Ekonomi och samhälle
Economics and Society

Skrifter utgivna vid Svenska handelshögskolan
Publications of the Hanken School of Economics

Nr 197

Arto Thurlin

Essays on Market Microstructure

Price Discovery and Informed Trading

Helsinki 2009

Essays on Market Microstructure: Price Discovery and Informed Trading

Keywords: market microstructure, price discovery, informed trading, tick data, cross-listing, market transparency, event study, macroeconomic data, equities.

Arto Thurlin
Hanken School of Economics
Department of Finance and Statistics
P.O.Box 479, 00101 Helsinki, Finland

Distributor:

Library
Hanken School of Economics
P.O.Box 479
00101 Helsinki, Finland

Telephone: +358-40-3521 376, +358-40-3521 265
Fax: +358-40-3521 425
E-mail: publ@hanken.fi
http://www.hanken.fi

ISBN 978-952-232-032-2 (printed)
ISBN 978-952-232-033-9 (PDF)
ISSN 0424-7256

Edita Prima Ltd, Helsinki 2009

PREFACE

Most of this research was carried out at the Department of Finance and Statistics at the Hanken School of Finance in Helsinki. I would like to express my deepest gratitude to the Head of the Department, Professor Eva Liljeblom, for providing the resources to conduct this research, including extensive intraday databases and the computing power needed to analyze the data. I am also indebted to the two other members of my Thesis Advisory Committee, Assistant Professor Anders Ekholm and Professor Mats Hansson, who provided me with innumerable valuable comments and suggestions. I am particularly pleased that many of our discussions with Anders took place in the soothing environment of a traditional Finnish smoke sauna, and during invigorating bicycle rides.

I also conducted part of this research as a Visiting Scholar at HEC in Paris and the Australian School of Business at UNSW in Sydney. I would therefore also like to thank Professor Thierry Foucault, Professor Peter Swan and Senior Lecturer Joakim Westerholm for stimulating talks, comments, and ideas during these research visits. I also wish to thank Mikko Leppämäki, Director of the Graduate School of Finance, for organizing excellent opportunities for presenting my work at seminars and workshops. Mikko also arranged for me to meet with several excellent researchers in my field, and in this way also made my first overseas visit possible.

I am also greatly indebted to the external examiners of the thesis, Professor Erik Theissen and Professor Thierry Foucault. This thesis benefited greatly from their perceptive, constructive and detailed comments.

My colleagues at the Department of Finance and Statistics have been a source of both inspiration and fun during my years at Hanken. I have had the pleasure to work with Anders Wilhelmsson, Andriy Andreev, Annika Sandström, Benjamin Maury, Bernard Ben Sita, Doureige Jurdi, Erkki Nikoskelainen, Hanna Westman, Hans-Kristian Sjöholm, Jan Antell, Karl Felixson, Magnus Blomkvist, Mari Hintikka, Niklas Ahlgren, Nikolas Rokkanen, Olga Karakozova, Peter Nyberg, Sheraz Ahmed, and many others.

I gratefully acknowledge financial support from Stiftelsen Svenska handelshögskolan, CEFIR, Näringslivets fond, Stiftelsen för främjandet av värdepappersmarknaden i Finland, Bröderna Ernst och Lars Krogius forskningsfond, Säästöpankkien tutkimussäätiö, Nordea pankin säätiö, and OP-Ryhmän tutkimussäätiö.

Finally, I want to thank my parents, Airi and Juhani, *sine quibus non*.

CONTENTS

APPENDICES

Essay 1

TABLES

Essay 1

Essay 2

Essay 3

FIGURES

Part 1

THEORETICAL BACKGROUND AND MAIN FINDINGS

1 INTRODUCTION

In this thesis, I study several aspects of market microstructure. Market microstructure has been defined as "the study of the process and outcomes of exchanging assets under explicit trading rules" (O'Hara (1995)), and as "the study of the trading mechanisms used for financial securities" (Hasbrouck (2007)). Hasbrouck (2007) also lists the following three features as setting the study of market microstructure apart from a more classical view of financial economics. First, microstructure seeks to understand the sources of value and reasons for trade, in a setting with different types of traders, and different private and public information sets. Second, the actual mechanisms of trade are a continually changing object of study. These include continuous markets, auctions, limit order books, dealer markets, or combinations of these operating as a hybrid market. And third, microstructure has to allow for the possibility of multiple prices. At any given time an investor may be faced with a multitude of different prices, depending on whether he or she is buying or selling, the quantity he or she wishes to trade, and the required speed for the trade. The price may also depend on the relationship that the trader has with potential counterparties.

I touch upon all of the above issues in this research. I do this by studying three specific areas, all of which have both practical and policy implications. First, I study the role of information in trading and pricing securities in markets with a heterogeneous population of traders, some of whom are informed and some not, and who trade for different private or public reasons. Second, I study the price discovery of stocks in a setting where they are simultaneously traded in more than one market. Third, I make a contribution to the ongoing discussion about market design, i.e. the question of which trading systems and ways of organizing trading are most efficient. A common characteristic throughout my thesis is the use of high frequency datasets, i.e. tick data. These datasets include all trades and quotes in a given security, rather than just the daily closing prices, as in traditional asset pricing literature.

This thesis consists of four separate essays. The main topics and contributions of the essays are the following. In the first essay I study price discovery for European companies cross-listed in the United States. I also study explanatory variables for differences in price discovery. In my second essay, I contribute to earlier research on two issues of broad interest in market microstructure: market transparency and informed trading. After the changes made in the trading system at the OMX Helsinki Stock Exchange, no market participant was able to see the identity of other traders posting buy and sell orders in the limit order book. I examine the effects of this change to market quality. I broaden my focus slightly in the third essay, to include releases of macroeconomic data in the United States. I analyze the effect of these releases on European cross-listed stocks. The fourth and last essay examines the uses of standard methodologies of price discovery analysis in a novel way within one single market. Specifically, I find that local exchange members dominate price discovery, even if a majority of all trading is done by remote traders from abroad.

In this Introduction to the PhD Thesis, I wish to accomplish several things. I start out in Section II by describing the literature pertinent to my research. By no means do I intend to present an exhaustive and complete survey of the subject. Rather, I wish to convey an idea of which earlier research, both theoretical and empirical, is relevant to my research, and to which existing literature I want to bring my "incremental contribution", as the saying goes. In Section III my objective is to build a bridge from first principles in asset pricing to the methodologies and ideas used in studying price discovery and informed trading. This is of course a very bold aim. My intention should be understood as an intuitive overview of the steps leading from early theoretical work in market microstructure in the 1980s, to research done during the last few years, including this Thesis. In Section IV I describe my datasets, as well as the exchanges and markets included in my essays. I summarize my essays in Section V, and offer some concluding remarks in Section VI.

2 RELEVANT EARLIER RESEARCH

Market microstructure has seen a lot of growth in the last two or three decades. The term is usually attributed to Garman (1976), who used it in the title of a paper studying inventory costs and market making. Many of the most important theories in microstructure were formulated and presented during the 1970s and 1980s. A partial list would have to include at least the following. Roll (1984) presents the distinction between the fundamental value of a security, and the observed market prices, which may differ from the fundamental value due to market organization and the trading process[1]. His model is based on the random walk for asset prices, and homogeneous traders. Another very important group of models allow for heterogeneous traders, who enter the market sequentially and independently. An early example of these models was presented by Glosten and Milgrom (1985). Once you allow for a trader to enter the market several times, the need for strategic optimization arises, as in the Kyle (1985) model. I discuss sequential trading models in Section III below.

There are several general surveys of market microstructure. For the theory of market microstructure, the standard reference is O'Hara (1995). Two newer book references are Hasbrouck (2007) and Vives (2008), the former more empirical and the latter more theoretical in its focus. Harris (2003) is the standard reference for a practical description of how security markets work, with a thorough description of trading practices in different asset classes globally. Two relatively recent general survey papers are Madhavan (2000), and Biais, Glosten, and Spatt (2005). Lyons (2001) focuses on the microstructure of the foreign exchange market, where most trading is conducted away from exchanges, on an over-the-counter basis. A study focused on the volatility of asset prices, and stochastic volatility in particular is Shephard (2005). Dacorogna, Gencay, Mueller, Olsen, and Pictet (2001) offer a more practitioner-oriented view on the analysis of high-frequency data, with an emphasis on the foreign exchange markets.

Many theoretical papers study informed trading. An empirical method of studying the amount of informed trading in a market was formulated in a number of papers by Easley, Kiefer, O'Hara, and Paperman (1996), and Easley, Kiefer, O'Hara (1996, 1997). This model, the Probability of Informed Trading (PIN), is based on the Glosten and Milgrom (1985) sequential trading model.

The PIN model has been used to examine numerous issues in finance. Easley, O'Hara, and Paperman (1998) study the relationship between international analyst coverage and informed trading, without finding a significant relationship. Easley, O'Hara, and Saar (2001) find that informed trading increases after a stock split. Easley, Hvidkjaer, and O'Hara (2002) find that within the Fama – French (1992) asset pricing framework, there is a significant positive relation between informed trading, measured by PIN, and the expected returns of a stock. Duarte and Young (2007) take the results of Easley et al. (2002), and decompose the PIN into asymmetric information and illiquidity components. They find that illiquidity is priced, whereas information is not. Grammig, Schiereck, and Theissen (2001) compare the anonymous automated Xetra market and the non-anonymous trading floor of the Frankfurt Stock Exchange. They find that information based trading is significantly lower in floor trading.

With the growing internationalization of markets, cross-listings have risen to prominence during the last two decades. Karolyi (2006) is a comprehensive survey of

[1] Roll (1984) was not the first to present these ideas; see e.g. Demsetz (1968).

the literature. Price reactions to the news of a cross-listing have generally been found to be positive, both for U.S. companies seeking a listing abroad, and for non-U.S. companies listing in the U.S. The most comprehensive studies are Miller (1999), and Foerster and Karolyi (1999). A lower cost of capital is made possible in most cases by a broader investor base, and getting across possible technical, legal, or tax barriers between countries (see, e.g. Errunza and Miller (2000)). Notable theoretical models of trading in a multiple markets setting are presented and discussed, among others, by Chowdhry and Nanda (1991), Biais (1993), Madhavan (1995), de Frutos and Manzano (2002), and Yin (2005).

Early empirical studies on multi-market trading, arbitrage, and price discovery use low frequency daily data. Lau and Diltz 1994 study opening and closing prices of Japanese companies cross-listed in New York. They find that both markets have an impact on the pricing of these stocks. Hauser, Tanchuma, and Yaari (1998), and Lieberman, Ben-Zion, and Hauser (1999) study Israeli stocks listed in New York, and find that in most cases pricing is efficient in the sense that there are no arbitrage profits. They also find that the home market dominates price discovery. Kim, Szakmary, and Mathur (2000) consider the exchange rate as well as the U.S. market index for a group of stocks from five different markets. They find that the U.S. market, as well as the exchange rate, has an influence on the pricing of these cross-listed stocks. Wang, Rui, and Firth (2002) study pricing and volatility between the London and Hong Kong markets. They find that both markets influence asset returns as well as volatility.

The availability of high frequency data enables researchers to study the pricing of a stock, when it is simultaneously traded on two or more exchanges. Hedvall, Liljeblom, and Nummelin (2000) conduct one of the first studies using high frequency data. They study the most liquid stock in the Finnish market, Nokia, and find that the New York market is dominant for price discovery. Ding, Harris, Lau, and McInish (1999) find that the pricing of a Malaysian company cross-listed in Singapore mainly occurs in the home market. Eun and Sabherwal (2003) study price discovery for Canadian stocks cross-listed in New York. They find huge differences in shares of price discovery. Grammig, Melvin, and Schlag (2004) perform a comprehensive study of price discovery for stocks from Canada, Germany, France, and the UK, cross-listed at the NYSE. Using data for 1999, they find that the home exchange typically dominates, but that there are great differences between companies. Pascual, Pascual-Fuster, and Climent (2006) and Phylaktis and Korczak (2007) study Spanish and U.K. stocks cross-listed in the U.S., respectively. Sapp (2002) studies the foreign exchange market, and the contributions of five large banks acting as dealers in this largely over-the-counter market, i.e. a market operating outside the confines of a centralized exchange. Menkveld, Koopman, and Lucas (2007) study price discovery for Dutch stocks cross-listed in the U.S., using state-space methods, and the entire combined trading day, instead of focusing exclusively on the simultaneous trading hours of the two exchanges.

3 ECONOMETRIC METHODOLOGY

Econometric analysis of high frequency financial data plays a central role in this thesis. In this section, I discuss the most important methodologies that I use in the essays. I also give an intuitive account of the derivation of the models used, starting from the basic tenets of asset pricing.

I do not attempt to present an overview of the entire field. Rather, I will focus on the two most important areas of research for this thesis: informed trading and multiple prices.

The main methodological tool used in the second essay is the maximum likelihood estimation of the Probability of Informed Trading (PIN). The Glosten and Milgrom (1985) model is a starting point for this method of empirical analysis of informed trading.

Price discovery is an essential function of a securities market. According to the efficient markets hypothesis (see, e.g. Fama (1965, 1970)), prices reflect all available information in a quick and accurate manner. It is not clear, however, how this process occurs in practice. Theoretical models suggest that the acquisition of information is costly, which means that not all market participants may be able or willing to do so at all times. The literature on price discovery studies these issues. Price discovery has been described as the impounding of new information into the security price (Hasbrouck (1995)), and as consisting of the efficient and timely incorporation of the information implicit in investor trading into market prices (Lehmann (2002)). In the first, third, and fourth essays I estimate shares of price discovery in a multiple markets setting, i.e. when a security trades in more than one market. A generalized version of the Roll (1984) model is a starting point for an analysis of multiple prices and price discovery. I discuss this model in Part 3.2 below.

In all my papers, I also use many standard econometric techniques, such as Ordinary Least Squares, and unit root and cointegration analysis. I do not discuss these methods in this Introduction.[2]

3.1. Informed trading

3.1.1. The Glosten and Milgrom model

In this chapter I present an intuitive outline of the model presented by Glosten and Milgrom (1985). This discussion is based on Glosten and Milgrom (1985), O'Hara (1997), and Hasbrouck (2007). This model starts with an analysis of different types of traders. There is a risk neutral and competitive market maker, who sets prices and trades with a population of traders. The traded asset has an eventual true value, toward which the trading price converges. There are two types of traders, informed and uninformed traders. In this multiple period model, the asset has an eventual value V, which is known to the informed traders. That there are uninformed traders should not be understood as meaning that these traders are necessarily irrational. Rather, their

[2] Standard econometrics texts provide excellent references; See e.g. Greene (2007) and Hayashi (2000).

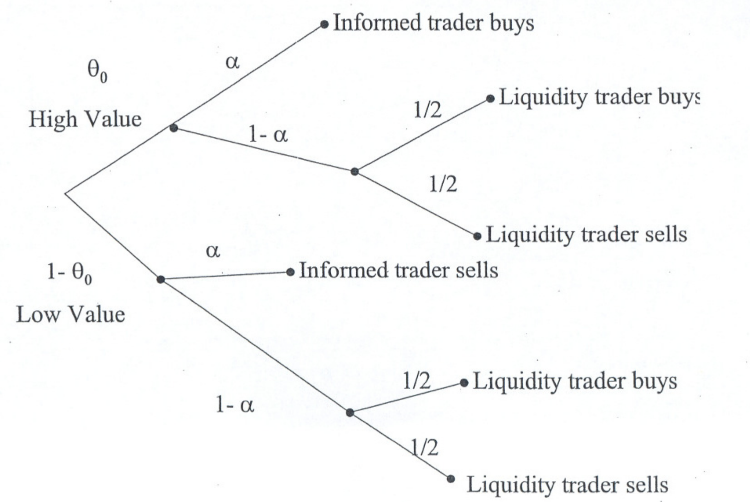
motives for trading are exogenous to the model, such as portfolio rebalancing for an institutional investor.

Trade takes place sequentially, meaning that only one trader is allowed to transact with the market maker at any given time. The uninformed traders face a problem in that the informed traders know the true value of the traded asset, and will thus profit at the expense of the uninformed traders. The goal of the informed traders is to maximize their profits from the information they possess.

