot 41 = 1,476$.

Figure 1.1 shows the comparative statics for average sell and buy herding intensity with respect to changes in information risk μ. The simulation results clearly indicate that $\overline{SHI}$ and $\overline{BHI}$ symmetrically increase with information risk. The boxplots demonstrate that the simulation results are very stable. The variation of average herding intensity for a given level of information risk is relatively low, whereas its increase is rather steep as μ goes up.[13] Only as μ approaches 1 do $\overline{SHI}$ and $\overline{BHI}$ level out and exhibit higher variations.

The model simulation shows that the increasing effects of a rise in information risk on herding intensity dominate the decreasing effects. Only as the share of informed traders surpasses 80%, does the adverse selection problem of the market maker begin to impair market liquidity severely enough that trading among the potential herders breaks down. The ambiguity of their signal prevents them from paying the high premiums now demanded by the market maker via large bid-ask

[13] This particularly applies to the empirically relevant range of $\mu \in [0.2, 0.7]$ studied in Kremer and Nautz (2013a).

Table 1.1: The effects of market stress on average herding intensity

	$\overline{SHI}$	$\overline{BHI}$
Low market stress	0.0351 (0.0029)	0.0306 (0.0020)
High market stress	0.0382 (0.0023)	0.0635 (0.0038)

Notes: This table reports the simulated average sell ($\overline{SHI}$) and buy herding intensity ($\overline{BHI}$) for stocks under high market stress and stocks under low market stress. Standard deviations are in parentheses. Welch's t-test reveals that $\overline{SHI}$ as well as $\overline{BHI}$ increase significantly during times of high market stress for usual significance levels. Out of the 13,284 simulated stocks, 1,368 classify as high market stress and 1,008 as low market stress. Average herding intensities are calculated as the weighted cross-sectional averages of the simulated SHI and BHI for stocks in each respective class. The figures in the table are the weighted average and the weighted standard deviation of 2,000 iid simulated outcomes of $\overline{SHI}$ and $\overline{BHI}$ under high and low market stress, respectively. For all calculations, the weights correspond to the observed number of informed trades.

spreads. We summarize the simulation-based insight from Figure 1.1 as follows:

Hypothesis 1.1 (Information Risk and Herding Intensity).
Average sell and buy herding intensity increase in information risk.

1.4.4 Simulation Results: Market Stress and Average Herding Intensity

For the analysis of the effects of market stress we define two distinct classes of stocks and compare the average herding intensity of each. The first class comprises of all stocks that have high market stress characteristics, the second class includes all stocks that show low market stress characteristics. In line with the definition of market stress developed in Section 1.3.2, a simulated stock $\omega \in \Omega$ is subject to high market stress if it exhibits both, above-average uncertainty and below average $E[V]$. Correspondingly, low market stress stocks are defined by below-average uncertainty and above-average $E[V]$. The averages are the respective medians of the simulated model parameterizations.[14] We compare the cross-sectional average $\overline{SHI}$ and $\overline{BHI}$ over all high market stress stocks with the $\overline{SHI}$ and $\overline{BHI}$ obtained for all low market stress stocks.

The simulation results for the impact of market stress on average sell and buy herding intensity are shown in Table 1.1. As expected, both sell and buy herding

[14]Specifically, we obtain the median degree of pessimism (public uncertainty) by calculating $E[V]$ ($\mathrm{Var}(V)$) for each of the 36 simulated prior distributions $P(V) \in \mathcal{P}$ and then determine their median. Correspondingly, we calculate the median informed uncertainty over the set of simulated signal structures $\mathcal{C}$.

Table 1.2: The effects of uncertainty on average herding intensity

	$\overline{SHI}$	$\overline{BHI}$
Low uncertainty	0.0373 (0.0018)	0.0340 (0.0016)
High uncertainty	0.0557 (0.0022)	0.0555 (0.0022)

Notes: This table reports the simulated $\overline{SHI}$ and $\overline{BHI}$ for stocks with high and low uncertainty respectively. Standard deviations are in parentheses. Welch's t-test reveals that $\overline{SHI}$ as well as $\overline{BHI}$ increase significantly during times of high uncertainty for usual significance levels. Out of the 13,284 simulated stocks, 3,078 exhibit high and, 2,268 low, uncertainty. Average herding intensities are calculated as the weighted cross-sectional averages of the simulated SHI and BHI for stocks in each respective class. The figures in the table are the weighted average and the weighted standard deviation of 2,000 iid simulated outcomes of $\overline{SHI}$ and $\overline{BHI}$ under high and low uncertainty, respectively. For all calculations, the weights correspond to the observed number of informed trades.

are more pronounced during times of high market stress. Interestingly, however, the rise in buy herding intensity is greater than that of sell herding intensity. This puzzling asymmetry can be explained by disentangling the effects of an increase in uncertainty and pessimism.

Table 1.2 shows that $\overline{SHI}$ and $\overline{BHI}$ symmetrically increase with uncertainty. High public uncertainty is associated with lower prior probabilities for the middle state of the risky asset. Since informed traders receiving U-shaped signals discount the probability for the middle state anyway, high public uncertainty amplifies their tendency to form strong beliefs that only the extreme states of the risky asset can be true. As they rule out one of the extreme states based on the observed trading history, they quickly alter their trading decisions toward that of the crowd. This effect is intensified if private uncertainty is also high since such leads to a larger share of U-shaped traders. Since this argument applies equally to sell and buy herding, the increasing effect of uncertainty on herding intensity is symmetric.

In contrast, Table 1.3 reveals that a reduction in $E[V]$ affects $\overline{SHI}$ and $\overline{BHI}$ in opposite ways. While increased pessimism contributes to buy herding, it significantly reduces sell herding. This result is driven by the fact that during times of grim economic outlook, most informed traders sell anyway. Herd behavior, however, requires a trader to *alter* her initial trading decision. For *sell* herding to be possible, for instance, the trader has to be initially inclined to *buy* the asset. Only informed traders receiving U-shaped signals with strong biases toward the high state of the risky asset (i.e. $p^{21} << p^{23}$) may still be inclined to buy initially for low $E[V]$. As $E[V]$ drops, so does the number of simulated signal structures in $\mathcal{C}$ that exhibit a sufficiently strong positive bias of the U-shaped trader for sell

Table 1.3: The effects of economic outlook on average herding intensity

	$\overline{SHI}$	$\overline{BHI}$
High $E[V]$	0.0502 (0.0010)	0.0357 (0.0010)
Low $E[V]$	0.0370 (0.0016)	0.0504 (0.0016)

Notes: This table reports the simulated $\overline{SHI}$ and $\overline{BHI}$ for stocks where traders show high and low degrees of pessimism respectively. Standard deviations are in parentheses. Welch's t-test reveals a highly asymmetric effect for sell and buy herding. Indeed, $\overline{SHI}$ decreases as pessimism increases while $\overline{BHI}$ increases with the degree of pessimism. The results are significant at all usual significance levels. Out of the 13,284 simulated stocks, 5,904 stocks exhibit high and low degrees of pessimism. Average herding intensities are calculated as the weighted cross-sectional averages of the simulated SHI and BHI for stocks in each respective class. The figures in the table are the weighted average and the weighted standard deviation of 2,000 iid simulated outcomes of $\overline{SHI}$ and $\overline{BHI}$ under high and low uncertainty, respectively. For all calculations, the weights correspond to the observed number of informed trades.

herding to be possible. By the same line of reasoning, $\overline{BHI}$ increases with low $E[V]$.

We emphasize that the results in Table 1.3 do not contradict strong accumulations of traders on the sell side during times of deteriorated economic outlook. The Park and Sabourian (2011) model predicts that such a consensus in trade behavior is not driven by a switch in traders' opinion toward that of the crowd but results from a high share of equally pessimistic traders all acting on similar information. Such correlation of trade behavior is called spurious or unintentional herding in the literature, compare e.g. Kremer and Nautz (2013a) and Hirshleifer and Hong Teoh (2003).

The simulation shows that the positive effect of increased uncertainty on sell herding dominates the negative effect of increased pessimism. This leads to an overall slight increase in $\overline{SHI}$ during times of high market stress. In contrast, the complementary effect of uncertainty and pessimism on buy herding results in a surge of $\overline{BHI}$ during times of high market stress. We consolidate these simulation results in the following

Hypothesis 1.2 (Market Stress and Herding Intensity).
In times of high market stress, the increase in buy herding is more pronounced than that of sell herding.

1.5 Conclusion

Predictions from herding models are often used to inform to what extent the results obtained from empirical measures are in fact consistent with a particular theory of herding and whether the empirical findings are more in line with intentional rather than spurious herding. This is done because empirical measures of herding are typically only proxies of the particular type of herding the researcher is interested in. The theoretical predictions, however, are not rigorously derived from a particular model, but instead only loosely inferred. This is problematic, because herding models are, in fact, highly complex and non-linear such that even seemingly simple prediction cannot be easily derived by just "eyeballing" the respective model.

In this paper, therefore, we show how theory-based predictions can be derived from a particular herding model by means of numerical simulations focusing on the effects of information risk and market stress on herding intensity for a heterogeneous set of assets. The model predicts that both buy and sell herding increase symmetrically with information risk. The effects of market stress on herding intensity are more complicated. We show that buy and sell herding both increase with market stress, however, they do so in an asymmetric fashion. Interestingly, the model-implied hypothesis is that the increase of *buy* herding is more pronounced in times of high market stress than the one of sell herding. This is because the model-based measure of aggregate herding intensity only detects intentional herding as opposed to unintentional one. Traders may very well accumulate on the sell side of a market during downturns. Such coordination of traders, however, tends to be unintentional since they all follow their own private information that advises them to sell and, hence, is not reflected in the aggregate herding intensity. Conversely, the shortage of good news during crisis periods causes investors to be particularly susceptible to signals that the market rebounds. A temporary increase in stock prices due to trader accumulation on the buy-side of the market is such a signal. Consequently, investors are prone to intentionally follow others into buying stocks.

While an empirical counterpart to the simulation exercise is not part of this thesis, in Boortz, Kremer, Jurkatis, and Nautz (2014) we present how such simulation results may be used explicitly to inform empirical herding measures. Chapter 2 of this thesis, however, cautions that further steps on the empirical measurement part may be needed to fully gain from a combination of such predictions derived from a herding model and an empirical proxy of herding.

Appendix 1

1.A Demonstrating the need of numerical simulations

Financial market herd models including the model of Park and Sabourian (2011) are not designed to provide closed-form solutions for expected herding intensity. In this Appendix, we use two examples to demonstrate why numerical simulations are required for obtaining model-based results regarding the impact of information risk and market stress on herding intensity.

1.A.1 The History Dependence of Herding Intensity

Even for a given parameterization model complexity prevents deriving a closed-form analytical formula for herding intensity. The herding definition depends on the market maker's quotes, ask_t and bid_t, as well as the informed traders' expectations regarding the asset's true value $E[V \mid S, H_t]$. These quantities, in turn, depend on the whole history of trades until t. In fact, not only the number of observed buys, sells and holds but also their order affects expectations and quotes at time t. As a consequence, even for a given model parameterization, each history path would need to be analyzed separately to derive results on expected herding intensity.[15]

Let us illustrate this issue with a concrete numerical example. Assume the conditional signal matrix $P(S \mid V)$ to be

[15]Given the sheer number of possible trading histories alone, an analytical derivation of SHI and BHI is not feasible even for relatively small T. For any length T of the history H_T, there are 3^T different history paths.

$P(S \mid V)$	$V_1 = 0$	$V_2 = 1$	$V_3 = 2$
S_1	0.6	0.5	0.1
S_2	0.3	0.1	0.4
S_3	0.1	0.4	0.5

The distribution of the risky asset is $P(V) = [0.3\ 0.4\ 0.3]$. Multiplying $P(S \mid V) \cdot P(V)$ yields the unconditional probabilities $P(S) = [0.41\ 0.25\ 0.34]$ that a trader receives a signal S given that she is informed. Finally, the share of informed traders is set to be $\mu = 0.5$. Only informed traders receiving the U-shaped signal S_2 can herd. Given that $E[V] = 1 < 1.12 = E[V \mid S_2]$, the U-shaped trader can engage in sell herding only if she is inclined to buy initially.

We discuss two distinct trading histories consisting of 100 trades and the exact same number of buys and sells. The only difference is the order in which the trades are observed. Let $H_1^{100} = \{25\ buys,\ 50\ sells,\ 25\ buys\}$ and $H_2^{100} = \{25\ sells,\ 50\ buys,\ 25\ sells\}$. Figure 1.A1 shows how a U-shaped trader would decide to trade at every time $t = 1, ..., 100$ for the respective trading histories.

Note that the number of trades for which S_2 sell herds differs for the two histories. Under H_1^{100}, S_2 potentially sell herds between periods 51 and 85, i.e. 35 times.[16] Under H_2^{100}, S_2 potentially sell herds only 30 times. The share of U-shaped traders among the population of all traders is $\mu P(S_2) = 0.5 \cdot 0.25 = 0.125$. Consequently, we expect to observe a total number of $s_{T,1}^h = 0.125 \cdot 35 = 4.375$ herding sells under H_1^{100}. Correspondingly, under H_2^{100}, we only have $s_{T,2}^h = 0.125 \cdot 30 = 3.75$ expected herd sells.

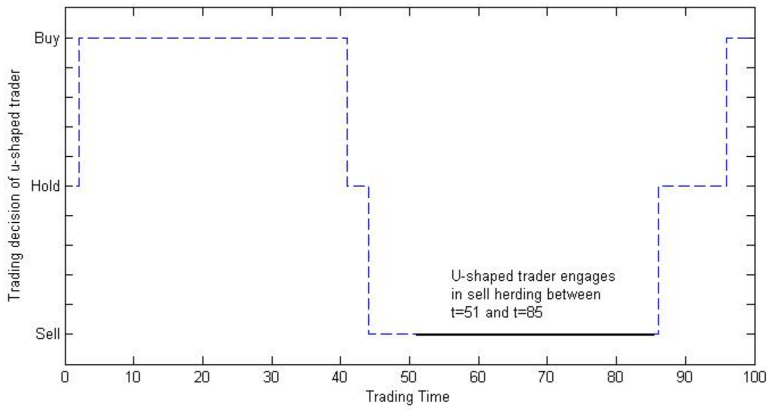
Moreover, since $\mu = 0.5$ and $T = 100$, we expect that both histories contain 50 informed trades. For an arbitrary history, calculation of the expected number of informed trades is much less straight forward since there is the possibility that informed traders hold and we hence have fewer informed trades than 50. Since H_1^{100} and H_2^{100} do not contain any holds, however, this is not an issue here.

According to Definition 1.2, the sell herding intensity is $SHI = s_T^h/(b_T^{in} + s_T^{in})$. Plugging in the expected values for numerator and denominator that we just calculated, we obtain an expected sell herding intensity $SHI_1 = 4.375/50 = 0.0875$ under H_1^{100} and $SHI_2 = 3.75/50 = 0.075$ under H_2^{100}.[17]

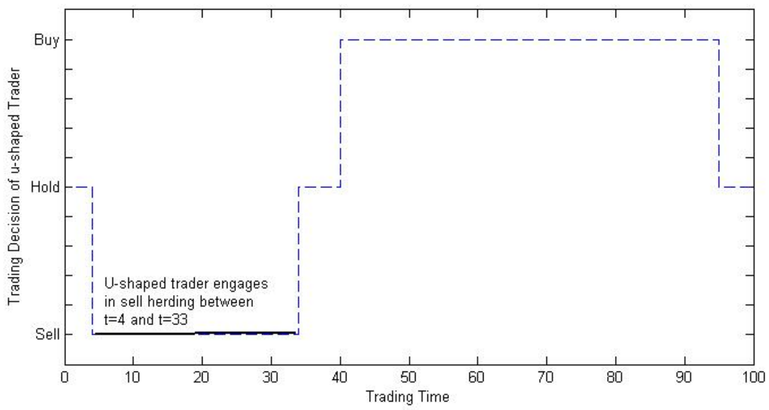
[16]Note that S_2 does in fact start herding only in period 51, although she would already have decided to sell in period 44. This is because the complete history does not contain more sells than buys until period 51, which we demand in order to ensure that S_2 actually follows the majority in the market.

[17]Note that since numerator and denominator are clearly correlated, we have that $E[\frac{X}{Y}] \neq \frac{E[X]}{E[Y]}$. A Taylor approximation of order 1, however, yields that the expectation of a ratio can be consistently estimated by the ratio of the expectations. As a consequence, all equations should be understood as approximations. An exact calculation of expected herding intensity would be even more complicated.

Figure 1.A1: Trading decisions of U-shaped trader for $\mu = 0.5$



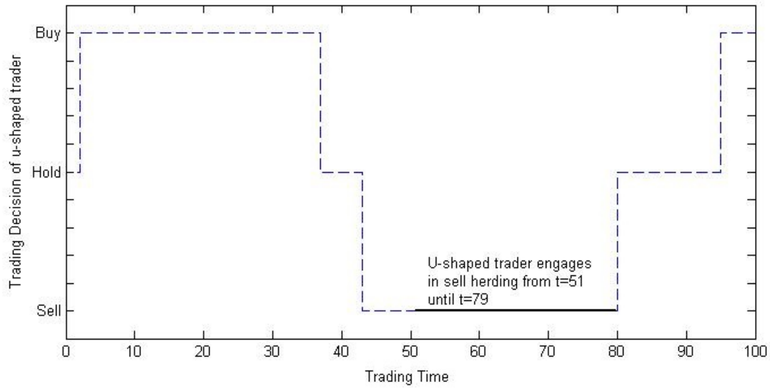
(a) $H_1^{100} = \{25\ buys,\ 50\ sells,\ 25\ buys\}$



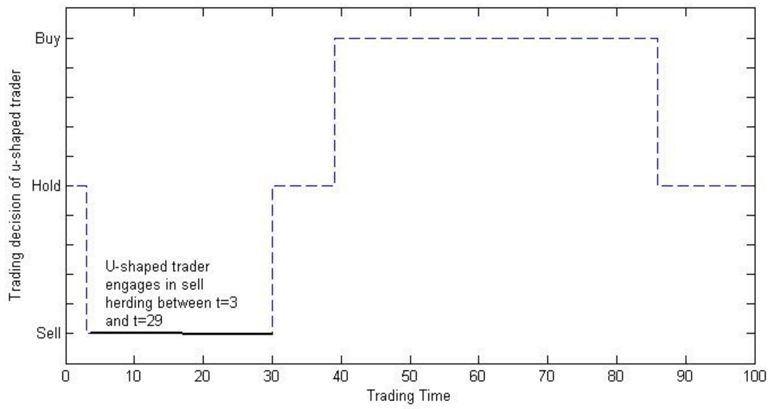
(b) $H_2^{100} = \{25\ sells,\ 50\ buys,\ 25\ sells\}$

Finally note that the probability of observing these histories $P(H_i^{100})$ is also different for $i = 1, 2$, since the probability of observing a certain trade (i.e., buy or sell) in t depends on the trading decisions of the informed traders at t. This means that in order to calculate an overall expected herding intensity for the model parameterization above, we would need to analyze SHI and $P(H^{100})$ for all 3^{100} possible history paths separately, a task well beyond our current computational capacity. Even if we were able to calculate that number, we still would not have a formula that tells us how SHI would react to changes in certain model parameters such as μ. Indeed, one can illustrate the many counteracting effects of a change in μ that result in quite different outcomes for specific trading histories and thus also prevent the derivation of analytical comparative static results.

Figure 1.A2: Trading decisions of U-shaped trader for $\mu = 0.6$



(a) $H_1^{100} = \{25\ buys,\ 50\ sells,\ 25\ buys\}$



(b) $H_2^{100} = \{25\ sells,\ 50\ buys,\ 25\ sells\}$

1.A.2 The Impact of a Change in μ on Herding Intensity: An Analytical Approach

Let us now assume that $\mu = 0.6$ and see how SHI changes for H_i^{100}, for $i = 1, 2$.

Figure (1.A2) shows that the increase in μ causes the number of potential sell herd trades to drop from 35 to 28 and from 30 to 27 for H_1^{100} and H_2^{100} respectively. Given that now $\mu P(S_2) = 0.15$, we expect $SHI_1 = 0.07$ and $SHI_2 = 0.0675$ for the respective histories. In other words, an increase in μ causes a drop in SHI for the above two trading histories.

The effects that drive this result are higher bid-ask-spreads quoted by the market maker in conjunction with a higher average information content of each single trade. Both effects contribute towards a stronger preference of S_2 of holding the asset. In particular, the sell herds are broken much faster than before: While

Table 1.A1: Probabilities of selected histories

$\mu = 0.5$	Number of herd trades	$P(H.)$	$(P(H_1) + P(H_2))/P(H_3)$
H_1^{100}	35	$7.62 \cdot 10^{-38}$	
H_2^{100}	30	$3.75 \cdot 10^{-38}$	$6.72 \cdot 10^{-7}$
H_3^{100}	97	$1.69 \cdot 10^{-31}$	
$\mu = 0.6$			
H_1^{100}	28	$4.15 \cdot 10^{-36}$	
H_2^{100}	27	$1.95 \cdot 10^{-36}$	$7.02 \cdot 10^{-8}$
H_3^{100}	97	$8.69 \cdot 10^{-29}$	

Notes: This table reports the probabilities of three different histories for the previously specified model parameterizations with $\mu = 0.5$ and $\mu = 0.6$ respectively. It also compares the probability ratio of observing histories H_1 or H_2 with observing history H_3 for each scenario. H_1 and H_2 are as before, H_3 is a history consisting of 100 sells.

for $\mu = 0.5$, the sell herding U-shaped traders had to observe 9-10 consecutive buys before switching back into holding the asset, the observation of merely 5 consecutive buys already triggers this switch in trading behavior of S_2 when $\mu = 0.6$.

The results in Section 1.4, however, suggest that $\overline{SHI}$ increases with μ. The reason for this is yet another effect of a change in μ. An increase in μ alters the probability with which a certain history is observed. Indeed, an increase in μ shifts probability mass from histories with low or decreasing herding intensity to histories with persistently high herding.

This effect is documented in Table 1.A1. Consider the previously introduced histories H_1^{100} and H_2^{100}. Also consider history H_3^{100} consisting of 100 sells. Under H_3, S_2 sell herds from $t = 4$ until $t = 100$ resulting in 97 potential herd sells regardless of μ. Yet, the probabilities for each of the histories changes as μ changes. More specifically, the probability to observe H_1 or H_2 relative to the probability to observe H_3 decreases.

This can be attributed to the self-enforcing nature of herd behavior. Once investors start herding, it is on average more likely that they keep herding than that their herd is broken.

We emphasize that this is not a complete comparative static analysis. For that we would have to consider all 3^{100} different histories. As outlined before, this is beyond current computational capabilities. Also note that the discussed examples are only for a single stock. The calculations further complicate if one aims at calculating average herding intensities for a heterogeneous stock market as we do in Section 1.4.

Paper 2

How to Measure Herding in Financial Markets*

2.1 Introduction

Investor herding describes the behavior of individual investors that follow the decision of the majority although they hold private information that advises them to act differently (Brunnermeier, 2001, p. 148). There is strong consensus in the literature that herding has the potential to cause informational inefficiencies, distort prices and ultimately destabilize financial markets altogether.

Consequently, empirical studies have been putting great efforts into detecting herd behavior by assessing whether groups of investors coordinate and to gauge the effect of their coordination on asset prices (for an overview see e.g. Bikhchandani and Sharma, 2001). The empirical literature on investor coordination has been strongly influenced by the seminal work of Lakonishok, Shleifer, and Vishny (1992). Their well-known LSV measure has long become a benchmark to test for the presence of investor coordination, see e.g. Wermers (1999), Dorn et al. (2008), Barber et al. (2009) and Brown et al. (2014).

This paper shows, however, that the LSV measure captures a very specific empirical notion of investor coordination that neglects an important aspect of the theoretical definition provided above. Consequently, the LSV measure fails to provide an empirical link of investor behavior to the aforementioned inefficiencies.

We provide a new measure by adjusting the LSV measure in accord with general implications from the market microstructure literature on herd behavior. Using simulated trade data we quantify the differences between the two approaches. We show that our measure accurately distinguishes between the different types of investor coordination, i.e. herding, contrarianism and independent trading, whereas

*This paper was written in collaboration with my co-author Christopher Boortz. Note that there have been some significant changes, especially in the simulation section, when compared to the presentation in Boortz (2016, chap. 3).

the LSV measure generally fails to do so if the trade data do not fulfill the rather restrictive assumptions associated with its approach.[1]

The LSV measure uses transaction data of a subset of investors to measure their tendency to buy and sell stocks in *crowds*. To do so, it assesses the deviation of the subgroup's observed buy propensity in each stock from its average buy propensity across all stocks.

The fundamental criticism of this approach is that it does not account for the information set of investors (Bikhchandani and Sharma, 2001; Cipriani and Guarino, 2014). Under the null hypothesis of independent trading the LSV measure assumes that the probability to observe a buy is equal to the average buy propensity for all stocks. Coordination detected by the LSV measure can thus arise simply because traders hold similar information sets that incline them to buy one stock and sell another. Such coordination is dubbed spurious herding, as opposed to intentional herding. Hence the LSV measure is commonly viewed as a necessary but not sufficient signal of herd behavior.

We show, however, that the consequences are more severe. Not only is the LSV measure a biased measure of the true deviation from independent trading, but it also fails to correlate with intentional investor coordination.

Our adjustment to the LSV approach is a response to the fundamental criticism. While we adhere to the comparison of buy propensities under actual and independent trading, we allow the buy propensities under independent trading to be stock-specific to account for the event that traders hold similar information sets that recommend them to buy or sell the same stocks.

With this relaxing assumption the estimation of the independent buy propensities becomes an empirical challenge, which we show can be dealt with by the following two assumptions. We argue in line with microstructure models that the first few trades after the start of trading are carried out independently conditional on traders' information sets.[2] Secondly, we assume that the buy propensities under independent trading come from a common distribution. We show that this distribution can be accurately estimated from the few independent trades even in small cross-sections. This estimated distribution then provides the proper benchmark of independent trading to compare the observed buy propensities with.

Due to these adjustments our measure brings the empirical approach to detecting herding (and contrarianism) closer to its theoretical counter-part from the

[1] Contrarianism can be seen as the counter-part of herding. Instead of following the crowd, contrarians act against it although they have information that tells them to trade in the same direction as the majority of the traders, (see e.g. Park and Sabourian, 2011).

[2] See, for example, Proposition 7 in Avery and Zemsky (1998).

observational learning literature (Bikhchandani et al., 1992; Avery and Zemsky, 1998; Park and Sabourian, 2011) in two respects. First, as we obtain the buy propensities under independent trading from the early trades and compare it to the subsequent trading behavior we capture the notion of switching behavior that underlies herding and contrarianism (see Park and Sabourian, 2011). Second, as we account for the distribution of information under the independent benchmark, we stress that herding and contrariansm involve a change against the behavior based only on private information.

Other modifications of the LSV measure have been proposed in the literature to improve its performance. Frey et al. (2014) modify the LSV measure by taking the squared instead of the absolute difference between the observed buy propensities and the average one. Wylie (2005) corrects the LSV measure to account for possible biases that can arise from short-selling constraints. Yet, since both maintain the assumption of a constant buy propensity across all stocks under the null our arguments apply to their approaches as well.[3]

Another measure related to LSV is the one proposed by Sias (2004). Like Lakonishok et al. (1992), Sias (2004) uses the buy propensity of investors as the underlying statistic. Yet, the Sias measure assesses whether buy propensities are persistently high or low over time by measuring the correlation of buy propensities between adjacent time periods for a fixed cross-section of stocks. Though we will not compare our approach to the one of Sias (2004) directly, our arguments are valid for his measure as well. By assessing the correlation of buy propensities, the cross-sectional averages of the buy propensities constitute a part of the Sias measure and, therefore, our arguments in favor of an approach that accounts for the idiosyncrasy of these propensities apply here as well.

The disconnect of empirical measures on coordinated trading with the theoretical literature has also been noted by Devenow and Welch (1996) and Cipriani and Guarino (2014). To provide a rigorous measurement in line with the theoretical definition of herding, the latter fit the parameters of a specific model of herd behavior to the data. Though we attempt to bring the empirical literature closer to the theoretical idea of herding and contrariansm, we do not go as far as estimating a specific model of herding.[4] Our measure is not designed to explain why investors

[3]Statistically put, the assumption of equal buy propensities under independent trading stems from the fact that the LSV measure tests whether the observed number of buys are more (or less) dispersed than suggested by a binomial distribution. Consequently, our arguments generally apply to any test on binomial dispersion (e.g. Cochran (1954), Tarone (1979)) that is applied for the purpose of finding deviations from independent trading.

[4]Note that the model of Cipriani and Guarino (2014) does not allow their agents to engage in contrarian behavior and, thus, rules out the empirical possibility of contrariansm. In contrast, due to our more general empirical approach our measure allows us to distinguish between herding,

coordinate, but to provide a rather model-independent statistical tool that signals when herding or contrarianism is present in the data and to provide an indirect assessment of traders' observational learning strategies.[5]

The rest of the paper is organized as follows. Section 2.2 briefly reviews models of observational learning that set the scene for the type of investor coordination that we want to measure. Section 2.3 presents the measure of investor coordination proposed by Lakonishok et al. (1992) and pinpoints its shortcomings to be applied as a measure of the type of herding and contrarianism that we are interested in. It follows the introduction of our new measure that derives from two adjustments of the LSV measure. Section 2.4 provides an evaluation of our new measure and LSV by means of simulations and confirms both the ability of our measure to detect herding and contrarian behavior and our criticism of the LSV measure to not achieve the same. Finally, Section 2.5 concludes with a few remarks on how to best apply our measure to the data.

2.2 Models of Observational Learning, Contrarian and Herd Behavior

We rely on a particular class of models that is rooted in the observational learning literature to define the specific terms of dependent trading that we want to measure. We will, therefore, briefly introduce the general model framework to help us understand why existing measures of coordinated trading may not be suited to measure the particular type of herding and contrariansm that is discussed in this literature and how we may be able to adjust these measures accordingly to achieve that goal.

The literature on herding has been sparked by the seminal works of Banerjee (1992), Bikhchandani et al. (1992) and Welch (1992) where agents learn an unknown value from the observed decisions of others and a private signal. Particularly the model of Bikhchandani et al. (1992) has found numerous implementations in experimental studies (see Weizsäcker, 2010, for a meta-study) and has been advanced by a Glosten and Milgrom (1985) type trading mechanism to extend the

contrarianism and independent trading.

[5]For a better understanding of potential drivers for investor coordination, we refer the reader to the rich theoretical herding literature. The seminal works of Bikhchandani et al. (1992) and Banerjee (1992) demonstrate that herding is triggered by information externalities that a decision by one agent imposes on the decisions of the subsequent agents. Reputational concerns of financial decision makers are identified as cause of herd behavior by, e.g., Scharfstein and Stein (1990), Graham (1999) and Dasgupta et al. (2011). So-called investigative herding has been discussed by Froot et al. (1992) and Hirshleifer et al. (1994).

observational learning literature to a financial market context where prices aggregate information (see, e.g., Avery and Zemsky, 1998; Cipriani and Guarino, 2014; Park and Sabourian, 2011). The general setup of these types of models is as follows.

Traders trade an asset of unknown fundamental value with a market maker. Traders' decisions to buy or sell the asset is based on their private information on the asset's value and the actions and prices they observe before they trade. The market maker who posts competitive bid and ask quotes learns the asset's value from the order flow as well, but is not endowed with private information.