It is usually assumed that traders are chosen probabilistically, for example according to a Poisson process. In this way, the market maker always faces the same probability of trading with an informed trader. Please refer to Figure 1 for a graphical presentation of the trading process.

Figure 1 The Glosten Milgrom (1985) Model

Θ_0 is the probability of positive news and a high final value for the security and α is the proportion of informed traders. An informed trader will always buy if the news is good, and always sell when the news is bad. An uninformed liquidity trader is equally likely to buy or sell.



The market maker then sets prices in such a way, that the expected profit on each trade is zero. Thus, the market maker will expect to lose on average when trading with an informed trader, and to gain when faced with an uninformed (liquidity) trader. The rationale for this is that since market makers are competitive and risk neutral, any positive average profit would disappear due to competition from other market makers. This implies that prices are set equal to the conditional expectation of the value of the asset, given the type of the trade (buy or sell). This means that the market maker's bid price is equal to the expected value of the asset, given that someone is willing to sell the asset. With a certain probability the seller is informed, and the trade will incur a loss for the market maker. It may also be that the seller is uninformed, and that the trade results in a profit for the market maker.

In this model buys and sells are not equally likely. This is because once there is at least one informed trader, who possesses information about the fair value of the asset being higher (lower) than the prevailing price, this informed trader will only be buying (selling) the asset. An important result is that the size of the bid-ask spread, i.e. the difference between the prevailing bid and ask prices, arises independently of any inventory or trading costs. It only depends on factors such as the number of informed traders, and the nature of the underlying information. Another important result of the model is that transaction prices form a martingale. Simply put, this means that the market maker's best prediction of future prices is the current price. Using the concepts of Fama (1965), this suggests that markets converge towards strong-form efficiency, reflecting all information, including private information.

Various extensions to this model have been proposed, allowing for order of different sizes (Easley and O'Hara (1987)), and orders of other types than market orders, including stop orders (Easley and O'Hara (1991)). Easley and O'Hara (1992) introduce event uncertainty to the basic sequential model. In their model it is not certain whether an information event has occurred or not. This is an important step towards the modeling of informed trading (PIN),

3.1.2. Information-based Trading and the PIN measure

The Probability of Informed Trading (PIN) measures information asymmetry between investors and traders in a stock. The model was introduced by Easley, Kiefer, O'Hara, and Paperman (1996), and Easley, Kiefer, O'Hara (1996, 1997). The model is based on the Glosten and Milgrom (1985) sequential trading model presented above.

The following is a presentation of the model, in the form it is widely used in the literature. There are three types of traders in our model: informed traders, uninformed liquidity traders, and risk neutral market makers. At time zero an information event may occur. This event has either a positive ('high'), or negative ('low') effect on the value of the security. An informed trader is risk neutral, takes prices as given, and does not engage in any strategic behavior.

If an information event has occurred in period zero, in period one an informed trader acts upon this information, which only he or she possesses. If an informed trader has seen a high signal, he or she buys the stock if the current quote is below the value given the signal; if he has seen a low signal, he will sell if the quote is above the value given the signal. If there is no information event, an informed trader does not trade.

The uninformed trader's behavior is more complex. In most microstructure models, the presence of traders with better information dictates that an uninformed trader trading for speculative reasons would always be better off not trading at all. To avoid this no-trade equilibrium, at least some uninformed traders must transact for nonspeculative reasons such as liquidity needs or portfolio considerations. We assume that when an uninformed trader checks the quote, the probability that he will trade is strictly positive.

Before trading starts each day, an information event occurs with probability α. There is thus a probability of $(1 - \alpha)$ that there is no information, in which case only the uninformed traders are active in the market. If there is information, it is negative with probability δ and positive with probability $(1 - \delta)$. In the former case, informed traders sell, in the latter, they buy. Liquidity traders are equally likely to buy or sell shares. We model the arrival of informed and liquidity traders as Poisson processes, with intensity parameter μ for informed trader, and ε_s and ε_b for sell and buy orders from liquidity traders.

Once we have estimated the above parameters we are in a position to calculate the Probability of Informed Trading, PIN, as

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_s + \varepsilon_b}, \tag{1}$$

where $\alpha\mu + \varepsilon_s + \varepsilon_b$ is the arrival rate of all orders, and $\alpha\mu$ is the arrival rate of informed orders.

There is an ongoing discussion about the uses of the PIN model, as well as estimation and improvements to the model. Based on both simulation results and empirical findings, Boehmer, Grammig, and Theissen (2007) show that inaccurate input in the form of buy and sell data lead to downward-biased estimates of PIN. Easley, Hvidkjaer, and O'Hara (2005), and Yan and Zhang (2006) suggest improved numerical methods and extensions for the maximum likelihood estimation procedure. A recent paper by Easley, Engle, O'Hara, and Wu (2008) allows for time-varying PIN measures, incorporating the GARCH family of volatility specifications.

Hasbrouck (2007) points out some of the weaknesses of the PIN measure. Although the model is a reasonable depiction of order flow one-sidedness, there are obvious shortcomings. The input dataset only includes the number of buy and sell trades over an entire trading day. This means that the model cannot distinguish between events that occur only once during the day, as opposed to several times, or during different times of a day.

3.2. Price discovery

In this section, we present the steps needed to go from a simple Roll (1984) model, which takes into account the bid-ask spread, to a multivariate description of the price process. This is an apt description of the pricing of a cross-listed stock, simultaneously traded in more than one exchange. The following discussion is based on Lehmann (2002), Harris et al. (2002a, 2002b), Baillie et al. (2002), de Jong (2002), Hasbrouck (2002, 2007).

3.2.1. The Roll (1984) model and its extensions

A random walk process is the sum of independently and identically distributed (i.i.d.) random variables. Stock prices, which, unlike fixed-income securities, have neither a maturity nor boundary conditions, are plausibly approximated by a random walk. A random walk has the property that the first differences (p_t – p_{t-1}) are a stationary process, whereas the price process itself is not[3].

[3] A random walk, or more formally, an integrated process *I(1)*, is nonstationary. However, it has the property that the differences of that process are stationary. It is a special case of a martingale. Two random walk processes are said to be cointegrated, when a linear combination of them is stationary. For a more detailed explanation, see e.g. Engle and Granger (1987), or Hamilton (1994). Johansen (1995) discusses cointegration and VAR models in depth.

Roll (1984) starts with the following random walk process:

$$m_t = m_{t-1} + \mu_t + u_t, \tag{2}$$

where t is the transaction time, m_t is the price, μ_t is the expected return, and u_t is an identically and independently distributed (i.i.d.) random variable. In order to be able to model transaction prices, however, it is necessary to take the bid-ask spread into account. This can be modeled in the following way:

$$p_t = m_t + q_t c, \tag{3}$$

where q_t is the trade indicator: +1 for a buy, and -1 for a sell; c is the half spread (1/2 of the difference of the bid and the ask prices). A succinct result of the classic Roll model is that it is possible to relate the cost of trading, defined as the bid-ask spread, to the first-order autocorrelation of the price series:

$$Cov(\Delta p_t, \Delta p_{t-1}) = -c^2, \tag{4}$$

where Δp_t and Δp_{t-1} are the price changes in periods t and t-1, respectively. In the absence of quote data, this is a way to estimate the cost of trading, based solely on the autocorrelation of the observed prices.

The model can be extended to include asymmetric information. Hasbrouck (2007) uses the term Generalized Roll model to describe this extension. In this extension, the trade indicator q_t is unobserved. The structural model is now the following:

$$m_t = m_{t-1} + w_t \tag{5}$$

$$w_t = \lambda q_t + u_t, \tag{6}$$

where w is an *i.i.d.* random variable, and λ is the information content of the trade (if there are no informed traders, λ=0). If we now add the order processing cost c (the half spread), we get the following structural model:

$$ask_t = m_{t-1} + u_t + c + \lambda \tag{7}$$

$$bid_t = m_{t-1} + u_t - c - \lambda, \tag{8}$$

where m_{t-1} is the quote midpoint in the previous period, u_t is an i.i.d. random variable, c is the half spread, and λ is the information content of the trade.

Once we have defined this generalized Roll model, we can take the analysis one step further, and consider a multivariate linear model.

3.2.2. Price discovery in a multiple markets setting

The above discussion forms a basis for a study of price discovery of a stock in a multiple markets setting. There are two main methodological approaches to measuring price discovery: the error correction approach, also known as "Permanent-Temporary decomposition" (PT), of Harris, McInish, Shoesmith, and Wood (1995) and the so-called "Information Shares" (IS) method of Hasbrouck (1995). A thorough discussion of the two methodologies, including their differences and similarities, is the topic of a special edition of the Journal of Financial Markets (Issue 3, 2002). The papers published in this issue include Lehmann (2002), Harris et al. (2002a, 2002b), Baillie et al. (2002), de Jong (2002), and Hasbrouck (2002). The main conclusion of these papers is that the models are directly related and provide broadly similar results in most cases. This overview of the methodologies is based on the above papers and Hasbrouck (2007).

Continuing from our presentation above, we have a structural model, with an efficient price m_t that is common to both price series. Let us consider the following model

$$m_t = m_{t-1} + u_t \tag{9}$$

$$p_{1t} = m_t + cq_t \tag{10}$$

$$p_{2t} = m_{t-1}, \tag{11}$$

where m_t is the efficient price, p_{1t} and p_{2t} are two price series, and q_t is the trade type indicator, *-1* or *+1*. The first price reflects the Roll model with a bid-ask spread, whereas the second price is based on the lagged efficient price.

We are now in a position to construct a VMA representation of this model. According to the Wold Theorem, any zero-mean covariance stationary process can be represented in the form of an infinite moving average process[4]. However, the moving average process is not invertible, which means that it is not possible to construct an autoregressive (AR) process out of the MA representation. This poses a problem for estimation. The solution to this dilemma is a so-called error correction model in vector form: a vector error correction model (VECM). In this way, we treat the vector of prices as having a common stochastic trend, which is a random walk. The actual prices are then based on this common trend, adding only trade related noise to form the actual observed market prices. Below I present the main steps of this analysis.

The autocovariances of the price differences Δp_t are:

$$\Gamma_0 = \begin{bmatrix} 2c^2 + \sigma_u^2 & 0 \\ 0 & \sigma_u^2 \end{bmatrix} \tag{12}$$

$$\Gamma_1 = \begin{bmatrix} -c^2 & 0 \\ \sigma_u^2 & 0 \end{bmatrix}. \tag{13}$$

4 Wold's decomposition: Any zero-mean covariance-stationary process Y_t can be represented in the form

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \kappa_t ,$$

where $\psi_0 = 1$ and $\sum_{j=0}^{\infty} \psi^{\infty}_{j=0} < \infty$. (Hamilton (1994), page 109).

The higher autocovariances for k > 1 are zero. Using these autocovariances, we get the general form of a vector moving average (VMA) model:

$$\Delta p_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} \tag{14}$$

Where we have the following terms:

$$\theta_1 = \left(c^2 + \sigma_u^2\right) \begin{bmatrix} -c^2 & c^2 \\ \sigma_u^2 & -\sigma_u^2 \end{bmatrix} \tag{15}$$

$$\Omega = Var(\varepsilon_t) = \begin{bmatrix} 2c^2 + \sigma_u^2 - \dfrac{c^4}{c^2 + \sigma_u^2} & c^2 \sigma_u^2 \\ \dfrac{c^2 \sigma_u^2}{c^2 + \sigma_u^2} & \dfrac{c^2 \sigma_u^2}{c^2 + \sigma_u^2} \end{bmatrix} \tag{16}$$

Usually, a vector autoregressive (VAR) representation, which exists for most VMA models, is very convenient for estimation. We wish to write the coefficients for the VAR representation for the following model:

$$\varphi(L)\Delta p_t = \varepsilon_t \tag{17}$$

We get the following coefficients:

$$\varphi(L) = I + \left(c^2 + \sigma_u^2\right)^{-1} \left(\begin{bmatrix} c^2 & -c^2 \\ -\sigma_u^2 & \sigma_u^2 \end{bmatrix} L + \begin{bmatrix} c^2 & -c^2 \\ -\sigma_u^2 & \sigma_u^2 \end{bmatrix} L^2 + \ldots \right) \tag{18}$$

This series does not converge. However, there is a way around this. We can write the above equation in so-called error correction form, as a Vector Error Correction Model (VECM)[5]. These models were first used in modeling dynamic systems in disequilibrium; therefore the name error correction.

In a situation with *n* prices for the same security in different exchanges, we can write the error correction model in the following way

$$\Delta p_t = \phi_1 \Delta p_{t-1} + \phi_2 \Delta p_{t-2} + \ldots + \beta\left(z_{t-1} - b\right) + \varepsilon_t \tag{19}$$

The error terms in the above expression have the following form:

$$z_{t-1} = \begin{bmatrix} p_{1,t-1} - p_{2,t-1} \\ p_{1,t-1} - p_{3,t-1} \\ \vdots \\ p_{1,t-1} - p_{n-1,t-1} \end{bmatrix} = A' p_{t-1} \tag{20}$$

---

[5] This follows by the Granger Representation Theorem. If there exists a nonstationary autoregressive representation, then by the Theorem there exists an error correction representation. See e.g. Engle and Granger (1987) and Hamilton (1994), page 582 for a formal definition.

Where p_t is an *n x 1* vector of prices, z_{t-1} is a *(n-1) x 1* vector, with each component equal to the difference relative to the first price. From a statistical point of view, the ordering is arbitrary. In the above error correction model, the term *b* is a column vector of mean errors. The adjustment process can be thought of as having the mean error as its target. In the case of a cross-listed stock, the mean error would converge toward zero, supposing that the prices of the two stocks are largely the same.

The above Error Correction Model has a Stock and Watson (1988) common trends representation (see, e.g. Hasbrouck (1995), Johansen (1988, 1991)):

$$p_t = p_0 + C(1)\sum_{s=1}^{t}\varepsilon_s + C^*(L)\varepsilon_t \tag{21}$$

where p_0 is a constant vector of size *n*, and *C*(L)* is a matrix polynomial in the lag operator. The first term on the right hand side of Equation (21) is a vector of initial values. The second term captures the random walk component that is common to all *n* prices in the model. The last term on the right hand side is a zero-mean covariance stationary process.

Relying on a Cholesky decomposition of the covariance matrix Ω, Hasbrouck (1995) assigns information shares to the market places represented in the price vector:

$$S_j = \frac{c_j^2 \Omega_{jj}}{c\Omega c'} \tag{22}$$

where *c* is the common row vector of the impact matrix in the common trends representation of Equation (21). Applying different orderings will now yield lower and upper bounds for the information shares.

4 DATA AND EXCHANGES

4.1. High frequency data

The rise of high frequency datasets has made it possible for researchers in financial economics to explore market microstructure in new ways empirically. As is often the case, however, the use of high frequency data is not without its problems. Engle (2000) discusses the issues associated with studying what he calls *ultra high frequency data*. By this he means a full record of all data points, either all trades or all quotes in a certain security. There are a number of issues that need to be taken into account, from the random arrival of trades and the non-synchronous nature of trading to the practical problems associated with working with huge databases. In my case the use of ultra high frequency data has meant hundreds of gigabytes of data.

Trade data are likely to suffer from an autocorrelation problem, or bid-ask bounce, originally analyzed by Roll (1984). Quote midpoints are not entirely unaffected by this problem, but they are affected to a much lesser degree. In this sense, midquotes are more precise price estimates (see, e.g. Hasbrouck (2007), and Eun and Sabherwal (2003)). Quote data is therefore more conducive to my research. Since quotes are updated more often than trades occur, they may also reflect the pricing of a security more quickly. A trade price may become stale in that it no longer reflects the real market price, or all available information, whereas quotes may be updated at any time, even with no trading taking place. Since anybody posting a bid or ask quote runs the risk of being picked off by a more informed trader, there is a strong motivation to keep quotes up to date. I use logarithmic midquotes, which is to say that I take the average of the bid and ask prices, and then the natural logarithm of the thus obtained midquote.