Three scenarios are possible. (1) Prices update with each incoming trade at a similar rate as traders update their valuation of the asset when they learn from a new trade. In this case, traders always follow their private signal such that each signal gets incorporated into the price. (2) Prices update too sluggishly relative to the rate at which traders update their valuation into the direction of the preceding trades. In this case, traders may engage in herd behavior defined as a state where traders buy (sell) irrespective of their signal when the majority of traders bought (sold) the asset. (3) Prices update too quickly relative to the rate at which traders update their valuation into the direction of the preceding trades. In this case, traders may engage in contrarian behavior where traders buy (sell) irrespective of their signal when the minority of traders bought (sold) the asset.[6]

The importance of the definitions of herding and contrarian behavior in these models is that they go beyond a mere reference to the action of the majority or minority of traders. By linking the investor behavior to actions that involve private information the literature stresses the potential of temporary states of informational inefficiency caused by herding or contrarianism.

Implementing an equivalent empirical definition is inherently difficult due to private information being unobservable, which is why empirical measures typically equate herd behavior with the action of a crowd. We will argue, however, that we can make use of the predictions of these trading models to bring the empirical approach closer to the above definitions of herding and contrariansm. If neither herding nor contrarian behavior arises and if there are no fundamental changes to the information sets of investors, the fraction of buyers and sellers are stationary variables. Under herd behavior, on the other hand, the probability of extreme numbers of buys or sells increases relative to a state without herding due to its self-

[6]Park and Sabourian (2011) show in an import contribution that the existence of herding, as well as contrarian behavior, depends on the shape of traders' signals. Private information that weights the likelihood of extreme outcomes more than that of moderate ones encourages herd behavior, while private information that favors moderate outcomes encourages contrarian behavior. Their theoretical predictions have been confirmed experimentally by Park and Sgroi (2012, 2016).

reinforcing effect, whereas under contrarian behavior the probability of extreme numbers of buys and sells decreases due to the self-defeating feature of contrarian behavior.

2.3 Measuring Investor Coordination

2.3.1 LSV — A Measure of Crowd Behavior

The most prominent measure of investor coordination that is based on transaction data is the LSV measure of Lakonishok et al. (1992). The measure is computed for a subset of investors and selection of stocks over any desired time-horizon, which, for our purpose, we set to a day. For each stock, the LSV measure is given by

$$LSV_i = |br_i - p| - AF_i^{LSV}, \tag{2.1}$$

where $br_i = B_i/T_i$ is the buy-ratio in stock $i = 1, \ldots, I$, i.e. the number of buys over the number of trades of the set of investors in stock i, and p is the expected proportion of traders buying estimated by $\hat{p} = \sum_i B_i / \sum_i T_i$. AF_i^{LSV} is an adjustment factor given by

$$AF_i^{LSV} := \mathbb{E}_{\zeta_k} |\frac{k}{T_i} - p| = \sum_{k=0}^{T_i} \zeta(k|T_i, p) |\frac{k}{T_i} - p| \tag{2.2}$$

where $\zeta(k|\cdot)$ is the binomial distribution.[7]

Hence, LSV measures investor coordination by the deviation of buy-ratios from the average one. The adjustment factor accounts for the random deviation of the buy-ratios that we would expect even if investors do not coordinate. The design of the adjustment factor, thereby, essentially entails the view that the number of buys under independent trading is binomially distributed with the same success probability p for each stock.

The empirical definition of herding offered by the LSV measure is, thus, one of crowd behavior. If a subgroup of investors buys one stock 60% of the time and another one 40% of the time, they tend to buy and sell the same stocks and the LSV measure would accordingly indicate that 10% of the traders engaged in herd behavior (ignoring the adjustment factor for the sake of simplicity). As such, however, the LSV measure does not account for the coordination of investors that

[7]A closed form solution for the expectation is given by $\frac{1}{T_i}\mathbb{E}_{\zeta_k}|k - pT_i| = \frac{1}{T_i}2(1 - p)^{T_i - \lfloor T_i p \rfloor} p^{\lfloor T_i p \rfloor + 1}(\lfloor T_i p \rfloor + 1)\binom{T_i}{\lfloor T_i p \rfloor + 1}$ (see Diaconis and Zabell, 1991).

we would expect if traders base their decisions on information that they acquired for different stocks and which is likely to be correlated across investors.

Yet, accounting for the information that investors have prior to the trading process is an integral part of the herding definition provided above. Only by benchmarking the actual trading behavior against the behavior that would result from trading solely on private information, the definition explicitly links herding behavior to informational inefficiencies.

Trading behavior that leads to price inefficiencies can, of course, also arise from trading that is based solely on private information.[8] However, by not controlling for either source of the observed coordination, private information or information inferred from the trading process, it is not possible to link the empirically measured coordination to one of the sources and, hence, to point to the root of a potentially destabilizing behavior.

Finally, note that the empirical notion of investor coordination offered by LSV does not leave room for the counter-part of herding, contrarian behavior. To best see this, consider the case where we have extremely high trading activity such that the adjustment factor goes to zero.[9] Then, the minimum attained by the LSV measure is zero, while any deviation from the average buy-ratio is interpreted as herding.

If, on the other hand, we would account for coordination due to private information, we can distinguish between herding and contrarianism in line with the theoretical notion. Taking up the above example again, had we known that initially 70% of traders intended to buy one asset and to sell the other one, the final coordination of 60% of traders buying the former and selling the latter, is rather in line with contrarian tendencies instead of herd behavior.

2.3.2 A New Measure that Accounts for Investors' Stock-Specific Information

Based on the previous discussion, we propose an inconspicuous but profound modification to the LSV measure by allowing the expected proportion of traders buying to be stock-specific:

$$\widetilde{LSV}_i = |br_i - \tilde{p}_i| - AF_i^{\widetilde{LSV}} \tag{2.3}$$

[8]See, for example, Froot et al. (1992) for a model of herding on information and Enke and Zimmermann (2018) for an experimental account of neglecting correlation between sources of information.

[9]The mean absolute deviation of a binomially distributed random variable is upper bounded by its standard deviation (Blyth, 1980). It follows that $AF_i^{LSV} = \mathbb{E}_{\zeta_k}|k - pT_i|/T_i \leq \sqrt{p(1-p)/T_i}$, which goes to zero for $T_i \to \infty$.

where $AF_i^{\widetilde{LSV}}$ is modified accordingly.

This modification bears an empirical challenge as the stock-specific buy-ratio under independent trading is not observable. A second adjustment to the LSV measure, however, makes our measure operational and allows us to distinguish between herding and contrarianism. The adjustment follows from two assumptions, the first of which defines the range of data that we can use to estimate the independent buy-ratios and the second one allows us to use that data in a meaningful fashion.

Assumption 2.1. *The first few $\tau_i \in \mathbb{N}$ trades in any stock i on a particular day are carried out independently.*[10]

In a non-trivial sense Assumption 2.1 is always true. Starting with the first trade, no other trade could have been observed that could have influenced the decision of the first trader.[11] Moreover, market microstructure theory tells us that at the outset of the trading process deviations from independent trading are less likely to occur (see Proposition 7 in Avery and Zemsky, 1998), which is also intuitive. For a trader to change her opinion, there has to be a sufficient amount of information that she can infer from preceding trades. This is unlikely to be the case if only a few trades have been executed.

Statistically put, this means that given the number of trades in stock i, T_i, there exists a τ_i, $1 \leq \tau_i \leq T_i$, such that the number of buys in stock i until τ_i is $B_i^{\tau_i}|_{\tilde{p}_i} \sim \text{Bino}(\tau_i, \tilde{p}_i)$. An estimator of $\tilde{p}_i$ would thus be given by B^{τ_i}/τ_i. However, τ_i should be chosen as low as possible. If based on small τ_i, estimators such as B^{τ_i}/τ_i are too noisy to conduct meaningful inference on them. Hence, we add the following assumption.

Assumption 2.2. *On each day, the buy-ratios under independent trading $\tilde{p}_i$ are drawn from a common distribution. More specifically, we assume $\tilde{p}_i \overset{iid}{\sim} \text{Beta}(\alpha, \beta)$.*

This assumption is equivalent to the number of buys under independent trading being iid beta-binomially distributed, i.e. $B_i^{\tau_i} \overset{iid}{\sim} \text{Beta-Bino}(\tau_i, \alpha, \beta)$. The importance of assuming a common distribution for the independent buy-ratios is that we can now utilize the size of the cross-section to obtain a proper benchmark of the trading behavior that we would expect under independent trading. We will later show numerically that the distribution of independent buy-ratios can be estimated

[10] Be reminded that independence here means independence conditional on the private information set of the traders.

[11] The first trade is always the one that the researcher defines to be the first trade. Anything observable that happened before that trade will enter the information sets of the traders and, thus, be part of their priors.

accurately, even though it is based on only a small amount of data considering the individual stock.

The particular choice of the Beta distribution for the probability of a buy under independent trading is not strictly necessary for the definition of our measure of dependent trading. Any distribution that can be consistently estimated will work (subject to small sample performance). The Beta distribution, however, presents itself as a natural candidate. In Bayesian inference on a Bernoulli distributed random variable, for example, the Beta distribution is often chosen as a prior over the success probability due to its conjugacy property (Bishop, 2009, p. 71) and the Beta-Binomial distribution and its generalization the Dirichlet-Multinomial distribution if often chosen to model overdispersion in count data, similarly to our application, (e.g. Neerchal and Morel, 1998). More importantly, however, we find empirical support for our assumption which is presented in the Appendix and in more detail in Boortz (2016).

Using these assumptions we can now add our second modification to the LSV measure by which we obtain our new measure for investor coordination, the expected deviation from independent trading, defined by

$$\begin{aligned} H_i &= \mathbb{E}_{f_p}|br_i - p| - AF_i \\ &= \int_0^1 f(p|\alpha,\beta)|br_i - p|\,\mathrm{d}p - AF_i \end{aligned} \tag{2.4}$$

where $f(\cdot|\alpha,\beta)$ is the Beta density. AF_i is an adjustment factor to center H_i over zero if in fact all trades were carried out under independent trading. It is given by

$$\begin{aligned} AF_i &= \mathbb{E}_{f_p}\mathbb{E}_{g_k}|\frac{k}{T_i} - p| \\ &= \int_0^1 f(p|\alpha,\beta)\sum_{k=0}^{T_i} g(k|T_i,\alpha,\beta)|\frac{k}{T_i} - p|\,\mathrm{d}p \\ &= \int_0^1 f(p|\alpha,\beta)\int_0^1 f(\tilde{p}|\alpha,\beta)\sum_{k=0}^{T_i}\zeta(k|T_i,\tilde{p})|\frac{k}{T_i} - p|\,\mathrm{d}\tilde{p}\,\mathrm{d}p, \end{aligned} \tag{2.5}$$

where $g(k|\cdot)$ is the Beta-Binomial density. Note that AF_i corrects for *two* sources of randomness. First, as for the LSV measure, we observe only a finite number of trades. That is, even if each single trade has been drawn from a Bernoulli distribution with the success probability equal to the independent buy-ratio, there is a positive change that the observed buy-ratio is not equal to the independent one. In addition, the true independent buy-ratio $\tilde{p}$ is itself a random variable and will, therefore, deviate from most hypothesized $p \in (0,1)$.

2.3.3 Estimating H_i

To estimate H_i we need to estimate α and β. According to our assumptions we can obtain these estimates from a maximum-likelihood estimation of the Beta-Binomial distribution on the data $\{B_i^{\tau_i}, \tau_i\}_{i=1}^{I}$. Since the maximum-likelihood estimator of the Beta-Binomial distribution is consistent (see Garren, 2004, p. 240) and because H_i is a composition of continuous functions in α and β, the continuous mapping theorem implies that H_i is consistently estimated as well.[12]

Finally, to put our method into action requires a choice of τ_i. Pointing to the precise moment when traders start to go against their private information amounts to uncovering the latent private information itself. In line with Assumption 2.1, however, a conservatively small choice of τ_i, but large enough for the Beta-Binomial estimation to make any sense should suffice.[13] By means of simulation, we find that $\tau_i = 10$ is already large enough even for relatively small cross-sections to provide good estimates. Details are presented below.

2.3.4 Interpretation of H and a Comparison to LSV

The LSV measure is typically interpreted at the cross-sectional level as it is basically a test on binomial over-dispersion of the cross-sectional buy-ratios. If the buy-ratios are more dispersed than suggested by the Binomial distribution, the result is interpreted as a sign of herd behavior. As we argued before, binomial over-dispersion, however, can arise simply because traders use similar information to make their trading decisions, and a measure of binomial over-dispersion does not leave room for detecting contrarian behavior.

Similarly, given our distributional assumptions, our approach measures beta-binomial over- or under-dispersion and should be interpreted at the cross-sectional level as well. In fact, our approach can be seen as a generalization of the LSV approach that collapses to LSV if the assumptions underlying its approach are fulfilled, but provides generally very different results if the assumptions of LSV

[12]Note that a similar consistency result can be derived if one does not want to restrict oneself to a particular family of distributions for the independent buy-ratios and uses kernel density estimation instead. A multitude of consistency results is available for kernel density estimators, see e.g. Parzen (1958), Silverman (1978) and Epanechnikov (2006).

[13]The precise meaning of "small", hereby, depends on the empirical context regarding, e.g. sampling frequency and the definition of a trade. One may be interested in counting each transaction as a single trade, others may be interested in aggregating single transactions into the orders that induced them, or even aggregating transactions of single traders into their net-positions over a certain time interval. Those choices affect the amount of data available at any point after the start of trading and, thus, after information starts to accumulate in the market. Note that these choice may not only affect the proper choice of τ_i, but also the proper choice for the distribution of the independent buy-ratios.

are to restrictive.[14]

Importantly, our modifications to the LSV measure allow for profoundly different interpretations that are more in line with our understanding of herding and contrarian behavior. The dispersion under the null-hypothesis is estimated from the early independent trades and provides the benchmark for what dispersion should be expected if trading continued independently and under similar information sets as the early trades. Overly dispersed buy-ratios compared to the benchmark distribution ($H > 0$) are interpreted as herd behavior, while under-dispersed buy-ratios as contrarian behavior ($H < 0$).

These interpretations follow from general implications of these types of behavior. Under herd behavior the opinions of traders are updated more quickly in the direction of the trades of their predecessors than is the price, hence increasing the proportion of buyers if there were more buyers than sellers and increasing the proportion of sellers if there were more sellers than buyers. That is, we would expect the buy-ratios to become on average more extreme under herding than under independent trading. In contrast, under contrarian behavior opinions are updated more sluggishly in the direction of preceding trades than is the price. This sluggishness then leads to a decrease in the proportion of buyers if the majority of traders bought the stock, and to an decrease of the proportion of sellers if the majority of traders sold the stock. Hence, we would expect the buy-ratios to be on average less extreme under contrarian behavior than under independent trading.

As the LSV measure uses all trades of the respective trading day to estimate the expected buy-ratio under independent trading, the observed buy-ratios are always dispersed around the center of the hypothesized Binomial distribution.[15] An equivalent observation does not hold for our measure. Because we use only the first few trades of a day for the estimation of the distribution of the independent buy-ratios, the observed buy-ratios computed from all trades over the day may not be dispersed around the center of the estimated Beta distribution.

For example, we might have estimated a Beta distribution that is centered over 0.5 (i.e. $\alpha \approx \beta$), yet the observed buy-ratios have all increased to a higher value, say, 0.8. Or conversely, we might have estimated a Beta distribution that is centered over 0.8, but the observed buy-ratios cluster around 0.5. In both cases we would probably find an average H greater than zero, but these examples do not readily fit into the above description of herd behavior. While one might interpret the former example as an extreme case of market-wide buy-herding and the latter

[14] A formal proof this claim is provided in the Appendix.

[15] Note that this implies that the LSV measure assumes that positive and negative deviations from independent trading have to cancel each other out, otherwise the estimated independent buy-ratio is biased towards the direction, buying or selling, of the dependent trades.

as a case of contrarian behavior, we would caution to follow such interpretations. Instead, we would recommend to treat such examples as special cases. Such shifts in the distribution of buy-ratios suggest the occurrence of an external event that changed the general information structure. While it is interesting in its own right to analyze such cases in order to study the reaction of traders' behavior to changes in the overall environment, they say probably little about the general information updating procedure of traders.

2.4 Estimation Performance of $\hat{H}$ and LSV

2.4.1 Estimation Accuracy

To evaluate the estimation accuracy of our new measure and to provide a quantitative comparison to the LSV measure we make use of Monte Carlo simulations. Note that we do not need to simulate a fully-fledged market microstructure model with decision rules and a price mechanism. Instead we can simply make use of the general concept of herding and contrarian behavior, namely that the former leads to an increase and the latter to a decrease in the probability of observing the same action as the one in the past. Doing so not only greatly reduces the simulation complexity compared to simulating a herding model such as the one of Park and Sabourian (2011), but also provides us with more control over the degree of the deviation from the independent trading benchmark.

To generate the trade data we use Friedman's urn model (Friedman, 1949), which bears a close analogy to the concepts of herding and contrarian behavior. Imagine an urn that contains S_t silver balls and B_t blue balls at time t. One ball is drawn at random and then replaced, while h balls of the same color as the one that was drawn and c balls of the opposite color are added to the urn.

For $h, c = 0$, the fraction $p_t \equiv B_t/(B_t + S_t)$ remains constant for all t and the number of blue balls drawn after t trials follows a Binomial distribution, i.e. $B_t \sim \text{Bino}(t, p)$. This setup matches our independent trading scenario where B_t and S_t are the number of buys and sells after t trades. A single urn represents a single stock and the stock-specific probability to observe a buy, p, is drawn from a Beta distribution. Hence, the number of buys follows a Beta-Binomial distribution.

For $h > 0$ and $c = 0$, the fraction of blue balls increases when a blue ball is drawn and the fraction of silver balls increases when a silver ball is drawn. This matches our concept of herd behavior where the probability to observe a particular action, buy or sell, increases with the number of the same actions in the past. Freedman (1965) shows that p_t converges almost surely to a limiting random variable which has a Beta distribution with parameters B_0/h, S_0/h, where B_0 and S_0 are the initial number of blue and silver balls, respectively. In our simulation, B_0 and S_0 will be determined by the distribution of the independent buy-ratios. Hence, with the number of buys, B_t, in each stock being governed by a Beta-Binomial distribution with an additional Beta prior on the initial condition, we see over-dispersion in B_t compared to the Beta-Binomial distribution.

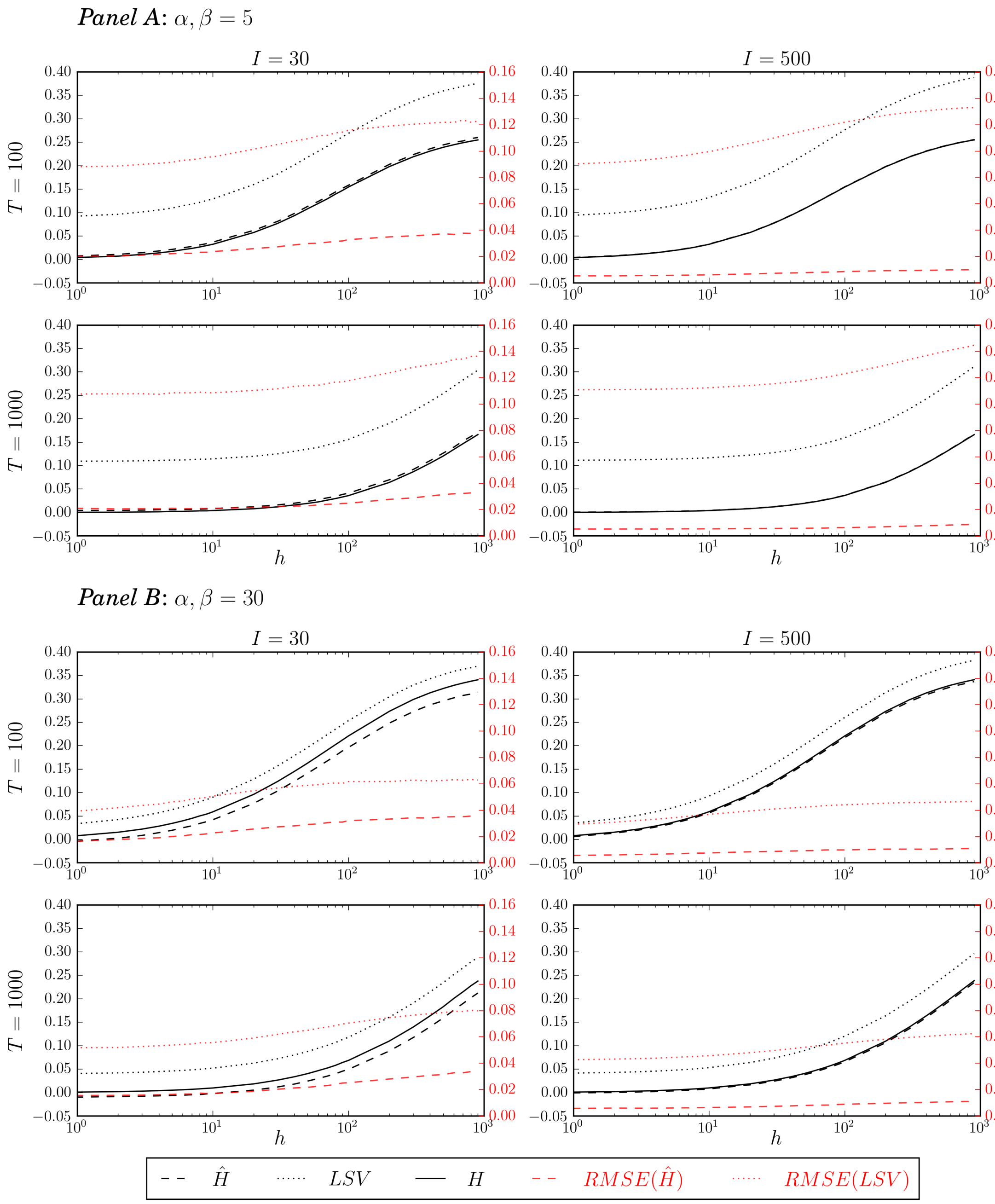
For $c > 0$ and $h = 0$, on the other hand, the fraction of blue balls decreases when a blue ball is drawn matching our idea of contrarian behavior. Freedman (1965) shows that p_t converges almost surely to $\frac{1}{2}$. That is, no matter what the initial probability to observe a buy, in the limit this probability converges to 0.5 leading to under-dispersion in the number of buys after t trades compared to the independent trading setup.

The precise simulation setup is as follows. We choose two different sizes for the cross-section, $I \in \mathcal{I} = \{30, 500\}$, and two different numbers of total trades in each stock, $T \in \mathcal{T} = \{100, 1000\}$. We draw the stock-specific probabilities of observing a buy when there is independent trading from a Beta distribution with parameters $(\alpha, \beta) \in \mathcal{P} = \{(5, 5), (30, 30)\}$. With $\alpha = \beta$ we ensure that the Beta distribution is centered over 0.5 following our idea of having a very heterogeneous set of stocks without a particularly strong market trend of buying or selling. The larger α and β the less dispersed the Beta distribution.

We let the trades evolve independently for $\tau = 10$ trades, after which trades are drawn according to Friedman's urn model. Friedman's urn model requires that we set initial B_0 and S_0. These are set to pT and $(1-p)T$, respectively, where p is the probability to observe a buy if traders trade independently. Hence, the probability to observe a buy under dependent trading ($t > \tau$) evolves according to $p_t = (pT + h\sum_{j=0}^{t} \mathbb{1}_{\{a_j = buy\}} + c\sum_{j=0}^{t} \mathbb{1}_{\{a_j = sell\}})/(T + t(h+c))$, where a_j is the action, buy or sell, at time j. For simulating herd behavior we set $c = 0$ and $h = nk$ for $n \in \mathcal{N} = \{1, 2, \ldots, 9\}$ and $k \in \mathcal{K} = \{1, 10, 100\}$. To simulate contrarian behavior we set $h = 0$ and $c = nk$, where $n \in \mathcal{N} = \{\frac{1}{29}\}$ and $k \in \mathcal{K} = \{0, 1, \ldots, 29\}$. c in the contrarian trading setup is chosen to be rather small compared to h in the herding setup, because p_t converges quickly towards $\frac{1}{2}$ for larger c. Each setup in $\mathcal{I} \times \mathcal{T} \times \mathcal{N} \times \mathcal{K}$ is repeated 5000 times for both herding and contrarian behavior. For each repetition we compute the cross-sectional mean of H, its estimator, $\hat{H}$, and the cross-sectional LSV.

Figure 2.1 and 2.2 present the results. On the left axis, the figures display the means across the 5000 repetitions. On the right axis, the figures show the root-mean-squared error of the cross-sectional means $\hat{H}$ and LSV across the 5000 simulations using H as the target value.

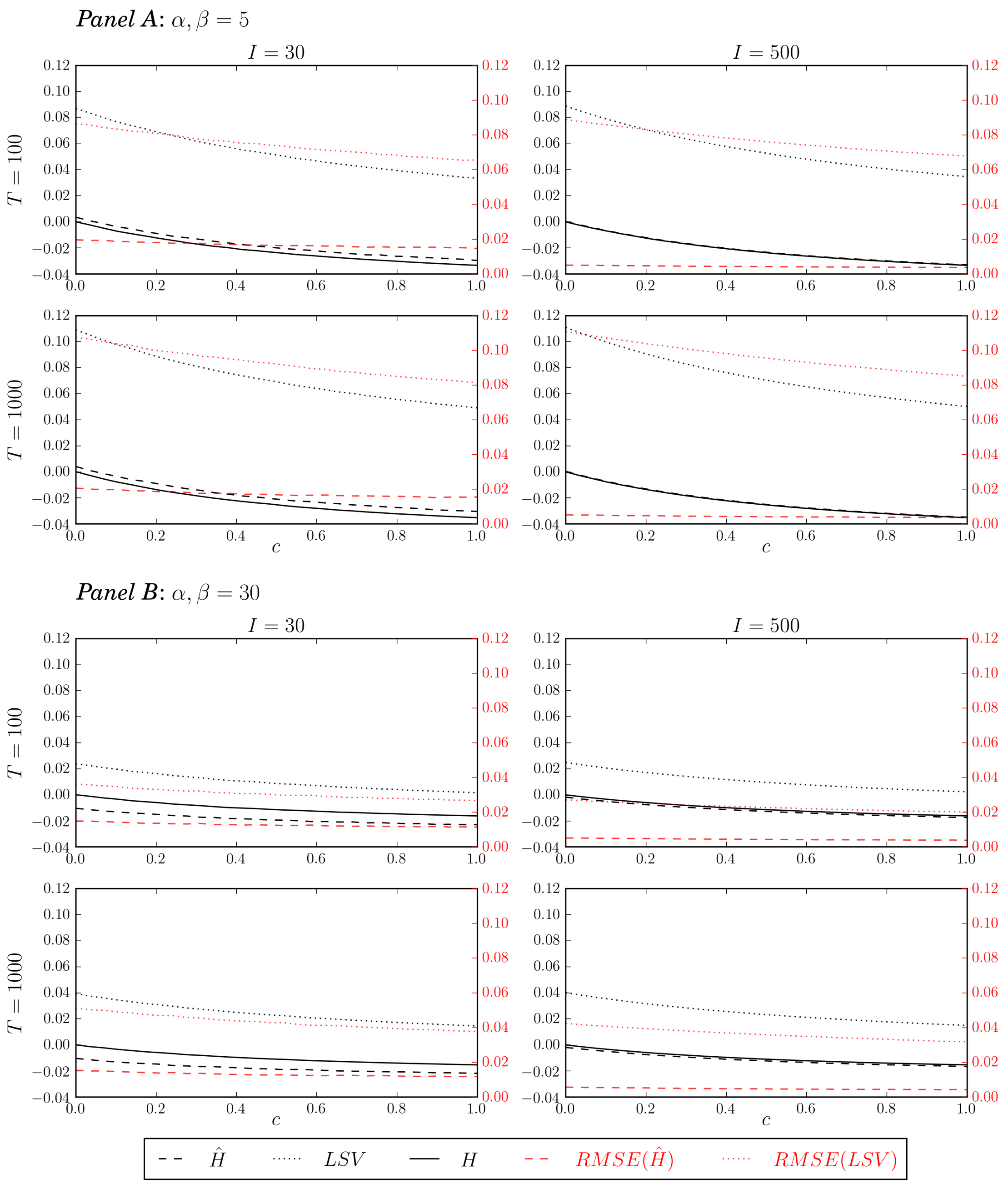
Figure 2.1: Estimation accuracy — herding



Notes: We simulate trade data by using Friedman's urn model. The first $\tau = 10$ trades evolve as under independent trading where the probability to observe a buy, p, is drawn from a Beta distribution with parameters $(\alpha, \beta) \in \{(5, 5), (10, 10)\}$. For $t > \tau$, the probability to observe a buy evolves according to $p_t = (pT + h\sum_{j=0}^{t} \mathbb{1}_{\{a_j = buy\}})/(T + th)$, where a_j is the action, buy or sell, at time j and $h = nk$ for $n \in \mathcal{N} = \{1, 2, \ldots, 9\}$ and $k \in \mathcal{K} = \{1, 10, 100\}$. Each setup in $\mathcal{I} \times \mathcal{T} \times \mathcal{N} \times \mathcal{K}$ is repeated 5000 times and for each repetition we compute the cross-sectional mean of H, its estimator, $\hat{H}$, and the cross-sectional LSV. The root-mean-squared error uses the true cross-sectional H as target value.

We see that the LSV measure shows a considerable upward bias across the board increasing slightly for stronger herd behavior and decreasing slightly for stronger contrarian behavior. Naturally, the performance of the LSV measure im-

Figure 2.2: Estimation accuracy — contrarian behavior



Notes: We simulate trade data by using Friedman's urn model. The first $\tau = 10$ trades evolve as under independent trading where the probability to observe a buy, p, is drawn from a Beta distribution with parameters $(\alpha, \beta) \in \{(5,5),(10,10)\}$. For $t > \tau$, the probability to observe a buy evolves according to $p_t = (pT + c\sum_{j=0}^{t} \mathbb{1}_{\{a_j=sell\}})/(T + tc)$, where a_j is the action, buy or sell, at time j and $c = nk$ for $n \in \mathcal{N} = \{\frac{1}{29}\}$ and $k \in \mathcal{K} = \{0, 1, \ldots, 29\}$. Each setup in $\mathcal{I} \times \mathcal{T} \times \mathcal{N} \times \mathcal{K}$ is repeated 5000 times and for each repetition we compute the cross-sectional mean of H, its estimator, $\hat{H}$, and the cross-sectional LSV. The root-mean-squared error uses the true cross-sectional H as target value.

proves as the Beta-Binomial distribution moves towards the Binomial distribution, i.e. increasing α and β.

Our measure shows only a small bias (slightly upwards for $\alpha, \beta = 5$ and slightly downwards for $\alpha, \beta = 30$), even for cross-sections as small as 30 stocks. Even though the bias worsens for larger α and β, our measure consistently outperforms the LSV. For already moderately sized cross-sections, the estimated measure is almost indistinguishable from the true one.

The total number of trades does not have a decisive effect on the accuracy of our measure, since its accuracy is determined first and foremost by the few early independent trades. The total number of trades, however, has an effect on the accuracy of the LSV measure due to its effect on the adjustment factor. Since the adjustment factor decreases with increasing number of trades, the upward bias increases for more intensively traded sets of stocks.