Since market prices, whether quotes or trades, arrive at random intervals, there is a need to construct interval data. I do this in the following way. I use intervals from one minute all the way to thirty minutes. For each interval I wish to use the prevailing quote at the time. To accomplish this I compare the time stamps of all available quotes, and choose the most recent one. The start of the trading day is a special case. At the NYSE the specialist performs an initial matching of all available buy and sell orders before the beginning of continuous trading. Therefore, the start of the simultaneous trading period between European markets and the NYSE may be delayed by a few minutes. I exclude this time period from our analysis.

I use a number of sources for my high-frequency datasets. All equities data are obtained directly from the exchanges: TAQ for the New York market is from NYSE Data; I also use the equivalent databases for Euronext Paris, the Frankfurt Stock Exchange, and OMX Helsinki. These datasets (with the exception of the German data) include a complete set of trades and quotes for the exchanges in question. My foreign exchange data are aggregated interval observations provided by Olsen Associates.

4.2. The exchanges

The NYSE (New York Stock Exchange) is the oldest and largest stock exchange in the United States. It is a hybrid market, whose trading protocols are fairly complex. The trading systems include the exchange floor, as well as the increasingly important electronic trading system. The hybrid trading system has been under heavy development during the last ten years, and it has indeed seen many changes during these years. Mostly the changes have been towards more electronic trading, and the trading floor has diminished in importance. Many papers discuss the characteristics of the NYSE trading system, see e.g. Hasbrouck, Sofianos, and Sosebee (1993), Harris and Panchapagesan (2005), and Hasbrouck (2007). However, there is no up to date, comprehensive presentation of the trading system, except for the official rules and regulations of the exchange. These, on the other hand, may be complete and detailed, but not easy to interpret for an outsider (Hasbrouck (2007)).

All three European exchanges included in the study are quite similar. Most importantly for this study, all are electronic limit order markets. Euronext Paris is a subsidiary of the Euronext Group, formed in 2001 as a merger between the Paris, Brussels, Amsterdam, and Lisbon stock exchanges. It is the main marketplace for French stocks. The Frankfurt Stock Exchange is owned and operated by the Deutsche Börse Group. In addition to the Frankfurt exchange, there are a number of local stock exchanges in Germany. However, the Frankfurt Stock Exchange is by far the largest and most important one. The Helsinki Stock Exchange merged with the Stockholm Stock Exchange to form OM HEX. The group later merged with The Copenhagen, Tallinn, Riga, Vilnius, and Reykjavík Stock Exchanges. The group has been a part of NASDAQ OMX Group since February 2008.

Euronext trades liquid stocks continuously from 9.00 a.m. to 5.25 p.m. CET (Central European Time), with call auctions at the opening and at the closing. Within prices restricted by the order book, cross trades and block trades may be negotiated outside the system. The trading hours at the electronic Xetra system of Deutsche Börse are 9.00 a.m. to 5.30 p.m. CET. There are auctions at the opening and at the closing, but also two other auctions during the trading day. OMX Helsinki starts the trading day with a morning call auction at 9.45 a.m. local time (8.45 a.m. CET). The market opens for continuous trading at 10 a.m. local time (9 a.m. CET). Continuous trading ends and a closing auction starts at 6.20 p.m. local time (5.20 p.m. CET). The trading systems of the European exchanges are described on the exchanges' websites.[6]

Most European stocks cross-listed in the U.S. are traded as ADRs (American Depositary Receipts). ADRs are claims against home-market common shares issued by a U.S. depositary bank. They are quoted, traded and settled in U.S. dollars, and dividends are paid in dollars. It is possible to convert ADRs into common stocks and vice versa through the custodian bank for a small fee. For most practical purposes they are equivalent to common stocks. In particular, they are close enough substitutes for an arbitrage relationship to keep the prices of a company's common stock in the home market and the equivalent ADR very closely matched[7].

[6] Please see the following websites for additional information. OMX: http://www.nasdaqomxnordic.com/, Deutsche Börse: http://deutsche-boerse.com/ , NYSE-Euronext: http://www.euronext.com/ , and the NYSE: http://www.nyse.com/.

[7] See e.g. Gagnon and Karolyi (2004) for a detailed discussion of ADRs.

5 SUMMARY OF THE ESSAYS

5.1. Essay 1: Price discovery and cross-listed stocks

In my first essay, I study price discovery for European stocks cross-listed in the U.S. My sample includes 20 of the most liquid companies from the French and German markets. I study the time of simultaneous trading in the home exchange, Euronext Paris or Deutsche Börse Frankfurt, and the NYSE (New York Stock Exchange). The time of simultaneous trading spans the time from the opening of the New York market to the closing of the European markets, approximately two hours later.

My main results and contribution to the literature are as follows. First, the estimation results of an Error Correction Model indicate that the home exchange dominates price discovery. This means that both the home exchange and the cross-listed exchange react to price changes, but that the cross-listed exchange adjusts more to price changes on the home exchange than vice versa. Second, cross-sectional regression results indicate that the U.S. share of price discovery is directly related to the U.S. trading volume, inversely related to the relative bid-ask spread, and directly related to market capitalization. I also examine two ownership variables, the size of the holding of the largest shareholder, and the share of U.S. based investors. These variables are not a significant determinant of shares of price discovery. Third, I study the effects of converting prices from one currency to the other. I find that this conversion introduces a systematic bias to shares of price discovery. This finding lends empirical support to simulation findings of Grammig, Melvin, and Schlag (2005).

5.2. Essay 2: Transparency and informed trading

My second paper is an event study. I study changes made to the trading system of the OMX Helsinki Stock Exchange in March 2006. These changes made a previously fully pre-trade transparent limit order book less transparent. By pre-trade transparency I mean that the information about the identities of other traders posting bid and ask prices in the limit order book is available to all market participants.

My main contribution is using the Probability of Informed Trading (PIN) as a proxy for the participation rate of informed traders. I find that stocks with a high PIN exhibit a higher bid-ask spread after a change to anonymous trading. In practical terms, this means that the more informed traders there are in a stock, the more market quality suffers from a move to anonymous trading.

Contrary to previous studies about market transparency, I do not find unambiguous evidence of an improvement in market quality, measured by bid-ask spreads, intraday volatility, and trading volume. I also study the effects of the change for informed trading, using Probability of Informed Trading (PIN), which I estimate using a Maximum Likelihood method. The evidence here is mixed, with no significant change in the number of informed traders. Also, contrary to my expectations, there is no change in the proportion of upstairs trades.

5.3. Essay 3: Macroeconomic announcements and cross-listed stocks

In my third essay, I examine the reactions of European stocks cross-listed in the United States to scheduled U.S. macroeconomic announcements. My main result is that the U.S. share of price discovery is lower on announcement days than on days with no announcement. I also find that daily returns are higher on announcement days, both in absolute values and in logarithmic returns. This result applies to both stock prices and the EUR/USD exchange rate. As expected on the basis of a large literature on international stock market correlations, both cross-correlations within the sample of 20 stocks, and correlations between the stocks and two broad market indices are higher when macroeconomic data are released. The indices are STOXX 600, which includes large capitalization stocks in 18 European countries, both within and outside the euro area, and the S&P 500, which includes large capitalization common stocks in the United States.

In addition, I observe an increase in daily return volatility during days with macroeconomic announcements, as opposed to days with no announcements. My results also indicate that the U.S. share of price discovery is lower on announcement days. The most marked difference is in correlations between the markets, which are significantly higher during release days. These results contribute to existing literature on information asymmetry.

I also compare the changes in shares of price discovery, between days with and days without macroeconomic data releases. I find that the number of years that a company has a cross-listing, and the size of the largest shareholding are significant explanatory variables.

5.4. Essay 4: Intramarket price discovery

In my fourth and last essay, I use a novel method to study the unresolved question of whether local or foreign equity investors are more informed. Two examples of opposing views are Hau (2001) and Grinblatt and Keloharju (2000). Hau (2001) studies proprietary traders active on the Frankfurt Stock Exchange. He finds that non-German speaking traders located outside Frankfurt have lower trading profits than German speakers located in Frankfurt. The opposite result is obtained by Grinblatt and Keloharju (2000)), for the Finnish market. The authors find that foreign investors are mostly momentum traders, whereas domestic investors are mainly contrarian. They also find that foreigners outperform local investors.

I use an error correction model to study price discovery between local and foreign traders in a small limit order market, OMX Helsinki. I conclude that local brokers dominate the price discovery process for most stocks, even if the market share of the foreign members is on average approximately 70% of all trading volume. Also, I perform a cross-sectional analysis in order to explain differences in shares of price discovery. Market capitalization and a cross-listing abroad are directly related to a higher foreign share of price discovery. However, trading volume, measured by the turnover and the number of trades, is inversely related to the foreign share of price discovery. This is surprising, given that foreign investors are generally expected to be unwilling to trade illiquid assets.

6 CONCLUSIONS

I study several topics in market microstructure. Price discovery, determining the prevailing price of a security at any given time, is a fundamental function of a market. My first, third, and fourth essays offer new insights into price discovery, both in a multiple markets setting, and within one market. I use an Error Correction Model to estimate shares of price discovery.

Another main topic in this thesis is informed trading. The distinction between informed and uninformed trading is a starting point in most theoretical and empirical literature in market microstructure. I study informed trading using the so-called Probability of Informed Trading (PIN) measure. This is a maximum likelihood estimate of the number of informed traders in a market.

As I am writing this, in December 2008, we seem to be in the midst of what looks like the worst financial and economic crisis that the industrialized nations have seen, at least since the Great Depression. A solution to all these problems is, fortunately, not expected of me. However, I would like to point out that markets, both financial and otherwise, exist for a reason. Two very good motivations, one might even say raisons d'être, for a market are bringing buyers and sellers together and determining a price for whatever it is the people want to exchange. And thus, as long as there is a need for markets, there is also a need for *efficient* and *well functioning* markets. And, arguably, in times of trouble the need to ensure that unnecessary frictions are eliminated, is all the greater. Market microstructure seeks an answer to these questions, both by theoretical inquiry and empirical investigation. In other words, making sure that markets work as efficiently as possible, and in a fair manner, is in the public interest, irrespective of the direction that prices of stocks are taking in the market.

A final point I would like to make is that the rapid development of information technology in the last two or three decades has also been reflected in dramatic changes in financial markets. This also means that empirical and practical answers to many questions in market microstructure may change over time. Trading practices which have at some point been considered state-of-the-art are simply later discarded as arcane and inefficient. A case in point is the move from open outcry markets to electronic trading. A more recent development is the huge rise in computer-driven algorithmic trading. These questions will provide both researchers and regulators many interesting research topics in the years to come.

REFERENCES

Baillie, Richard T., G. Geoffrey Booth, Yiuman Tse, and Tatyana Zabotina, 2002, Price discovery and common factor models, *Journal of Financial Markets* 5, 309–322.

Biais, Bruno, 1993, Price formation and equilibrium liquidity in fragmented and centralized markets, *Journal of Finance* 48, 157–185.

Biais, Bruno, Lawrence R. Glosten, and Chester Spatt, 2005, Market microstructure: A survey of microfoundations, empirical results, and policy implications, *Journal of Financial Markets* 8, 217–264.

Boehmer, Ekkehart, Joachim Grammig, and Erik Theissen, 2007, Estimating the probability of informed trading – does trade misclassification matter?, Journal of Financial Markets 10, 26–47.

Chowdhry, Bhagwan, and Vikram Nanda, 1991, Multimarket trading and market liquidity, *Review of Financial Studies* 4, 483–511.

Copeland, Thomas E., and Dan Galai, 1983, Information Effects and the Bid-Ask Spreads, *Journal of Finance* 38, 1457–69.

Dacorogna, Michel M., Ramazan Gencay, Ulrich A. Muller, Richard B. Olsen, and Olivier B. Pictet, 2001, High-frequency Finance, Academic Press, New York, NY.

Demsetz, Harold, 1968, The Cost of Transacting, *The Quarterly Journal of Economics* 82, 33–53.

Ding, David K, Frederick H. deB. Harris, Sie Ting Lau, and Thomas H. McInish, 1999, An Investigation of Price Discovery in Informationally-Linked Markets: Equity Trading in Malaysia and Singapore, *Journal of Multinational Financial Management* 9, 317–329.

Duarte, Jefferson and Lance A. Young, Why is PIN Priced? (May 18, 2007). http://ssrn.com/abstract=971197, forthcoming in the Journal of Financial Economics.

Easley, David, Robert F. Engle, Maureen O'Hara, and Liuren Wu, 2008, Time-varying arrival rates of informed and uninformed trades, *Journal of Financial Econometrics* 6, 171–207.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2002, Is information risk a determinant of asset returns?, *Journal of Finance* 57, 2185–2221.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2005, Factoring information into returns, (July 27, 2005). *EFA 2004 Maastricht Meetings Paper No. 4118*, http://ssrn.com/abstract=556079

Easley, David, Nicholas M. Kiefer, Maureen O'Hara and Joseph B. Paperman, 1996. Liquidity, Information, and Infrequently Traded Stocks, *Journal of Finance* 51, 1405–1436.

Easley, David, Nicholas M. Kiefer, and Maureen O'Hara, 1996, Cream-Skimming or Profit-Sharing? The Curious Role of Purchased Order Flow, *Journal of Finance* 51, 811–833.

Easley, David, Nicholas M. Kiefer, and Maureen O'Hara, 1997, One Day in the Life of a very Common Stock, *Review of Financial Studies* 10, 805–835.

Easley, David, Maureen O'Hara, and Joseph Paperman, 1998, Financial analysts and information-based trade, *Journal of Financial Markets* 1, 175–201.

Easley, David, Maureen O'Hara, and Gideon Saar, 2001, How Stock Splits Affect Trading: A Microstructure Approach, *Journal of Financial and Quantitative Analysis* 36, 25–51.

Engle, Robert F., 2000, The Econometrics of Ultra-High-Frequency Data, Econometrica 68, 1–22.

Engle, Robert F., and Clive W. J. Granger, 1987, Co-integration and error correction: representation, estimation and testing, Econometrica 55, 251–276.

Errunza, Vihang R., and Darius P. Miller, 2000, Market segmentation and the cost of capital in international equity markets, *Journal of Financial and Quantitative Analysis* 35, 577–600.

Eun, Cheol S., and Sanjiv Sabherwal, 2003, Cross-Border Listings and Price Discovery: Evidence from U.S.-Listed Canadian Stocks, *Journal of Finance* 58, 549–575.

Fama, Eugene, 1965, The behavior of stock market prices, *Journal of Business* 38, 38–105.

Fama, Eugene, 1970, Efficient Capital Markets: A Review of Theory and Empirical Work, *Journal of Finance* 25, 383–417.

Fama, Eugene, and Kenneth French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

Foerster, Stephen R., and G. Andrew Karolyi, 1999, The Effects of Market Segmentation and Investor Recognition on Asset Prices: Evidence from Foreign Stocks Listing in the United States, *Journal of Finance* 54, 981–1013.

Foucault, Thierry, Sophie Moinas, and Erik Theissen, 2007, Does anonymity matter in electronic limit order markets?, *Review of Financial Studies* 20, 1707–1747.

Froot, K.A., and T. Ramadorai, 2008, Institutional Portfolio Flows and International Investments, Review of Financial Studies 21, 937–971.

de Frutos, M. Angeles, and Carolina Manzano, 2002, Risk aversion, transparency, and market performance, *Journal of Finance* 57, 959–984.

Gagnon, Louis Joseph and G. Andrew Karolyi, Multi-Market Trading and Arbitrage (June 17, 2004), *Working paper*, Dice Center No. 2004–9, http://ssrn.com/abstract=577004.

Garman, Mark, 1976, Market microstructure, *Journal of Financial Economics* 3, 257–275.

Glosten, Lawrence R., and Paul R. Milgrom, 1985, *Journal of Financial Economics* 14, 71–100.

Grammig, Joachim, Dirk Schiereck, Erik Theissen, 2001, Knowing me, knowing you: Trader anonymity and informed trading in parallel markets, *Journal of Financial Markets* 4, 385–412.

Grammig, Joachim, Christian Schlag, and Michael Melvin, The Role of U.S. Trading in Pricing Internationally Cross-Listed Stocks, EFA 2004 Maastricht Meetings Paper No. 3267, http://ssrn.com/abstract=567066.