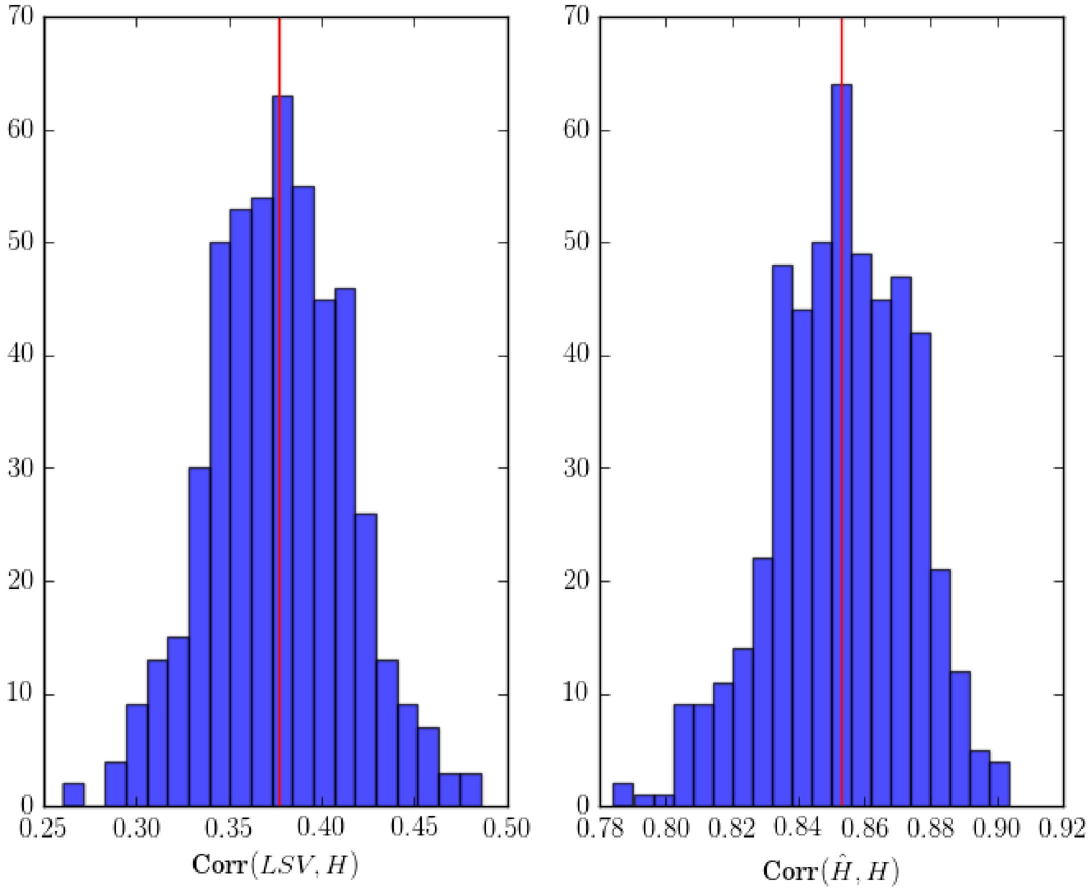
The overall impression obtained from Figure 2.1 and 2.2 suggests that while biased the LSV measure at least favorably correlates with H. This is indeed the case for the stark simulation setup presented here. The severity of the bias in the LSV measure, however, becomes more apparent in a more flexible simulation setup, in particular, when the degree of herding or contrarian behavior correlates with the dispersion in the independent buy-ratios.

2.4.2 Correlation between $\hat{H}$, LSV and H

Herding and contrarian behavior are usually viewed as events that relate to the overall trading environment. For example, herding models emphasize the role of uncertainty for the existence of herding and contrarian behavior and it is, therefore, often hypothesized that "greater uncertainty" leads to increased herding. In this section, we, therefore, want to analyze the performance of our measure and that of the LSV measure in terms of their correlation with the true (expected) deviation from independent trading when the overall level of herding and contrarian behavior depends on the simulation setup. The dependence is chosen in order to demonstrate the potential severity of the bias in the LSV measure for what might be seen as a minimum requirement of a measure of coordinated trading, namely that it correlates with the degree of the deviation from independent trading.

We simulate 250 trading days for a cross-section of 200 stocks. For each stock-day the number of trades is drawn uniformly from $\mathcal{T} = \{50, 51, \ldots, 500\}$. The independent buy-ratios are drawn from a Beta distribution with the parameters α, β being drawn each day from a Uniform distribution on $[2, 30]$. Again, we set $\alpha = \beta$ to have an expected independent buy-ratio of $\frac{1}{2}$. On each day, trading evolves according to the independent buy-ratios with probability 0.5. If a trading day is affected by dependent trading, the first $\tau = 10$ trades are conducted as under independent trading, after which trades are generated by the Friedman urn

Figure 2.3: Correlation with H



Notes: We simulate 250 trading days for a cross-section of 200 stocks. For each stock-day the number of trades is drawn uniformly from $\mathcal{T} = \{50, 101, \ldots, 500\}$. The independent buy-ratios are drawn from a Beta distribution with $\alpha \sim \mathrm{U}[2, 30]$ and $\alpha = \beta$. With probability 0.5 the day evolves under independent trading. If a trading day is affected by dependent trading, with equal change of herding or contrarianism, the first $\tau = 10$ trades are conducted as under independent trading, after which trades are generated by the Friedman urn model. In case of a herding day we set $h = 10(\alpha - 2)/(30 - 2)$ and $c = 0$, whereas for a contrarian day we set $h = 0$ and $c = 1 - (\alpha - 2)/(30 - 2)$. This figure plots the histograms of the correlation between the cross-sectional average of LSV and H, as well as of $\hat{H}$ and H for 500 simulations. The vertical red line indicates the mean correlation.

model with equal chance that the day will be governed by herding or contrarian behavior.

The parameters h and c of the Friedman urn model governing the deviation from independent trading will depend on the dispersion of information. The degree of herding will negatively correlate with the dispersion of the Beta distribution, capturing the idea that less precise information on the part of traders regarding the value of an asset (i.e. buy-ratios under independent trading close to 0.5) are associated with a stronger tendency to engage in herd behavior. The degree of contrarian behavior, on the other hand, positively correlates with the dispersion of the Beta distribution.[16] More precisely, in case of a herding day we set $h = 10(\alpha - 2)/(30 - 2)$ and $c = 0$, whereas in case of a contrarian day we set $h = 0$ and $c = 1 - (\alpha - 2)/(30 - 2)$.

For each of the 250 trading days we compute the cross-sectional average of LSV, $\hat{H}$ and H. We then compute the correlation between the daily measures of LSV and H, as well as between those of $\hat{H}$ and H. This is repeated 500 times. Figure 2.3 plots the histograms of these 500 correlations.

The figure shows that the correlation between LSV and the true measure of dependent trading is generally weak. The mean correlation across the 500 simulations is 0.38. That is, high LSV measures are only a weak indication of increased herd behavior. The correlation between $\hat{H}$ and H, on the other hand, is expectedly strong with a mean correlation of 0.85. A perfect correlation is hampered by the estimation uncertainty of α, β for finite samples.

2.5 Conclusion

We propose a new measure of herding and contrarian behavior to provide an empirical account of these terms as they are discussed especially in the market microstructure literature. This literature makes the link of coordinated behavior to potential financial market inefficiencies explicit by benchmarking the trading behavior that is subject to informational externalities of preceding trades against a counter-factual environment where traders would not have observed the trades of others. An empirical approach to this definition of investor coordination is, therefore, particularly worthwhile.

We build on the classical measure of investor coordination proposed by Lakonishok, Shleifer, and Vishny (1992) and adjust it in accord with general implication from the market microstructure literature. We show that our measure of buy-ratio dispersion accurately signals deviations from independent trading in the directions of herding ($H > 0$) and contrarian behavior ($H < 0$). Contrasting the performance of our measure to that of LSV, we find that the LSV measure provides very different results and should not be applied for measuring coordinated investor behavior of the type defined in the microstructure literature.

Since we do not provide an empirical application of our measure in this paper, a few remarks in that direction are in order. In principal, our measure can be

[16] Note that this is, in a sense, anyway the case, because contrarian behavior is bounded the closer to zero the less dispersed the Beta distribution. If we would let c correlate positively with (α, β), the correlation between LSV and H would be strongly negative, because it is determined by the lower bound of H and the behavior of the LSV bias in relation to α and β.

applied following the various applications of the LSV measure as long as one has an appropriate sample from which one can estimate the distribution of independent buy-ratios. Given, however, our focus on a different type of investor coordination than the one that underlies the LSV measure, we would recommend that the empirical application should differ from the typical example of the LSV approach (e.g. Wermers, 1999; Brown et al., 2014) as well.[17]

Our measure derives from the observational learning literature to capture the systematic impact, if any, of preceding traders on the trading decisions of subsequent traders. To that end, our measure is best applied at high frequency (e.g. daily), as in the simulated examples provided above, using the transactions of the active side of the complete order flow.[18] Ideally, the data would allow to identify the parties behind a transaction to mitigate effects of order-splitting on the measurement. An application of our measure following these suggestions is left for future research. We present, however, a first application that goes into a similar direction in Boortz (2016, chap. 4).

[17]The typical example is one of low-frequent transaction data of a specific investor group (e.g. quarterly equity holdings of mutual funds).

[18]That is, all trades that drive the price. These are typically assumed to be market orders or marketable limit orders, which are the counter-part of what is usually modeled as a market-maker.

Appendix 2

2.A Empirical Evidence on the Distribution of Independent Buy-Ratios

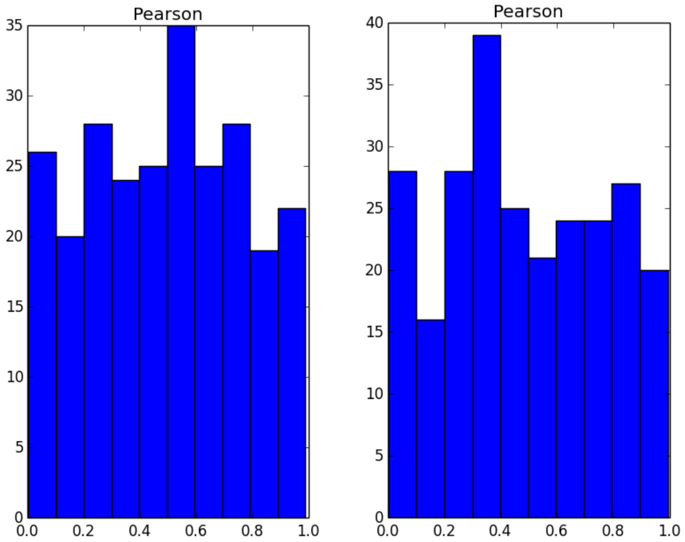
A first empirical application of our measure in collaboration with co-author Puriya Abbassi is presented in Boortz (2016, chap. 4). We will borrow some of our results presented there to underline the validity of our assumption on the Beta-Binomial distribution for the early independent buys.

We apply our measure to transaction data of Prime Standard stocks traded on XETRA during 2008.[19,20] The data allow us to distinguish between transactions conducted by trading institutions (i.e. institutions permitted to trade directly on any German exchange) for their own account and for their customers. Accordingly, we apply our measure to both groups separately. We use for both groups their market- and marketable limit-orders only, obtained by trade classification (see chap. 3), since non-marketable limit-orders enter the order book before they are executed and, thus, cannot be influenced by the trades of others that were executed in the mean time.

Because we are able to identify the trading institution behind each transaction, we summarize the trades for their own accounts into net-positions. For the customer trades, on the other hand, we count each trade individually. While one might debate the sensibility of one approach versus the other, it is important to note that our evidence on the validity of the Beta-Binomial distribution is robust to these choices.

[19]Prime Standard is a class of stocks that have to fulfill certain liquidity requirements and transparency standards set up by Deutsche Börse. For example, to be listed in the Prime Standard, companies have to submit quarterly reports in addition to half-year and year-end reports. The Prime Standard contains the most prominent German indices such as DAX, MDAX or TecDAX.

[20]XETRA is an electronic trading system that allows investors from all over the world to trade all stocks listed on the Frankfurt Stock Exchange from 9.00h to 17.30h CET. It is the largest stock exchange in Germany accounting for a market share of $> 90\%$.



(a) Customer trades **(b)** Institutional trades

Figure 2.A1: P-values from Pearson's Goodness-of-Fit test

Notes: This figure shows histograms of the p-values from Pearson's GoF tests. The test is applied to 252 estimated Beta-Binomial distributions. Each test tests whether the observed distribution of the number of buys from the first 10 trades fits to what we should expect under the estimated distribution.

For each of the 252 trading days in 2008 we take the first 10 trades (individual transactions or net-positions depending on the investor group) in each stock to estimate the Beta-Binomial distribution in order to uncover the distribution of the expected independent buy-ratios over the cross-section. To test the validity of our distributional assumption we apply to each estimated distribution the Pearson Goodness-of-Fit test (GoF) with the usual rule of thumb for the minimum number of observations in each bin. If all 252 test-statistics were drawn from the null hypothesis of beta-binomial distributed buys, the p-vales from the GoF tests would be uniformly distributed.

Figure 2.A1 shows histograms for the p-values from the GoF tests for both investor groups. A Kolmogorov-Smirnov test cannot reject the null that the 252 p-values of each group are uniformly distributed. The p-values from the Kolmogorov-Smirnov tests are 0.28 for the group of customer trades and 0.31 for the group of proprietary trades. This provides strong support for our assumption that the early buys follow a Beta-Binomial distribution. Testing the fit of the Binomial distribution, on the other hand, generally rejects the binomial distribution as an appropriate description of the data. More than 95% of the 252 tests from each group reject the null at a significance level of 0.05.

2.B H — A Generalization of LSV

PROPOSITION 2.1. *If the expected buy-ratios under independent trading are the same for all stocks and deviations under dependent trading cancel each other out over the considered cross-section of stocks, then our approach and the LSV approach asymptotically render the same degree of dependent trading, i.e.:*

If $\tilde{p}_i \equiv p^* \; \forall i$ *and* $\sum_{i=1}^{I} \varepsilon_i = 0$ *where* $\varepsilon_i = br_i - \tilde{p}_i$, *then* $\underset{I\to\infty}{plim}\ \hat{H}_i = \underset{I\to\infty}{plim}\ \widehat{LSV}_i$, *where* $\widehat{LSV}_i$ *is the estimated LSV measure.*

Proof. If $\tilde{p}_i \equiv p^* \; \forall i$, then in distributional terms, we have $\tilde{p}_i \sim \delta_{p^*}$ iid, where $\delta_.$ is the dirac-measure. Noting that $\lim\limits_{\alpha,\beta\to\infty} \text{Beta}(\alpha,\beta) = \delta_.$ and re-invoking the consistency of the maximum-likelihood estimator, we infer that $\hat{\alpha}, \hat{\beta} \underset{I\to\infty}{\to} \infty$, and

$$\lim_{I\to\infty} \frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}} = \lim_{I\to\infty} \frac{\sum_{i=1}^{I} B_i^{\tau_i}}{\sum_{i=1}^{I} \tau_i} = p^*.$$

As a consequence, we have that $f_p(\hat{\alpha},\hat{\beta}) \underset{I\to\infty}{\to} \delta_{p^*}$ and, hence,

$$\begin{aligned} \underset{I\to\infty}{plim}\ \hat{H}_i &= \mathbb{E}_{\hat{f}_p}|br_i - p| - \hat{AF}_i \\ &= |br_i - p^*| - \sum_{k=0}^{T_i} \binom{T_i}{k} (p^*)^k (1-p^*)^{T_i-k} \left|\frac{k}{T_i} - p^*\right|. \end{aligned} \tag{2.6}$$

Noting that $\sum_{i=1}^{I} \varepsilon_i = 0$ impliest $\underset{I\to\infty}{plim}\ \frac{\sum_i B_i}{\sum_i T_i} = p^*$, we conclude that the last line of Equation (2.6) equals $\underset{I\to\infty}{plim}\ \widehat{LSV}_i$, which is the desired result.

□

Proposition 2.1 states that our measure is equal to the LSV measure if the LSV assumptions hold. The reverse of the statement is also true for most of the cases: If the LSV assumptions do not hold, our measure H is generally very different from the LSV measure.[21]

Two additional remarks are in order. First, the result of Proposition 2.1 generalizes to any distributional assumption for the $\tilde{p}_i$, as long as we can estimate the

[21] One could, however, construct unlikely scenarios, where the reverse is not true. To see this, consider some T_i and $\alpha, \beta < \infty$. H_i attains it's minimum if $br_i = Median(\tilde{p}_i)$. This minimum is less than minus the adjustment factor of the LSV measure, i.e. $< -AF_i$. Now note that $LSV_i = -AF_i$ if for any $c \in (0;1)$, the observed buy ratios are $br_i \equiv c$ for all i. Moreover, $\exists br_i \in (0;1)$ such that $H_i > 0$. Since H_i is also continuous in br_i, the intermediate value theorem implies that $\exists c^* \in (0;1)$ such that $H_i = -AF_i$ if $br_i = c^*$ and, thus, $H_i = LSV_i$ even though the conditions of Proposition 2.1 are not met.

distribution consistently. Second, Proposition 2.1 also holds for the cross-sectional averages of $\hat{H}_i$ and $\widehat{LSV}_i$.

2.C Beta Distribution Fact Pack

This section provides some facts about the Beta distribution. For full details see Gupta and Nadarajah (2004).

The Beta distribution is a continuous distribution with support $[0;1]$. It is, thus, well-suited to model the realization of buy-ratios. The Beta distribution has two parameters $\alpha > 0$ and $\beta > 0$ that determine the shape of it's density. The Beta density is given by $p^{\alpha-1}(1-p)^{\beta-1}/\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\mathrm{d}u$.

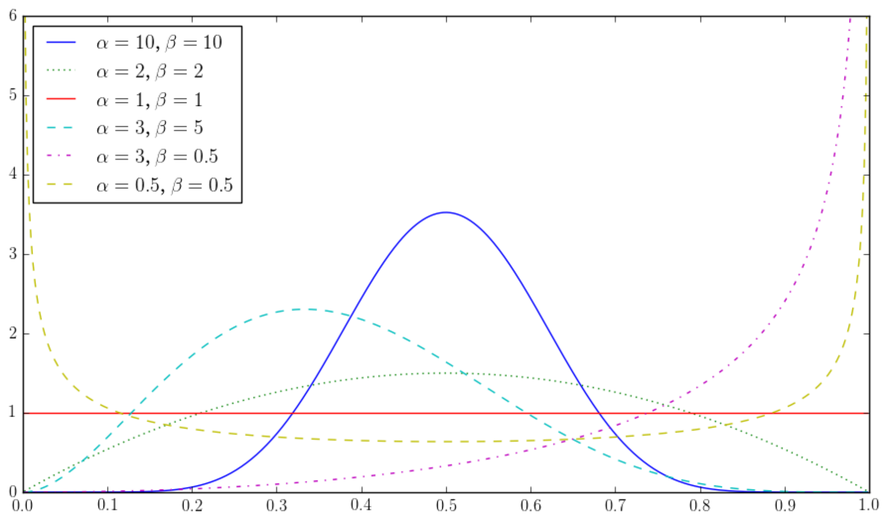
The expected value of the Beta distribution is given by $\alpha/(\alpha+\beta)$. The variance is equal to $(\alpha\beta)/[(\alpha+\beta)^2(\alpha+\beta+1)]$ and the skewness is given by $2(\beta-\alpha)\sqrt{\alpha+\beta+1}/[(\alpha+\beta+2)\sqrt{\alpha\beta}]$.

Figure 2.C1 illustrates how different parameters α, β affect the distributional shape. The larger α, β, the less disperse the distribution and vice versa. For $\alpha = \beta$, the distribution is symmetric around its mean 0.5. For $\alpha = \beta = 1$, the Beta distribution is identical with the Uniform distribution on $[0;1]$.

The cyan dashed graph ($\alpha = 3, \beta = 5$) shows a right skewed distribution, while the purple dotted-dashed line ($\alpha = 3, \beta = 0.5$) shows a strongly left skewed Beta distribution. If both parameters are less than 1, the density becomes u-shaped.

Finally note, that if the success probability of a binomially distributed random variable X is beta distributed, then X is beta-binomially distributed.

Figure 2.C1: Different beta densities



Paper 3

Inferring Trade Directions in Fast Markets

3.1 Introduction

The separation of securities market transactions into the orders of the liquidity demanding and supplying side is central to many financial research topics. For each buyer there is a corresponding seller and vice versa. Yet, only one side of the transaction holds the relevant information to which the price adjusts on its path to its efficient level. Typically, it is assumed that the liquidity demanding, impatient party of the trade is the informed side.[1] Market microstructure models, including their experimental counter parts, are designed such that traders learn from the trade direction of preceding liquidity demanders (Glosten and Milgrom, 1985; Park and Sabourian, 2011; Park and Sgroi, 2012). The imbalance of buyer- and seller-initiated trades is thus a prominent indicator of informed trading (Hasbrouck, 1991; Easley et al., 1996b; Cipriani and Guarino, 2014; Hu, 2014; Bernile et al., 2016). Furthermore, market quality is measured by the costs that the liquidity demanders incur relative to the efficient price, or by the price changes subsequent to their trades (Huang and Stoll, 1996). The sign of these measures is determined by the trade direction of the liquidity demander.

Information on the trade direction of the liquidity demander, however, is not readily available in common data sets. Instead, one has to rely on so called trade classification algorithms to infer the trade direction from the data. The radical changes of financial markets over the past 15 years, however, pose profound difficulties for the established methods. These methods base their classification on the proximity of the transaction price to the quotes in effect at the time of the trade.[2] Knowing the actual quotes, however, is difficult with today's high order

[1] Today, this assumption is more controversial than it used to be (see e.g. O'Hara, 2015; Easley et al., 2016). I discuss this topic in more detail in Section 3.9.

[2] This is true for the well-known algorithms of Lee and Ready (1991), Ellis et al. (2000) and Chakrabarty et al. (2007), but also for the less popular methods of Rosenthal (2012) and Blais and Protter (2012). The classification algorithm of Easley et al. (2012, 2016) is an exception. Their algorithm is, however, not directly comparable to the ones presented in this paper. I

submission and cancellation rates (Easley et al., 2016). Choosing the wrong quotes leads to erroneous classifications, which, in turn, can compromise the conclusions regarding the information content of trades, price efficiency or market quality.

In this paper, I propose a new method to classify transactions into the orders of liquidity demanders and suppliers and show that it outperforms the traditional alternatives, particularly under the conditions of fast markets that make trade classification so difficult for the established methods.

The established methods classify trades by first matching trades to their corresponding quotes based on the timing of the two. This is problematic for at least two reasons. First, with the increased frequency of order submissions and cancellations, the data often shows several quote changes occurring at the same time as the trade. It is then not clear which quote to select for the decision rule of the algorithm, and the wrong choice impedes its accuracy. For example, the Monthly Trade and Quote data (MTAQ[3]), which provides intraday trade and quote data from the consolidated tape of NYSE, AMEX, Nasdaq NMS and more listed stocks, is timestamped to seconds.[4] I find a median of 17 quote changes at the time of trades recorded at a precision of seconds, even for the data from Nasdaq alone.[5,6] The high market fragmentation characteristic of today's equity markets, however, makes it more than ever necessary to study data from the consolidated tape to obtain a fair view of the entire market (Holden and Jacobsen, 2014).

The MTAQ remains a popular database for research in equities (Hu, 2014; Bernile et al., 2016; Chordia et al., 2017, 2016). The Daily Trade and Quote data (DTAQ), however, which is timestamped to the millisecond, provides a now common alternative.[7] Still, with order submission and cancellation rates taking place at microseconds or faster even data timestamped to milliseconds will not be sufficiently precise (O'Hara, 2015).

Second, when trade and quote data is collected by different sources, there is a potential for lagging timestamps in one data set relative to the other. Consider

comment on that in more detail in Section 3.9.

[3]"MTAQ is the most popular intraday database for academic research in U.S. equities." (Holden and Jacobsen, 2014, p. 1748)

[4]The same is true for other internationally used databases, e.g. the equity transaction data of the German Financial Supervisory Authority studied in Kremer and Nautz (2013a), which contain all trades conducted on German exchanges.

[5]Angel et al. (2015) record an average of almost 700 quote changes per minute (i.e. more than 11 per second) for all stocks in the TAQ dataset at the peak in 2012. This number includes infrequently traded stocks and intra-day periods of low traffic. The number can be expected to be much higher if one would only count the seconds at which trades occurred.

[6]Low timestamp precisions is not exclusive for equity data. Bernile et al. (2016) classify futures transaction data that is timestamped to the second.

[7]From August 2015 onwards DTAQ data is timestamped to microseconds, from October 2016 onwards to nanoseconds for Nasdaq.

again the trade and quote data from the Consolidated Tape Association. While quotes are collected via the Consolidated Quotation System in Brooklyn, trades are collected via the Consolidated Tape System located in lower Manhattan. As the timestamp is added after the processing of the data at the respective data center, with additional data error checking, there is room for misalignment between trades and their respective quote changes. Although, since they were first noted by Lee and Ready (1991), technical advances have decreased the potential for such misalignments, there is still disagreement by how much quote times should be lagged in the MTAQ.[8,9] As the degree of misalignment is typically unknown and probably varying over time and across security, there is again a risk of using the wrong quotes in the classification procedure by simply matching trades and quotes by their timing, especially with many quote changes occurring over the time delay.

The algorithm proposed in this paper takes a new approach to the issues of imprecise and misaligned timestamps that deviates from the previous methods in two fundamental ways. Instead of selecting a single pair of ask and bid quotes *before* the actual classification step, it matches the transaction to its corresponding quote at the same time as it is classified. The idea is that a trade executed against the ask must leave its footprint on the ask-side, while a trade against the bid must leave its footprint on the bid-side. Finding these footprints is equivalent to *simultaneously* finding the quote corresponding to a trade *and* classifying it. Second, the algorithm uses more than the information contained in prices in classifying a trade. The algorithm considers all quotes that are potential candidates for a match based on their timing and then reduces the potential candidates based on price and volume information. The first step circumvents the problem of not knowing the actual trade-quote correspondence and the second one allows for many unambiguous assignments despite the potentially high number of quotes considered.

To evaluate the new algorithm against the alternatives routinely applied in the literature, the Lee and Ready (1991) (LR), the Ellis et al. (2000) (EMO) and the Chakrabarty et al. (2007) (CLNV) algorithm, I use data from Nasdaq's electronic

[8]Chakrabarty et al. (2012) recommend to lag quotes by 1 second, Henker and Wang (2006) recommend to use the last quote from the second before the trade, Piwowar and Wei (2006) and Vergote (2005) find optimal delay times for quotes between 1 and 2 seconds, Peterson and Sirri (2003) and Bessembinder (2003) recommend a 0 lag for quotes, though they consider only 5 seconds intervals ranging from 0 to 30 seconds. Reviewing all published papers between 2006 and 2011 in the Journal of Finance, Journal of Financial Economics and the Review of Financial Studies, Holden and Jacobsen (2014) find that in 28 articles using the MTAQ data 7 used the prior-second rule, 3 the same-second rule, 5 the five-second rule and rest did not provide information on the timing-rule used to match quotes and trades.

[9]I also find some indirect evidence that the misalignment problem is not fully amended even in the DTAQ data. I elaborate on this issue in Section 3.5.

limit order book. The data contain all transactions against standing (visible or hidden) limit orders as well as the development of the best bid and ask prices (including the depths at the quotes) for the three month May to July 2011 with a total of over 134 million transactions. The important feature of this data set is that it contains the trade direction of the executed standing order in the limit order book. Hence, the liquidity supplying and demanding side for each transaction is known, which allows us to evaluate the ability of the algorithms to recover this information from the trade and quote data.

The Nasdaq data, of course, do not contain the same number of trades and quote changes as, for example, the consolidated tape and possibly other high-frequency databases.[10] This is, however, not a problem *per se* as we are interested in the effect of high order submission and cancellation rates *relative* to the data timestamp precision. To simulate this problem I simply truncate the timestamp precision at frequencies ranging from nanoseconds to seconds. This corresponds to a median number of quote changes during the time of trades ranging from 1 to 17. To analyze the problem of lagging transaction timestamps (relative to the timestamps of the quote changes), on the other hand, I add exponentially distributed noise to the original trade times.

The results provide a clear message: the new algorithm outperforms the traditional trade classification algorithms. First, at every considered timestamp precision the new algorithm does not perform worse than the others and it offers considerable improvement in classification accuracy at lower timestamp precisions. For example, for the data with a median of 17 quote changes during the time of a trade (i.e. timestamped to the second) the new algorithm correctly classifies the trade initiator for 95% of the trading volume, whereas the best competitor, the EMO algorithm, classifies 90% of the trading volume correctly.

Second, the ability of the new algorithm to provide accurate classifications at low timestamp precisions provides a simple and effective way to counteract the adverse effects of delayed trade times. Applying the algorithm to the data timestamped at seconds still yields 94% correctly classified volume for an average delay of up to one centisecond. The traditional algorithms, on the other hand, achieve an accuracy of only around 89% correctly classified volume under the same setup.

Third, the improved accuracy of the new algorithm translates into considerable improvements in the estimation of the dollar effective spread, the dollar price impact, the dollar realized spread and order imbalances—common measures where

[10] For example, Angel et al. (2011) report that Nasdaq's market share in Nasdaq-listed stocks decreased from 53% in April 2005 to around 30% in April 2009.

knowledge of the trade direction of the liquidity demander is key. For example, with a median of 17 quote changes at the time of a trade, the average deviation of the daily estimate of the realized spread from the true value is 0.9¢ per share compared to 1.41¢ for the best competitor. With each stock-day split into 10 equal volume bins, the average deviation of the estimated order imbalance around the true one is only 6.3%-points compared to 10.1%-points of the best competitor.

The main results are derived under the assumption that each transaction against a visible order is reflected in the order book by a corresponding change in volume at the respective quote. Moreover, trades and quotes are assumed to be in correct order. These assumptions may not hold in other data sets. In particular, due to the different latencies of the exchanges to the consolidated tape, trades and quotes from the consolidated tape can be out of order over short intervals. Changes in the data structure, however, affect the information content of volume that the algorithm can use in the classification procedure.

Therefore, I relax one-by-one the initial assumptions on the data structure, and present the appropriate adjustments to the new algorithm. The empirical exercises are carried out equivalently to the main section. The general conclusion does not change. Although the algorithm offers less improvement the less information we can draw from the data, it generally outperforms the competitors. Even under the minimum of data structure, the new algorithm improves the mean classification accuracy by 3 percentage points for the data timestamped at seconds. In particular, the improvement in the estimation of the liquidity measures and order imbalance remains robust.

The remainder of the paper is structured as follows. Section 3.2 introduces the established algorithms of Lee and Ready (1991), Ellis et al. (2000) and Chakrabarty et al. (2007), followed by Section 3.3, which introduces the new algorithm proposed in this paper. Section 3.4 presents the data used to evaluate the algorithms. Section 3.5 presents the main results. The results from the estimation of measures of liquidity and order imbalances are presented in Section 3.6. To get more insight into the determinants of misclassification by the new algorithm, Section 3.7 presents the results from a logistic regression of the event of a correct classification on a number of covariates. Section 3.8 presents robustness check against varying assumption on the data structure. Finally, Section 3.9 discusses the relation of the order imbalance constructed from the classification results with informed trading, and Section 3.10 concludes.

3.2 The LR, EMO and CLNV Algorithm

3.2.1 The Decision Rules

The LR algorithm (Lee and Ready, 1991) is the most popular choice to classify trade data into the orders of the liquidity demanding and supplying sides. It compares the transaction price to the mid-point of the ask and bid quote at the time the trade took place. If the transaction price is greater (smaller) than the mid-point the liquidity demanding side is the buyer (seller), i.e. the trade is buyer (seller) initiated. If the transaction price is equal to the mid-point, the trade initiator is assigned according to the tick-test. That is, if the transaction price is greater (smaller) than the last price that is not equal to the current transaction price, the trade was buyer (seller) initiated.

The algorithm can be rationalized by the market structure where marketable buy-orders trade against the standing offer at the ask, and marketable sell-orders against the standing bid. There is also, however, an underlying economic motivation. The liquidity demanding party requires immediate execution of the order. This impatience comes at a price, the "immediacy premium" (Asquith et al., 2010), which should put the transaction price above the mid-point for an impatient buyer and below for an impatient seller. From the view of the uninformed patiently providing liquidity, the bid-ask spread compensates for the risk of trading against the informed.

The most notable alternatives to the LR algorithm are the algorithms proposed by Ellis, Michaely, and O'Hara (2000) and Chakrabarty, Li, Nguyen, and Van Ness (2007). The EMO algorithm classifies a trade as buyer (seller) initiated if the transaction price is equal to the ask (bid) price. For all trades off the quotes the tick-test is used. The CLNV algorithm assigns the liquidity demander to the buying (selling) side if the transaction price is equal to the ask (bid) or up to 30% of the spread below (above) the ask (bid). For all trades above (below) the ask (bid) or within a 40% range of the spread around the mid-point the tick-test is used. Table 3.B1 in the Appendix summarizes the classification algorithms in terms of pseudo codes.

3.2.2 Quote-Matching Rules

The LR, EMO and CLNV algorithms require assigning one bid and ask quote to each trade in order to classify it. In an ideal data environment where at the time of the trade we record only one quote change, we know that the quotes in effect at the time of the trade are the last ones recorded before the time of the trade.

With several quote changes occurring at the same time as the trade, however, it is not clear which quotes to select for the classification procedure. For example, with one trade and three quote changes recorded at the same millisecond, the quotes corresponding to the trade could be the last quotes from before the millisecond or one of the first two recorded at the millisecond.

The convention in such a case is to take the last ask and bid price from before the time of the trade.[11] An alternative, recently suggested by Holden and Jacobsen (2014), advises transforming the data first to correspond to the ideal environment. This is achieved by interpolating the recorded times according to the number of trades or quotes during that time. For example, for trades recorded at seconds, the interpolated time t is computed by

$$t = s + \frac{2i-1}{2I}, \quad i = 1, \ldots, I$$

where s is the recorded time, and I is the number of trades at time s. The algorithms then use the last ask and bid price from before the time of the trade according to the interpolated time.

Another reason that the assignment of trades to quotes is difficult, is a misalignment between the timing of trades and their corresponding quote change. In particular, for the trade and quote data from the consolidated tape of NYSE and AMEX listed stocks, it was found that quote changes were recorded ahead of the trades that triggered them (Lee and Ready, 1991). The delay was caused by a different use of floor reporters and an electronic display book in reporting trades and quotes (see Lee and Ready, 1991, p. 737 and Vergote, 2005). With changes in the reporting procedure and the full reliance on an automated electronic procedure the potential for a reporting delay in trade times diminished. However, Vergote (2005) finds that even after the abolishment of the floor reporters and a full reliance on the automated Display Book quotes seem to lead trades by 2 seconds. With a geographic separation of the processing stations of the quote and trade data and timestamps that reflect the end of the processing of the data at the respective processing station, both of which applies to the Consolidated Tape

[11] Another suggestion is to take the first ask and bid price at the time of the trade. For two reasons, however, I do not recommend such a matching-rule. First, if there is only one quote change recorded at the time of the trade, the quotes corresponding to the trade are the ones recorded before the time of the trade. The quotes recorded at the time of the trade are new quotes that resulted from the trade. Second, the mid-point needed for the LR and CLNV algorithms can be invalid using the first bid and ask quote if the order book data is not symmetrically constructed in the sense that for each ask entry there is also a bid entry and vice versa. For example, if the ask price is updated before the bid price, the correct mid-point would be the average of the first ask at the time of the trade and the last bid from before the time of the trade, but not the average of the first ask and first bid.

Association (see Holden and Jacobsen, 2014, p. 1753) which provides the MTAQ and DTAQ data, there remains a potential for misaligned timestamps.

The common procedure to account for the misalignment, is to lag the time of quotes by the amount of the suspected delay in the reporting of the trades and then to match each trade with the last quote from before the time of the trade. The delay is usually inferred from indirect evidence, e.g., by choosing the timing-rule that minimizes the occurrence of trades off the quotes (Bessembinder, 2003; Piwowar and Wei, 2006), or by observing the frequency of quote revisions around isolated trades (Lee and Ready, 1991; Henker and Wang, 2006).[12] For trade and quote data from the nineties the typical choice is to lag quotes by 5-seconds. For data from more recent periods there is considerable disagreement about how much to lag quotes (Chakrabarty et al., 2012; Henker and Wang, 2006; Piwowar and Wei, 2006; Peterson and Sirri, 2003; Bessembinder, 2003 and see footnote 27 in Holden and Jacobsen, 2014).

3.3 The Full-Information Classification Algorithm

Selecting a single ask and bid quote to be used in classifying a transaction is likely to induce errors in the classification results under the described data deficiencies. The algorithm proposed here aims to reduce the number of erroneous classifications by allowing for more than one ask and bid quote to be considered in the classification of a trade. When there are, for example, three ask prices that could have been in effect at the time of the trade, either because of imprecise timestamps or because they are all in the potential range of the reporting delay, we may want to consider all three of them and use the full information provided by the transaction price and volume to derive the classification.

To understand how we can use price and volume information to determine the trade-quote correspondence we need to make some assumptions about the data structure.

Data Structure 1.

(i) *Each transaction against a visible order leads to a corresponding reduction in volume available at the respective quote.*

(ii) *Trades and quotes are reported in the correct order.*

At first glance, assumption (i) seems probably harmless. We would certainly expect the order book to display that kind of information when a market order

[12] More involved solutions to the timing problem consist of parametric estimations of the optimal delay time (Vergote, 2005; Rosenthal, 2012).

trades against a single limit order. Consider, however, a market order that is too large to be filled by a single limit order. The assumption then states that the order book displays the successive steps in the completion of the market order. That is, if a market buy order for 100 shares trades against two limit orders for 50 shares each, the order book will first show a reduction of 50 shares at the bid and then another reduction of the same size, even though these changes happen basically instantaneously. Though this degree of detail is provided in the data set that I use here, we may not expect the same of every other data set.

Assumption (ii) is certainly harmless when we use data from a single exchange. For the data from the consolidated tape, however, due to the different latencies of the exchanges to the tape, trades and quotes can be out of order over short intervals of time.

I will later relax the assumptions made here and discuss the adjustments to the Full-Information algorithm. For now, assumptions (i) and (ii) mean that we can use the exact transaction volume to determine whether a trade could have been executed at a particular quote. That is, if, from among the available bid quotes that could have been in effect at the time of the trade, we cannot find any quote that matches the transaction price and where the decrease in volume matches the transaction volume, we can confidently ascertain that the transaction did not execute against the bid. If, on the other hand, we find such a quote among the candidate asks, we would conclude that the trade executed against at the ask side and is, thus, buyer-initiated.

Having established a rough idea of the algorithm, let me now describe the exact procedure. For that, I will introduce some additional notation for the transaction and ask data. The notation for the bid data follows analogously to that of the ask.

Notation

- Transaction index: $i \in \{1, \ldots, I\}$
- Transaction price and volume: p_i and v_i
- Recorded transaction time: s_i
- Ask quote index: $j \in \mathcal{J}_a = \{1, \ldots, J_a\}$
- Ask price and volume: a_j and v_j^a

- Change in volume at the j-th ask to the next:

$$\Delta v_j^a = \begin{cases} v_j^a - v_{j+1}^a & \text{if } a_j = a_{j+1} \\ v_j^a & \text{if } a_j < a_{j+1} \\ -1 & \text{otherwise,} \end{cases}$$

  Explanation: If the ask price increases from j to $j+1$ ($a_j < a_{j+1}$), then all of the volume at that the j-th ask must have disappeared (either because of a trade or because the order was canceled), and hence $\Delta v_j^a = v_j^a$.[13] If the ask price decreases from j to $j+1$ ($a_j > a_{j+1}$), then a new sell order must have been submitted with a better limit price than that of the j-th quote. So a trade cannot have taken place at the j-th quote. This is indicated by -1.

- Recorded time of an ask: s_j^a. This indicates the time from which point on the j-th ask price and volume determine the best visible ask.
- The collection of ask quote indices with the same timestamp s: $\mathcal{N}_s^a = \{j \in \mathcal{J}_a \colon s_j^a = s\}$.
- Interpolated time of an ask: $t_j^a = s_j^a + n_j^a/(|\mathcal{N}_s^a| + 1)$ with $n_j^a \in \{1, \ldots, |\mathcal{N}_s^a|\}$
- Auxiliary variable: l^a. This will be used to approximate the ask quote at the time of an execution against a hidden order
- Trade direction of the liquidity demanding party: o_i (1 for buy, -1 for sell)

The algorithm works as follows (see Figure 3.1 for a graphic representation):

Step 1 – Quote Selection and Matching: Starting with the first trade $i = 1$, we collect all ask and bid quotes against which the trade could have been executed only by considering the timing, i.e. for the ask

$$\mathcal{C}_a = \max\{j \in \mathcal{J}_a \colon s_j^a < s_i\} \cup (\mathcal{N}_{s_i}^a \setminus \max \mathcal{N}_{s_i}^a).$$

These are the last quotes from before the time of the trade and all but the last quote at the time of the trade. We initialize the variable $l^a = a_k$ with $k = \min \mathcal{C}_a$. Analogously, we obtain $\mathcal{C}_b$ and l^b for the bid.

Using transaction price and volume, search for the first match among the selected ask and bid quotes:

$$\boldsymbol{\alpha} = \min\{j \in \mathcal{C}_a \colon p = a_j \text{ and } v = \Delta v_j^a\}.$$

Analogously we obtain the first match among the bid quotes denoted $\boldsymbol{\beta}$.

[13] For the bid quote, the second case reads "if $b_j > b_{j+1}$".

Step 2 – Unique Match: If we find a match among the ask quotes, but not for the bid quotes the trade has been executed on the ask and the liquidity demander is the buyer. Go back to Step 1 and proceed with the next trade. If the next trade is recorded at the same time as the current one we use the same collection of ask and bid quotes and update l^a to a_α. (If we find that there is no match among the ask quotes, but at least one among the bids, all updates are made analogously.)

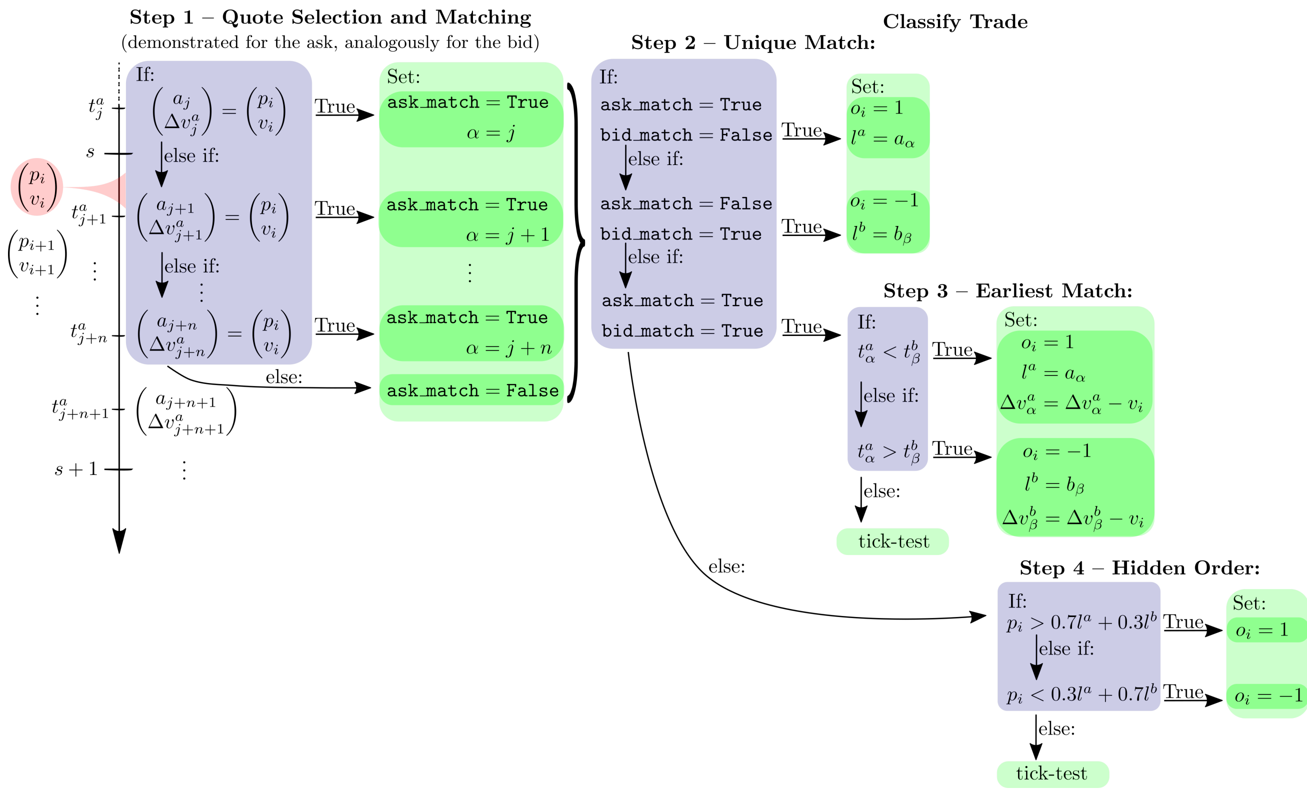
Step 3 – Earliest Match: If we find a match among both the ask and bid quotes, we classify the trade according to which quote seems to be affected first. That is, if $t^a_{\boldsymbol{\alpha}} < t^b_{\boldsymbol{\beta}}$, then the liquidity demander is the buyer. We go back to Step 1 and proceed with the next trade. Again, if the next trade is recorded at the same time, we update l^a to $a_{\boldsymbol{\alpha}}$ and use the same collection of ask and bid quotes, except that we omit the $\boldsymbol{\alpha}$-ask from further comparisons by subtracting from $\Delta v^a_{\boldsymbol{\alpha}}$ the size of the transaction v_i. If $t^a_{\boldsymbol{\alpha}} > t^b_{\boldsymbol{\beta}}$, then the liquidity demander is the seller and updates are made accordingly.

Step 4 – Hidden Order: If we cannot find a match among the ask and bid quotes, we are likely to face a trade against a hidden order. These are classified according to their position in the spread similar to Chakrabarty et al. (2007). If $p_i > 0.7l^a + 0.3l^b$ and $l^a > l^b$ the trade is buyer-initiated. If $p_i < 0.3l^a + 0.7l^b$ and $l^a > l^b$ the trade is seller-initiated. Go back to Step 1 and proceed with the next trade.

Step 5 – Tick-test: Any trade that could not be classified in **Step 2** to **Step 4** is classified by the tick-test.

Remarks The idea of using the interpolated time in **Step 3** to classify trades that match with both an ask and bid quote is as follows. Observing ask and bid quotes to equal each other within the same, say, second of a trade, may be due to the price impact of the trade. That is, quotes are updated in the direction of the trade initiator. This should be reflected in a relatively early interpolated time of the corresponding quote for the following reasons. First, in case of a buyer-initiated trade we may expect more activity on the ask side because of the information contained in the trade that leads to the price impact. Traders will either submit buy orders to take advantage of stale limit orders or cancel their stale limit orders in response to the trade. Either way, $|N^a_s|$ will be relatively large. Second, in case of a buyer-initiated trade, the trade executed *first* on the ask and then bid quotes were updated subsequently upwards. That is, $\boldsymbol{\alpha}$ will be relatively small while $\boldsymbol{\beta}$ relatively large. In total, this means that $t^a_{\boldsymbol{\alpha}}$ will be smaller than $t^b_{\boldsymbol{\beta}}$.

Figure 3.1: The Full-Information classification algorithm



Notes: This Figure shows the process of the Full-Information algorithm to classify a trade. The variables are defined in the **Notation** list. In **Step 1** we collect all ask and bid quotes against which the trade could have executed considering only the timing of the trade and the quotes. Starting with the first ask and bid quote, respectively, from these collections we search for an exact match of the quote and its volume change with the transaction price and volume. If a match could be found, we set an indicator variable to `True` and memorize the index of the respective quote. In **Step 2**, if only the ask/bid side matches the trade, we classify it as buyer-/seller-initiated and assign the respective quote to the auxiliary variable l^a/l^b, which is used to construct the spread in case of hidden order executions. In **Step 3**, if both sides match the trade, it is classified according to the interpolated time of the matched quotes. The corresponding quote is then omitted from further proceedings by subtracting the transaction volume from the volume change at the quote. In **Step 4** the trade is classified by the position of the price within the spread, which is approximated by the auxiliary variables. Trades not classified in any of these steps are classified by the tick-test.

To avoid further conflicts between ask and bid quotes with the price and volume characteristics, the quote to which the trade is matched is omitted from assignments of subsequent trades. This assumes that a quote can only be hit once, which means that we eliminated the counter-party to each trade. The counter-party is easily identified by the neighboring trade with the same price and volume, but opposite trade direction. Alternatively, if the counter-party is not omitted from the data for the classification process, we drop the corresponding quote after it has been assigned twice to a trade.

The spread in Step 4 is constructed from the auxiliary variables l^a and l^b. They serve to approximate the ask and bid valid at the time of the execution of the hidden order. They are initialized to the first ask and bid valid during the time of the trade. If we are able to classify a trade involving a visible order, the corresponding auxiliary variable is updated. In that way, we obtain a better approximation of the spread at the time of the hidden order execution due to the correct order of the trades.

I follow the design of the traditional algorithms and use the tick-test to classify the most ambiguous cases. This can be motivated by the finding of Perlin et al. (2014) who show that the misclassification rate of the tick-test is upper-bounded by 50% so that for large enough samples one is not worse off than by using a coin flip, but may have a chance to do better.

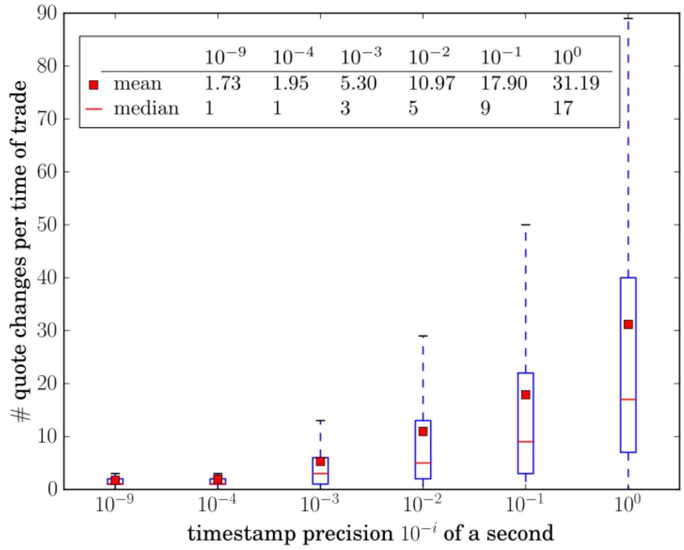
3.4 Data

The evaluation of the algorithms is based on equity trading in Nasdaq's electronic limit order book. The sample is constructed from Nasdaq's TotalView-ITCH data.[14] The trade data contain all transactions against visible and hidden limit orders with information on the price and volume of the transaction. The order book data contain the development of the order book. That is, whenever a visible limit order that affects the best quotes is submitted, canceled (partially or completely) or executed, the order book contains an entry of the best bid or ask indicating the new price and volume available. Changes regarding hidden orders are not displayed in the order book.

The data covers the continuous trading phase from 9:30 am to 4 pm for all trading days during the 3 month period May to July 2011. I selected the 30 largest

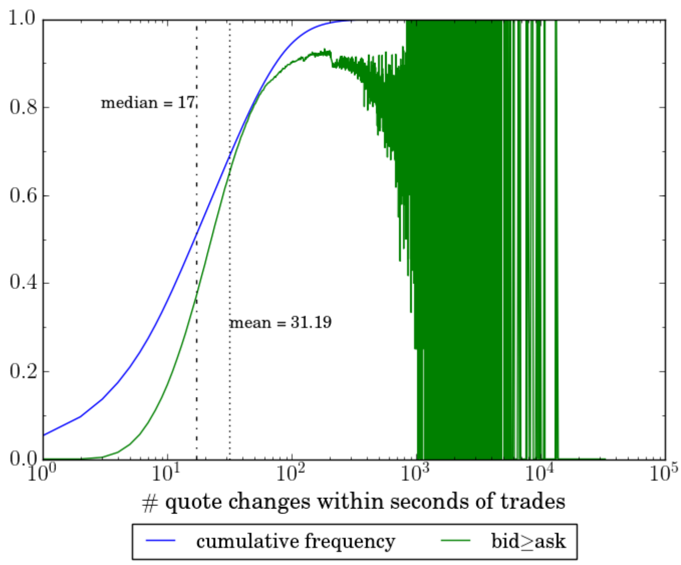
[14]The reconstruction from the TotalView-ITCH data is done by the software LOBSTER, which in turn produces the order book data and messages files containing the information on the events causing the changes in the order book. A detailed description of how I obtain the trade and quote data from these files is given in the Appendix.

Figure 3.2: Distribution of the number of quote changes at different frequencies



Notes: For each time where at least one trade takes place, I count the number of quote changes with the same timestamp. This figure shows the distribution of these counts in terms of boxplots for different timestamp precisions. For example, 50% of the milliseconds with at least one trade also display 3 or more quote changes.

Figure 3.3: Distribution of the number of quote changes at seconds and the frequency of crossing quotes



Notes: At each second where at least one trade takes place, I count the number of quote changes with the same timestamp. The blue line shows the cumulative frequency of these counts. For example, 68% of the seconds with at least one trade experience 31 quote changes or less. The green line displays the fraction of cases where one of the bids is at least as high as one of the asks for a given number of quote changes at the second of the trade.

stocks (by market capitalization) in 2015 from the 11 Nasdaq industry sectors.[15] Following Chakrabarty et al. (2015), I drop stock-days with an end-of-day price of less than one dollar or with less than 10 trades, which leaves me with a total of 19842 stock-days. A list with all ticker names studied in this paper is provided in Table 3.B2 in the Appendix. Table 3.A1 in the Appendix provides some summary statistics.

The quotation frequency, i.e. the number of quote changes at a given time of a trade at a given timestamp precision, is important for the main analysis, but not the timestamp precision *per se*. Figure 3.2, therefore, plots the distribution of the number of quote changes at the times of trades for different timestamp precisions: the original precision of nanoseconds (10^{-9} of a second), as well as 10^{-4} to 10^{0} of a second.

At a precision of 10^{-4} or less, most of the trade times have only a small number of quote changes with the same timestamp allowing us to match trades to their quotes based only on the timing of the two. With decreasing timestamp precision, however, the frequency of high numbers of quote changes at trade times increases quickly. At a precision of seconds, the median number of quote changes with the same timestamp as that of a trade is 17. With 17 quote changes occurring at the same time as a trade, we cannot deduce, just from the timing of the two, which quote belongs to which trade.

Figure 3.2 omits extreme values for illustrative purposes. Figure 3.3, therefore, gives a closer account of the distribution of the number of quote changes at the time of trades for the data timestamped to the second. We can see that a number of quote changes as high as 100 or more over an interval of one second occur more than 5% of the time. The figure also displays the fraction of cases where, during the second of a trade with a given number of quote changes, one of the bid quotes is at least as high as one of the ask quotes. We see that for the median number of quote changes at a given trade time, in 38% of the cases one of the bid quotes is at least as high as one of the asks. In all these cases, the wrong choice of a quote can easily lead to a wrong classification of the trade.

[15]These sectors are: Basic Industries, Finance, Capital Goods, Healthcare, Consumer Durables, Consumer Non-Durables, Public Utilities, Consumer Services, Technology, Energy, Transportation, Miscellaneous.

3.5 Results

3.5.1 Classification Accuracy at Different Timestamp Precisions

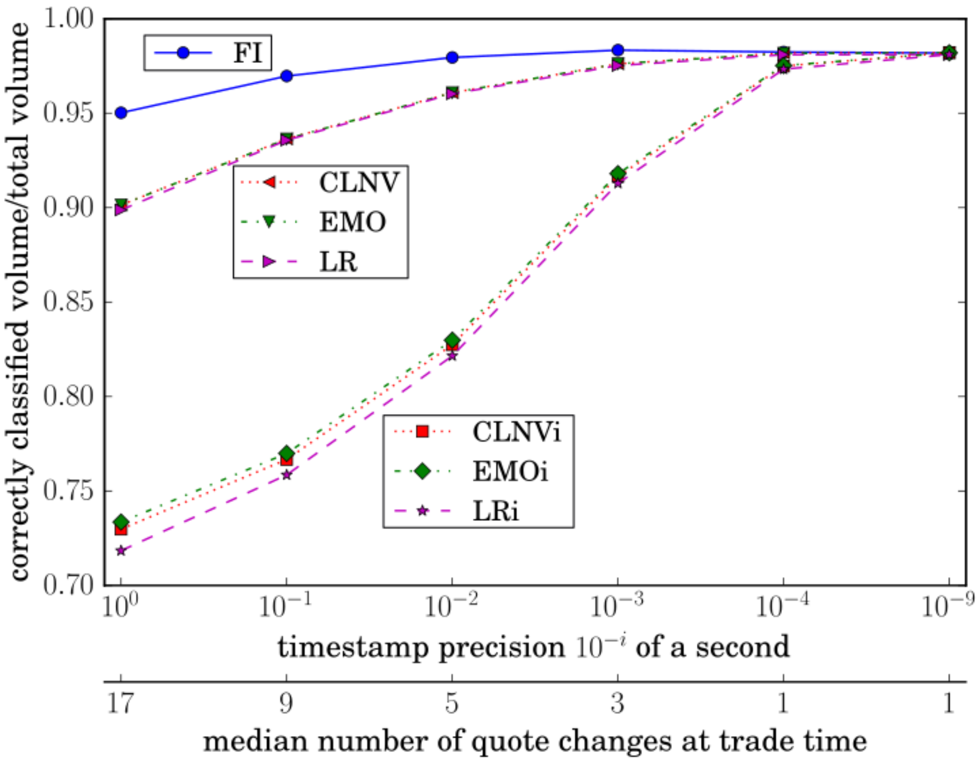
I analyze the improvements we can achieve over the traditional algorithms by applying them and the Full-Information algorithm (FI) to the data with varying timestamp precisions. The traditional algorithms are used in combination with the quote matching rule of using the last quotes from before the time of the trade (denoted by LR, EMO and CLNV), and using the interpolated time of trades and quotes (denoted by LRi, EMOi and CLNVi). The timestamp precisions are chosen to be of 10^{-i} for $i = 0, 1, 2, 3, 4$ of a second, as well as the original data precision of nanoseconds (10^{-9} of a second). These precisions correspond to a median (average) number of quote changes at the time of trades of 1 (1.73), 1 (1.95), 3 (5.3), 5 (10.97), 9 (17.9) and 17 (31.19) going from nanosecond to second precision. I evaluate the quality of the algorithms on the basis of correctly classified trading volume.

Figure 3.4 presents the sum of correctly classified trading volume over total trading volume of the entire sample for each algorithm and each of the timestamp precisions. Table 3.C1 in the Appendix provides the corresponding numbers, along with the means and standard deviations across the sample.

The results show that the FI algorithm dominates the others. At the original timestamp precision of nanoseconds all of the algorithms correctly classify around 98% of trading volume. Approaching the timestamp precision of seconds, however, the performance of the traditional algorithms falls off more quickly than that of the FI algorithm. At the precision of seconds the traditional algorithms correctly classify around 90% of trading volume (around 73% when using the interpolated time), in contrast to 95% correctly classified volume by the FI algorithm. That is, the FI algorithm reduces the number of misclassified shares by half.

Table 3.C1 in the Appendix shows that the FI algorithm also dominates in terms of the variation in classification accuracy. While the standard deviation of the stock-day classification accuracy barely moves for the FI algorithm (from 2.09%-points at nanosecond to 2.4 at second precision), the standard deviation of the traditional algorithms increases from the same level of around 2.1%-points to more than 3.4.

Figure 3.4: Classification accuracy at different timestamp precisions



Notes: This Figure depicts the fraction of correctly classified trading volume by the FI algorithm and the traditional algorithms using the last quotes from before the time of the trade (EMO, CLNV and LR), and using the interpolated time of trades and quotes (EMOi, CLNVi and LRi). The algorithms are applied to the data with reduced timestamp precisions (10^{-i} of a second for $i = 0, \ldots, 4$), and using the original precision of nanoseconds (10^{-9} of a second). These correspond to median number of quote changes at the time of trades ranging from 17 (for $i = 0$) to 1 (for $i = 9$).

3.5.2 Comparison to Previous Studies

The traditional algorithms, using the common quote matching rule of taking the last quote from before the time of the trade, perform better than documented in most of the past studies. Traditionally, the classification accuracies are in the range of 75-90% (see e.g. Chakrabarty et al., 2007; Theissen, 2001; Finucane, 2000; Ellis et al., 2000). The differences in classification accuracies result, in large part, from applications of the algorithms under different market structures and different ways of identifying the trade initiator. For example, Ellis et al. (2000) study a dealer-market and identify the trade initiator by the relationship of the parties involved in the trade, e.g., in a customer-dealer trade the customer is the trade initiator, because the dealer is supposed to cater to the demand of the customer. In markets where dealers play a larger role, however, there is more room for individual deviations from a standard procedure of matching orders, which are not captured by the algorithms, and dealers may not always trade passively to manage their inventory.

In contrast to many other studies, the liquidity demanding and supplying parties are unambiguously identified in the present data set, and the results presented here show that the traditional classification algorithms are well suited to distinguish these parties in the environment of the simple and consistent mechanisms of an electronic limit order book, as long as the number of quote changes at the time of trades is not too high.

The study that is closest to the analysis in this paper is Chakrabarty et al. (2015) who study the accuracy of the LR algorithm for transactions from Nasdaq's ITCH data over the same time span. The comparison with their study is interesting, because they use the same data source for the transactions, but the quotes from the DTAQ. That is, a comparison of the classification accuracy of the LR algorithm can tell us something about the quality of the DTAQ quote data.

The classification accuracy of the LR algorithm is 10 to 15%-points lower than reported here, even though Chakrabarty et al. (2015) aggregate the classification results over time intervals such that opposite misclassifications cancel each other out.[16] Importantly, these results do not seem to be merely driven by possibly asynchronous timing of the transactions from the ITCH data and the quotes from the DTAQ. Chakrabarty et al. (2015) match the trades from the ITCH data to the trades from the DTAQ data and repeat the classification exercise solely for the trade and quote data from the DTAQ, but the classification accuracy remains overall low. That is, the DTAQ data apparently still pose substantial problems for accurate trade classification by the traditional algorithms.

3.5.3 Explaining the Performance of the Interpolation Method

The results show that interpolating the trade and quote times to circumvent the problem of imprecise timestamps is not a fruitful alternative to the traditional quote matching approach. The idea behind the interpolation of trade and quote time is that trades and quotes are equally distributed over a given interval, e.g. a second. In that case, the best guess to when these trades and quote changes took place would be to distribute them equally over the interval.

[16]See Panel A and B of Table 1 in Chakrabarty et al. (2015, p. 60). They aggregate the classification accuracy over different time intervals, because they compare the accuracy of the LR algorithm to that of the algorithm of Easley et al. (2016), which is not meant to classify single trades and does not serve quite the same purpose as the traditional classification algorithms. The numbers best comparable to those presented here are in the first row of the "Time bars" columns.

Conditioned on the event of a trade, however, this reasoning may not be valid. Given the event of a trade, the reason to observe a quote change is the trade itself. Moreover, a single trade may not only lead to a single quote change, but to several quote changes in response to the information contained in the trade. That is, the number of quote changes to the right of the quote change that was triggered by the trade is likely to be greater than the number of quote changes to its left. This implies that a trade is likely to be placed behind the quote change that was triggered by the trade if we interpolate the times of trades and quotes according to their number of occurrences. If we further conjecture that the price impact follows the direction of the trade, misclassification is often the result.