Grammig, Joachim, Michael Melvin, and Christian Schlag, 2005, Internationally Cross-Listed Stock Prices during Overlapping Trading Hours: Price Discovery and Exchange Rate Effects, *Journal of Empirical Finance* 12, 139–164.

Greene, William H., 2007, Econometric Analysis (6th Ed.), Prentice Hall.

Grinblatt, Mark, and Matti Keloharju, 2000, The investment behavior and performance of various investor types: a study of Finland's unique data set, *Journal of Financial Economics* 55, 43–67.

Hamilton, James D., 1994, Time Series Analysis, Princeton University Press, Princeton, NJ.

Harris, Frederick H. deB., Thomas H. McInish, Gary L. Shoesmith, and Robert A. Wood, 1995, Cointegration, Error Correction, and Price Discovery on Informationally Linked Security Markets, *Journal of Financial and Quantitative Analysis* 30, 563–579.

Harris, Frederick H. deB., Thomas McInish, and Robert A. Wood, 2002a, Security price adjustment across exchanges: an investigation of common factor components for Dow stocks, *Journal of Financial Markets* 5, 277–308.

Harris, Frederick H. deB., Thomas McInish, and Robert A. Wood, 2002b, Common factor components versus information shares: a reply, *Journal of Financial Markets* 5, 341–348.

Harris, Lawrence E., and Venkatesh Panchapagesan, 2005, The information content of the limit order book: evidence from NYSE specialist trading decisions, *Journal of Financial Markets* 8, 25–67.

Harris, Lawrence, 2003, Trading and Exchanges, Oxford University Press, New York, NY.

Harris, Lawrence and Venkatesh Panchapagesan, 2004, The information content of the limit order book: evidence from NYSE specialist trading decisions, Journal of Financial Markets 8, 25–67.

Hasbrouck, Joel, George Sofianos, and Deborah Sosebee, 1993, New York Stock Exchange systems and trading procedures, NYSE Working Paper #93–01.

Hasbrouck, Joel, 1995, One security, many markets: determining the contributions to price discovery, *Journal of Finance* 50, 1175–1199.

Hasbrouck, Joel, 2002, Stalking the "efficient price" in market microstructure specifications: an overview, *Journal of Financial Markets* 5, 329–339.

Hasbrouck, Joel, 2007, Empirical Market Microstructure, Oxford University Press, New York, NY.

Hau, Harald, 2001, Location matters: An examination of trading profits, *Journal of Finance* 56, 1959–1983.

Hauser, Shmuel, Yael Tanchuma, and Uzi Yaari, 1998, International transfer of pricing information between dually listed stocks, *Journal of Financial Research* 21, 139–157.

Hayashi, Fumio, 2000, Econometrics, Princeton University Press.

Hedvall, Kaj, Eva Liljeblom, and Kim Nummelin, 1997, Dual listing and price discovery: The role of the domestic market, Working Paper 355, Swedish School of Economics and Business Administration.

Johansen, Soeren, 1988, Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* 12, 231–254.

Johansen, Soeren, 1991, Estimation and hypothesis testing of cointegrated vectors in Gaussian vector autoregressive models, *Econometrica* 59, 1551–1580.

Johansen, Soeren, 1995, Likelihood-based inference in cointegrated vector autoregressive models, Oxford University Press, Oxford.

de Jong, Frank, 2002, Measures of contribution to price discovery, a comparison, *Journal of Financial Markets*, 5, 323–327.

Karolyi, G. Andrew, 2006, The world of cross-listings and cross-listings of the world: Challenging conventional wisdom, *Review of Finance* 10, 99–152.

Kim, Minho, Andrew C. Szakmary, and Ike Mathur, 2000, Price transmission dynamics between ADRs and their underlying foreign securities, *Journal of Banking & Finance* 24, 1359–1382.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1336.

Lau, Sie Ting, and J. David Diltz, 1994, Stock returns and the transfer of information between the New York and Tokyo stock exchanges, *Journal of International Money and Finance* 13, 211–222.

Lehmann, Bruce N., 2002, Some desiderata for the measurement of price discovery across markets, *Journal of Financial Markets* 5, 259–276.

Lieberman, Offer, Uri Ben-Zion, and Shmuel Hauser, 1999, A characterization of the price behavior of international dual stocks: an error correction approach, *Journal of International Money and Finance*, 18, 289–304.

Lyons, Richard K., 2001, The Microstructure Approach to Foreign Exchange Rates, MIT Press, Cambridge, MA.

Madhavan, Ananth, 1995, Consolidation, fragmentation, and the disclosure of trading information, *Review of Financial Studies* 8, 579–603.

Madhavan, Ananth, 2000, Market microstructure: A survey, *Journal of Financial Markets* 3, 205–258.

Menkveld, Albert J., Siem Jan Koopman, and André Lucas, 2007, Modeling around-the-clock price discovery for cross-listed stocks using state space methods, *Journal of Business & Economic Statistics* 25, 213–225.

O'Hara, Maureen, 1995, Market Microstructure Theory, Blackwell, Cambridge, MA.

Pascual, Roberto, Bartolomé Pascual-Fuster, and Francisco Climent, 2006, Cross-listing, price discovery and the informativeness of the trading process, *Journal of Financial Markets* 9, 144–161.

Phylaktis, Kate, and Piotr Korczak, 2007, Specialist Trading and the Price Discovery Process of NYSE-Listed Non-US Stocks, *Working paper,* http://ssrn.com/abstract=567104.

Roll, Richard, 1984, A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Finance* 39, 1127–1139.

Sapp, Stephen G., 2002, Price leadership in the spot foreign exchange market, *Journal of Financial and Quantitative Analysis* 37, 425–448.

Shephard, Neil, 2005, Stochastic Volatility, Oxford University press, Oxford, UK.

Vives, Xavier, 2008, Information and Learning in Markets. The Impact of Market Microstructure, Princeton University press, Princeton, NJ.

Wang, Steven Shuye, Oliver Meng Rui, and Michael Firth, 2002, Return and volatility behavior of dually-traded stocks: the case of Hong Kong, *Journal of International Money and Finance* 21, 265–293.

Yan, Yuxing and Shaojun Zhang, 2006, An Improved Estimation Method and Empirical Properties of the Probability of Informed Trading, (July 27, 2006), http://ssrn.com/abstract=890486.

Yin, Xiangkang, 2005, A Comparison of Centralized and Fragmented Markets with Costly Search, *Journal of Finance* 60, 1567–1590.

Part 2

THE ESSAYS

Essay 1

PRICE DISCOVERY FOR EUROPEAN CROSS-LISTED STOCKS

Abstract

We study price discovery for liquid French and German stocks cross-listed in the U.S. Our findings are the following. First, the home exchange plays a dominant role in price discovery. Second, the home market share of price discovery is inversely related to U.S. trading volume, directly related to the relative bid-ask spread, and inversely related to market capitalization. Ownership, measured as the size of the largest shareholding and U.S. holdings, is not a significant determinant of price discovery. Third, the conversion of prices to a common currency introduces a systematic bias to shares of price discovery.

Keywords: cross-listing, price discovery, high-frequency data

JEL classification: G12, G14, G15.

1 INTRODUCTION

Fragmentation of trading to multiple exchanges and nonconventional trading platforms has emerged as a dominant trend (see, e.g. Pagano, Roell, and Zechner (2002)). As of late 2007, the number of non-U.S. companies listed at the New York Stock Exchange was 420. The issue of price discovery then arises in such a multiple markets setting. Price discovery has been described as "the search for an equilibrium price" (Schreiber and Schwartz (1986)), and as "the efficient and timely incorporation of the information implicit in investor trading into market prices" (Lehmann (2002)). It is a key function of a stock exchange.

Possible explanations for the increase in cross-listings world-wide include a reduced cost of capital, and improving the quality of the market in the company's stock, including liquidity and the cost of trading[8]. Alexander, Eun and Janakiramanan (1987) find that the cost of capital is lower for cross-listed companies. This is because the need to compensate for the risk of investing across borders disappears. Another explanation is increased investor recognition, as suggested by Foerster and Karolyi (1999). On the other hand, Smith and Sofianos (1997), and Foerster and Karolyi (1998) find that after a stock is cross-listed, it generally becomes substantially more liquid in the home market. Domowitz, Glen, and Madhavan (1998) suggest that a degree of pre-trade transparency and the absence of foreign ownership restrictions are essential for the beneficial effects of a cross-listing on the cost of trading to arise.

Early studies on price discovery in the context of multiple listings used low frequency daily data. Lau and Diltz (1994) study opening and closing prices of Japanese companies cross-listed in New York. They find that both markets have an impact on the pricing of these stocks. Hauser, Tanchuma, and Yaari (1998), and Lieberman, Ben-Zion, and Hauser (1999) study Israeli stocks listed in New York, and find that in most cases pricing is efficient in the sense that there are no arbitrage profits. They also find that the home market dominates price discovery. Kim, Szakmary, and Mathur (2000) consider the exchange rate as well as the U.S. market index for a group of stocks from five different markets. They find that the U.S. market, as well as the exchange rate, has an influence on the pricing of these cross-listed stocks. Wang, Rui, and Firth (2002) study pricing and volatility between London and Hong Kong. They find that both markets influence asset returns as well as volatility.

Hedvall, Liljeblom, and Nummelin (2000) perform one of the first studies using high frequency data. They study the most liquid stock in the Finnish market, Nokia, and find that the New York market is dominant for price discovery. Eun and Sabherwal (2003) study price discovery for Canadian stocks cross-listed in New York. They find huge differences in shares of price discovery. The authors define shares of price discovery as the proportion of the adjustment that occurs at one of the marketplaces to the total adjustment of both marketplaces. Grammig, Melvin, and Schlag (2004) perform a comprehensive study of price discovery for stocks from Canada, Germany, France, and the UK, cross-listed at the NYSE. Using data for 1999, they find that the home exchange typically dominates, but that there are great differences between companies. Additionally, Pascual, Pascual-Fuster, and Climent (2006) and Phylaktis and Korczak (2007) study Spanish and U.K. stocks cross-listed in the U.S., respectively.

8 For a comprehensive survey of international cross-listings, see Karolyi (2006).

Grammig, Melvin, and Schlag (2005) focus on the role of the exchange rate as a third variable in their analysis of price discovery for three German firms traded at the NYSE and the Frankfurt Stock Exchange. They also conclude that most price discovery is influenced by the home market. They include the USD-EUR exchange rate in their analysis, and pose two additional questions. First, they examine how stock prices in the two markets adjust to shocks in the EUR-USD exchange rate. Second, they explore the effects of converting the price in one market into a price in the other currency using the prevailing exchange rate. Using a simulation study, they find that the conversion of prices introduces a significant bias, in that the share of the market whose price is converted into the other currency is overstated.

Menkveld (2008) conducts an empirical test of the Chowdhry and Nanda (1991) model of multimarket trading. Using a sample of stocks from two European markets, the UK and the Netherlands, cross-listed in the U.S, he finds evidence of order splitting during the simultaneous trading hours. Traders who are able to split their orders across markets gain at the benefit of traders bound to their local market. He also finds evidence that the traders who are able to split their orders are informed traders, since the common component in order imbalance has some long-term price impact.

It is not immediately clear what to expect when studying price discovery for an international stock cross-listed in the United States. On the one hand, the home exchange of a company is likely to dominate because it is situated where the company operates, and a lot of information is generated there. On the other hand, the U.S. equity market is the largest and most liquid in the world, which may lead one to expect it to contribute to price discovery. An additional variable is the exchange rate in the case of international cross-listings. In the case of European stocks cross-listed in the United States, such as in this paper, the relevant exchange rate is of course the EUR/USD exchange rate. Assuming that the market is efficient enough that any deviation from the law of one price is only temporary, the question then becomes how share prices adjust to external shocks in the exchange rate.

The first objective of this paper is to examine the relative importance of the two exchanges in a cross-listed setting. Specifically, we look at the relative shares of price discovery for a sample of ten French and ten German stocks, cross-listed at the NYSE. We find that the contribution of the cross-listed market to price discovery during simultaneous trading hours is significant, but smaller than the role of the home market for all stocks. The shares of price discovery for NYSE, the cross-listed market, range from 4% to 41%. This means that the cross-listed market (U.S. market) does most of the adjusting to price innovation in the home market (Paris or Frankfurt), rather than vice versa. This finding differs from Eun and Sabherwal (2003), and Grammig, Melvin, and Schlag (2005), who both find that for some stocks the majority of price discovery occurs in the U.S. market.

Another objective of this paper is to study a broader set of factors affecting the relative shares of price discovery. In a domestic U.S. setting, Hasbrouck (1995) finds that trading volume in medium sized trades is a significant factor in explaining differences in shares of price discovery between the NYSE and regional exchanges. Harris, McInish, and Wood (2002) find that the tighter the bid-ask spreads at the NYSE, the bigger the contribution of the NYSE to price discovery. In an international context, Eun and Sabherwal (2003) find that for Canadian stocks cross-listed in the United States, the share of price discovery of the U.S. market is related to the share of trading volume, the relative bid-ask spread, and the amount of mid-sized trades. Grammig, Melvin, and

Schlag (2005) find that the difference in bid-ask spreads and the ratio of turnover of the two markets determine most of the differences in price discovery.

In addition to the variables mentioned above, we use logarithmic market capitalization, the number of years cross-listed as an ADR in the U.S. market, and the non-domestic share of a company's revenue. We also employ two firm-specific ownership variables, the U.S. ownership share, and the holdings of the largest shareholder as a percentage of all shares. We conclude from a cross-sectional analysis that the U.S. share of price discovery is directly related to the U.S. trading volume, inversely related to the relative bid-ask spread, and directly related to market capitalization. This finding is in line with Eun and Sabherwal (2003), and Grammig, Melvin, and Schlag (2005); however, market capitalization is an added significant variable in our study. The ownership variables do not add any explanatory power to the analysis.

The third objective of this paper is to study the significance of the exchange rate in a setting with cross-listed stocks traded in two different currencies. We lend empirical support to the simulation based findings of Grammig, Melvin, and Schlag (2005), in that conversion of prices from one currency to another introduces a systematic bias in the shares of price discovery, exaggerating the share of the market, whose prices are converted to the other currency. In other words, if prices in the European home market, where trading takes place in euros, are converted to U.S. dollars, this makes the U.S. share of price discovery lower, when all dollar prices are converted into euros.

Our sample of stocks includes the ten most liquid cross-listed stocks for both the German and French markets. The sample provides an ideal opportunity to study price discovery for two markets that are geographically separated, but still share a period of simultaneous trading. Also, it gives us an opportunity to study the differences between two major European markets. The shares are listed in the U.S. as ADRs (American Depositary Receipts), which makes them close, but not perfect substitutes for the ordinary shares traded in the home market (see Gagnon and Karolyi (2004) for a detailed discussion of ADRs). ADRs trade mostly within fairly narrow arbitrage bounds, dictated by the home market price, the exchange rate, and a conversion cost of 5 U.S. cents a share.

The rest of this paper is organized as follows. Chapter 2 discusses the main features of the markets included in this study. Chapter 3 describes the data used in estimation. In Chapter 4 we perform the preliminary analyses: unit root and cointegration tests. Chapter 5 includes the actual price discovery analysis, and Chapter 6 the cross-sectional analysis on the basis of results obtained in the previous chapter. Chapter 7 offers some concluding remarks.

2 THE EXCHANGES

All European stocks in our sample are traded as ADRs (American Depositary Receipts) on the NYSE (the New York Stock Exchange). These are claims against home-market common shares issued by a U.S. depositary bank. They are quoted, traded and settled in U.S. dollars, and dividends are paid in dollars. They are directly exchangeable with common stocks; issuances and cancellations are possible every day through the custodian bank; this costs approximately 5 U.S. cents a share.

The trading protocols and practices of NYSE are fairly complex, since it is essentially a hybrid market[9]. The NYSE market opens for trading at 9.30 a.m. local time with an opening auction. The market closes at 4 p.m. Euronext trades liquid stocks continuously from 9.00 am to 5.25 pm, with call auctions at the opening and at the closing of trading. The market is fully transparent, with the exception of the hidden part of iceberg orders[10]. Within prices restricted by the order book, cross trades and block trades may be negotiated outside the system. The trading hours at the electronic Xetra system of Deutsche Börse are 9.00 am to 5.30 pm. Similarly to the Paris Exchange, there are auctions at the opening and at the closing, but the Frankfurt exchange also has two intra-day auctions during the trading day. The trading hours of the exchanges are such that from the opening of the New York market until the closing of the Paris market there is one hour and 55 minutes of simultaneous trading hours. In this paper we study these hours of simultaneous trading.