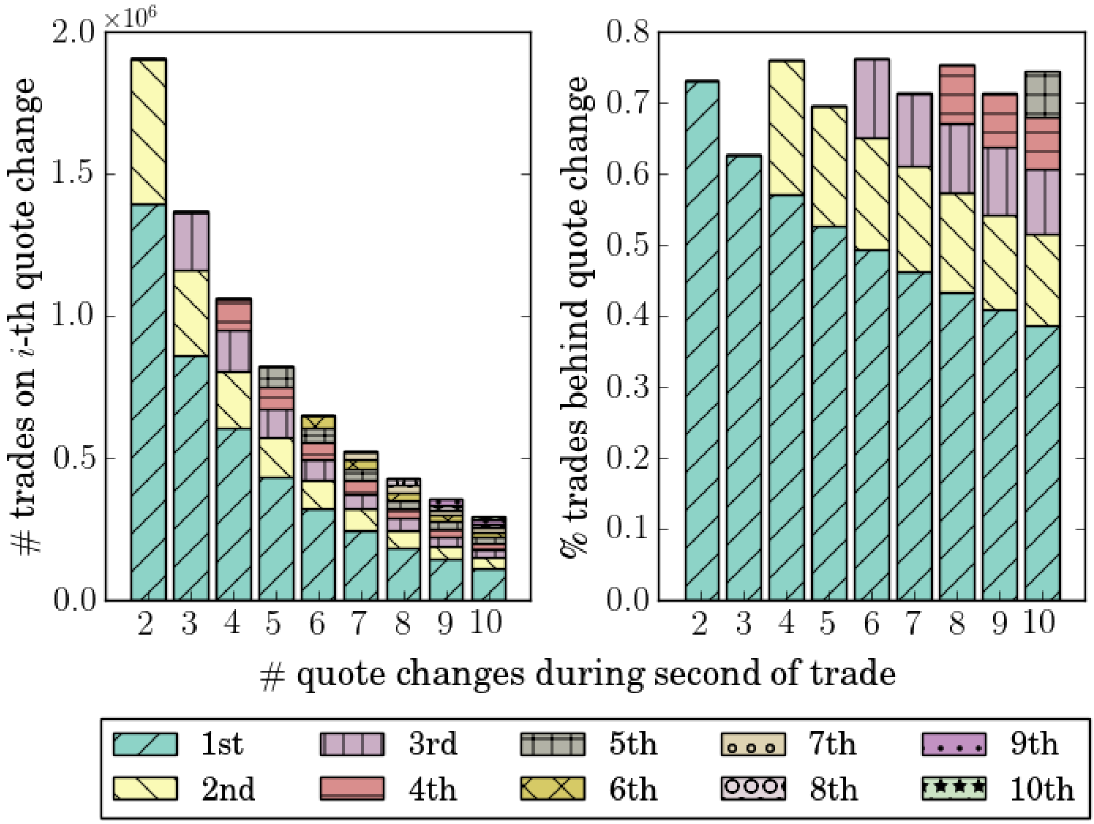
Figure 3.5 confirms these considerations. The left panel shows how often a trade triggered the first, second, third etc. quote change (y-axis) for a given number of quote changes during the second of the trade (x-axis).[17] For example, the first bar shows that almost 1.5 million trades triggered the first quote change, whereas only around 0.5 million were responsible for the second quote change in all cases where we observe 2 quote changes during the second of the trade. We see that even at seconds with a large number of quote changes, trades most often account for the first or second quote change.

The right panel of Figure 3.5 shows the percentage of trades that are moved behind their corresponding quote changes by the interpolation method. It shows that in more than 70% of the seconds with a single trade, the trade is moved behind the quote change that was triggered by that trade. Moreover, the panel shows that the fraction of trades that triggered the first quote change out of all quote changes during the same second remains large. Even for seconds with 10 quote changes, in around 40% of the cases the first quote change was due to a trade, while the other 9 quote changes were due to other reasons. So the quote matching rule using the interpolated times is relatively unsuccessful because it moves trades behind the quotes that were triggered by the trade.

Interestingly, Holden and Jacobsen (2014) report that by interpolating trade and quote times they increase the agreement of trade classification from the MTAQ data with the results obtained from the DTAQ data that are timestamped at higher precision than the MTAQ. In light of the results presented here this has important implications. Since the results show that by interpolating trade and quote times transactions tend to be moved behind the quotes which they triggered, the result of Holden and Jacobsen (2014) indicates that the problem of quotes being reported ahead of their corresponding trades prevails even today in the more accurate DTAQ

[17] I used only those seconds with a single trade.

Figure 3.5: Single trade seconds and quote changes



Notes: The LEFT panel presents the number of times a trade is executed on the first, second, third, . . ., 10-th quote (y-axis) within seconds of 2, 3, . . ., 10 quote changes (x-axis). The RIGHT panel shows the percentage of trades that are placed behind the quote change that was triggered by the trade if trade and quote times are interpolated following Holden and Jacobsen (2014). Only seconds containing a single trade are used.

data.[18] This adds further doubt on the ability to readily use the DTAQ data in combination with the traditional classification algorithms to infer the trade direction.

3.5.4 A Closer Look at the Classification Accuracy at the Individual Classification Steps

The Full-Information algorithm classifies trades at different steps, depending on the criteria that apply to the specific trade. The ambiguity of the classification

[18]Without a closer look into the DTAQ data, which is currently not available to this author, the assertion of a possible time delay of reported trades in the DTAQ must, of course, be viewed with caution. It is quite possible that the increase in agreement in the classification results between the DTAQ and MTAQ by using the interpolation method reported in Holden and Jacobsen (2014) is only due to chance.

increases with each step in the algorithm and we would, thus, expect the accuracy to differ with the different classification criteria.

Therefore, to study the performance of the individual classification steps, Table 3.1 presents the individual classification accuracies. Panel A shows the percentage of correctly classified volume at the individual classification step summed over the entire sample, while Panel B shows the percentage of trading volume that is classified at the respective classification step. The Table differentiates between trades executed against visible (visible = YES) and hidden (visible = NO) orders. The column "cl. step" refers to the classification steps (2 to 5) at which the trade initiator is assigned. "Cl. step 0" refers to cases where the trade direction of the liquidity demander could not be derived.

Trades against visible orders are almost exclusively classified during Step 2 or 3 of the classification process. That is, any trade that executed against a visible order must have at least one match among the quotes with the corresponding change in volume. Matches between quotes and trades that actually executed against hidden orders, on the other hand, are only accidental and occur rarely. With decreasing timestamp precision the number of hidden orders classified in Step 2 or 3 increases as the number of quotes that we consider during the classification of a trade increases. Overall, however, the number of hidden orders classified in Step 2 and 3 remains relatively small. Trades involving hidden orders are predominantly classified in Step 4 or 5 of the algorithm, as they are supposed to.

The accuracy of the assignments of visible orders in Step 2 of the algorithm is almost 100% at any timestamp precision. That is, an unambiguous match at one side of the order book leads almost always to the correct classification of the trade. With decreasing timestamp precision, however, the number of unambiguous assignments decreases and the algorithm refers to the interpolated times of the matched quotes more often. Though the interpolated time is a suitable indicator for the assignment with accuracies between 90 to 95% for timestamp precisions between seconds and milliseconds, it does not provide the same certainty as a classification at the second step. In fact, the decrease in overall classification accuracy going from nanoseconds to seconds is largely driven by the substitution of assignments between Step 2 and 3.

The classification of trades involving hidden orders is inherently more difficult than for visible orders. At nanosecond precision, the number of misclassifications can almost entirely be attributed to trades involving hidden orders. Even though the position of the transaction price within the spread is informative as we can see from the classification accuracies of around 94% at Step 4, the economic reasoning that is behind the decision rule employed at that step does not apply to all cases.

Table 3.1: Accuracy of individual classification criteria

		timestamp precision: 10^{-i} for $i =$					
visible	cl. step	0	1	2	3	4	9
Panel A: % correctly classified volume							
YES	0	–	–	–	–	–	–
	2	99.85	99.91	99.96	99.99	100.00	100.00
	3	90.31	94.18	95.98	95.11	68.96	–
	4	79.67	80.32	80.70	80.57	80.55	80.57
	5	55.97	55.90	56.63	56.31	31.74	29.80
NO	0	–	–	–	–	–	–
	2	67.25	74.49	84.52	94.50	99.79	99.98
	3	83.44	87.66	87.17	79.64	66.67	–
	4	92.59	94.79	95.75	95.58	94.39	93.86
	5	64.96	65.21	65.94	67.20	68.60	68.81
Panel B: % classified volume							
YES	0	0.00	0.00	0.00	0.00	0.00	0.00
	2	61.79	71.44	81.20	88.55	90.42	90.42
	3	28.55	18.93	9.19	1.86	0.00	0.00
	4	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.07	0.05	0.03	0.01	0.00	0.00
NO	0	0.00	0.00	0.00	0.00	0.00	0.00
	2	1.87	1.84	1.70	1.33	0.80	0.72
	3	0.65	0.30	0.09	0.01	0.00	0.00
	4	4.01	4.23	4.34	4.29	3.87	3.79
	5	3.05	3.20	3.45	3.95	4.92	5.07

Notes: Panel A shows the classification accuracy of the different classification criteria applied by the FI algorithm. Panel B shows the percentage of trading volume that is classified by the respective criterion. The column "cl. step" refers to the step in the classification process at which the trade initiator is assigned (with 0 referring to cases which could not be classified).

With decreasing timestamp precision the classification accuracy of trades involving hidden orders, however, does not change much. The informativeness and the number of cases assigned by the position of the transaction price is almost the same whether for data timestamped at nanoseconds or seconds. Most of the change in hidden orders classification accuracy is due to a shift from classifications by the tick-test to assignments at Step 2 of the algorithm.

3.5.5 Classification Accuracy under Randomly Delayed Timestamps

Similar to the problem of imprecise timestamps relative to quotation frequency, the problem of delays in reported trades is not knowing the exact trade-quote correspondence. Consider a trade timestamped at 9:45:50.9 and three quote changes timestamped at 9:45:50.1, 9:45:50.3 and 9:45:50.5. Given the uncertainty revolving around the degree of the report delay we may want to consider all three of them instead of picking only one quote for the classification procedure. In this particular case, the Full-Information algorithm would allow us to do so by simply decreasing the timestamp precision to that of seconds. The results from the previous section show that we would lose little in terms of classification accuracy if, in fact, a higher timestamp precision would suffice, but we would ensure that the results are not driven by the noise in the timestamps.

To explore the effect of random delays in reported trade times on the classification performance of the traditional algorithms and the FI algorithm, I add exponentially distributed noise to the original trade timestamp at nanosecond precision. That way the time of the trade will lag behind the reported time of its corresponding quote change but to a varying degree from trade to trade.

The exponential distribution is given by $F(x;\beta) = 1-\exp\{-x/\beta\}$ for $x \geq 0$ and I choose $\beta = 10^{-j}$ for $j = 1, \ldots, 4$.[19] I also choose different timestamp precisions at which the algorithms are applied to the data. The precision s ranges from 10^{-4} of a second to 2.5 seconds. For example, if $s = 10^{-3}$, the FI algorithm will consider all quotes that are valid during the millisecond at which the trade is reported and the traditional algorithms use the last quotes reported before the millisecond of the trade.[20]

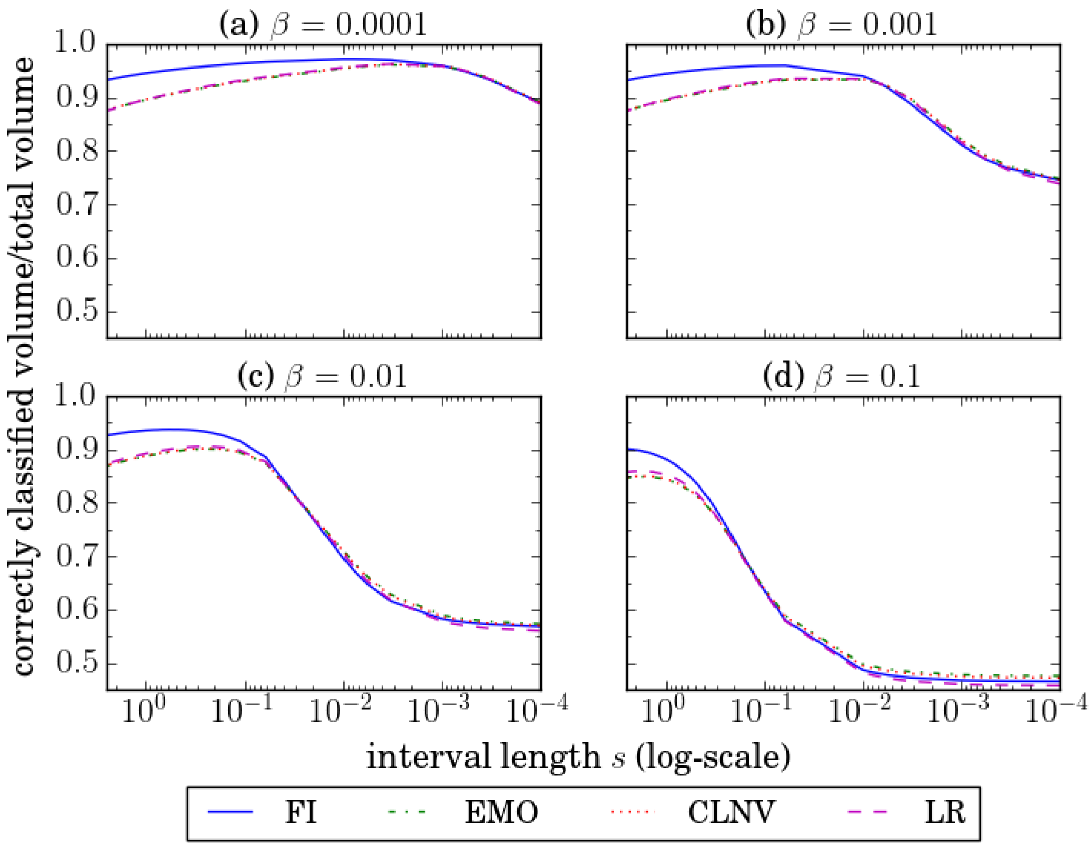
The Appendix presents a brief derivation of how we can expect the classification accuracy to be affected by noisy timestamps. To give a quick idea, consider applying the algorithms at a precision of seconds. The reduction in classification accuracy compared to the situation without noise is only determined by the number of trades that are shifted outside the second at which they actually occurred. The average classification accuracy of these trades will tend towards 0.5, while the classification accuracy is unaffected for those trades that remain in the same second as they were in the absence of noise.[21] Choosing the optimal timestamp precision

[19]The mean of the exponentially distributed variable is given by β and the q-th percentile by $-\ln(1-q)\beta$. For example, if $\beta = 1/10^3$, we expect a delay in the reported trade time of one millisecond and 99% of all trades to have a delay of less than 5 milliseconds.

[20]Note that I do not report the results of the traditional algorithms using the interpolated time due to their relatively unsuccessful performance in the absence of noise.

[21]Since the FI algorithm makes use of the correct order of trades at least to some extent, the

Figure 3.6: Classification accuracy and noisy timestamps



Notes: This figure shows the fraction of correctly classified trading volume (y-axis) for the data with delayed trade times. The trade time equals the actual time plus ε, with $\varepsilon \sim \text{Exp}(1/\beta)$ and $\beta \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$. To counteract the effect of the noise on the classification accuracy the algorithms FI, EMO, CLNV and LR are applied to the data with reduced timestamp precision (s) ranging from 10^{-4} of a second to 2.5 seconds presented on a $\log_{10}$-scale (x-axis).

is, thus, a trade-off between a reduction in accuracy due to imprecise timestamps on the one hand, and a reduction in accuracy due to trades being reported outside their actual time interval at high timestamp precision on the other. Since the classification accuracy of the FI algorithm is greater-equal to the accuracy of the traditional algorithms at any timestamp precision, we would expect that the FI algorithm dominates the traditional algorithms under noise as well.

Figure 3.6 presents the numerical results. Note that the timestamp precision s is presented on a $\log_{10}$-scale to obtain a better impression at high precisions.

As expected, we find that the FI algorithm dominates the others. It is only at relatively high precision timestamps that the performance of the algorithms align, trending towards an accuracy that is not different from a random classification of the trade initiator. We observe, however, that the FI algorithm is better able to

classification accuracy for those trades not shifted outside their actual time interval can still be affected if the order of the trades changes due to the noise. I discuss this issue in more detail below.

profit from the decrease in timestamp precision. The classification accuracy of the FI algorithm keeps increasing with growing interval length, where the accuracy of the other algorithms stagnates or falls off.

Importantly, the FI algorithm offers high and robust accuracies across the different noise intensities, especially for the probably more relevant range (a)-(c). For example, choosing a precision of a bit more than 0.5 of a second, the FI algorithm achieves accuracies of 94-95% for β between 10^{-4} and 10^{-2}. Even for relatively strong noise of $\beta = 0.1$, the FI algorithm correctly classifies more than 90% of trading volume if the timestamp precision is reduced below 1.9 of a second.

For any level of accuracy of the traditional algorithms under a given noise intensity, we can find the same accuracy for the FI algorithm at lower timestamp precision, which then offers robustness against higher levels of noise. That is, by choosing the FI algorithm at decreased timestamp precision, one gains robustness against unknown degrees of noise without forfeiting classification accuracy against the alternative algorithms if the noise intensity is, in fact, smaller than suspected.

3.6 Application to Measuring Liquidity and Order Imbalances

3.6.1 Measuring Liquidity

To analyze how the plus in classification accuracy translates into dollar values, I apply the different algorithms to the measurement of liquidity. The liquidity measures I consider are the dollar effective spread, the dollar price impact and the dollar realized spread. The estimation of these measures is a typical application where the knowledge of the trade initiator plays a crucial role.

The effective spread is defined as

$$DES_k = 2D_k(P_k - M_k)$$

where P_k is the price per share of the k-th trade, M_k is the spread mid-point associated with the k-th trade and D_k is the direction of the trade initiator, that is $+1$ in case of buyer-initiated trades and -1 in case of seller-initiated trades. The effective spread measures the costs incurred by liquidity demanders relative to the ideal environment where trades execute at the mid-point.

Opposite to the costs of liquidity demanders are the profits of liquidity suppliers. These gains are usually measured by the realized spread which subtracts the price impact from the effective spread. If prices move in the direction of the trade, the price impact is detrimental to the liquidity provider's profits.

The price impact given by

$$DPI_k = 2D_k(M_{k+\Delta} - M_k)$$

where $M_{k+\Delta}$ is the mid-point at Δ units, here chosen to be 10 minutes, after the k-th trade and the realized spread is, thus,

$$\begin{aligned} DRS_k &= DES_k - DPI_k \\ &= 2D_k(P_k - M_{k+\Delta}). \end{aligned}$$

For each stock-day I compute the volume weighted averages $L = \sum_k V_k L_k / V$ for $L_k \in \{DES_k, DPI_k, DRS_k\}$. I compare the measures computed from the true trade initiator label and the knowledge of which mid-point belongs to which trade, with the measures computed using the estimated trade initiator label and the associated mid-point provided by the algorithms.

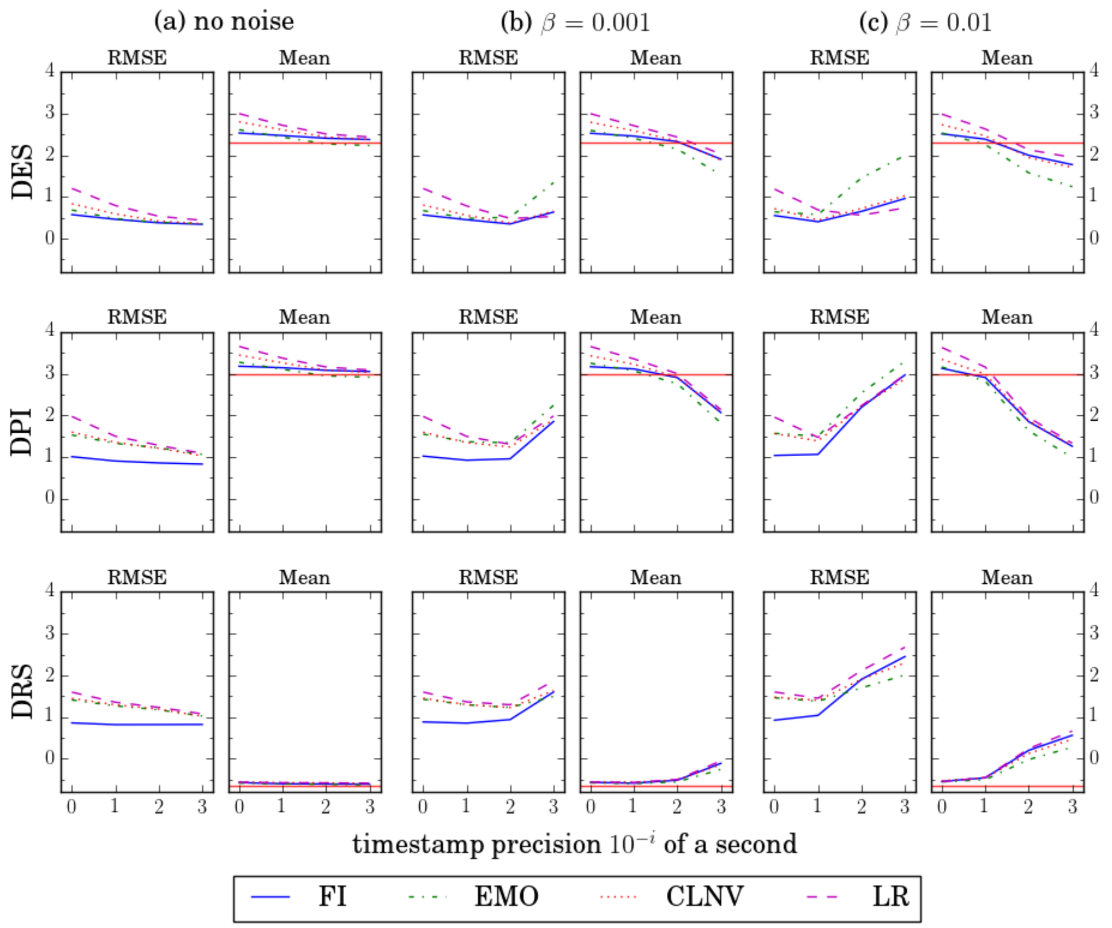
Figure 3.7 shows the root-mean-square error and the estimated mean across all stock-days measured in cents.[22] The red line in the graphs for the mean (2nd, 4th and 6th column) indicates the mean measured from the true trade-initiator label. The algorithms are applied to the data where the timestamp of trades is not affected by noise (column (a)) and where it is delayed by exponential noise with intensity $\beta = 10^{-j}$ with $j = 2, 3$ (columns (b) and (c)). The precision of the timestamp is reduced to 10^{-i} for $i = 0, \ldots, 3$ (which corresponds to a median number of quote changes per trade time of 17 to 3).

The results are again clearly in favor of the FI algorithm. The estimates based on the FI algorithm generally provide the smallest root-mean-square error. The improvements over the traditional algorithms are strongest for the dollar price impact and the dollar realized spread at the timestamp precisions of seconds and 10-th of a second. For example, for the data timestamped at seconds, the average deviation of the stock-day estimate of the realized spread is 0.9¢ for the FI algorithm, while it is 1.41¢ for the best competitor, the EMO algorithm. Even at high timestamp precision, the estimates of the traditional algorithms do not provide the same precision as the ones of the FI algorithm at lower timestamp precision.

Also the overall sample means are generally estimated closer to the true ones for the FI algorithm than for other algorithms. Only the EMO algorithm provides very similar mean estimates. The EMO algorithm is, however, more strongly affected at high timestamp precisions than the other algorithms.

[22] The root-mean-square error is given by $\text{RMSE}(L, \hat{L}) = \sqrt{\sum_{i,d}(L_{i,d} - \hat{L}_{i,d})^2 / |I \times D|}$.

Figure 3.7: Estimating liquidity



Notes: This Figure shows the sample averages and the root-mean-square error between the stock-day estimates and the true values of the dollar effective spread (DES), the dollar price impact (DPI) and the dollar realized spread (DRS) as defined in the text displayed in cents. The true values are computed from the true trade-initiator label and the knowledge of the ask and bid in place at the time of any given trade. The estimates are constructed from the classification results of the different algorithms and the ask and bid quotes that they assume to be in effect at the time of the trade. The algorithms are applied to the data with and without delayed trade times, where the delay is given by $\varepsilon \sim \text{Exp}(1/\beta)$ with $\beta = 10^{-3}, 10^{-2}$, and with varying timestamp precision ranging from seconds to milliseconds. These timestamp precisions correspond to a median number of quote changes at the time of trades of 17 to 3.

The advantage of applying the FI algorithm at lower timestamp precision is visible in columns (b) and (c), where trade times are affected by noise. The performance of the algorithms deteriorates at high timestamp precision if trades are reported with even mild delay, which translates into poor estimates of the liquidity measures. In the absence of noise the estimates of the FI algorithm at low timestamp precision, however, are barely different from the ones at high timestamp precision, but offer strong robustness against the report delay.

3.6.2 Order Imbalance

Another frequent application where the initiator label enters the analysis is the estimation of the order imbalance. The order imbalance over a given interval is defined as

$$OI = \frac{|V_B - V_S|}{V},$$

where $V = V_B + V_S$ is the total trading volume, V_B is the volume of buyer initiated trades and V_S the volume of seller initiated trades. The order imbalance provides the underlying statistic for measures such as the probability of informed trading (PIN) (Easley et al., 1996b), or more directly for its volume synchronized version (VPIN) (Easley et al., 2012). It is also the driving force behind estimations of deviations from price efficiency (Cipriani and Guarino, 2014). Often, the order imbalance is itself the variable of interest (e.g. Chordia et al., 2002; Dorn et al., 2008; Chordia et al., 2016).

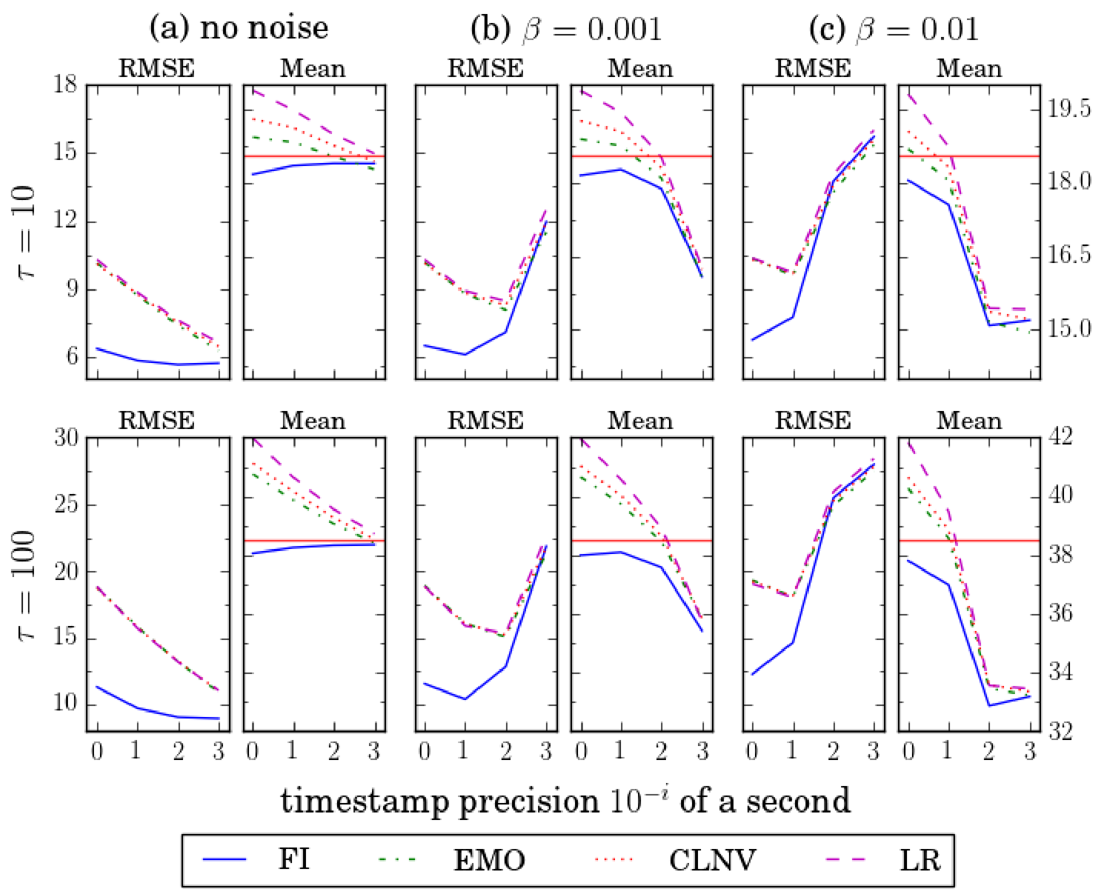
The analysis follows the same procedure as the previous section. The order imbalance is estimated from the classification results of the different algorithms applied to the data at varying timestamp precisions (from seconds to milliseconds) and for different degrees of report delays in trade times.

The order imbalance is estimated as follows. Each stock-day is split into τ bins of equal volume size. If necessary, the last trade in a bin is split between the two successive bins to ensure equal volume. Thus, we have τ estimates of the order imbalance for each stock-day and $\tau \times 19842$ order imbalance estimates in total, with varying volume sizes across the stock-days. Within each volume bin, the order imbalance is computed according to the above formula using the true trade initiator label and the labels provided by the classification algorithms.

The main variable of interest is again the root-mean-square error between the vector of true order imbalances and their estimates provided by the algorithms. Figure 3.8 presents the results. The red line in the columns 2, 4, and 6 indicates the mean order imbalance computed from the true trade initiator label. The first row presents the results where each stock-day is split into 10 equal volume bins ($\tau = 10$) and the second row for stock-days split into 100 equal volume bins ($\tau = 100$). The numbers are displayed as percentages.

The results mirror those for the estimation of the liquidity measures. The root-mean-square errors of the estimates based on the classifications of the FI algorithm are generally the smallest. Again, the largest improvements occur at low timestamp precision. For the data split into 10 volume bins at each stock-day the average deviation of the estimated order imbalance based on the FI algorithm

Figure 3.8: Estimating order imbalance



Notes: This Figure shows the sample averages and the root-mean-square error between estimates of the order imbalance and the true order imbalance displayed as percentages. The true values are computed from the true trade-initiator labels and the estimates from the classification results of the different algorithms. For the computation of the order imbalance each stock-day is split into equally sized volume bins. The number of bins is chosen to be $\tau = 10, 100$. The algorithms are applied to the data with and without delayed trade times, where the delay is given by $\varepsilon \sim \text{Exp}(1/\beta)$ with $\beta = 10^{-3}, 10^{-2}$, and with varying timestamp precision ranging from seconds to milliseconds. These timestamp precisions correspond to a median number of quote changes at the time of trades of 17 to 3.

is 6.34%-points, while that of the best competitor, the CLNV algorithm, is 10.1%-points. It is only at high timestamp precision (millisecond) that the estimation results of the traditional algorithms can compete with those of the FI algorithm at low timestamp precision. However, only at low timestamp precision are the results robust against moderate delays in reported trade times.

So again, we do not forfeit any estimation accuracy by using the FI algorithm at low timestamp precision compared to applying the other algorithms at higher timestamp precision, but we gain robustness against various degrees of noise.

3.7 Determinants of Misclassification

In an analysis that necessitates the estimation of the trade initiator, a variation of the classification accuracy in tandem with the variation of other variables entering the analysis can impact the statistical inference and, in the worst case, compromise the researcher's conclusions. For this reason, I analyse the determinants of misclassification by the FI algorithm in more detail using a logistic regression following Finucane (2000), Ellis et al. (2000) and Chakrabarty et al. (2007).

The previous studies focused on the LR algorithm (only Chakrabarty et al. (2007) also analyzed the determinants of misclassification for the EMO and the CLNV algorithm) and found that the most important determinant for correct classification is the execution of a trade against the quotes. The influence of other explanatory variables like trade size, spread, stock volume, firm size and the speed of trading does not always agree across the studies and is generally small in terms of their marginal effects. Still, slow trading and larger spreads seem to help the LR algorithm to infer the trade initiator.