The minimum tick size differs slightly between the three exchanges. NYSE and Frankfurt have a uniform minimum tick size of 0.01 dollars and euros, respectively. Euronext, however, applies a tick size of 0.01 up to a price of 50 euros, after which the tick size rises to 0.05, and to 0.10 after 100, and finally to 0.50 for prices over 500.

Both Euronext and Deutsche Börse also have liquidity providers, with slightly different specifications. But these differences do not affect our study, since they are only used for smaller stocks, not the liquid ones included in our study.

[9] One particular aspect of the NYSE is relevant for this study. The 30 second rule that applies to treatment of limit orders by the specialist is important for this research, since it directly influences the ease of implementing arbitrage strategies between the NYSE and another market. The significance of this rule in the present context is that an arbitrageur, trying to exploit a possible price discrepancy between the two markets, must be prepared to wait up to 30 seconds for the specialist to update the order book, match orders, and facilitate trades. In contrast, the fully electronic limit order book trading systems of both the Paris and Frankfurt exchanges provide near-instantaneous trades and fills. This relative slowness of the NYSE trading system makes the work of an arbitrageur riskier and more complicated.

[10] An Iceberg order (also known as a hidden-size order) allows the trader to show other participants only part of the total quantity of the order entered.

3 DATA

We use high-frequency trade and quote data for the U.S., French, and German markets. Our U.S. dataset is from the TAQ (Trade and Quote) database. This is a complete intraday data set which contains all quotes and trades for all U.S. stock exchanges in 2004. French market data are provided by Euronext Information Services. This dataset also consist of a complete set of trades and quotes for the markets in question. Data for the German market was provided by the Karlsruher Kapitalmarktdatenbank (KKMDB) at the Universität Karlsruhe. This dataset consists of trade data for the selected stock sample in our study. The original source of this dataset is Deutsche Börse AG.

Foreign exchange data for the USD - EUR exchange rate are provided by Olsen Associates. This is a data set of indicative bid and asks quotes with 1-minute intervals. According to Olsen Data, their data collection software interfaces with several data feeds, such as Reuters, Tenfore, and Bloomberg. See Müller (2001) for a detailed description of the data collection and filtering process.

The data for the cross-sectional analysis are from several sources. Data for market capitalization and trading volume are from the Amadeus database; the number of years listed as ADRs is from the New York Stock Exchange. The ownership data, largest shareholders and the share of U.S. holdings in a stock, are collected from the companies' annual reports and 20F forms, as well as the Thomson ONE Banker database.

In accordance with the general practice in the literature, we use midquotes, rather than transaction prices[11]. The reason for doing so is the following. The use of trade prices is generally known to suffer from autocorrelation bias. This was first analyzed by Roll (1984), when he introduced the concept of the bid-ask bounce. Also, trade prices may become stale. Quotes, however, can be updated even in the absence of trading; the number of quotes during a trading day is usually several times greater than the number of trades. Quote data do not suffer from autocorrelation bias to the same degree as trade data.

For the above reasons, we construct interval data with regular intervals in the following way. With regular intervals (1, 2, and 5 minutes) and calculate the midpoint of these prices. The results presented in this paper are calculated using the highest frequency, i.e. one-minute intervals. We use the 2- and 5-minute intervals as a robustness check. We then use this price series in our analyses. Following usual practice, we use logarithmic midquotes, i.e. logarithms of the average of the bid and offer prices.

The choice of the one-minute intervals at which we sample the price process was made with many considerations in mind. First, with an extremely high frequency, microstructure noise effects, such as nonsynchronous reporting of trades, may influence the results. Also, few of the stocks in our sample are liquid enough to trade several times a minute on average. On the other hand, with an extremely wide interval it is difficult to capture the actual price discovery process.[12]

[11] We have only trade data for the German market. Therefore the analysis is based on midquotes for the French and U.S. markets, and on trade data for the German market.

[12] There is considerable variation in choosing an appropriate interval length when studying price discovery in previous literature. Hasbrouck (1995) uses extremely short, 1-second intervals when studying price discovery between different U.S. exchanges. Grammig, Melvin and Schlag (2005) use a ten-second interval

The Appendix to this paper presents a test of the liquidity of the stock sample. We calculate the percentage of intervals with a zero return, meaning that there is no change in the midquote from the previous interval. The average share of zero-return intervals is approximately 40% for the one-minute intervals, and around 14% – 18% for the five-minute intervals. We do not feel that our sample is lacking in liquidity, however. The results presented in this paper are fairly robust to changes in the estimation interval, i.e. there are no great differences in results between using the one-minute, two-minute and five-minute interval data.

The selection criteria for the stock sample are that the stocks be cross-listed, and liquid. Thus, out of a sample of all French and German companies listed at the NYSE, we pick the stocks with the greatest turnover in the sample year (2004). Our sample of stocks and some descriptive statistics are listed in Table 1.

The descriptive statistics show that, with the exception of the French Thomson and the German Infineon, all companies in our sample are very large, with a market capitalization greater than 15 Billion USD as of the end of 2004. There are ten stocks from each exchange, Euronext Paris and Deutsche Börse Frankfurt. The U.S. ADRs mostly trade at a 1:1 ratio to the common stocks traded in the home market. However, for some stocks, such as Allianz, the ratio can be as high as 10, which means that 10 ADRs are equivalent to one common stock. Bid-ask spreads are generally lower in the home market, with an average percentage spread of 0.07% in the home market, compared with 0.13% in the cross-listed market. Total and Sanofi-Aventis are the two exceptions, with a smaller average U.S. bid-ask spread. The home market is also dominant as far as average trading volume is concerned. The situation is more varied, however: there are seven stocks with a higher average U.S. trading volume, compared to the home market. We measure trading volume as the number of shares traded, corrected, as needed, for the ADR ratio.

We filter our price series for errors, requiring among other things that the price be a multiple of the tick size for the exchange in question. We also exclude quotes flagged as odd lots, pre-opening, opening, closing, or halted.

---

in their study on German companies. Huang (2002) studies price discovery among the various Nasdaq quote participants using one-minute intervals. Eun and Sabherwal (2003) use 10-minute intervals in their study on Canadian stocks cross-listed in the U.S.

Table 1 Summary statistics of sample companies

Industry classification is the Industry Classification Benchmark (ICB) by FTSE and Dow Jones. *Market cap* is in millions of USD, as of 31st December, 2004, as reported by Reuters. *Exchange* is either Euronext Paris, or Deutsche Börse Frankfurt. *Last price* is the last traded price of the year 2004. *ADR Ratio* gives the number of ADRs that are equivalent to one common share. *U.S. spread* and *Home spread* are average bid-ask spreads. *Trading volume U.S.*, and *Trading volume, home* are the average daily number of shares traded in the U.S. and home markets, respectively, in 000s of shares. *# trades, U.S.*, and *#trades, home*, are the average daily number of trades in the U.S. and home markets, respectively.

Company name	Industry classification	Market cap, USD	Exchange	Last price	ADR Ratio	U.S. spread	Home spread	Volume, U.S.	Volume, home	#trades, U.S.	#trades, home
Alcatel	Telecommunications Equipment	21,020	Paris	11.41	1	0.12%	0.10%	1686	1302	823.0	6268.1
Allianz	Life Insurance	48,295	Frankfurt	97.13	10	0.21%	0.05%	122	32	226.5	4409.4
AXA	Full Line Insurance	45,669	Paris	18.17	1	0.12%	0.07%	295	828	580.2	5378.8
BASF	Specialty Chemicals	39,518	Frankfurt	52.81	1	0.11%	0.05%	36	25	303.6	4128.4
Bayer	Specialty Chemicals	24,817	Frankfurt	24.9	1	0.13%	0.08%	67	44	205.0	2779.5
Daimler Chrysler	Automobiles	48,665	Frankfurt	35.31	1	0.07%	0.06%	274	65	942.0	5450.5
Danone	Food Products	23,112	Paris	67.7	5	0.24%	0.09%	35	121	179.3	3416.9
Deutsche Bank	Banks	40,189	Frankfurt	65.29	1	0.09%	0.05%	51	412	300.3	4501.2
Deutsche Telekom	Fixed Line Telecommunications	95,143	Frankfurt	16.58	1	0.10%	0.08%	298	175	641.5	10656.0
E.ON	Multiutilities	59,840	Frankfurt	66.89	1	0.11%	0.05%	22	27	219.8	6702.1
France Telecom	Fixed Line Telecommunications	80,849	Paris	24.48	1	0.12%	0.06%	133	917	252.3	10377.6
Infineon	Semiconductors	7,645	Frankfurt	7.94	1	0.14%	0.12%	384	1684	542.4	856.2
Lafarge	Building Materials & Fixtures	16,253	Paris	71.25	4	0.22%	0.09%	36	89	113.9	3117.7
Sanofi-Aventis	Pharmaceuticals	69,022	Paris	59.05	2	0.08%	0.09%	535	424	1114.1	4970.9
SAP	Software	54,933	Frankfurt	131.83	4	0.06%	0.05%	822	960	1240.2	6432.3
Siemens	Electronic Equipment	65,671	Frankfurt	62.16	1	0.08%	0.04%	111	603	590.2	7355.2
Suez	Multiutilities	26,619	Paris	19.43	1	0.25%	0.08%	28	485	83.6	4843.7
Thomson	Broadcasting & Entertainment	7,218	Paris	19.29	1	0.22%	0.08%	47	216	89.3	3195.1
Total	Integrated Oil & Gas	133,610	Paris	161.7	2	0.04%	0.06%	629	1159	1508.2	6806.8
Vivendi	Broadcasting & Entertainment	34,389	Paris	23.54	1	0.10%	0.06%	331	651	433.3	5230.0
Mean		47,124		52	2	0.129 %	0.071 %	297.1	511.0	519.4	5,343.8
Median		42,929		44	1	0.114 %	0.069 %	127.5	418.0	368.4	5,100.4
Std. deviation		31223.6		41.2	2.2	0.063%	0.021 %	397.5	491.1	413.5	2,374.2
Min		7,218		8	1	0.038%	0.044 %	22.0	25.0	83.6	856.2
Max		133,610		162	10	0.247%	0.120 %	1,686.0	1,684.0	1,508.2	10,656.0
Skewness		1.166		1.318	2.757	0.762	0.693	2.5	0.9	1.0	0.7
Kurtosis		1.786		1.565	8.367	-0.631	-0.009	7.6	0.0	0.3	1.0

4 UNIT ROOT AND COINTEGRATION TESTS

A necessary condition for our later price discovery analysis is that the price series be cointegrated (see e.g. Eun and Sabherwal (2003)). We therefore first test whether the home market price and the cross-listed price are indeed cointegrated. A necessary condition for two price series to be cointegrated is that they individually contain a unit root.

We begin by determining the appropriate lag length for each stock, using the Schwarz Information Criterion. We also employ the rule of thumb of a maximum lag length of $T^{1/3}$ (where T is the number of observations), resulting in a maximum lag length of 20. The optimum lags are 1 or 2 for most stocks, with a maximum value of 3. We do not report these results.

4.1. Unit root tests

We perform unit root tests for all equity data series, using the augmented Dickey-Fuller (1981) test, following the standard procedure in the literature. This test checks for the presence of a unit root by including lagged first differences of the price series. We consider the following three standard formulations of the model: a random walk, a random walk with a drift term, and a random walk with a drift term and a time trend.

$$\Delta x_t = \delta x_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta x_{i-1} + e_t \tag{1}$$

$$\Delta x_t = c_0 + \delta x_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta x_{i-1} + e_t \tag{2}$$

$$\Delta x_t = c_0 + \delta x_{t-1} + c_1 t + \sum_{i=1}^{p} \varphi_i \Delta x_{i-1} + e_t \tag{3}$$

In each case, the null hypothesis is that δ equals zero, and that the price series thus contains a unit root. This means that the price series is nonstationary, but that the changes are stationary.

Since the augmented Dickey-Fuller tests have a low power to reject the null of *I(1)*, we complement the tests with Phillips-Perron tests. Using a significance level of 5%, we accept the null hypothesis of *I(1)* for all stocks. These results are available on request.

4.2. Cointegration tests

Simple arbitrage bounds for a cross-listed stock imply that the quotes in the home and cross-listed markets cannot be expected to diverge significantly from each other. In other words, we can expect the two price series to be cointegrated. If the series are indeed cointegrated, then the following is also true: series are *I(1)*, i.e. they contain a unit root, and there exists a so-called cointegration vector $\beta = \left(\beta^{EU}, \beta^{US}\right)$, such that $\beta^{EU} P_t^{EU} + \beta^{US} P_t^{US}$ is *I(0)*.

To test for cointegration of the price series we use the Johansen (1988) method and the number of cointegration vectors.

The Johansen method is as follows. We rewrite a *p*th order autoregressive process,

$$x_t = A_1 x_{t-1} + A_2 x_{t-2} + \ldots + A_p x_{t-p} + \varepsilon_t \tag{4}$$

as

$$\Delta x_t = \sum_{i=1}^{p-1} \Pi_i \Delta x_{t-i} + \Pi x_{t-p} + \varepsilon_t \,, \tag{5}$$

where Δ is the first-difference lag operator, x_t is a vector of *I(1)* time series, ε_t is a zero-mean n-dimensional white noise vector, Π_i are *(n x n)* matrices of parameters, and Π is a *(n x n)* matrix of parameter vectors.

The tested hypothesis is that the number of cointegrating vectors is equal to or less than *r*. It is tested using one of two statistics: $\lambda_{trace}(r)$ or $\lambda_{max}(r,r+1)$. These statistics are given by

$$\lambda_{trace}(r) = -T \sum_{i=t+1}^{n} \ln(1-\hat{\lambda}_t) \tag{6}$$

and

$$\lambda_{\max}(r, r+1) = -T \ln(1-\hat{\lambda}_{r+1}) \,, \tag{7}$$

where T is the number of observations, and $\hat{\lambda}$ is the estimated value of the characteristic root, obtained from the estimated *Π* matrix.

The two statistics test different hypotheses. The null hypothesis for $\lambda_{trace}(r)$ is that the number of distinct cointegrating vectors is less than or equal to *r*. The statistic $\lambda_{max}(r,r+1)$ tests for the null hypothesis of *r* cointegrating vectors against an alternative of *r+1* vectors. Johansen and Juselius (1990) provide critical values for both statistics, which do not follow any standard distributions.

We reject the null hypothesis of no cointegrating vector for all stocks in the sample. These results are available on request.[13]

[13] We also employ a direct test for the stationarity of the logarithmic price difference between the prices in the two markets. These results lend support to the above conclusion, in that the price differential is indeed stationary, and very close to zero at all times.
This is indeed the expected results, since all stocks are very liquid, and there is no reason for the two prices to deviate. An analysis of deviations from the law of one price indicates that any deviation above the arbitrage bounds imposed by the conversion cost of USD 0.05 per share is very short-lived. These results are available on request.

5 PRICE DISCOVERY DURING HOURS OF SIMULTANEOUS TRADING

5.1. The “Permanent-temporary” and “Information shares” models

There are two main methodological approaches to measuring price discovery for cross-listed stocks: the error correction approach, also known as permanent-temporary decomposition (PT), and the so-called information shares (IS) approach. The Permanent-Temporary decomposition method was first used by Gonzalo and Granger (1995), and Harris, McInish, Shoesmith, and Wood (1995); the Information Shares (IS) method was developed by Hasbrouck (1995). The models are directly related and provide similar results in most cases, see e.g. de Jong (2002), Baillie, Booth, Tse, and Zabotina (2002).