3.7.1 Variable Selection

The logistic model is given by

$$P(y_i = 1 \mid x_i) = \Lambda(x_i'\beta)$$

where $y_i = 1$ is the event of a correct classification (and $y_i = 0$ the event of a misclassification) and $\Lambda(w)$ is the logistic distribution function, $\Lambda(w) = 1/(1 + \exp\{-w\}$. The explanatory variables, x_i, for the regression exercise are chosen as follows.

Section 3.5.4 reveals that hidden orders are particularly difficult to classify. I will, thus, include a dummy variable, labeled Hidden, that takes the value 1 if the transaction involved a hidden order and 0 otherwise. Mid-point trades received a lot of attention in previous studies due to the reliance of the LR algorithm on the tick-test for such trades and their lack of a clear economic sign of the direction of the trade initiator. I, therefore, include a variable that captures the distance of the transaction price to the mid-point at the time of the trade constructed as Mid $= 1 - 2|p_t - m_t|/(a_t - b_t)$, where p_t is the execution price, m_t is the corresponding mid-point and a_t, b_t are the corresponding ask and bid quote, respectively. That is, Mid takes the value 1 if the trade executed at the mid-point and decreases towards 0 with the price approaching one of the quotes. Despite the difficulty of classifying mid-point trades, what might be generally more important for accurate classification is the proximity of buyer-initiated trades to the ask and that of seller-initiated trades to the bid. Hence, I include the variable $\mathsf{Q\text{-}Dist} = |D_t - p_t|/(a_t - b_t)$ where $D_t = a_t$ if the trade is buyer-initiated and $D_t = b_t$ if the trade is seller-initiated.

The consideration of the mentioned variables is motivated by the classification criteria of the FI algorithm, and the obvious difficulty of classifying trades that deviate from the reasoning behind these criteria. To consider variables that could interact with the classification accuracy, though in a less obvious way, and play a role in more general economic and financial analyses I choose the following. I include the squared return of each transaction defined as $R^2 = (\log(p_i) - \log(p_{i-1}))^2$, where i is the i-th transaction, the size of each transaction in 100 shares (Size), the absolute spread size at the time of a trade measured in dollars (Spread), the total trading volume of the stock-day in 10^5 shares (Vol), the 5-minute realized variation (see Liu et al., 2015) over each stock-day (RV), the distance of each transaction to the previous trade in seconds (Δt-Trade), the distance of each transaction to the last quote change in seconds (Δt-Q), the number trades during the second of each trade (#Trades), the number of quote changes during the second of each trade (#Q) and a dummy variable indicating whether a transaction was part of a trade involving more than one counter party (MultiTrade). The latter is identified by observing more than one execution on the same side of the order book during the same nanosecond.

The Appendix provides summary statistics of the explanatory variables, a bivariate correlation analysis, as well as a description of the filtering of the data before the actual estimation procedure. For numerical stability in the optimization procedure and to allow for a better comparison of the marginal effects across the variables, R^2 to #Q (i.e. all variables not ranging in $[0, 1]$) are standardized to have zero mean and unit variance.[23]

3.7.2 Estimation Results

Due to computational constraints, I did not use the full sample in the logistic regression. Instead, I selected all observations where the FI algorithm misclassified the transaction (almost 7.6 million observations) and randomly selected (without replacement) a sample of equal size from the observations where the FI algorithm correctly classified the trade. Maximum likelihood estimation of the logistic model

[23]Due to the large number of zeros in R^2 (caused by many multi-party or successively placed, small trades), zero-observations have been left out in the standardization procedure for R^2. That is, they did not enter the computation for the mean and variance. Otherwise one would divide by near-zero and inflate the non-zero observations.

Table 3.2: Logistic regression result

No. Observations:				15190680
McFadden R-squ.:				0.2205
Log-Likelihood:				-8.2072e+06
LL-Null:				-1.0529e+07
			marginal effects	
	β	std err	at the mean	overall
const	3.1719	0.001		
Hidden	-2.3120	0.003	-0.2055	-0.1130
Mid	0.8656	0.008	0.0299	0.0423
Q-Dist	-3.6268	0.015	-0.1254	-0.1773
R^2	0.0479	0.001	0.0017	0.0023
Size	0.1714	0.001	0.0059	0.0084
Spread	0.3403	0.001	0.0118	0.0166
Vol	0.3388	0.001	0.0117	0.0166
RV	-0.0912	0.001	-0.0032	-0.0045
Δt Trade	0.3880	0.001	0.0134	0.0190
Δt Q	0.2853	0.001	0.0099	0.0140
# Trades	0.2157	0.001	0.0075	0.0105
# Q	-0.7683	0.001	-0.0266	-0.0376
MultiTrade	0.5896	0.001	0.0223	0.0307

Notes: This table shows the regression results of a maximum likelihood estimation of the model

$$P(y_i = 1|x_i) = \Lambda(x_i'\beta)$$

with $y_i = 1$ being the event of a correct classification by the FI algorithm applied to the data timestamped to the second, and x_i containing the explanatory variables described in Table 3.E1. $\Lambda(\cdot)$ is the logistic distribution function. The variables R^2 to #Q have been standardized. The estimates are based on a sub-sample containing all observations where $y_i = 0$ and a random selection of equal size of observations where $y_i = 1$. This yields consistent estimates except for the coefficient of the constant term, β_0. To obtain a consistent $\hat{\beta}_0$ one simply subtracts $\log((1-p)\bar{y}/p(1-\bar{y}))$, where p is the frequency of $y_i = 1$ in the full sample and $\bar{y}$ the corresponding frequency in the sub-sample. $\hat{\beta}_0$ in the Table is the bias corrected estimate. The marginal effects are evaluated at the sample mean, as well as evaluated at each data point of the standardized data and then averaged, i.e.:

$$\text{at the mean:} \quad \frac{\partial P(y = 1|\bar{x})}{\partial x_k} = \Lambda(\bar{x}_i'\hat{\beta})(1 - \Lambda(\bar{x}_i'\hat{\beta}))\hat{\beta}_k$$

$$\text{overall:} \quad \frac{\partial P(y = 1|x)}{\partial x_k} = \sum_i \Lambda(x_i'\hat{\beta})(1 - \Lambda(x_i'\hat{\beta}))\hat{\beta}_k/N.$$

For the dummy variables the effects are computed analogously using

$$P(y_i = 1|x_{ik} = 1, x_i) - P(y_i = 1|x_{ik} = 0, x_i).$$

still yields consistent estimates, except for the constant, which can be easily corrected to obtain consistency (see e.g. Eq. (7) and Appendix B in King and Zeng, 2001).[24] Table 3.2 presents the regression results. The FI algorithm is applied to the data timestamped to the second.

We see that the single most important determinant for misclassification is the execution against a hidden order. On average, a trade that executes against a hidden order as opposed to a visible order decreases the estimated probability of a correct classification by 11%-points. Once controlled for the impact of a hidden order, the effect of the distance to the quotes or the proximity to the midpoint seems less important. For example, moving 10% of the spread size away from the quote against which we would expect the trade to execute decreases the estimated probability of a correct classification by approximately 1.8%-points on average.

The variables that may play a more decisive role in more general economic and financial studies involving the estimation of the trade initiator, like total trading volume, the realized variation or the speed of trading, do not strongly impact the classification accuracy. For example, a one standard deviation increase in the realized variation decreases the estimated probability of correct classification by only about 0.45%-points. Among these variables, frequent quote changes during the second of the trade (#Q) exhibit the strongest effect on the classification accuracy: a one standard deviation increase in the number of quote changes during the second of the trade decreases its probability of being correctly classified by around 3.8%-points on average.[25]

3.8 Adjusting the FI Algorithm to Different Data Structures

So far, we assumed the same level of data granularity (summarized in Data Structure 1) that is provided by the reconstructed limit order book from the NASDAQ TotalView-ITCH data. The advantage of the FI algorithm over the traditional approaches feeds on the use of information offered from this granularity. In this section, I will relax the assumptions in Data Structure 1 and present appropriate adjustments to the algorithm. With less information at hand we cannot expect to

[24]The only requirement for consistency is that the conditional densities of the subsampled data $(x|y)$ matches the conditional density of the full sample $(X|Y)$, i.e. $P(x \mid y = 1) = P(X \mid Y = 1)$ and $P(x \mid y = 0) = P(X \mid Y = 0)$. The latter is trivially satisfied in my case as I select all observations where $y_i = 0$ and the former should be satisfied by the random subsampling scheme.

[25]Note that despite the predictive content of the explanatory variables for the probability of a correct classification, this does not imply a predictive power of these variables for the trade initiator label. The analysis only identifies environments under which it is more difficult to arrive at the true initiator label, but it did not identify the direction of the misclassification.

achieve the same level of accuracy as presented above, but we may still be able to obtain more accurate results than those provided by the traditional algorithms.

3.8.1 Relaxing Assumption (i) of Data Structure 1

Assumption (i) of 1 states that each transaction against a visible standing order is reflected by a corresponding decrease in volume at the respective quote. This includes transactions that are part of an order too big to be filled by a single standing order. Even though the transactions between the parties involved are carried out almost instantaneously in the order book, we assumed that the data displays the successive steps in the execution according to its order precedence rules.[26] In this section, we assume instead that at the time of a trade the order book displays the state of the order book after the completion of the order that led to the trade.

Data Structure 2. *Aggregated Quote Changes*

(i) At the time of a trade, the order book displays the new state of the order book after the completion of all transactions that were carried out due to the same buy or sell order.

(ii) Trades and quotes are reported in the correct order.

The FI Algorithm under Data Structure 2

The change in the data structure means that we cannot use the strict equality between the transaction volume and the change in volume at the quote to eliminate potential matches. If an order for 100 shares trades against two limit orders for 50 shares each, posted at the same price, the trade data record two transactions for 50 shares each, while the order book data shows a decrease in volume at the respective quote by 100 shares.

Hence, we change the search for a match among the ask quotes in Step 2 of the algorithm to

$$\boldsymbol{\alpha} = \min\{j \in \mathcal{J}_a \colon p_i = a_j \text{ and } v_i \leq \Delta v_j^a\},$$

and analogously for the bid.

[26] Order precedence rules determine the order in which standing orders are executed when a marketable order enters the order book. Usually, the order offering the best price is executed first. If several visible orders offer the same price, the one that was submitted earliest is executed first (visible orders are usually precedented over hidden orders at the same price, even if the hidden order was submitted first), and so on.

If we find a match at one or both sides of the order book, the algorithm proceeds as before. If, however, we cannot find a match on either side of the order book, we need to insert two additional steps before we can conclude that we apparently face a transaction involving a hidden order, which would be classified under Step 4.

Consider a market order for a number of shares greater than what is available at the best quote. The trade data will show the corresponding transaction at the next-best quote, but the order book data will not show any decrease in volume at that quote. In the extreme case, where the market order is so large that it will go through several levels of the order book, the order book data will not even show the quotes against which the order executed on its way to the last quote.

To accommodate these cases the adjusted algorithm injects two additional searches for a match at the ask or bid side before it proceeds with Step 4. The first search (demonstrated for the ask) under Step 4a is conducted as

$$\widetilde{\boldsymbol{\alpha}} = \min\{j \in \mathcal{J}_a \colon p_i = a_j \text{ and } a_{j-1} < a_j\}.$$

In case we find a match on one or both sides of the order book we proceed as prescribed by Step 2.[27]

The second additional search for a match among the quotes if we cannot find one under Step 4a, is conducted under Step 4b (again demonstrated for the ask) as

$$\hat{\boldsymbol{\alpha}} = \min\{j \in \mathcal{J}_a \colon p_i > a_j \text{ and } a_{j+1} < p_i\},$$

and analogously for the bid, proceeding exactly as under Step 4a if a match on one or both sides can be found. If again neither a match at the ask side nor the bid side can be found, we are likely facing a hidden order and the classification is derived under Step 4 as before.

Results for Varying Timestamp Precisions

The evaluation of the adjusted FI algorithm (FI_{DS2}) is conducted as before. The order book data, however, has been changed to reflect the new data structure. That is, all intermediate changes in the order book due to a trades that are executed against more than one counter-party are neglected. The results are presented in Appendix 3.F.

[27]Note that if we classify the transaction according to the interpolated time, we do not adjust the volume at the matched quote, as there was no corresponding volume change to begin with.

The conclusions from Section 3.5.1 do not change. The classification accuracies at relatively high timestamp precisions are not different to the accuracies under more granular data. The classification accuracy suffers a bit from the loss of information in the order book data only for the data timestamped at seconds. Still, the FI_{DS2} algorithm improves the classification accuracy of the traditional algorithms by more than 3%-points.[28] The loss in classification accuracy is due to the inability to use the trading volume as an exact match to changes in the volume at the quotes. This leads to more transactions that have to be classified using the interpolated time of quote changes.

Results for Noisy Timestamps

The results from the application of the FI_{DS2} algorithm to the data with delayed transaction timestamps are presented in Figure 3.F1 in Appendix 3.F.[29] Again, the conclusions from the earlier exercise using the original data structure do not change. At high timestamp precision the classification accuracy of all algorithms is strongly affected even by relatively moderate noise intensities. Decreasing the timestamp precision, however, helps to counteract this adverse effect. Importantly, the FI_{DS2} algorithm is more accurate than the traditional ones over the range of timestamp precisions that show stable results over the different noise intensities, albeit slightly less accurate than the version for the original data structure.

Estimating Liquidity and Order Imbalances

The results for the estimation of various liquidity measures and the order imbalance under the new data structure are presented in Figures 3.F2 and 3.F3 in the Appendix. The conclusions regarding the improvements achieved by the FI_{DS2} algorithm do not change. By and large, the estimates based on the FI_{DS2} algorithm provide the smallest root-mean-square errors.[30] That is, the FI_{DS2} algorithm provides the most precise estimates over the sample of stock-days. For example, for the dollar realized spread under no noise and a timestamp precision of seconds, the average deviation of the estimate based on the FI_{DS2} algorithm around the true dollar realized spread is 1¢ per share, compared to that of the EMO estimate

[28] The results of the traditional algorithms are the same as the ones from the previous section, as they are not affected by the new data structure.

[29] Due to the more granular data structure in the previous sections, we were basically treating each transaction as a single trade. Each transaction was therefore shocked by a separate noise realization. Here, since we count transactions that belong to the same marketable order as a single trade, I shock transactions belonging to the same order by the same noise realization.

[30] Note that the measures computed from the true trade initiator label are also computed under the new data structure. Therefore, the results for the true measures deviate slightly from the ones presented earlier.

of 1.5¢ per share. Importantly, applying the methods to the data timestamped at seconds, or 10-th of a second, stabilizes the estimates against noise in the trade times. At those timestamp precisions the improvement of the FI_{DS2} algorithm in terms of the root-mean-square error is the greatest. The same applies to the estimates of the order imbalance.

3.8.2 Relaxing Assumption (ii) of Data Structure 1 and 2: Randomized Order of Trades

In certain datasets, trades may not follow the actual order in which they were executed (e.g. Easley et al., 2016). That may be due to two reasons. First, the legal framework may allow for some delay in reporting trades. Depending on whether the timestamp of the data reflects the time of the report or the time of the actual trade and depending on the extent to which trading institutions exploit their right of delayed reporting, trades may be out of order. Second, for data from a consolidated tape, which timestamps trades when the corresponding data are processed, trades are out of order due to different latencies for sending information from different market places to the same data processor. These latencies can be expected to be small, but large enough to affect trades that are executed over small intervals.[31]

Nevertheless, we can expect that the FI algorithm will be little, if at all, affected by trades being out of order. The correct order of trades plays a role for the FI algorithm only if it uses the tick-test, which it rarely does.[32]

We already examined the consequences of trades being out of order in the sections where we delayed the trade times by exponential noise (though we did not mentioned it explicitly). When we add to each trade time an independently distributed exponential variable, the order of trades can change. For example, for two trades with the second trade following one millisecond after the first trade, the probability that the first trade will be shifted behind the second one if both trades

[31] For example, the speed of light in a vacuum is roughly $300 * 10^6$ m/s. Sending data from Chicago to New York (a distance of around 1300km) at the speed of light would thus take 4ms. So even at this physically lower limit of transmission time the report delay of a Chicago trade is 4ms compared to a trade at the NYSE where the consolidated tape is located.

[32] To a lesser extent, the correct order of trades also plays a role if the algorithm uses the interpolated time of quotes to classify a trade. If a trade is classified using the interpolated time, the volume at the quote that is matched to the trade is reduced by the size of the transaction. This is done because two different trades cannot cause the same quote change and to avoid that this quote causes further conflicts between an assignment of trades to either the ask or the bid side. If there are several trades with the same price and volume and these trades are out of order, it is possible that the FI algorithm assigns these trades to the conflicting ask and bid quotes in the exact opposite order in which they actually occurred. However, for statistics like the order imbalance such errors are irrelevant.

are affected by exponential noise with $\beta = 1/10^3$ is 0.1839.[33] We saw from the previous exercises under noisy trade times that the randomization of the order of trades did not greatly influenced the accuracy of both versions of the FI algorithm and did not greatly affect the performance against the traditional algorithms.

3.8.3 Relaxing Assumption (ii) of Data Structure 2: Randomized Order of Trades and Quotes

The reasons for the possibility that trades could be out of order apply to the recorded quotes, at least for a consolidated tape, just as well. If, indeed, quote changes are out of order the decision criteria of the FI algorithm have to be adjusted. The definition of the change in volume used to find matches between transactions and quotes is only meaningful if the order of the quotes is correct. Since we assume here that the order of quotes is incorrect, we need to adjust the search of trade-quote correspondences.

Data Structure 3. *Aggregated Quote Changes and Random Trade and Quote Order*

(i) At the time of a trade the order book displays the new state of the order book after the completion of all transactions that were carried out due to the same buy or sell order.

(ii) Trades and quotes can be out of order.

The FI Algorithm under Data Structure 3

Instead of the change in volume we can now rely only on the absolute volume displayed at the respective quote. For a transaction to be executed at a particular quote the volume of the transaction cannot exceed that of the volume available at the quote. Therefore, the search of a match between a transaction and a quote in Step 2 is changed to (demonstrated for the ask)

$$\boldsymbol{\alpha} = \min\{j \in \mathcal{J}_a : p_i = a_j \text{ and } v_i \leq v_j^a\},$$

[33] More formally, for two trades at time t_1 and t_2 with $t_2 = t_1 + \Delta$ and $\Delta \geq 0$ the probability that the first trade is shifted behind the second one due to noise is given by

$$P(t_1 + \varepsilon_1 > t_2 + \varepsilon_2) = \int_\Delta^\infty f(\varepsilon_1) F(\varepsilon_1 - \Delta)\, \mathrm{d}\varepsilon_1,$$

with $\varepsilon_i \in \mathbb{R}_{\geq 0}$ and $\varepsilon_i \overset{\text{iid}}{\sim} F$ for some distribution function F with density f. For F being the exponential distribution $\text{Exp}(1/\beta)$ this is given by $\exp\{-\Delta/\beta\}/2$.

and analogously for the bid. From here, the adjusted algorithm proceeds as the baseline version.[34] If the classification is derived under Step 3 the algorithm subtracts the transaction volume from the volume available at the matched quote. Step 4a and Step 4b from the previous adjustment to the algorithm do not apply here because they rely on the correct order of the quotes.

Results for Varying Timestamp Precisions

Table 3.G1 in Appendix 3.G shows the results of the application of the FI algorithm adjusted for the new data structure (FI_{DS3}) for the same data as under Data Structure 2 not being affected by noise in trade or quote times. Although the order of trades and quotes will thus not be affected, the exercise demonstrates the loss in classification accuracy we have to incur for not being able to use the full amount of information. We see that the algorithm, albeit precise at high timestamp precisions, reacts with greater sensitivity to the reduction in timestamp precision than the previous versions, as it is more difficult to resolve situations where both ask and bid quotes seem to provide a match to the transaction. However, the FI_{DS3} algorithm still achieves a 1.6 to 3.0%-points improvement over the traditional algorithms at a timestamp precision of seconds.

Results for Noisy Timestamps

To study the effect of random trade and quote order, I add exponential noise to the timestamps of both trade and quote data. Note that in doing so, not only will the order of trades and quotes change, but trades may now also be reported before their corresponding quote change. We may view this section as an examination of classification accuracy under a minimum of data structure. All that we require is that prices and volumes of trades and quotes are correctly recorded and that trades are executed in a reasonable interval around their corresponding quote change.

Figure 3.G1 in Appendix 3.G shows that, contrary to the above setups, there are regions of timestamp precisions and noise where the traditional algorithms outperform the FI_{DS3} algorithm. At these regions of higher timestamp precision, however, classification accuracy is quite low for all the algorithms and not stable across the different degrees of noise. At the timestamp precisions that ensure that the algorithms are not too strongly affected by noise, the FI_{DS3} algorithm again outperforms the others.

[34] Note, however, that the auxiliary variables l^a and l^b are not updated after a classification.

Estimating Liquidity and Order Imbalances

With respect to measuring liquidity and order imbalances, Figures 3.G2 and 3.G3 in Appendix 3.G show that despite the decrease in classification accuracy due to the changes in the data structure, the FI_{DS3} still achieves considerable improvements in terms of the root-mean-square error.

3.9 Discussion

3.9.1 The Bulk Volume Classification Algorithm, the Aggressor Flag and Informed Trading

Sharing the motivation that high-frequency quoting and possible inaccuracies in timestamps pose difficulties for established classification algorithms to generate reliable results, Easley et al. (2016, 2012) present an alternative classification algorithm (BVC).[35] Their motivation, however, goes one step further questioning the equality of liquidity demanders and informed traders in today's markets (one reason for the frequent application of trade classification algorithms), a claim supported by studies like Collin-Dufresne and Fos (2015). The BVC algorithm is, therefore, proposed to discern the underlying information from trade data.

Even though this paper is not directly concerned with the topic of informed trading, the claims by Easley et al. (2016), of course, affect the range of applicability of the algorithm proposed here, because it falls into line with the traditional approaches. Hence, I want to briefly comment on this subject.

In market microstructure models, information usually refers to private signals of traders regarding the liquidation value of an asset at some (terminal) point in the future. Traders in possession of such information choose liquidity demanding orders, while the uninformed side provides liquidity. This assumption about the relation of order and trader types, informed or uninformed, is built into many empirical applications. Yet, Collin-Dufresne and Fos (2015) show that traders with private information frequently choose passive orders.

The concept of information in empirical studies is, however, far less uniformly defined than its theoretical counterpart. The private information in Collin-Dufresne and Fos (2015), for example, refers to an investor's intention to increase her stake in a publicly traded company up to some critical limit, at which point the stake of that investor becomes public information. Until that point, the investor

[35] They classify buckets of trading volume into fractions of buyer- and seller-initiated trades by multiplying trading volume by the density function of a t-distribution at the standardized price change over the trading interval.

can carefully time her transactions without revealing her intentions and her influence on the future of this company. Many recent studies, on the other hand, show that the order imbalance constructed from the classification results of the traditional algorithms signals fundamental information for specific events where time is crucial (Bernile et al., 2016; Chordia et al., 2017; Hu, 2014, 2017; Muravyev, 2016). In fact, Collin-Dufresne and Fos (2015) also show that investors use liquidity demanding orders more frequently approaching the day when their stake becomes public information.[36]

The BVC algorithm also has its own notion of what constitutes information. The BVC algorithm classifies trade data based on the price movement over an interval preferably defined by trading volume instead of time. The normalized price change over the interval is plugged into the probability density function of a t-distribution which yields the fraction of buyer-initiated trades.[37] Large order imbalances from the BVC algorithm thus signal large, historically abnormal price movements over a given amount of trading volume, which do not necessarily have to be driven by liquidity demanding orders, but instead by, e.g., a flight of liquidity providing orders. Such order imbalances put certain types of high-frequency traders under stress and, thus, certainly signal valuable information to them or to regulators concerned with short-term market distortions, such as the flash crash of 2010. These price movements, however, can be of only temporary pressure and completely unrelated to fundamental information.

We see that the appropriateness of using one classification algorithm rather than another to discern the direction and strength of information from the order flow cannot be discussed without reference to a specific concept of information that may include investor preferences and strategies, the type of information event and the half-life of the information. While one should not invariably equate informed traders with liquidity demanders, it clearly depends on the definition of information to decide whether one should do so for the specific context of the analysis.

3.9.2 The Speed of the Full-Information Algorithm

An important aspect of the FI algorithm is that it utilizes more information than the traditional algorithms. This raises the question of computational feasibility. I implemented the FI algorithm in Python using the Cython hybrid language for the computational intense parts. On the entire data set consisting of 19842 stock-days

[36] Note that all studies mentioned in this paragraph use classical methods of trade classification and encounter some of the problems which are dealt with in this paper. Their analysis could have thus profited from the algorithm presented here.

[37] One minus the pdf of the t-distribution yields the fraction of seller-initiated trades, correspondingly.

with a total of 134,449,578 trades the classification function needs 20.91 minutes on an Intel Core i7 CPU with 3.6 GHz (*not* using the different cores for parallel computing which can easily be done) to sign all trades for the data timestamped to a second, which is on average 9.3 microseconds for a single trade. The majority of time is actually spent reading the data from a SQLite database: 0.34 seconds for a stock-day of order book data and 0.1 seconds for a stock-day of transaction data. That is, the use of the additional information to arrive at an improved classification accuracy does not come at any noteworthy computational costs.

3.10 Conclusion

This paper proposes a new trade classification algorithm that improves the classification of trades into the liquidity demanding and supplying side under the characteristics of today's markets and data records. In particular, the high frequency of quote submission and cancellation pose a problem for established classification algorithms. Under a median of 17 quote changes at the time of a trade, for example, the new algorithm manages to reduce misclassification rates by half. The improvements in classification rates translate into considerable improvements in the estimation of transaction costs and order imbalances. The evidence presented in this paper also raises some concern about using the DTAQ data in combination with the traditional classification algorithms without worrying about data quality.[38]

[38]Unfortunately, not having access to the DTAQ data I cannot address to which extent the proposed algorithm is able to improve the classification of trades from the consolidated tape. To the extent that the DTAQ suffers from the different data deficiencies analysed in this paper, however, we saw that the improvement can be sizeable.

Appendix 3

3.A Extracting Trade and Quote Data from LOBSTER's Message Files

The software LOBSTER reconstructs from the original Nasdaq TotalView-ITCH data feed the full limit order book, as well as a message file containing information on the events causing the changes in the order book.[39]

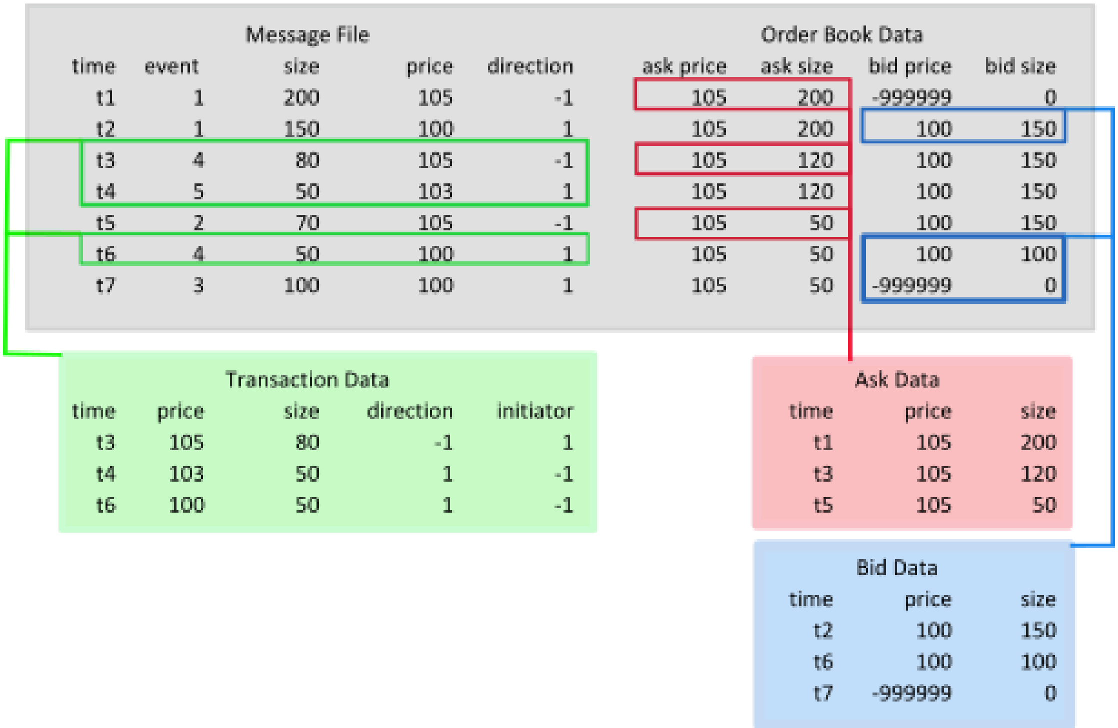
The gray shaded area in Figure 3.A1 provides an example of the design of LOBSTER's message and order book files. The k-th row of the message file describes the cause of the change in the order book from the $(k-1)$-th row to the k-th row. The events 1, 2 and 3 refer to the submission, partial cancellation and total deletion of a limit order. The events 4 and 5 refer to the execution of a visible and hidden limit order, respectively. The direction indicates whether a buy (+1) or sell (-1) limit order is affected. If a hidden order is executed, the order book is not visibly affected. In that case, to maintain a symmetric output, the LOBSTER order book data displays the order book's state after the execution of the hidden order.

As an example take the first row of the order book and message file. We start here with an empty order book indicated by negative quotes. At `t1` the message file indicates a submission of a limit sell order for a price of `105` per share for a total of `200` shares. In the same row, the order book displays its new state. The bid side is still empty and the ask side is now displaying the price and volume of the limit sell order.

Below the gray shaded area in Figure 3.A1, it is illustrated how I extract the trade and quote data from the order book and message file. I construct the trade data by extracting all visible and hidden executions of limit orders (events 4 and 5) from the message file, with the respective information on the price and volume

[39] For more information on the TotalView-ITCH data feed and the order book reconstruction by LOBSTER, see Hautsch and Huang (2012) and Huang and Polak (2011).

Figure 3.A1: Data construction



Message File

time	event	size	price	direction
t1	1	200	105	-1
t2	1	150	100	1
t3	4	80	105	-1
t4	5	50	103	1
t5	2	70	105	-1
t6	4	50	100	1
t7	3	100	100	1

Order Book Data

ask price	ask size	bid price	bid size
105	200	-999999	0
105	200	100	150
105	120	100	150
105	120	100	150
105	50	100	150
105	50	100	100
105	50	-999999	0

Transaction Data

time	price	size	direction	initiator
t3	105	80	-1	1
t4	103	50	1	-1
t6	100	50	1	-1

Ask Data

time	price	size
t1	105	200
t3	105	120
t5	105	50

Bid Data

time	price	size
t2	100	150
t6	100	100
t7	-999999	0

Notes: Each row in the Message File describes the cause of the change in the order book from the previous row to the next. Event types 1, 2 and 3 refer to the submission, partial cancellation and total deletion of a limit order, events 4 and 5 to the execution of a visible and hidden limit order, respectively. The direction 1 (-1) refers to a buy (sell) limit order.

of the transactions. As the direction in the message file refers to the limit order, the initiator is given by the opposite party to the trade. Note that I omit the active counter-party to each trade from the trade data. In doing so, however, I do not omit any relevant information as the counter-party simply mirrors the passive trade with opposite trade direction. I remind the reader of that decision in the main text at any point where it is relevant, and discuss the alternative of including the counter-party to each trade.

The data for the ask side of the order book is constructed by extracting the state of the order book at any point the ask side is affected by the submission, cancellation, deletion or execution of a visible sell limit order (events 1, 2, 3, 4). Any event that is related to a hidden order is omitted. The construction of the bid side follows analogously.

Table 3.A1: Summary statistics of Nasdaq's transaction and quote data

	mean	std	min	25%	median	75%	max
T	6776.01	7647.55	17	1673	4397	9247.75	106407
V	1131.88	2347.50	1.37	165.53	467.49	1160.76	58115.01
V/T	129.02	87.20	44.65	95.50	108.31	129.33	3573.15
$V \geq 100$	0.77	0.12	0.21	0.71	0.79	0.87	0.98
$V = 100$	0.62	0.12	0.19	0.55	0.63	0.70	0.92
P	59.01	54.25	4.48	32.07	48.22	69.96	623.37
$\#Q$	104.10	102.26	1.16	28.62	70.39	150.38	908.74

Notes: This table provides summary statistics to the following variables computed for each stock-day: T – Number of trades, V – Trading volume in 1000 shares, V/T Volume per trade (stock-day average), $V \geq 100$ – Percentage of trades with volume greater or equal to 100 shares, $V = 100$ – Percentage of trades with trading volume equal to 100 shares, P – Price per share (stock-day average), $\#Q$ – Number of quote changes in 1000.

Compared to Chakrabarty et al. (2015), who study ITCH-data from the same time span for a size stratified sample of 300 stocks, my sample displays slightly higher trading activity measured by the daily, cross-sectional average of the total number of shares traded. Also, trades in my sample tend to be smaller and higher priced. However, the sample retains a great deal of variability and as the purpose is to analyze trade classification under the problem of imprecise timestamps, a focus on slightly more frequently traded stocks seems only proper. The average number of transactions on a stock-day is 6776, which is less than a trade per second. The average number of quote changes, however, is substantially larger with 104,100 quote updates on a stock-day, which is almost 4.5 quote updates per second. It is the large number of quote changes that makes accurate trade classification with imprecise timestamps challenging.

3.B The Tick-Test, LR, EMO and CLNV Algorithm

Table 3.B1: Trade classification algorithms

Variables: p_i – transaction price of the i^{th} trade; a_i, b_i – ask, bid price corresponding to the i^{th} trade; o_i – trade initiator			
Tick-Test	**LR** (Lee and Ready, 1991)	**EMO** (Ellis et al., 2000)	**CLNV** (Chakrabarty et al., 2007)

for $i = 1 : I$ **do**
  if $p_i > p_{i-1}$ **then**
    o_i = buyer
  else if $p_i < p_{i-1}$ **then**
    o_i = seller
  else
    $o_i = o_{i-1}$
	for $i = 0 : I$ **do**
	  $m_i = (a_i + b_i)/2$
	  if $p_i > m_i$ **then**
	    o_i = buyer
	  else if $p_i < m_i$ **then**
	    o_i = seller
	  else
	    $j = 0$
	    while $i - j > 0$ **do**
	      $j = j + 1$
	      if $p_i > p_{i-j}$ **then**
	        o_i = buyer
	        break
	      else if $p_i < p_{i-j}$ **then**
	        o_i = seller
	        break
		for $i = 0 : I$ **do**
		  if $p_i = a_i$ **then**
		    o_i = buyer
		  else if $p_i = b_i$ **then**
		    o_i = seller
		  else
		    $j = 0$
		    while $i - j > 0$ **do**
		      $j = j + 1$
		      if $p_i > p_{i-j}$ **then**
		        o_i = buyer
		        break
		      else if $p_i < p_{i-j}$ **then**
		        o_i = seller
		        break
			for $i = 0 : I$ **do**
			  $\underline{a} = 0.7a_i + 0.3b_i$
			  $\bar{b} = 0.3a_i + 0.7b_i$
			  if $\underline{a} < p_i \leq a_i$ **then**
			    o_i = buyer
			  else if $b_i \leq p_i < \bar{b}$ **then**
			    o_i = seller
			  else
			    $j = 0$
			    while $i - j > 0$ **do**
			      $j = j + 1$
			      if $p_t > p_{i-j}$ **then**
			        o_i = buyer
			        break
			      else if $p_i < p_{i-j}$ **then**
			        o_i = seller
			        break

Table 3.B2: Ticker names

AAPL	AA	ABB	ABT	ACE	ACN	ADBE	ADM	ADP
ADS	AEP	AGN	AGU	AIG	AKAM	ALK	ALL	AME
AMGN	AMT	AMX	AMZN	AN	AON	AOS	APC	APD
APH	ASH	ASR	AVGO	AVY	AXP	AYI	AZN	BAC
BAM	BAX	BA	BBL	BBT	BCE	BEAV	BEN	BHI
BHP	BIDU	BIIB	BK	BLK	BLL	BMS	BMY	BP
BRFS	BR	BTI	BT	BUD	BX	CAJ	CAT	CCK
CELG	CF	CHA	CHL	CHRW	CHT	CHU	CLX	CL
CMCSA	CMCSK	CME	CMI	CM	CNI	CNQ	COF	COP
COST	CPA	CPRT	CP	CRH	CRM	CSCO	CSGP	CSX
CTRP	CTSH	CUK	CVS	CVX	C	DAL	DCM	DD
DEO	DE	DHR	DISH	DIS	DOW	DTV	DUK	DVN
D	EBAY	ECL	EL	EMC	EMR	ENB	ENR	EOG
EPD	ESRX	ETE	ETN	EXC	EXPD	E	FCX	FDX
FIS	FLT	FMX	F	GD	GE	GG	GILD	GIS
GLW	GMCR	GM	GOOG	GPK	GPN	GPRO	GSK	GS
GT	GWR	HAL	HD	HMC	HON	HPQ	HSY	IBM
IBN	IGT	ILMN	IMO	INFY	INTC	IP	IR	ITW
JAH	JBHT	JBLU	JCI	JPM	KAR	KMB	KMX	KO
KR	KSU	K	LBTYA	LBTYK	LEG	LFL	LLY	LMT
LNKD	LOW	LO	LUV	LVS	LYB	MA	MCD	MCK
MELI	MET	MGA	MHK	MJN	MMC	MMM	MON	MOS
MO	MPC	MRK	MSCI	MSFT	MS	MT	NCR	NEE
NGG	NKE	NLSN	NOC	NSC	NTES	NTT	NUE	NVO
NVS	ODFL	ORCL	OXY	PAC	PBR	PCAR	PCLN	PCP
PEP	PFE	PG	PHG	PH	PKG	PKX	PM	PNC
POT	PPG	PRU	PSA	PTR	PX	QCOM	RAI	REGN
RIO	RKT	ROP	RTN	RYAAY	SAP	SAVE	SBUX	SCCO
SCHW	SIAL	SLB	SNE	SNP	SNY	SON	SO	SPB
SPG	SRE	STO	STT	STZ	SU	SWFT	SYT	SYY
TEF	TEL	TEVA	TGT	TJX	TMO	TM	TOT	TRP
TRV	TSLA	TSM	TSS	TS	TTM	TWC	TWX	TXN
T	UAL	UL	UNH	UNP	UN	UPS	USB	UTX
VALE	VFC	VLO	VMW	VRX	VZ	V	WFC	WHR
WIT	WMB	WMT	WM	WPZ	WU	XOM	XRX	YHOO
YUM	Z							

Notes: This table provides the ticker names of all stocks included in the sample. However, not all of these stocks are analyzed over the whole range of the sample as some stock-days may not have fulfilled the criteria mentioned in the data section (day-end price ≥ 1 \$ and number of trades ≥ 10), or due to an initial public offering during the sample period (e.g. Z).

3.C Tables from the Results Section

Table 3.C1: Classification accuracy at different timestamp precisions

i	0	1	2	3	4	9
quote changes	17	9	5	3	1	1
Panel A: total volume						
FI	95.02	96.97	97.95	98.34	98.24	98.18
EMO	90.13	93.64	96.08	97.61	98.19	98.20
CLNV	90.14	93.63	96.08	97.61	98.17	98.19
LR	89.88	93.57	96.02	97.52	98.10	98.10
EMOi	72.98	76.67	82.74	91.66	97.49	98.20
CLNVi	73.36	76.99	82.98	91.80	97.52	98.19
LRi	71.84	75.85	82.16	91.28	97.33	98.10
Panel B: average volume						
FI	94.52	96.47	97.38	97.71	97.53	97.44
	(2.40)	(1.91)	(1.91)	(2.04)	(2.05)	(2.09)
EMO	89.49	92.75	94.93	96.52	97.46	97.43
	(3.41)	(2.99)	(2.86)	(2.60)	(2.06)	(2.11)
CLNV	89.39	92.69	94.92	96.55	97.42	97.42
	(3.40)	(3.01)	(2.85)	(2.54)	(2.07)	(2.10)
LR	89.31	92.66	94.79	96.31	97.26	97.21
	(3.55)	(3.08)	(3.00)	(2.77)	(2.26)	(2.32)
EMOi	73.04	77.07	82.70	90.60	96.52	97.43
	(7.51)	(6.98)	(5.65)	(3.77)	(2.53)	(2.11)
CLNVi	74.38	78.21	83.57	91.10	96.61	97.42
	(8.05)	(7.35)	(5.68)	(3.40)	(2.38)	(2.10)
LRi	70.54	75.12	81.25	89.71	96.15	97.20
	(6.81)	(6.42)	(5.54)	(4.20)	(2.90)	(2.33)

Notes: This Table shows the percentage of correctly classified trading volume by the FI algorithm and the traditional algorithms using the last quotes from before the time of the trade (EMO, CLNV and LR) and using the interpolated time of trades and quotes (EMOi, CLNVi and LRi). The algorithms are applied to the data with reduced timestamp precisions (10^{-i} of a second for $i = 0, \ldots, 4$) and using the original precision of nanoseconds (10^{-9} of a second). These correspond to a median number of quote changes at the time of trades ranging from 17 (for $i = 0$) to 1 (for $i = 9$). Panel A shows the percentage of correctly classified volume summed over the entire sample. Panel B shows the average of correctly classified volume over the 19842 stock-days with the standard deviations in brackets.

Table 3.C2: Quote changes caused by a trade

Quote change due to trade	Number of Quotes 2	3	4	5	6	7	8	9	10	$\sum$
1st	13.93	8.59	6.06	4.34	3.21	2.43	1.85	1.44	1.14	42.99
2nd	5.13	3.00	1.99	1.39	1.03	0.77	0.60	0.47	0.38	14.76
3rd		2.11	1.42	0.99	0.72	0.54	0.42	0.33	0.27	6.80
4th			1.12	0.81	0.61	0.46	0.35	0.27	0.22	3.84
5th				0.69	0.50	0.40	0.30	0.24	0.19	2.32
6th					0.44	0.34	0.28	0.22	0.18	1.45
7th						0.31	0.24	0.20	0.16	0.92
8th							0.22	0.18	0.15	0.55
9th								0.17	0.14	0.31
10th									0.13	0.13
$\sum$	19.06	13.70	10.60	8.21	6.50	5.25	4.27	3.53	2.95	74.07

Notes: This table shows the number of times a trade is executed on the first, second, third, ..., 10-th quote within seconds of 2, 3, ..., 10 quote changes. Only seconds with a single trade were used for the computation. All numbers are presented in 10^5. For example, entry (1st, 2) means that within seconds containing 2 quote updates and one trade (of which there are 1.9 million), 1.4 million of the first quote changes were due to a trade.

3.D Approximation of the Reduction in Classification Accuracy due to Noisy Timestamps

To get a feeling of how we can expect the results to be affected if the reported time of trades is delayed by a random amount, we can calculate the reduction in classification accuracy under a few distributional assumption. The approximation may also help the practitioner to choose the appropriate timestamp precision at which to apply the classification algorithm in her data set, if she has a rough idea of the degree of noise.

Let the timestamp of a trade reflect the actual trade time plus noise, $\varepsilon \in \mathbb{R}_{\geq 0}$, which follows a distribution $F(\varepsilon)$. Let the probability of a trade at some point x over an interval of length s be determined by the density $g(x)$, $0 \leq x < s$. The fraction of trades that is placed outside the interval in which they actually occurred is then given by $\int_0^s g(x)(1 - F(s-x))\,\mathrm{d}x$.

For example, if trades are equally distributed over the interval s and the delay in reported time follows the exponential distribution, $\varepsilon \sim \mathrm{Exp}(1/\beta)$, the fraction of trades placed to the right of the interval during which they actually occurred is $\beta(1 - \exp\{-s/\beta\})/s$. For an interval of the length of a second ($s = 1$) and an average delay of one 10-th of a second ($\beta = 0.1$), that would mean that 10% of trades are reported outside the second in which they occurred.

Denoting the classification accuracy of all trades that lie in the correct interval by $A(s)$, the overall classification accuracy is given by

$$A(s) \int_0^s g(x)F(s-x)\,\mathrm{d}x + 0.5 \int_0^s g(x)(1 - F(s-x))\,\mathrm{d}x$$

assuming that the average classification accuracy of trades outside their actual time interval is 0.5. That is, the reduction in the accuracy due to delayed report times is given by

$$\int_0^s g(x)(1 - F(s-x))\,\mathrm{d}x\ (A(s) - 0.5).$$

For example, given the above classification accuracy of 95% for data timestamped at seconds ($s = 1$, with a median of 17 quote changes at the time of trades), we would expect the reduction in accuracy to be around 4.5%-points due to noise of intensity $\beta = 0.1$. Note that the reduction in accuracy may exceed the 4.5%-points if the classification accuracy for trades not shifted outside their interval is affected by the permutation of trades or if the accuracy of trades shifted just behind the interval at which they actually occurred is less than 0.5.

Given that at any timestamp precision s we found that in the absence of noise $A(s;\text{FI}) \gtrsim A(s;j)$ for $j = \text{EMO, CLNV, LR}$, we would expect the FI algorithm to dominate under noise as well.

3.E Data Construction for Logistic Regression

The data used in the regression analysis is filtered as follows. Observations where either an ask or bid price is not quoted at the time of the trade are dropped because the spread is not defined in these cases. I also drop the first trade, because neither R^2 nor Δt-Trade are defined for the first trade. If a trade is not preceded by at least one quote change, it is also dropped from the sample. Due to several quote changes happening at the same nanosecond it is possible that trades appear to be executed at negative spreads. Though the number of these instances is small, these observations are dropped. Due to several transactions taking place at the same nanosecond, it is also possible that trades appear to be executed outside the spread such that Q-Dist and Mid become negative. In these cases their values are truncated to 0. Table 3.E1 presents summary statistics after this filtering process, which gives still over 134 million transactions to analyze.

To get a first idea of the influence of the explanatory variables on the event of a correct classification, as well as on possible cross-correlations among the regressors, Figure 3.E1 depicts the correlation matrix of the explanatory variables and the dependent variable (the event of a correct classification by the FI algorithm applied at a timestamp precision of seconds). We see that, indeed, the number of misclassified transactions is higher for hidden orders, trades close to the mid-point and trades that execute away from the quote against which we would expect them to execute in the absence of hidden orders. As only trades that execute against a hidden order can be executed inside the spread, we see a strong, positive correlation between Hidden and the distance to the mid-point (Mid) or the distance to the quotes (Q-Dist).

In the regression analysis one could be worried that the coefficient of the event of a hidden order might overestimate the true effect of such an event as the placement of hidden orders and the execution against those may be viewed as endogenous decisions. In times of larger spreads, there is more room for placing hidden orders inside the spread, and traders searching for cheap execution prices may place successively small orders to find those hidden orders. The correlation matrix suggests that such concerns can be neglected. In fact, the bivariate correlation analysis suggests that other than the variables Hidden, Mid and Q-Dist there is no strong, linear impact of the explanatory variables on the event of a correct

Table 3.E1: Summary statistics of explanatory variables

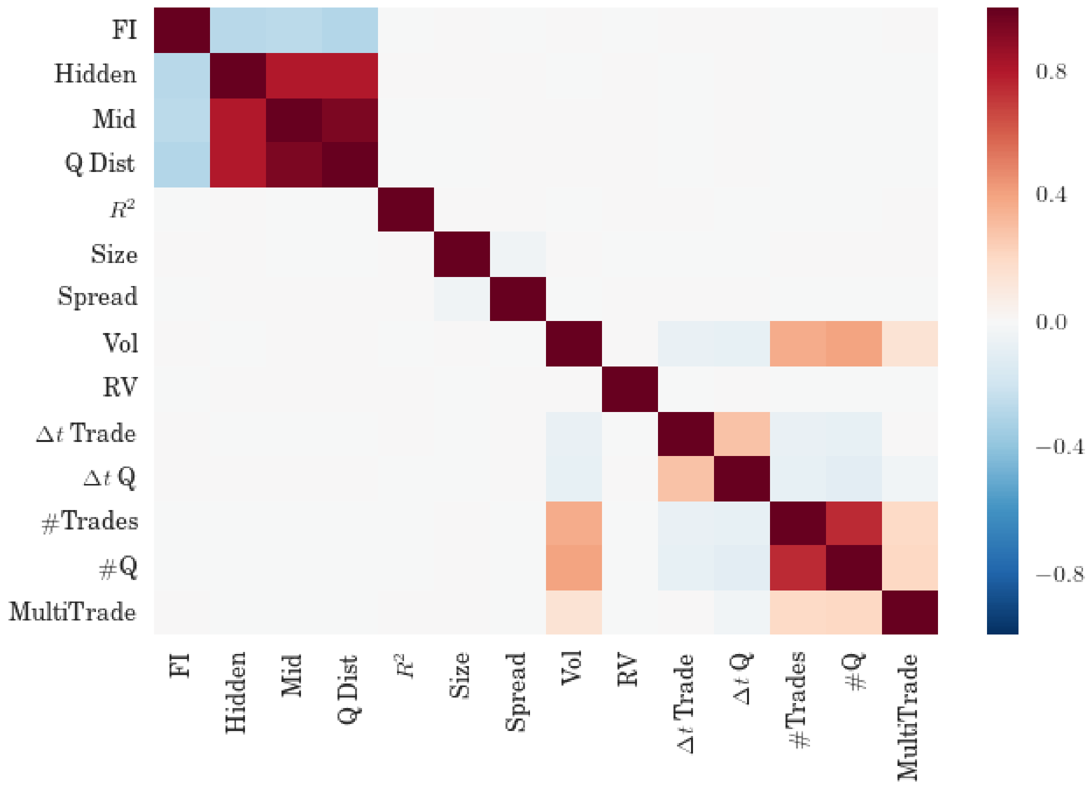
	mean	std	min	25%	50%	75%	max
Hidden	0.09	0.29	0	0	0	0	1
Mid	0.05	0.22	0	0	0	0	1
Q-Dist	0.03	0.12	0	0	0	0	1
$R^2 * 100$	0	0.11	0	0	0	0	1213.55
Size	1.67	4.50	0.01	1	1	1	4310.75
Spread	0.02	0.04	0.01	0.01	0.01	0.02	9.08
Vol	31.30	46.96	0.01	7.71	15.25	34.17	581.15
RV∗100	0.03	0.73	0	0.01	0.02	0.03	75.31
Δt Trade	6.62	23.01	0	0.01	0.48	5.21	12764.04
Δt Q	0.50	2.26	0	0	0.01	0.20	1098.91
# Trades	13.38	20.92	1	3	7	16	1020
# Q	76.70	104.44	0	18	46	98	32983
MultiTrade	0.65	0.48	0	0	1	1	1

Notes: This table shows the summary statistics of the explanatory variables. Hidden refers to a dummy variable taking the value 1 if the trade executed against a hidden order and 0 otherwise. Mid measures the distance of the execution price to the mid-quote, defined as $1 - 2|p_t - m_t|/(a_t - b_t)$, where p_t is the execution price, m_t is the corresponding mid-point and a_t, b_t are the corresponding ask and bid quote, respectively. Q-Dist measures the distance of the execution price to the quotes, defined as $|D_t - p_t|/(a_t - b_t)$ where $D_t = a_t$ if the trade is buyer-initiated and $D_t = b_t$ if the trade is seller-initiated. R^2 is the squared log-return of a transaction. Size is the number of shares exchanged in the transaction divided by 100. Spread is the absolute dollar spread at the time of the trade. Vol is the total trading volume of the stock-day divided by 10^5. RV is the 5-minute realized variation over the stock-day. Δt-Trade is the number of seconds since the previous trade. Δt-Q is the number of seconds since the last quote change. #Trades is the number of transactions during the same second of the trade. #Q is the number of quote changes during the second of the trade. MultiTrade is a dummy variable taking the value 1 if the transaction is part of a trade involving more than one counter-party and 0 otherwise.

classification.

To insure that we do not encounter problems with a few extreme outliers in the estimation procedure, I did not consider observations where one of the variables from R^2 to #Q exceeded their 99th-percentile. The summary statistics of this restricted sample used in the regression analysis are presented in Table 3.E2.

Figure 3.E1: Correlation matrix



Notes: This figure shows the correlations between the explanatory variables of the regression model, which are summarized in Table 3.E1, and the dependent binary variable of a correct/false classification of a trade by the FI algorithm, as well as the correlations between the explanatory variables themselves depicted as a heat map.

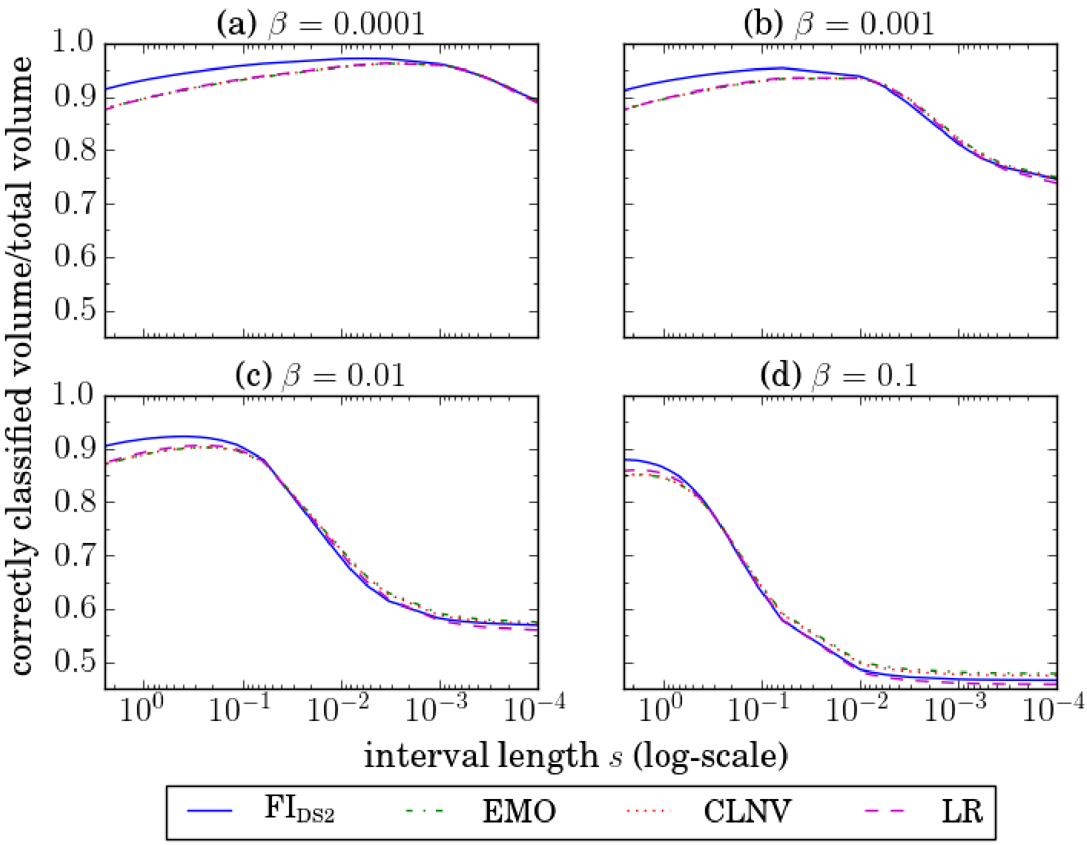
Table 3.E2: Summary statistics of explanatory variables for the restricted sample

	mean	std	min	25%	50%	75%	max
Hidden	0.09	0.28	0	0	0	0	1
Mid	0.05	0.21	0	0	0	0	1
Q-Dist	0.03	0.12	0	0	0	0	1
$R^2 * 10^6$	0.01	0.04	0	0	0	0	0.42
Size	1.38	1.46	0.01	1	1	1	13.99
Spread	0.02	0.02	0.01	0.01	0.01	0.01	0.15
Vol	27.52	34.51	0.01	7.83	15.12	31.78	233.70
RV*100	0.02	0.02	0	0.01	0.02	0.03	0.12
Δt Trade	4.98	10.46	0	0	0.45	4.83	81.36
Δt Q	0.34	0.89	0	0	0.01	0.18	7.75
#Trades	11.69	13.32	1	3	7	15	91
#Q	69.02	71.11	0	19	46	95	468
MultiTrade	0.65	0.48	0	0	1	1	1
N Obs.: 124254433							

Notes: This table shows the summary statistics of the sample presented in Table 3.E1 restricted to observations where the variables R^2 to #Q do not exceed their 99th-percentile. This sample provides the baseline for the logistic regression of the probability of a correct classification of a transaction by the FI algorithm.

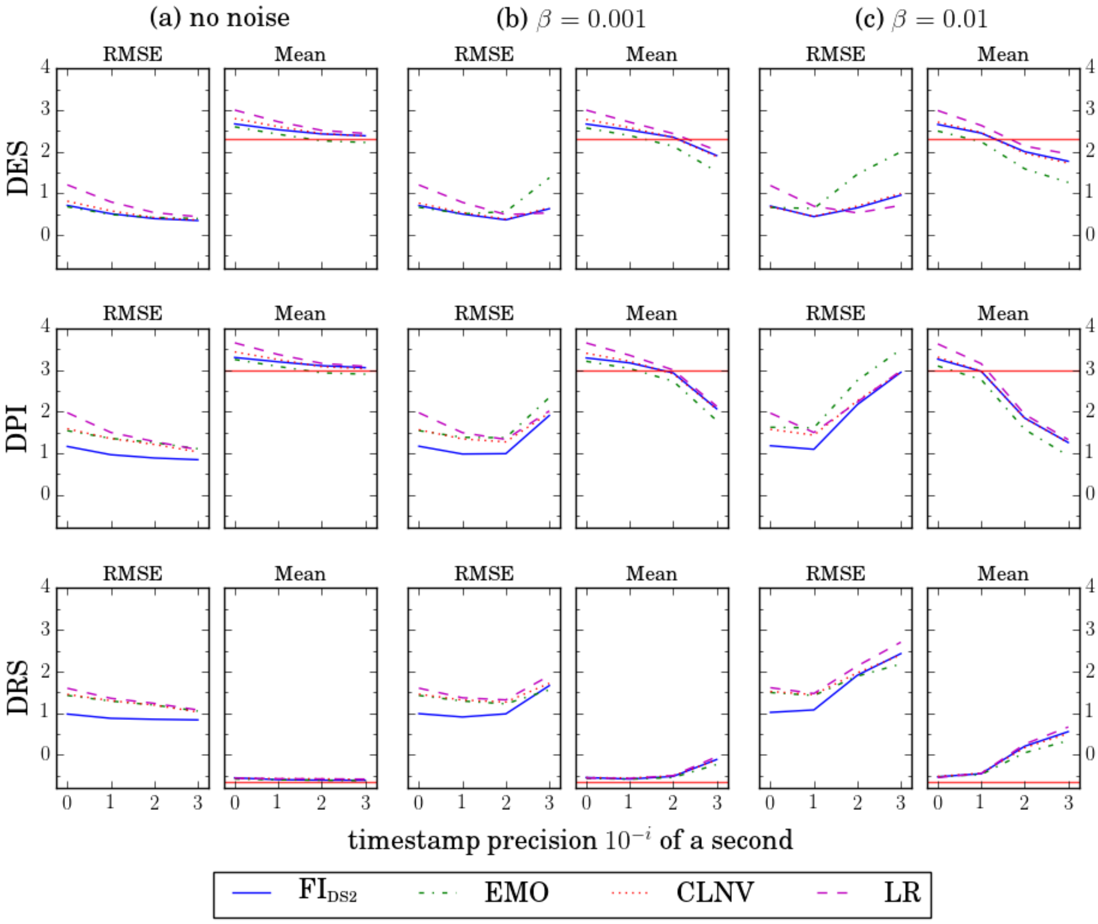
3.F The Full-Information Algorithm under Data Structure 2

Figure 3.F1: Classification accuracy under delayed trade times and Data Structure 2



Notes: This figure shows the fraction of correctly classified trading volume (y-axis) for the data with delayed trade times under Data Structure 2. The trade time equals the actual trade time plus ε, with $\varepsilon \sim \text{Exp}(1/\beta)$ and $\beta \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$. The classification algorithms FI_{DS2}, EMO, CLNV and LR are apply to the data with reduced timestamp precision (s) ranging from 10^{-4} of a second to 2.5 seconds presented on $\log_{10}$-scale (x-axis).

Figure 3.F2: Estimating liquidity under Data Structure 2



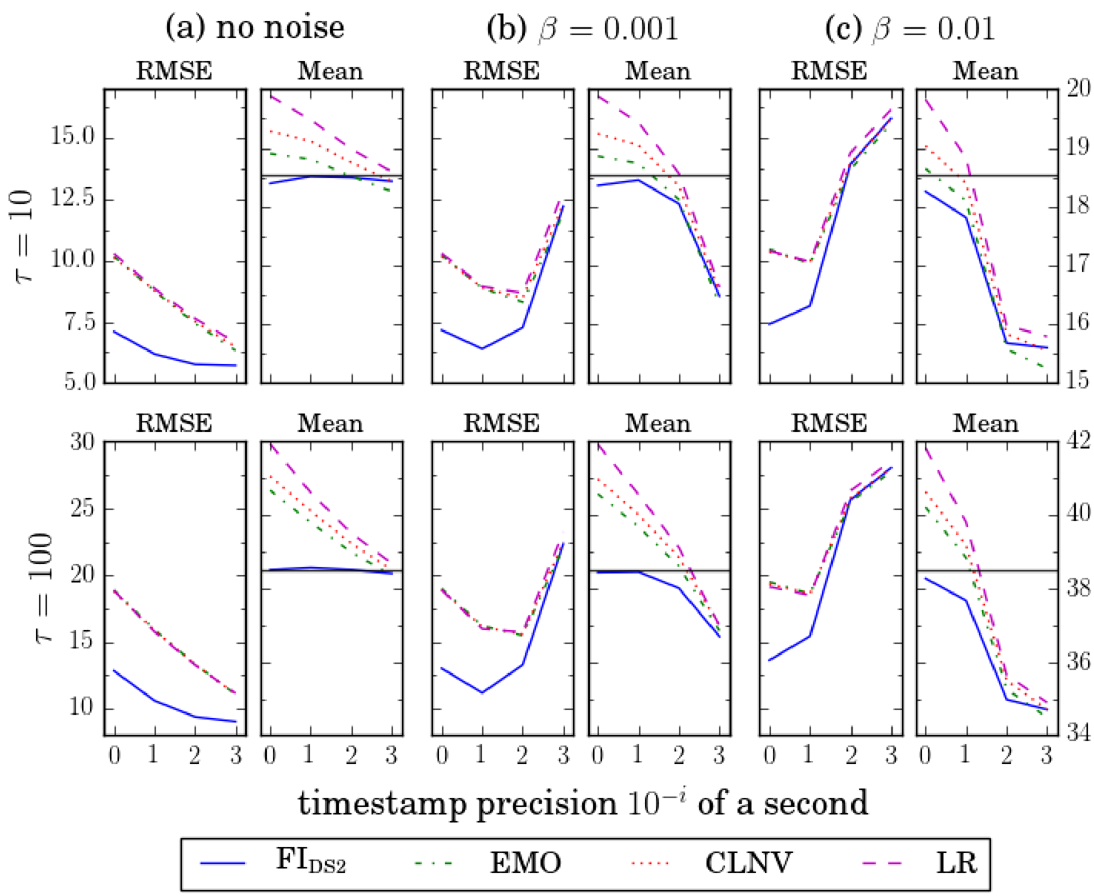
Notes: This Figure shows the sample averages and the root-mean-square error between the stock-day estimates and the true values of the dollar effective spread (DES), the dollar price impact (DPI) and the dollar realized spread (DRS) for the $\mathrm{FI}_{\mathrm{DS2}}$ algorithm and the traditional algorithms under Data Structure 2. The algorithms are applied to the data with and without delayed trade times, where the delay is given by $\varepsilon \sim \mathrm{Exp}(1/\beta)$ with $\beta = 10^{-3}, 10^{-2}$, and with varying timestamp precision ranging from seconds to milliseconds.