Fundamentally, both methods decompose the impact of a price innovation into permanent and temporary components. The Hasbrouck method employs Cholesky factorization of the covariance matrix of price innovations in the two exchanges, which requires that the prices be ordered. Unfortunately, this also means that the information shares obtained by this method are not unique. Rather, the methodology results in upper and lower bounds for the information shares, as they depend on the ordering of prices (see e.g. Hasbrouck (1995), Eun and Sabherwal (2003), and Lehmann (2002)). The difference between the upper and lower bounds is small when using extremely high frequency data, such as the one-second interval data in Hasbrouck (1995). On the other hand, Booth, Lin, Martikainen, and Tse (2002) and Huang (2002) find that using intervals in the range of a few minutes, the difference between the upper and lower bounds obtained for the information shares can be significant. Since not all stocks in our sample have a high trading frequency in the U.S. market, the Error Correction Methodology of Harris et al. (1995) is more conducive to our study.

Both methods are widely used. Eun and Sabherwal (2003), and Phylaktis and Korczak (2007) use the Harris (1995) error correction methodology; Hasbrouck (2002), Grammig, Melvin, and Schlag (2005), and Pascual, Pascual-Fuster, and Climent (2006) use the Hasbrouck (1995) information shares method.

5.2. The Error Correction Model

Cointegration between the European and U.S. prices of a stock implies the time series of both prices are influenced by any deviation from equilibrium. If there is a price innovation in one of the markets, one or both prices have to respond to the magnitude of this movement.

We estimate the following equations:

$$\Delta P_t^{EU} = \alpha_0^{EU} + \alpha^{EU}\left(P_{t-1}^{US} - P_{t-1}^{EU}\right) + \sum_{i=1}^{p} \gamma_i \Delta P_{t-i}^{EU} + \sum_{i=1}^{p} \delta_i \Delta P_{t-i}^{US} + \varepsilon_t^{EU} \qquad (8)$$

and

$$\Delta P_t^{US} = \alpha_0^{US} + \alpha^{US}\left(P_{t-1}^{US} - P_{t-1}^{EU}\right) + \sum_{j=1}^{p}\gamma_j \Delta P_{t-j}^{EU} + \sum_{j=1}^{p}\delta_j \Delta P_{t-j}^{US} + \varepsilon_t^{US}, \qquad (9)$$

where P^{EU} and P^{US} are the prices of the stock in the European market and the U.S. market, respectively; $\gamma_{i,j}$ are error correction parameters which reflect idiosyncratic adjustment of P_t to disparities in shocks to private value that cause the cointegrated series to diverge; p is the lag length determined earlier using the Schwarz Information Criterion; and ε_t is a zero-mean, covariance-stationary random disturbance, which is assumed to be identically distributed but may be autocorrelated[14].

The coefficients of main interest to us are α^{EU} and α^{US}. We have chosen to define the difference between the two prices in the error correction model as the U.S. price less the European price. We therefore expect α^{EU} to be positive, and α^{US} to be negative. The reason for this is the following. Consider a situation where the prices need to adjust, because the cross-listed price, P^{US}, is higher than the home market price, P^{EU}. In order to restore equilibrium, the home market price needs to rise, and/or the cross-listed price needs to decline. The adjustment coefficient α^{US} should therefore negative and α^{EU} should be positive. A similar result obtains in the opposite case.

Table 2 shows our results. First, all coefficients exhibit the expected sign, positive for α^{EU}, and negative for α^{US}. Second, the coefficients for U.S. price adjustment are greater in absolute value than the European coefficients. The interpretation of these coefficients is that they reflect the amount of adjustment that the price series in question makes to deviations from the law of one price. The U.S. adjustment coefficient is greater for all stocks in our sample. These results indicate that the response of the U.S. market to European price innovations is much greater than the European adjustment to price innovations in the U.S. market.

The reported results are based on one-minute interval data, the highest available to us. We also perform the same calculation using 2-minute and 5-minute intervals. These results are largely similar to the results presented here. These additional results are available on request. In judging the validity of these results, one of the main questions is liquidity. We therefore analyze the frequency of updates to the bid-ask quote data, which forms the basis of our analysis.

[14] We also use a trivariate specification, with a third error correction equation, in addition to Equations (8) and (9):

$$\Delta P_t^{FX} = \alpha_0^{FX} + \alpha^{FX}\left(P_{t-1}^{US} - P_{t-1}^{EU}\right) + \sum_{j=1}^{p}\gamma_j \Delta P_{t-j}^{EU} + \sum_{j=1}^{p}\delta_j \Delta P_{t-j}^{US} + \varepsilon_t^{FX},$$

where P^{FX} is the USD/EUR exchange rate; other parameters are as per the above discussion.

Table 2 Error Correction Model (ECM) estimation results

The coefficients listed in this table are α^{EU} and α^{US} from the error correction model. The interpretation for α^{EU} is the average adjustment of the home market price towards the cross-listed (US) price; α^{US} is the average adjustment of the U.S. market to the home market price. The numbers in parenthesis are t-values. * and ** denote significance levels of 5% and 1%, respectively.

French stocks			German stocks		
Company name	α^{EU}	α^{US}	Company name	α^{EU}	α^{US}
Alcatel	0.046	-0.351**	Allianz	0.105	-0.190*
	(0.531)	(-4.116)		-1.525	(-3.038)
AXA	0.027	-0.291**	BASF	0.046	-0.191
	(0.416)	(-4.013)		-0.681	(-2.860)
Danone	0.104	-0.164*	Bayer	0.062	-0.201
	(1.662)	(-2.804)		-0.781	(-2.578)
France Telecom	0.032	-0.234*	DaimlerChrysler	0.042	-0.209
	(0.460)	(-3.142)		-0.5	(-2.400)
Lafarge	0.097	-0.207*	Deutsche Bank	0.01	-0.242
	(1.371)	(-2.968)		-0.148	(-2.796)
Sanofi-Aventis	0.041	-0.234*	Deutsche Telekom	0.05	-0.203*
	(0.698)	(-3.468)		-0.808	(-3.047)
Suez	0.077	-0.213**	E.ON	0.041	-0.164
	(1.078)	(-3.352)		-0.631	(-2.233)
Thomson	0.128	-0.186	Infineon	0.132	-0.273*
	(1.622)	(-2.676)		-1.761	(-3.602)
Total	0.030	-0.169	SAP	0.041	-0.225
	(0.589)	(-2.431)		-0.609	(-2.779)
Vivendi	0.029	-0.289*	Siemens	0.035	-0.179
	(0.404)	(-3.597)		-0.489	(-2.008)

6 CROSS-SECTIONAL ANALYSIS

The share of price discovery varies considerably across firms. It is therefore a natural question to ask, whether these differences can be explained. For this purpose, we construct the variable USPD, which reflects the differences in the relative shares of price discovery between the home and cross-listed markets. This variable is the share of price discovery attributed to the cross-listed market by the Error Correction analysis described above.

6.1. The dependent variable

We estimate the coefficients α^{EU} and α^{US} above in the error correction model. These coefficients give us the average adjustment of each price series in the case of a deviation from the law of one price. We are now in a position to study the amount of adjustment occurring in one exchange as a proportion of total adjustment.

We measure the share of total adjustment that occurs in the cross-listed market, NYSE, as a share of the total adjustment, i.e. the sum the two adjustment coefficients. This measure was first proposed by Schwarz and Szakmary (1994); it has subsequently been used by e.g. Theissen (2002), and Eun and Sabherwal (2003). In other words, we simply take the European adjustment coefficient as a share of the summed adjustment coefficients:

$$USPD = \frac{|\alpha^{EU}|}{|\alpha^{EU}| + |\alpha^{US}|}. \tag{10}$$

The rest of the adjustment is then attributed to the home market:

$$HomePD = \frac{|\alpha^{US}|}{|\alpha^{EU}| + |\alpha^{US}|} \tag{11}$$

It is evident that *USPD = 1-HomePD*. A high value of the variable *USPD* means that the coefficient α^{EU} is high. The interpretation of this is that the European home market does most of the adjusting to price innovations. In the hypothetical case of no feedback from the U.S. to the European market, α^{EU} is zero (the home market does not adjust to changes in the U.S. price at all), and thus also USPD is zero.

Table 3 shows the values of USPD. As all values of USPD are greater than zero, but less than 50%, we conclude that there is price feedback both from the U.S. market into the home market and vice versa, but that the home market is much more significant in the price discovery process. Note that the middle column contains our main result; the first and third columns contain the same results, when converting all prices to either the U.S dollar, or the euro.

Table 3 Shares of price discovery

This table presents the U.S. share of price discovery (USPD) for each stock. We calculate USPD as follows:

$$USPD = \frac{\left|\alpha^{EU}\right|}{\left|\alpha^{EU}\right| + \left|\alpha^{US}\right|},$$

where the α coefficients are obtained from the Error Correction Model of Formulas 8 and 9. Since all numbers are less than 50%, we conclude that the reaction of the European home market prices to price shocks from the cross-listed market is smaller than the reaction of prices in the cross-listed market to European price shocks.
This table also presents differences in USPD, when converting all prices to USD / EUR / trivariate. This table gives the different results for different ways of treating the exchange rate: either converting all prices into U.S. dollars ('Prices in USD'), converting all prices into euros ('Prices in EUR'), or converting nothing and treating the exchange rate as a third variable in the error correction analysis ('Trivariate model').

Company name	Exchange	Prices converted to USD	Trivariate model, no conversion	Prices converted to EUR
Alcatel	Paris	10.0%	11.6%	16.5%
Allianz	Frankfurt	29.3%	35.6%	36.1%
AXA	Paris	4.7%	8.4%	14.0%
BASF	Frankfurt	7.3%	19.4%	21.6%
Bayer	Frankfurt	14.5%	23.6%	25.4%
DaimlerChrysler	Frankfurt	3.6%	16.8%	20.0%
Danone	Paris	31.8%	38.7%	42.8%
Deutsche Bank	Frankfurt	4.1%	4.0%	7.3%
Deutsche Telekom	Frankfurt	10.4%	19.9%	21.0%
E.ON	Frankfurt	3.0%	20.0%	22.9%
France Telecom	Paris	2.8%	12.2%	16.8%
Infineon	Frankfurt	30.3%	32.5%	32.6%
Lafarge	Paris	25.4%	31.8%	35.6%
Sanofi-Aventis	Paris	9.1%	15.1%	22.6%
SAP	Frankfurt	9.6%	15.4%	18.3%
Siemens	Frankfurt	2.5%	16.2%	19.9%
Suez	Paris	20.8%	26.4%	30.0%
Thomson	Paris	35.5%	40.8%	44.9%
Total	Paris	0.3%	15.2%	21.1%
Vivendi	Paris	2.2%	9.0%	14.0%
Average		12.9%	20.6%	24.2%
Median		9.4%	18.1%	21.4%
Min		0.3%	4.0%	7.3%
Max		35.5%	40.8%	44.9%

6.2. The effects of converting price data to another currency

In the case of a European stock cross-listed in the U.S., the stock is traded in two different currencies, in euros in the European home market, and in dollars in the U.S. Both the law of one price and the work of an arbitrageur involve the conversion of one price into the other using the prevailing exchange rate.

There is some disagreement as to the effects of converting price data from one currency to another. Before applying their error correction model to the price series, Eun & Sabherwal (2003) convert all U.S. prices to Canadian dollars, and then perform all analyses in this currency. They also perform their analyses using unconverted prices, but do not find any significant difference in the results. In contrast, Grammig, Melvin and Schlag (2005) specifically analyze the effects of the foreign exchange rate in the context of international cross-listings. Their simulations show that the conversion of prices from one currency to another introduces a bias into the analysis of price discovery. The relative share of price discovery of the market whose price is converted into the foreign currency is overstated. They also find that the more volatile the currency, the greater the bias. For this reason they recommend incorporating the exchange rate as an endogenous variable to the model.

We perform a systematic analysis of the three methods of studying prices in two different currencies: converting all prices into U.S. dollars, converting all prices into euros, and performing a trivariate analysis with the exchange rate as the third variable, and performing no conversion at all. We present the resulting differences in shares of price discovery in Table 3. There is a clear pattern in the prices. We find that the U.S. share of price discovery (USPD) is higher when we convert both price series into euros, and lower when we convert the prices into dollars. The trivariate model, advocated by Grammig, Melvin, and Schlag (2005), consists of converting nothing and including the exchange rate as a third variable. This model gives in every case a share of price discovery which is larger than the share obtained using prices in U.S. dollars, but smaller than when using prices in euros. We are thus able to lend empirical support to the simulation findings of Grammig, Melvin, and Schlag (2005). As far as our price discovery analysis is concerned, we consider these trivariate results our main result, and use them as our dependent variable in our cross-sectional analyses of the differences in shares of price discovery.

6.3. Explanatory variables

When seeking to explain the differences in shares of price discovery between the companies in our sample, we use a number of explanatory variables, presented in Table 4. We include both measures of liquidity (relative average trading volumes and bid-ask spreads of the two markets), variables that describe the company (market capitalization, the number of years listed as an ADR, and the international share of sales), as well as ownership variables (U.S. holdings and the size of the largest shareholding).

We define home market trading volume as the number of shares traded in the home market, Frankfurt or Paris, as a proportion of total shares traded in the home market and the U.S. market during simultaneous trading hours. We adjust for the ratio of ADRs to common shares as needed. As can be seen from table 4, relative trading volume between the two markets exhibits considerable variation. The home market share of total trading is ranges from a minimum value of 19% for AXA to a high of 94%

for BASF. However, on average the home market dominates, with an average share of 61% over all stocks.

In his seminal paper, Hasbrouck (1995) studies the relative contributions of the NYSE and regional exchanges to price discovery of the 30 stocks in the Dow Jones Industrial Average. He finds a positive and statistically significant relation between shares of trading volume and the shares of price discovery between the NYSE and U.S. regional exchanges. Foerster and Karolyi (1998) find that the more liquid the U.S. market for Canadian stocks is after a cross-listing, the more beneficial are the effects for the local, Canadian market. Their hypothesis is, that a liquid competing market in New York forces the Canadian market makers to lower bid-ask spreads and thus lower execution costs in the local market. Werner and Kleidon (1996) find much the same result for UK stocks cross-listed in the U.S. Dealers in the London market seem very responsive to the increased competition from the U.S. market.

The bid-ask spread is an important measure of trading costs. Following earlier studies, such as Fleming, Ostdiek and Whaley (1996), Harris, et al. (2002), and Eun and Sabherwal (2003), we expect the share of price discovery of the home exchange to be positively related to the ratio of bid-ask spreads of the cross-listed exchange and the home exchange. In other words, the greater the relative share of price discovery of the home exchange, the greater the bid-ask spread of the cross-listed exchange in relation to the spread of the home exchange. Since the bid-ask spread is a major component of trading costs, we can expect it to influence price discovery. Fleming, Ostdiek, and Whaley (1996) suggest that the market with a smaller bid-ask spread will have a greater share of informed trading. Harris et al. (2002) find evidence that the NYSE share of price discovery is negatively related to the size of its bid-ask spread. For our sample of twenty European stocks, the bid-ask spread at the home exchange is smaller than at the NYSE, with only two exceptions: Sanofi-Aventis, and Total.

We also include two variables to reflect the different ownership structures of the companies. The first variable is the share of the largest shareholder, as a percentage of total shares outstanding. The second ownership variable is the total share of U.S. based owners, as a percentage of the total number of shares.

We employ three further variables in the cross-sectional analysis. “Exchange” is a dummy variable, 1 for Paris and 0 for Frankfurt. This variable can potentially spot differences between the two exchanges, or the countries. “Years listed” is the number of years the company’s stock has been listed on the New York Stock Exchange as an ADR. “International sales” is the percentage share of the company’s total sales that take place outside the home market. We expect this to be positively related to the U.S. share of price discovery.

Table 4 The explanatory variables

Market cap is market capitalization, in U.S. dollars, the logarithm of which is used in the regression. *Home volume* is the share of trading volume of the home market, as a share of total trading value. *Spread ratio* is the ratio of the average U.S. bid-ask spread and the average home market bid-ask spread, where spread is (Ask-Bid)/((Ask+Bid)/2).*Years listed* is the number of years the company has been listed as ADRs in the U.S. *Int'l sales* is the share of the sales taking place outside the company's home market. *Largest shareholder* is the share of the largest shareholder as a percentage of total shares, and other ownership variable, *U.S. holdings,* is the share of U.S. based investors as a percentage of total shares.