Table 3.F1: Classification accuracy of FI_{DS2}

		timestamp precision: 10^{-i} of a second for $i =$					
		0	1	2	3	4	9
Panel A: overall correctly classified volume (in %)							
total		93.36	96.16	97.71	98.30	98.25	98.21
mean		93.42	95.93	97.18	97.65	97.54	97.45
std		2.64	2.02	1.96	2.06	2.05	2.09
Panel B: % correctly classified volume in each classification category							
visible	cl. step						
YES	0	–	–	–	–	–	–
	2	99.50	99.76	99.94	99.99	100.00	100.00
	3	90.80	94.52	96.78	97.34	67.51	–
	4a	94.79	98.34	99.72	99.98	100.00	100.00
	4b	98.73	99.57	99.90	99.99	100.00	100.00
	4	49.78	55.15	59.45	66.66	75.93	76.00
	5	53.39	52.45	51.94	50.54	38.99	36.09
NO	0	–	–	–	–	–	–
	2	71.68	79.77	88.60	95.57	99.77	99.96
	3	87.78	92.83	95.55	96.81	97.13	–
	4a	81.82	88.50	91.88	95.52	99.88	100.00
	4b	48.50	63.19	84.89	95.75	99.85	100.00
	4	90.11	92.52	93.41	93.16	92.15	91.53
	5	64.87	64.73	65.87	67.56	68.99	69.14
Panel C: % classified volume in each classification category							
visible	cl. step						
YES	0	0.00	0.00	0.00	0.00	0.00	0.00
	2	46.52	59.06	72.95	85.69	89.49	89.49
	3	43.48	30.85	16.84	3.94	0.00	0.00
	4a	0.22	0.27	0.32	0.39	0.46	0.46
	4b	0.12	0.19	0.29	0.40	0.46	0.46
	4	0.01	0.01	0.01	0.00	0.00	0.00
	5	0.07	0.04	0.03	0.01	0.00	0.00
NO	0	0.00	0.00	0.00	0.00	0.00	0.00
	2	2.89	3.04	3.06	2.67	1.80	1.67
	3	1.47	0.82	0.33	0.06	0.00	0.00
	4a	0.53	0.29	0.07	0.02	0.00	0.00
	4b	0.46	0.25	0.15	0.13	0.12	0.12
	4	1.82	2.35	2.69	2.79	2.74	2.74
	5	2.40	2.82	3.28	3.91	4.92	5.05

Notes: This Table shows the percentage of correctly classified trading volume for the FI algorithm adjusted to the Data Structure 2. "cl. step" refers to the accuracy at the corresponding step of the classification procedure.

Figure 3.F3: Estimating order imbalance under Data Structure 2



Notes: This Figure shows the sample averages and the root-mean-square error between estimates of the order imbalance and the true order imbalance displayed in percent for the data with aggregated quote changes. For the computation of the order imbalance each stock-day is split into equally sized volume bins. The number of bins is chosen to be $\tau = 10, 100$. The algorithms are applied to the data with and without delayed trade times, where the delay is given by $\varepsilon \sim \text{Exp}(1/\beta)$ with $\beta = 10^{-3}, 10^{-2}$, and with varying timestamp precision ranging from seconds to milliseconds.

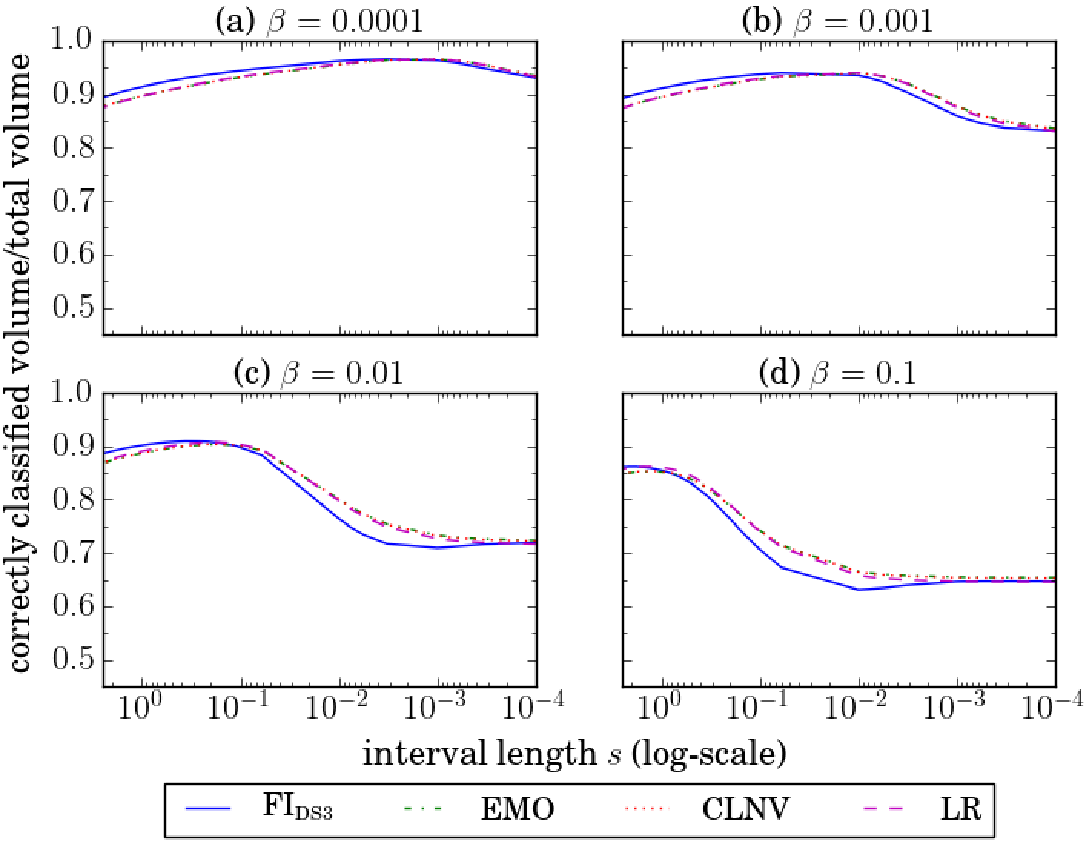
3.G The Full-Information Algorithm under Data Structure 3

Table 3.G1: Classification accuracy of FI_{DS3}

		timestamp precision: 10^{-i} of a second for $i =$					
		0	1	2	3	4	9
Panel A: overall correctly classified volume (in %)							
total		91.73	94.84	96.82	97.92	98.22	98.21
mean		92.33	94.86	96.43	97.40	97.52	97.45
std		2.97	2.24	2.00	2.05	2.05	2.09
Panel B: % correctly classified volume in each classification category							
visible	cl. step						
YES	0	–	–	–	–	–	–
	2	99.70	99.81	99.93	99.99	100.00	100.00
	3	89.36	94.14	97.24	98.88	98.15	–
	4	98.30	99.15	99.60	99.82	99.90	99.90
	5	51.77	50.85	48.37	42.03	41.39	38.94
NO	0	–	–	–	–	–	–
	2	67.64	66.03	71.96	83.44	97.17	99.81
	3	89.87	93.29	93.49	84.18	89.45	–
	4	83.29	89.46	92.33	93.00	92.50	91.95
	5	64.53	64.39	65.42	67.27	68.95	69.14
Panel C: % classified volume in each classification category							
visible	cl. step						
YES	0	0.00	0.00	0.00	0.00	0.00	0.00
	2	35.66	44.61	59.19	78.89	89.47	89.49
	3	54.56	45.54	30.84	10.95	0.05	0.00
	4	0.14	0.22	0.35	0.57	0.90	0.93
	5	0.06	0.06	0.04	0.01	0.00	0.00
NO	0	0.00	0.00	0.00	0.00	0.00	0.00
	2	2.78	3.13	3.39	3.04	1.87	1.68
	3	2.43	1.58	0.69	0.12	0.00	0.00
	4	1.95	2.11	2.34	2.65	2.84	2.85
	5	2.41	2.76	3.16	3.77	4.87	5.05

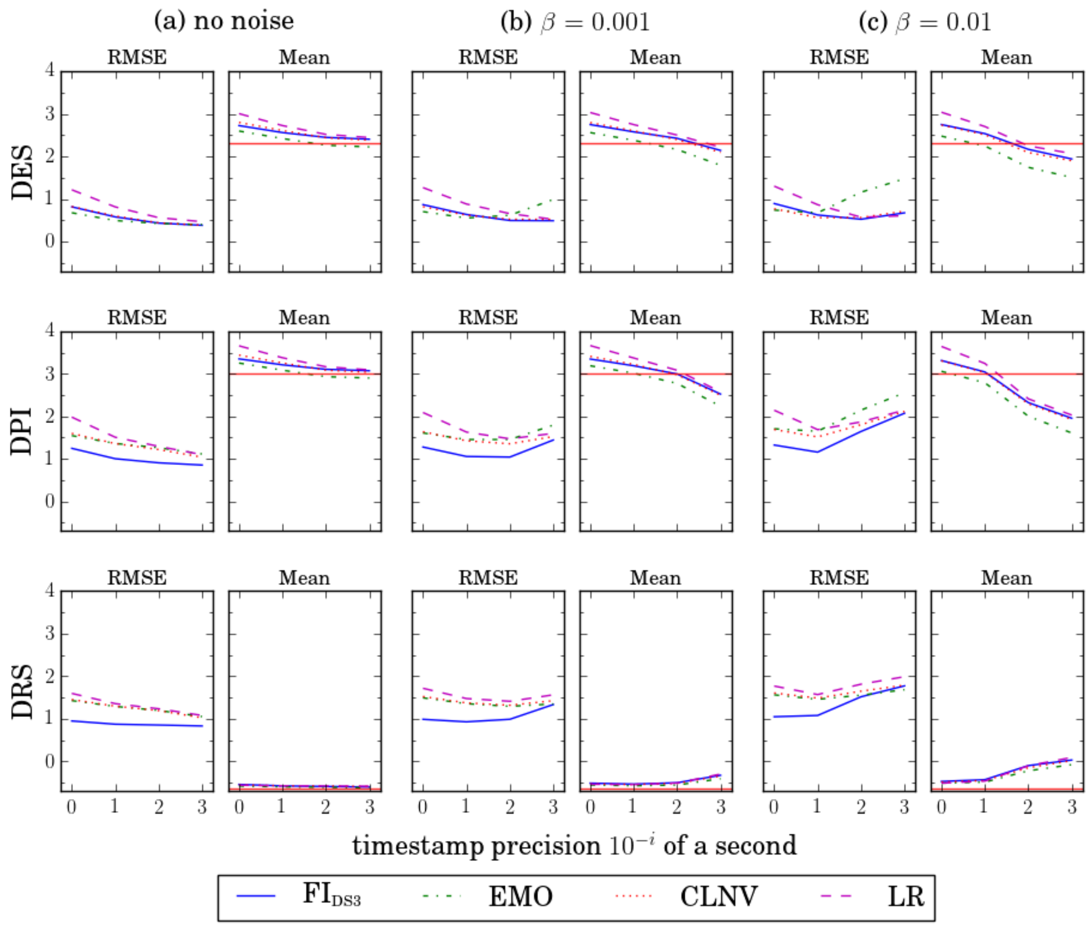
Notes: This Table shows the percentage of correctly classified trading volume for the FI algorithm adjusted to the Data Structure 3. “cl. step” refers to the accuracy at the corresponding step of the classification procedure.

Figure 3.G1: Classification accuracy under random trade and quote times and Data Structure 3



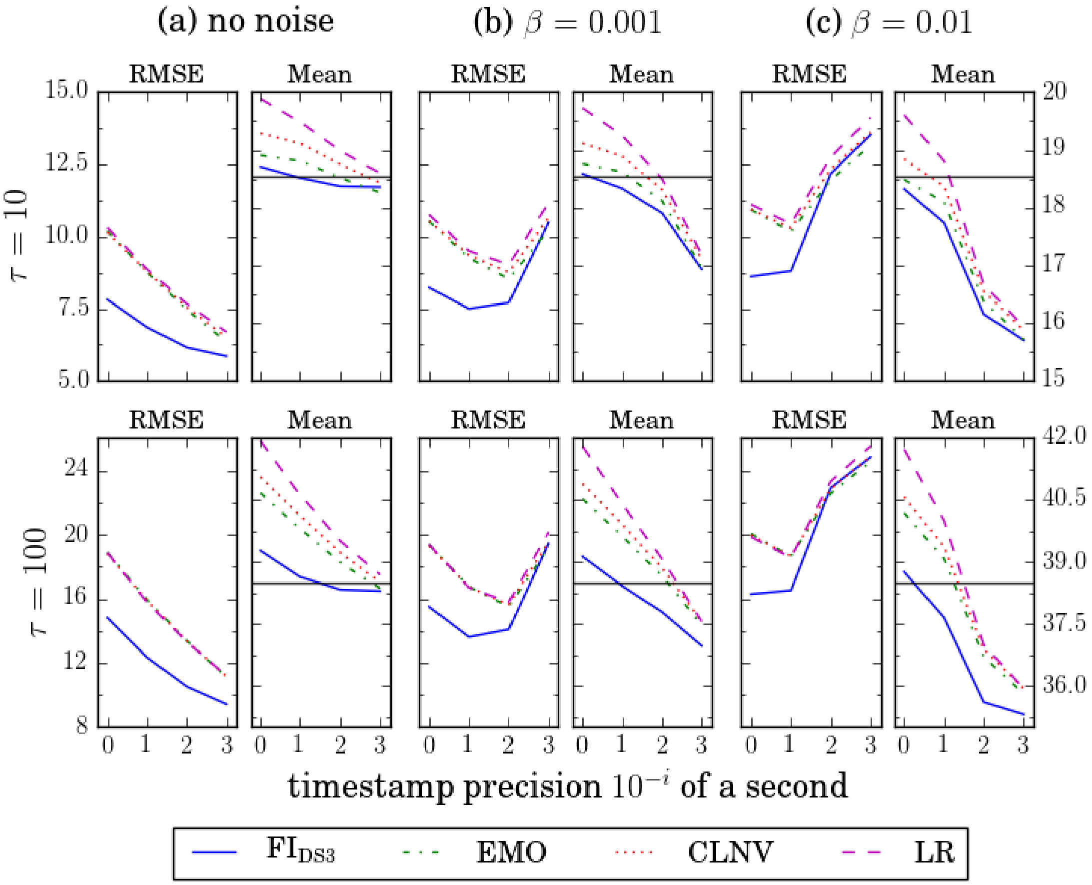
Notes: This figure shows the fraction of correctly classified trading volume (y-axis) for the data with noisy quote and trade times (Data Structure 3). The recorded time of trades and quotes equals the actual time plus ε, with $\varepsilon \sim \text{Exp}(1/\beta)$ and $\beta \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$. The classification algorithms FI_{DS3}, EMO, CLNV and LR are apply to the data with reduced timestamp precision (s) ranging from 10^{-4} of a second to 2.5 seconds presented on $\log_{10}$-scale (x-axis).

Figure 3.G2: Estimating liquidity under random trade and quote order



Notes: This Figure shows the sample averages and the root-mean-square error between the stock-day estimates and the true values of the dollar effective spread (DES), the dollar price impact (DPI) and the dollar realized spread (DRS) for the $\mathrm{FI}_{\mathrm{DS2}}$ algorithm and the traditional algorithms under the data structure with random trade and quote order (Data Structure 3). The algorithms are applied to the data with and without noise. The noise is applied to both trade and quote times, where the noise is given by $\varepsilon \sim \mathrm{Exp}(1/\beta)$ with $\beta = 10^{-3}, 10^{-2}$. The timestamp precision ranges from seconds to milliseconds.

Figure 3.G3: Estimating order imbalance under random trade and quote order



Notes: This Figure shows the sample averages and the root-mean-square error between estimates of the order imbalance and the true order imbalance displayed in percent for the data with random trade and quote order. For the computation of the order imbalance each stock-day is split into equally sized volume bins. The number of bins is chosen to be $\tau = 10, 100$. The algorithms are applied to the data with and without noise. The noise is applied to both trade and quote times, where the noise is given by $\varepsilon \sim \text{Exp}(1/\beta)$ with $\beta = 10^{-3}, 10^{-2}$. The timestamp precision ranges from seconds to milliseconds.

Bibliography

Angel, J. J., Harris, L. E., Spatt, C. S., 2011. Equity trading in the 21st century. The Quarterly Journal of Finance 1 (01), 1–53.

Angel, J. J., Harris, L. E., Spatt, C. S., 2015. Equity trading in the 21st century: An update. The Quarterly Journal of Finance 5 (01), 1–39.

Asquith, P., Oman, R., Safaya, C., 2010. Short sales and trade classification algorithms. Journal of Financial Markets 13 (1), 157–173.

Avery, C., Zemsky, P., 1998. Multidimensional uncertainty and herd behavior in financial markets. American Economic Review, 724–748.

Banerjee, A. V., 1992. A Simple Model of Herd Behavior. The Quarterly Journal of Economics 107 (3), 797–817.

Barber, B. M., Odean, T., Zhu, N., 2009. Systematic noise. Journal of Financial Markets 12 (4), 547 – 569.

Bernile, G., Hu, J., Tang, Y., 2016. Can information be locked up? Informed trading ahead of macro-news announcements. Journal of Financial Economics 121 (3), 496–520.

Bessembinder, H., 2003. Issues in assessing trade execution costs. Journal of Financial Markets 6 (3), 233–257.

Bikhchandani, S., Hirshleifer, D., Welch, I., 1992. A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. Journal of Political Economy 100 (5), 992–1026.

Bikhchandani, S., Sharma, S., 2001. Herd Behavior in Financial Markets. IMF Staff Papers 47 (3), pp. 279–310.

Bishop, C. M., 2009. Pattern Recognition and Machine Learning, 8th Edition. Springer.

Blais, M., Protter, P., 2012. Signing trades and an evaluation of the Lee–Ready algorithm. Annals of Finance 8 (1), 1–13.

Blyth, C. R., 1980. Expected absolute error of the usual estimator of the binomial parameter. The American Statistician 34 (3), 155–157.

Boortz, C., 2016. Herding in financial markets. Ph.D. thesis, Freie Universität Berlin.
URL http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000102374

Boortz, C., Kremer, S., Jurkatis, S., Nautz, D., 2014. Information risk, market stress and institutional herding in financial markets: New evidence through the lens of a simulated model. SFB 649 Discussion Paper 2014-029.

Brown, N. C., Wei, K. D., Wermers, R., 2014. Analyst recommendations, mutual fund herding, and overreaction in stock prices. Management Science 60 (1), 1–20.

Brunnermeier, M. K., 2001. Asset pricing under asymmetric information: Bubbles, crashes, technical analysis, and herding. Oxford University Press.

Chakrabarty, B., Li, B., Nguyen, V., Van Ness, R. A., 2007. Trade classification algorithms for electronic communications network trades. Journal of Banking & Finance 31 (12), 3806–3821.

Chakrabarty, B., Moulton, P. C., Shkilko, A., 2012. Short sales, long sales, and the Lee–Ready trade classification algorithm revisited. Journal of Financial Markets 15 (4), 467–491.

Chakrabarty, B., Pascual, R., Shkilko, A., 2015. Evaluating trade classification algorithms: Bulk volume classification versus the tick rule and the Lee-Ready algorithm. Journal of Financial Markets 25, 52–79.

Chari, V., Kehoe, P. J., 2004. Financial crises as herds: overturning the critiques. Journal of Economic Theory 119 (1), 128 – 150.

Chiang, T. C., Zheng, D., 2010. An empirical analysis of herd behavior in global stock markets. Journal of Banking and Finance 34 (8), 1911 – 1921.

Chordia, T., Goyal, A., Jegadeesh, N., 2016. Buyers versus sellers: Who initiates trades, and when? Journal of Financial and Quantitative Analysis 51 (5), 1467–1490.

Chordia, T., Hu, J., Subrahmanyam, A., Tong, Q., 2017. Order flow volatility and equity costs of capital. Management Science Forthcoming.

Chordia, T., Roll, R., Subrahmanyam, A., 2002. Order imbalance, liquidity, and market returns. Journal of Financial economics 65 (1), 111–130.

Christie, W. G., Huang, R. D., 1995. Following the pied piper: Do individual returns herd around the market? Financial Analysts Journal 51 (4), 31 – 37.

Cipriani, M., Guarino, A., 2005. Herd behavior in a laboratory financial market. American Economic Review, 1427–1443.

Cipriani, M., Guarino, A., 2008. Herd Behavior and Contagion in Financial Markets. The B.E. Journal of Theoretical Economics (Contributions) 8 (1), 1 – 54.

Cipriani, M., Guarino, A., 2009. Herd behavior in financial markets: an experiment with financial market professionals. Journal of the European Economic Association 7 (1), 206–233.

Cipriani, M., Guarino, A., 2014. Estimating a Structural Model of Herd Behavior in Financial Markets. American Economic Review 104 (1), 224 – 251.

Cochran, W. G., 1954. Some methods for strengthening the common χ^2 tests. Biometrics 10 (4), 417–451.

Collin-Dufresne, P., Fos, V., 2015. Do prices reveal the presence of informed trading? The Journal of Finance 70 (4), 1555–1582.

Dasgupta, A., Prat, A., Verardo, M., 2011. The price impact of institutional herding. Review of Financial Studies 24 (3), 892–925.

Devenow, A., Welch, I., 1996. Rational herding in financial economics. European Economic Review 40 (3), 603–615.

Diaconis, P., Zabell, S., 1991. Closed form summation for classical distributions: variations on a theme of De Moivre. Statistical Science, 284–302.

Dorn, D., Hubermann, G., Sengmueller, P., 2008. Correlated trading and returns. The Journal of Finance 63 (2), 885–920.

Drehmann, M., Oechssler, J., Roider, A., 2005. Herding and contrarian behavior in financial markets: An internet experiment. The American Economic Review 95 (5), 1403.

Easley, D., de Prado, M. L., O'Hara, M., 2016. Discerning information from trade data. Journal of Financial Economics 120 (2), 269–285.

Easley, D., de Prado, M. M. L., O'Hara, M., 2012. Flow toxicity and liquidity in a high-frequency world. Review of Financial Studies 25 (5), 1457–1493.

Easley, D., Hvidkjaer, S., O'hara, M., 2002. Is information risk a determinant of asset returns? The journal of finance 57 (5), 2185–2221.

Easley, D., Kiefer, N., O'Hara, M., 1997. One day in the life of a very common stock. Review of Financial Studies 10 (3), 805–835.

Easley, D., Kiefer, N. M., O'Hara, M., 1996a. Cream-skimming or profit-sharing? The curious role of purchased order flow. The Journal of Finance 51 (3), 811–833.

Easley, D., Kiefer, N. M., O'hara, M., Paperman, J. B., 1996b. Liquidity, information, and infrequently traded stocks. The Journal of Finance 51 (4), 1405–1436.

Easley, D., O'Hara, M., 1987. Price, trade size, and information in securities markets. Journal of Financial Economics 19 (1), 69 – 90.

Eguiluz, V. M., Zimmermann, M. G., 2000. Transmission of information and herd behavior: an application to financial markets. Physical review letters 85 (26), 5659.

Ellis, K., Michaely, R., O'Hara, M., 2000. The accuracy of trade classification rules: Evidence from Nasdaq. Journal of Financial and Quantitative Analysis 35 (04), 529–551.

Enke, B., Zimmermann, F., 2018. Correlation neglect in belief formation. The Review of Economic Studies forthcoming.

Epanechnikov, V. A., 2006. Non-parametric estimation of a multivariate probability density. Theory of Probability and Its Applications 14 (1), 153–158.

Finucane, T. J., 2000. A direct test of methods for inferring trade direction from intra-day data. Journal of Financial and Quantitative Analysis 35 (04), 553–576.

Freedman, D. A., 1965. Bernard friedman's urn. The Annals of Mathematical Statistics, 956–970.

Frey, S., Herbst, P., Walter, A., 2014. Measuring mutual fund herding–a structural approach. Journal of International Financial Markets, Institutions and Money 32, 219–239.

Friedman, B., 1949. A simple urn model. Communications on Pure and Applied Mathematics 2 (1), 59–70.

Froot, K. A., Scharfstein, D. S., Stein, J. C., 1992. Herd on the street: Informational inefficiencies in a market with short-term speculation. The Journal of Finance 47 (4), 1461–1484.

Garren, S. T., 2004. Handbook of Beta Distribution and Its Applications. Marcel Dekker, Ch. 9, pp. 237–253.

Glosten, L. R., Milgrom, P. R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14 (1), 71 – 100.

Graham, J. R., 1999. Herding among investment newsletters: Theory and evidence. The Journal of Finance 54 (1), 237–268.

Gupta, A. K., Nadarajah, S., 2004. Handbook of beta distribution and its applications. CRC press.

Hasbrouck, J., 1991. Measuring the information content of stock trades. The Journal of Finance 46 (1), 179–207.

Hautsch, N., Huang, R., 2012. Limit Order Flow, Market Impact and Optimal Order Sizes: Evidence from NASDAQ TotalView-ITCH Data. In: Abergel, F., Bouchaud, J.-P., Foucault, T., Lehalle, C.-A., Rosenbaum, M. (Eds.), Market Microstructure: Confronting Many Viewpoints. Wiley, pp. 136–161.

Henker, T., Wang, J.-X., 2006. On the importance of timing specifications in market microstructure research. Journal of Financial Markets 9 (2), 162–179.

Hirshleifer, D., Hong Teoh, S., 2003. Herd Behaviour and Cascading in Capital Markets: a Review and Synthesis. European Financial Management 9 (1), 25–66.

Hirshleifer, D., Subrahmanyam, A., Titman, S., 1994. Security analysis and trading patterns when some investors receive information before others. The Journal of Finance 49 (5), 1665–1698.

Holden, C. W., Jacobsen, S., 2014. Liquidity measurement problems in fast, competitive markets: expensive and cheap solutions. The Journal of Finance 69 (4), 1747–1785.

Hu, J., 2014. Does option trading convey stock price information? Journal of Financial Economics 111 (3), 625–645.

Hu, J., 2017. Option listing and information asymmetry. Review of Finance Forthcoming.

Huang, R., Polak, T., 2011. Lobster: Limit order book reconstruction system. Tech. rep., School of Business and Economics, Humboldt Universität zu Berlin. URL https://lobsterdata.com/info/docs/LobsterReport.pdf

Huang, R. D., Stoll, H. R., 1996. Dealer versus auction markets: A paired comparison of execution costs on NASDAQ and the NYSE. Journal of Financial economics 41 (3), 313–357.

Hwang, S., Salmon, M., 2004. Market stress and herding. Journal of Empirical Finance 11 (2), 585–616.

King, G., Zeng, L., 2001. Logistic regression in rare events data. Political analysis, 137–163.

Kremer, S., Nautz, D., 2013a. Causes and consequences of short-term institutional herding. Journal of Banking & Finance 37 (5), 1676–1686.

Kremer, S., Nautz, D., 2013b. Short-term Herding of Institutional Traders: New Evidence from the German Stock Market. European Financial Management 19 (4), 730 – 746.

Lakonishok, J., Shleifer, A., Vishny, R. W., 1992. The impact of institutional trading on stock prices. Journal of Financial Economics 32 (1), 23–43.

Lee, C., Ready, M. J., 1991. Inferring trade direction from intraday data. The Journal of Finance 46 (2), 733–746.

Lee, I. H., 1998. Market crashes and informational avalanches. The Review of Economic Studies 65 (4), 741–759.

Liu, L. Y., Patton, A. J., Sheppard, K., 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. Journal of Econometrics 187 (1), 293–311.

Lux, T., 1995. Herd behaviour, bubbles and crashes. The economic journal, 881–896.

Lux, T., 1998. The socio-economic dynamics of speculative markets: interacting agents, chaos, and the fat tails of return distributions. Journal of Economic Behavior & Organization 33 (2), 143–165.

Muravyev, D., 2016. Order flow and expected option returns. The Journal of Finance 71 (2), 673–708.

Neerchal, N. K., Morel, J. G., 1998. Large cluster results for two parametric multinomial extra variation models. Journal of the American Statistical Association 93 (443), 1078–1087.

O'Hara, M., 2015. High frequency market microstructure. Journal of Financial Economics 116 (2), 257–270.

Park, A., Sabourian, H., 2011. Herding and contrarian behavior in financial markets. Econometrica 79 (4), 973–1026.

Park, A., Sgroi, D., 2012. Herding, contrarianism and delay in financial market trading. European Economic Review 56 (6), 1020 – 1037.

Park, A., Sgroi, D., Januar 2016. Herding and contrarian behavior in financial markets: An experimantal analysis. CRETA DP17 (17).

Parzen, E., 1958. On asymptotically efficient consistent estimates of the spectral density function of a stationary time series. Journal of the Royal Statistical Society. Series B (Methodological) 20 (2), 303–322.

Patterson, D. M., Sharma, V., 2010. The incidence of informational cascades and the behavior of trade interarrival times during the stock market bubble. Macroeconomic Dynamics 14 (S1), 111–136.

Perlin, M., Brooks, C., Dufour, A., 2014. On the performance of the tick test. The Quarterly Review of Economics and Finance 54 (1), 42–50.

Peterson, M., Sirri, E., 2003. Evaluation of the biases in execution cost estimation using trade and quote data. Journal of Financial Markets 6 (3), 259–280.

Piwowar, M. S., Wei, L., 2006. The sensitivity of effective spread estimates to trade–quote matching algorithms. Electronic Markets 16 (2), 112–129.

Rosenthal, D. W., 2012. Modeling trade direction. Journal of Financial Econometrics 10 (2), 390–415.

Scharfstein, D. S., Stein, J. C., 1990. Herd Behavior and Investment. The American Economic Review 80 (3), pp. 465–479.

Schwert, G. W., 2011. Stock volatility during the recent financial crisis. European Financial Management 17 (5), 789–805.

Sias, R. W., 2004. Institutional Herding. Review of Financial Studies 17 (1), 165–206.

Silverman, B. W., 1978. Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. Ann. Statist. 6 (1), 177–184.

Tarone, R., 1979. Testing the goodness of fit of the binomial distribution. Biometrika 66 (3), 585–590.

Theissen, E., 2001. A test of the accuracy of the Lee/Ready trade classification algorithm. Journal of International Financial Markets, Institutions and Money 11 (2), 147–165.

Vergote, O., March 2005. How to match trades and quotes for nyse stocks? Center for Economic Studies Discussion Paper Series.

Vives, X., 1996. Social learning and rational expectations. European Economic Review 40 (3 to 5), 589 – 601.

Weizsäcker, G., 2010. Do We Follow Others When We Should? A Simple Test of Rational Expectations. The American Economic Review 100 (5), 2340.

Welch, I., 1992. Sequential sales, learning, and cascades. The Journal of finance 47 (2), 695–732.

Wermers, R., 1999. Mutual fund herding and the impact on stock prices. The Journal of Finance 54 (2), 581–622.

Wylie, S., 2005. Fund Manager Herding: A Test of the Accuracy of Empirical Results Using U.K. Data. The Journal of Business 78 (1), 381–403.

Zhou, R. T., Lai, R., 2009. Herding and information based trading. Journal of Empirical Finance 16 (2), 388–393.

Ehrenwörtliche Erklärung

Ich habe die vorgelegte Dissertation selbst verfasst und dabei nur die von mir angegebenen Quellen und Hilfsmittel benutzt. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen sind, sowie alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.

Berlin, den 4. Mai 2018