Company name	Market cap, MUSD	Home volume	Spread ratio	Years listed	Int'l sales	Largest shareholder	U.S. holdings
Alcatel	21,020	54%	1.15	12	36%	9	10.3
Allianz	48,295	66%	4.16	4	47%	12	7.2
AXA	45,669	19%	1.57	8	23%	17	13
BASF	39,518	94%	2.13	4	41%	0	14
Bayer	24,817	82%	1.68	2	49%	6	8
DaimlerChrysler	48,665	74%	1.19	6	16%	12	17
Danone	23,112	84%	2.77	7	68%	4	1.1
Deutsche Bank	40,189	44%	1.79	3	21%	3	8.5
Deutsche Telekom	95,143	71%	1.31	8	57%	26	29
E.ON	59,840	65%	2.2	7	64%	6	19.7
France Telecom	80,849	81%	1.98	7	59%	23	2
Infineon	7,645	87%	1.17	4	23%	28	3
Lafarge	16,253	37%	2.5	3	16%	9	2.3
Sanofi-Aventis	69,022	89%	0.91	2	49%	12	20
SAP	54,933	44%	1.17	6	27%	10	8.1
Siemens	65,671	78%	1.77	3	32%	9	12
Suez	26,619	41%	3.15	3	23%	7	9
Thomson	7,218	40%	2.64	5	23%	6	1
Total	133,610	55%	0.59	13	19%	5	23
Vivendi	34,389	21%	1.48	9	56%	10	2
Mean	47,124	61.3%	1.87	5.8	37.5%	10.7	10.5
Median	42,929	65.5%	1.73	5.5	34.0%	9.0	8.8
Standard deviation	31223.6	22.9%	0.86	3.1	17.3%	7.5	8.0
Min	7,218	19.0%	0.59	2.0	16.0%	0.0	1.0
Max	133,610	94.0%	4.16	13.0	68.0%	28.0	29.0
Skewness	1.166	-0.367	1.032	0.875	0.351	1.153	0.718
Kurtosis	1.786	-1.008	1.277	0.276	-1.384	0.792	-0.116

6.4. Cross-sectional results

We perform a cross-sectional regression to explain the differences in shares of prices discovery between the stocks included in our sample. We do this by performing an OLS regression, using *HomePD*, the home market share of price discovery, as our dependent variable.

Table 5 presents our regression results. The most important conclusion is that market capitalization, relative trading volume, and bid-ask spread are significant explanatory variables, whereas no other variable seems to add any significant explanatory power to the analysis.

Market capitalization has a negative coefficient in every regression, which is consistent with the conclusion that larger companies tend to be more diversified internationally. From the point of view of a U.S. based investor, they may be of more interest than a smaller company, thus making the role of the New York market more significant in price discovery.

The two trading related variables describe market quality and liquidity in the stock. The relationship between U.S. trading share of trading volume and U.S. price discovery is positive, as expected. This implies that the greater the U.S. share of trading, the greater the U.S. share of price discovery. The ratio of bid-ask spreads also has the expected sign. The greater the bid-ask spread in the U.S., relative to the European home market, the smaller the share of price discovery in the U.S. Market capitalization is also positively correlated with a high U.S. influence on price discovery.

Our two company-specific ownership variables do not seem to add any explanatory power to the analysis. One reason for this may be that the data set does not reflect the true composition of U.S. based shareholders, or that a number of those listed as U.S. based are able to choose to trade in the European market as well. In this case, their trading is not reflected in the U.S. share of price discovery.

Table 5 Cross-sectional results

The dependent variable is *HomePD*, the home market share of price discovery. It is equivalent to 1-USPD, the cross-listed share of price discovery, used in the analyses above. *Market cap* is the logarithmic market capitalization, in U.S. dollars. *Home volume* is the share of trading volume of the home market, as a share of total trading. *Spread ratio* is the ratio of U.S. bid-ask spread and the home market bid-ask spread. *Exchange dummy* is 1 for French stocks and 0 for German stocks. *Years listed* is the number of years the company has been listed as ADRs in the U.S. *Int'l sales* is the share of the sales taking place outside the company's home market. *Largest shareholder* is the share of the largest shareholder as a percentage of total shares. *U.S. holdings* is the share of U.S. based investors as a percentage of total shares. Numbers in parenthesis are adjusted t-statistics. Numbers in boldface are significant at a 5% significance level.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
intercept	0.726***	0.686**	0.700**	0.649**	0.625**	0.835**	0.808**	0.998**
	(3.123)	(2.804)	(2.815)	(2.434)	(2.232)	(2.698)	(2.772)	(2.913)
Market cap	-0.068***	-0.067***	-0.072***	-0.068**	-0.067**	-0.090**		
	(-3.255)	(-3.106)	(-3.148)	(-2.815)	(-2.674)	(-2.940)		
Home volume	0.144**	0.163**	0.178**	0.207**	0.205*	0.190*	-0.069**	-0.096**
	(2.135)	(2.179)	(2.270)	(2.240)	(2.145)	(2.055)	(-2.514)	(-2.741)
Spread ratio	0.060***	0.060***	0.066***	0.072***	0.074***	0.078***	0.057**	0.070**
	(3.251)	(3.221)	(3.234)	(3.158)	(3.091)	(3.375)	(2.295)	(2.631)
Exchange dummy		0.022	0.012	0.016	0.017	0.026	(0.011)	0.007
		(0.646)	(0.341)	(0.419)	(0.435)	(0.683)	(-0.264)	(0.175)
Years listed			0.005	0.006	0.006	0.005	0.001	0.000
			(0.765)	(0.895)	(0.850)	(0.719)	(0.127)	(-0.022)
Int'l sales				(0.001)	(0.001)	(0.001)	0.001	0.001
				(-0.636)	(-0.662)	(-0.434)	(0.499)	(0.551)
Largest shareholder					0.001			0.001
					(0.475)			(0.560)
U.S. holdings						0.004		0.005
						(1.145)		(1.312)
Adjusted R^2 (%)	59.8%	58.3%	57.1%	55.2%	52.3%	56.2%	42.3%	42.3%

7 CONCLUSIONS

We study price discovery for European stocks cross-listed at the NYSE. We address three main issues. First, we find that for all twenty companies in our sample the majority of price discovery occurs on the home exchange, either Paris or Frankfurt. The share of New York Stock Exchange is in all cases greater than zero, but smaller than that of the home exchange. The U.S. shares of price discovery range from 4% to 41%. This means that U.S. prices adjust more to the prices of the home exchange than vice versa.

Second, we explain the differences in shares of price discovery. We find that the bigger the company, measured by market capitalization, the more significant the U.S. market is for price discovery. The greater the share trading that takes place at the home exchange, the greater is the share of the home exchange in price discovery. Relative cost of trading, measured by the average bid-ask spread, is an additional significant explanatory variable. The greater the ratio of the average U.S. bid-ask spread and the average home market bid-ask spread, the greater is the role of the home market in determining price discovery. We also include two company-specific ownership variables: holdings by U.S. based investors and the size of the holding of the largest shareholder. However, these do not add explanatory power to the cross-sectional analysis.

Third, we explore the effects of the exchange rate for the price discovery process. Grammig, Melvin, and Schlag (2005) find, using a data generating process and simulations, that converting the prices of one market to the currency of the other market in a cross-listed setting introduces a bias to the shares of price discovery. We are able to lend empirical support to their findings. We find that converting home market prices of the European stocks to dollars raises the share of price discovery of the European market, and converting U.S. prices to euros raises the share of price discovery attributed to the U.S. exchange. An approach where we incorporate the exchange rate as a third variable without converting any prices to another currency yields unbiased results.

APPENDIX 1

Zero return intervals

In order to estimate the liquidity of the markets, we calculate the numbers of intervals with a zero return. In the context of our analyses, in which we use midquotes as our primary data, a zero return means that the midquote is unchanged from one interval to the next. An interval may have a zero return for two possible reasons. First, there have been updates to the bid–ask quotes, but the midquote is unchanged. The second possibility is that there have not been any updates, in which case we use the latest available quote.

These numbers are a measure of the liquidity of the market. It is clear from the results in Table 6 that for may stocks the one-minute intervals are of a higher frequency than the average time between quote changes. However, since our results for the analysis of price discovery are largely similar for all three intervals used, we do not believe that this affects our results too.

Table 6 The number of zero-return intervals

These figures represent the share of intervals with zero price return from the previous interval. The data are for the three intervals used in this article, 1, 2, and 5 minute intervals, and for the U.S. (cross-listed) market and the European (home) market. U.S. prices are in dollars and European prices in euros. The numbers are percentages of the total number of intervals.

	U.S. (cross-listed) market			European (home) market		
	1-minute intervals	2-minute Intervals	5-minute intervals	1-minute intervals	2-minute intervals	5-minute intervals
Alcatel	44.7%	30.9%	17.3%	47.3%	33.2%	21.4%
Allianz	64.6%	49.7%	32.3%	14.2%	8.0%	4.7%
AXA	44.3%	29.2%	14.5%	40.8%	28.4%	16.9%
BASF	33.3%	20.0%	9.1%	25.3%	15.9%	8.6%
Bayer	39.1%	24.3%	11.8%	39.3%	26.3%	15.4%
Daimler Chrysler	24.0%	14.0%	7.1%	26.2%	16.9%	9.8%
Danone	62.1%	46.7%	24.8%	65.7%	51.3%	33.6%
Deutsche Bank	18.6%	9.6%	4.9%	16.1%	9.9%	5.8%
Deutsche Telekom	52.3%	37.9%	23.3%	54.8%	40.7%	27.4%
E.ON	26.3%	14.5%	6.3%	21.2%	13.0%	8.1%
France Telecom	45.9%	29.5%	12.5%	42.1%	27.5%	16.3%
Infineon	52.4%	37.4%	22.0%	58.5%	45.4%	29.4%
Lafarge	54.3%	38.2%	18.2%	59.3%	44.9%	27.3%
Sanofi-Aventis	40.7%	26.7%	14.6%	66.9%	52.6%	35.1%
SAP	27.4%	16.1%	8.8%	11.5%	5.6%	2.8%
Siemens	17.9%	9.9%	4.7%	16.5%	10.3%	6.4%
Suez	55.7%	40.7%	22.9%	49.0%	35.0%	20.5%
Thomson	51.1%	35.8%	16.9%	56.1%	41.6%	25.9%
Total	15.9%	8.4%	4.6%	56.1%	41.8%	25.9%
Vivendi	38.0%	24.6%	12.1%	40.0%	26.7%	14.9%
Average	40.4%	27.2%	14.4%	40.3%	28.8%	17.8%

REFERENCES

Alexander, Gordon J., Cheol S. Eun, and S. Janakiramanan, 1987, Asset pricing and dual listing on foreign capital markets: A note, *Journal of Finance* 42, 151–158.

Baillie, Richard T., G. Geoffrey Booth, Yiuman Tse, and Tatyana Zabotina, 2002, Price discovery and common factor models, *Journal of Financial Markets* 5, 309–322.

Booth, Geoffrey, Ji-Chai Lin, Teppo Martikainen, and Yiuman Tse, 2002, Trading and pricing in upstairs and downstairs stock markets, *Review of Financial Studies* 15, 1111–1135.

Chowdhry, Bhagwan, and Vikram Nanda, 1991, Multimarket trading and market liquidity, *Review of Financial Studies* 4, 483–511.

Dickey, David A., and Wayne A. Fuller, 1981, Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica* 49, 1057–1072.

Domowitz, Ian, Jack Glen, and Ananth Madhavan, 1998, International Cross-Listing and Order Flow Migration: Evidence from an Emerging Market, *Journal of Finance* 53, 2001–2027.

Eun, Cheol S., and Sanjiv Sabherwal, 2003, Cross-Border Listings and Price Discovery: Evidence from U.S.-Listed Canadian Stocks, *Journal of Finance* 58, 549–575.

Fleming, Jeff, Barbara Ostdiek, and Robert E. Whaley, 1996, Trading costs and the relative rates of price discovery in stock, futures, and option markets, *Journal of Futures Markets* 16, 353–387.

Foerster, Stephen R., and G. Andrew Karolyi, 1998, Multimarket trading and liquidity: a transaction data analysis of Canada – US interlistings, *Journal of International Financial Markets, Institutions, and Money* 8, 393–412.

Foerster, Stephen R., and G. Andrew Karolyi, 1999, The Effects of Market Segmentation and Investor Recognition on Asset Prices: Evidence from Foreign Stocks Listing in the United States, *Journal of Finance* 54, 981–1013.

Foerster, Stephen R., and G. Andrew Karolyi, 2000, The Long-Run Performance of Global Equity Offerings, *Journal of Financial and Quantitative Analysis* 35, 499–528.

Gagnon, Louis Joseph and G. Andrew Karolyi, Multi-Market Trading and Arbitrage (June 17, 2004), *Working paper*, Dice Center No. 2004–9, http://ssrn.com/abstract=577004.

Grammig, Joachim, Christian Schlag, and Michael Melvin, The Role of U.S. Trading in Pricing Internationally Cross-Listed Stocks, *EFA 2004 Maastricht Meetings Paper No. 3267*, http://ssrn.com/abstract=567066.

Grammig, Joachim, Michael Melvin, and Christian Schlag, 2005, Internationally Cross-Listed Stock Prices during Overlapping Trading Hours: Price Discovery and Exchange Rate Effects, *Journal of Empirical Finance* 12, 139–164.

Harris, Frederick H. deB., Thomas H. McInish, Gary L. Shoesmith, and Robert A. Wood, 1995, Cointegration, Error Correction, and Price Discovery on Informationally Linked Security Markets, *Journal of Financial and Quantitative Analysis* 30, 563–579.

Harris, Frederick H. deB., Thomas McInish, and Robert A. Wood, 2002, Security price adjustment across exchanges: an investigation of common factor components for Dow stocks, *Journal of Financial Markets* 5, 277–308.

Hasbrouck, Joel, 1995, One security, many markets: determining the contributions to price discovery, *Journal of Finance* 50, 1175–1199.

Hasbrouck, Joel, 2002, Stalking the “efficient price” in market microstructure specifications: an overview, *Journal of Financial Markets* 5, 329–339.

Hauser, Shmuel, Yael Tanchuma, and Uzi Yaari, 1998, International transfer of pricing information between dually listed stocks, *Journal of Financial Research* 21, 139–157.

Hedvall, Kaj, Eva Liljeblom, and Kim Nummelin, 1997, Dual listing and price discovery: The role of the domestic market, Working Paper 355, Swedish School of Economics and Business Administration.

Huang, Roger D., 2002, The quality of ECN and Nasdaq market maker quotes, *Journal of Finance* 57, 1285–1319.

Johansen, Soren, 1988, Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* 12, 231–254.

Johansen, Soren, and Katarina Juselius, 1990, Maximum likelihood estimation and inference on cointegration with application to the demand for money, *Oxford Bulletin of Economics and Statistics* 52, 169–209.

de Jong, Frank, 2002, Measures of contribution to price discovery, a comparison, *Journal of Financial Markets*, 5, 323–327.

Karolyi, G. Andrew, 2006, The world of cross-listings and cross-listings of the world: Challenging conventional wisdom, *Review of Finance* 10, 99–152.

Kim, Minho, Andrew C. Szakmary, and Ike Mathur, 2000, Price transmission dynamics between ADRs and their underlying foreign securities, *Journal of Banking & Finance* 24, 1359–1382.

Lau, Sie Ting, and J. David Diltz, 1994, Stock returns and the transfer of information between the New York and Tokyo stock exchanges, *Journal of International Money and Finance* 13, 211–222.

Lehmann, Bruce N., 2002, Some desiderata for the measurement of price discovery across markets, *Journal of Financial Markets* 5, 259–276.

Lieberman, Offer, Uri Ben-Zion, and Shmuel Hauser, 1999, A characterization of the price behavior of international dual stocks: an error correction approach, *Journal of International Money and Finance*, 18, 289–304.

Menkveld, Albert, 2008, Splitting orders in overlapping markets: A study of cross-listed stocks, *Journal of Financial Intermediation* 17, 145–174.

Müller, Ulrich, 2001, The Olsen Filter for Data in Finance, A consulting document by Olsen Associates.

Pagano, Marco, Ailsa A. Roell, and Josef Zechner, 2002, The Geography of Equity Listing: Why do Companies List Abroad?, *Journal of Finance* 57, 2651–2694.

Pascual, Roberto, Bartolomé Pascual-Fuster, and Francisco Climent, 2006, Cross-listing, price discovery and the informativeness of the trading process, *Journal of Financial Markets* 9, 144–161.

Phylaktis, Kate, and Piotr Korczak, 2007, Specialist Trading and the Price Discovery Process of NYSE-Listed Non-US Stocks, *Working paper*, http://ssrn.com/abstract=567104.

Roll, Richard, 1984, A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Finance* 39, 1127–1139.

Schreiber, Paul S., and Robert A. Schwartz, 1986, Price discovery in securities markets, *Journal of Portfolio Management* 12, 43–48.

Schwarz, T., and A.C. Szakmary, 1994, Price discovery in petroleum markets: Arbitrage, cointegration, and the time interval of analysis, *Journal of Futures Markets* 14, 147–167.

Smith, Katherine, and George Sofianos, 1997, The impact of an NYSE listing on the global trading of non-U.S. stocks, *NYSE working paper* 97–02.

Theissen, Erik, 2002, Price discovery in floor and screen trading systems, *Journal of Empirical Finance* 9, 455–474.

Wang, Steven Shuye, Oliver Meng Rui, and Michael Firth, 2002, Return and volatility behavior of dually-traded stocks: the case of Hong Kong, *Journal of International Money and Finance* 21, 265–293.

Essay 2

PRE-TRADE TRANSPARENCY, MARKET QUALITY, AND INFORMED TRADING

Abstract

We study the effects of a change from a pre-trade transparent limit order book to an anonymous electronic limit order book. We estimate the Probability of Informed Trading (PIN) for the pre- and post-change periods. We do not find significant explanatory power for the PIN measure in explaining bid-ask spreads after a change to anonymous trading. More generally, we do not find unambiguous evidence of an improvement in market quality, measured by bid-ask spreads, intraday volatility, and trading volume. We also find no substantial change in upstairs trading.

Keywords: market transparency, informed trading, event study.

JEL Classification: G12, G14, G15.

1 INTRODUCTION

Should an exchange be as transparent as possible, revealing prices, market depth, and the identities of traders posting bid and ask prices to all market participants?[15] And if so, who gains and who loses if a transparent market is made less transparent? The question is timely, and it affects the ongoing competition and consolidation in global equities markets. In order to accommodate the needs of different kinds of investors, several exchanges, e.g. in Sydney, Seoul, Paris, Tokyo, and Helsinki have made changes in market transparency during the last few years[16].

We study the change from a transparent market to an anonymous market at the OMX Helsinki Stock Exchange in March 2006. We use the Probability of Informed Trading (PIN) to model the number of informed traders in the market. However, in a cross-sectional regression we do not find evidence of a significant explanatory power for the PIN variable.

We also study whether market quality has improved, measured in a smaller bid-ask spread, higher trading volume, or diminished intraday volatility. The evidence is mixed, with an unchanged average bid-ask spread, higher trading volume, and higher volatility. As a comparison to the electronic limit order book, we study the upstairs market, which is unaffected by the changes in the limit order book. We find no change in the numbers of upstairs trades, or internalization rates.

Theoretical research reaches mixed conclusions about the effects of transparency. Papers supporting the view that increased transparency enhances liquidity include Admati and Pfleiderer (1991), Pagano and Roell (1996), and Baruch (2005). The most common explanation is the mitigated effect of information asymmetry (see, e.g. Chowdhry and Nanda (1991), Madhavan (1995), and Bloomfield and O'Hara (1999)). Pagano and Roell (1996) differentiate between uninformed and informed traders, and find that the former benefit from greater transparency.

Empirical research on the effects of market transparency was initially conducted as experiments, due to a lack of data. Experimental studies with human subjects in laboratory conditions are a promising avenue of research. Early papers, such as Bloomfield and O'Hara (1999, 2000), and Flood, Huisman, Koedijk, and Mahieu (1999), find fairly complex results regarding the effects of transparency. Bloomfield and O'Hara (1999) find that price discovery is more efficient in an opaque market, while other measures of the quality of the market, such as bid-ask spreads and trading volume, deteriorate.

The results of the empirical study of the NASDAQ market by Harris and Schultz (1997) support market transparency. They find that the anonymous Small Order Execution System of the NASDAQ market makers show wider bid-ask spreads compared with the regular, non-anonymous dealer markets. Theissen (2003) finds a similar result for the German market. On the non-anonymous floor-based trading system of the Frankfurt

---

[15] Transparency with respect to prices and quantities is a different issue from transparency with respect to trader identities. These two issues may also have very different implications for market quality. In this paper our focus is on transparency with respect to trader identities.

[16] Euronext Paris, Tokyo Stock Exchange, and the Australian Stock Exchange removed the display of broker identifiers from the limit order book on April 23, 2001, June 30, 2003, and November 28, 2005, respectively. The Korea Stock Exchange made the opposite move, introducing broker IDs on October 25, 1999. OMX Helsinki, the object of the present study, removed the display of broker IDs on March 13, 2006.

Stock Exchange, specialists are able to credibly identify uninformed traders. When trading with an uninformed trader, the specialists are willing to offer price improvement. Transparency seems to enhance liquidity in this case.

Boehmer, Saar and Yu (2005) study the introduction of the electronic limit order market, OpenBook, at the NYSE. They find that this great increase in transparency to market participants outside the trading floor resulted in increased market depth and a reduced effective spread. Hendershott and Jones (2005) study a change to a less transparent trading system at Island, an Electronic Communications Network (ECN). Island stopped displaying the limit order book for the most liquid Exchange Traded Funds (ETFs), which resulted in higher trading costs, lost market share and decreased liquidity.

Madhavan, Porter, and Weaver (2005) examine an increase in pre-trade transparency on the Toronto Stock Exchange in 1990. In contrast to the above studies, they find that execution costs and volatility increased. The authors attribute the finding to the increased efficiency in order placement by market makers, and the decreased willingness of limit-order submitters to offer them free liquidity options. Other papers supporting less pre-trade transparency include Madhavan (1996), Garfinkel and Nimalendran (2003), and Foucault, Moinas, and Theissen (2007).

Most previous empirical research compares two exchanges with different characteristics. It is generally consistent with the notion that anonymity is associated with higher adverse selection costs. Huang and Stoll (1996) compare two markets, NYSE and NASDAQ, where a NYSE specialist can see the limit order book, and NASDAQ participants cannot. They find that spreads are generally higher on the more opaque NASDAQ. Chan and Lakonishok (1995) study institutional trading on NYSE and on NASDAQ. They find that smaller stocks have better execution on NASDAQ, and large stocks on NYSE.

Madhavan and Cheng (1997) find that, consistent with the model of Seppi (1990), the upstairs market is used by traders who can credibly signal that they trade for liquidity reasons. de Jong, Nijman, and Roell (1996) show that trades that are negotiated bilaterally (and thus non-anonymously) and are then executed through the Paris Bourse's CAC system have a lower price impact than regular CAC trades. Garfinkel and Nimalendran (2003) document that NYSE stocks exhibit larger increases in the bid–ask spread on insider trading days than NASDAQ stocks and conclude that the trading system of the NYSE is less anonymous. Grammig, Schiereck, and Theissen (2001) compare floor trading at the Frankfurt Stock Exchange with the electronic Xetra market. They find that informed traders tend to prefer the anonymous electronic market, whereas uninformed traders are drawn to the smaller adverse selection costs of the floor market.

Another related paper is Simaan, Weaver, and Whitcomb (2003). They compare the quotation behavior of Nasdaq market makers on different trading platforms. They find that when dealers are able to quote anonymously, on ECNs (Electronic Communications Networks), the bid-ask spread narrows. This implies collusion among the dealers on the non-anonymous Nasdaq platform.

A recent paper by Maher et al. (2008) corroborates the theoretical considerations of Fishman and Hagerty (1995). They find that anonymity significantly decreases market quality. The bid-ask spread increases, especially for small stocks. Additional effects of anonymity include a decrease in trade volume and an increase in intraday volatility.

As originally noted by Copeland and Galai (1983), limit orders have option-like features. Sell (buy) limit orders are similar to a free call (put) option with a strike price equal to the price of the limit order. As option prices are dependent on volatility, limit order traders should also be able to use volatility information when placing their orders.

This treatment of limit orders as options, whose pricing is of course dependent on the volatility of the underlying security, is the starting point of the analysis by Foucault, Moinas, and Theissen (2007). They construct a three period model with several types of participants. In period 0 two kinds of limit order traders, value traders and pre-committed traders, post limit orders. Value traders post limit orders only if it is profitable to do so, whereas pre-committed traders are committed to buying or selling a given number of shares. Value traders are either informed or uninformed about future volatility. Being informed about future volatility in this case simply means knowing whether there is an information event in the next period. In period 1, speculators and liquidity traders trade using market orders. If there is an information event, a speculator arrives with a given probability, and buys or sells according to the observed price innovation. He is informed about the asset value, and picks off buy or sell limit orders within the bounds given by the expected volatility. If there is no speculator, a liquidity trader is equally likely to buy or to sell, using a market order.

Foucault, Moinas, and Theissen (2007) model dealers whose limit orders may become stale and are subsequently picked off by faster speculators when volatility is high. The amount of informed trading is a key variable in this setting. In a transparent market with a low number of informed traders, uninformed traders learn from informed traders with volatility information. In this case bid-ask spreads widen in anticipation of adverse information events as dealers protect themselves. On the other hand, in an anonymous market spreads will narrow and the limit order book will become less informative. The bid-ask spread and its informativeness will both decline with the change to an anonymous market.

The opposite result obtains when there are lots of informed traders. In this case the uninformed traders are wary of trading with an informed trader. Thus the bid-ask spread will widen with anonymity. Intuitively this result can be understood as follows. In the case of few informed traders in the market, quotes are most likely posted by uninformed traders, and traders are not afraid to post aggressive limit orders, thus decreasing the bid-ask spread. When the number of informed traders increases, quotes are more likely to be informed. Anybody seeing a large bid-ask spread will be more cautious, and posts smaller orders, and at less aggressive prices. The bid-ask spread will either increase, or remain unchanged.

Another related paper is Rindi (2008), who studies the effect of pre-trade transparency on the informed traders' demand. She studies different types of markets, and concludes that factors such as the existence of potential insider information in a market can determine whether transparency is beneficial to a market. In a market where insider information is unlikely or non-existent, such as government bond and foreign exchange markets, and equity markets with strong insider regulations, less transparency would enhance market quality. The reason is that greater transparency reduces the incentive to acquire information. This reduces the number of informed traders, who are willing to accommodate liquidity demands from uninformed traders. The main difference between this model and the model of Foucault et al. (2007) is volatility, which is not analyzed by Rindi (2008).

Our main contribution is the study of the Probability of Informed Trading (PIN) as a measure of informed trading in the context of a change in the pre-trade transparency of trader identities. We first estimate the PIN measure for all stocks, for both the pre- and the post-change period. We then use the estimated values for PIN as an explanatory variable for the bid-ask spread, with a dummy variable to account for the change to an anonymous market. We do not find evidence PIN is a significant explanatory variable for the post-change spread.

We contribute to the existing literature on the effects of a change in market transparency by studying a recent change in pre-trade transparency. It is notable that the object of our study, the OMX Helsinki market in 2006, has full post-trade transparency. This means that he counterparties of each trade are immediately disclosed to all market participants. Our empirical analysis indicates that the change to a pre-trade anonymous market in not as clearly beneficial to market quality as many previous studies suggest. We find a significant increase in intraday volatility, and no significant change in average bid-ask spreads. Results from a pooled regression show that we can attribute part of the changes in quoted spreads to the change to an anonymous trading system.

The changes made to the transparency of the limit order book do not affect the upstairs market. This market consists of brokers personally talking to other brokers, either within or outside the brokerage. The information about counterparties is in this way automatically present. We do not find any change in upstairs trading, which is our expected result.

The rest of this study is organized as follows. Chapter 2 discusses the data and the exchange studied. Chapter 3 presents the methodology for studying informed trading, including details of estimation of PIN. Chapter 4 discusses our empirical results, and Chapter 5 concludes.

2 INSTITUTIONAL DETAILS AND THE DATASET

We study transparency of the order book at the OMX Helsinki Stock Exchange[17]. OMX Helsinki is an electronic limit order market. After an opening auction, continuous trading starts at 10 a.m., and ends in a closing auction starting at 6.20 p.m. Before the changes implemented on 13th March 2006, the trading system allowed all participants to see the identities of all brokers posting limit orders in the limit order book. After the change all identities are hidden. The only visible information is price and size for each limit order in the book. It is notable that these changes do not affect post-trade transparency. This means that the counterparties of each trade are known to all market participants immediately after a trade takes place.

We use a dataset provided by the Helsinki Stock Exchange. Our sample consists of the 35 most liquid stocks, measured by daily trading volume[18]. Our selection criteria are that the stocks be continuously listed during the entire sample period, and that the average daily trading volume be greater than one million euros.

Table 1 presents descriptive statistics of our sample companies. As is apparent both from the market capitalizations, and the trading volume statistics, our sample is very heterogeneous. Nokia is by far the largest and most liquid stock, both by market capitalization and trading volume. There is a second tier of companies, which could be said to include Fortum, Metso, Neste Oil, Outokumpu, Sampo, Stora Enso, and UPM-Kymmene. The rest of the sample companies are smaller and less liquid. All companies are liquid enough for the analysis of informed trading. The average daily number of trades is in the hundreds or thousands for almost all companies in our sample.

60 days is generally regarded as the minimum sample period for the reliable estimation of the Probability of Informed Trading, see e.g. Easley et al. (1997), Easley et al. (2005), and Aktas et al. (2007). On the other hand, most related event studies on exchange transparency use shorter sample periods, presumably to minimize the presence of other factors affecting the trading environment. With due consideration for these two factors, we use a total of six months of data in two sub-samples of equal size. There are 62 trading days before the changes of March 13, 2006, and 62 trading days after the changes. We exclude the day of the actual change, to give the market time to adjust to the new trading system. Our pre-change sample, which covers the period of non-anonymous trading, runs from December 13, 2005 to March 10, 2006. Our post-change (anonymous trading) sample period runs from March 14, 2006 to June 13, 2006.[19].

---

[17] HEX, the Helsinki Stock Exchange, merged with the Swedish OMX Group in September 2003 to form OM HEX. OM HEX later changed their name to OMX Group. The company has since then become a major operator in the Nordic area, after merging with or acquiring the following marketplaces: Copenhagen, Stockholm, Helsinki, and Iceland in the Nordic market, and Tallinn, Riga, and Vilnius in the Baltic states. There is also an alternative exchange for small companies, First North. In 2007, NASDAQ acquired OMX Group, to form the NASDAQ OMX Group, Inc.

[18] There are a number of changes in the exchange listings during the sample period. Ahlstrom Corporation Oyj (AHL1V) was listed on March 14, 2006. Also, SanomaWSOY Oyj combined their two share series, SWSAV and SWSBV into one series, SWS1V, on April 10, 2006. We therefore exclude both these stocks from our sample.

[19] Our sample periods of 62 + 62 days are longer than those used in many related studies, such as the sample periods of 20 + 20 trading days in Comerton-Forde et al. (2005), 25 + 29 days in Hendershott and Jones (2005), and 14 + 14 days in Foucault et al. (2007). Foucault et al. (2007) eliminate two weeks of

Table 1 Descriptive statistics

This table presents descriptive statistics for our sample of stocks. The sample period is December 13, 2005 – May 13, 2006. *Ticker code* is the official stock ID in the trading system. *Market cap* is the market capitalization as of the end of 2005. *Number of trades* is the average daily number of trades. *Price* is the average price over the entire sample period. *Trading volume* is the average daily trading volume, in millions of euros. *Trade size* is the average number of shares per trade.
For each descriptive variable, we also report average, median, standard deviation, minimum, and maximum values, the skewness coefficient, and excess kurtosis.

Stock name	Ticker code	Market cap	Number of trades	Price	Trading volume	Trade size
Alma Media Oyj	ALN1V	878	56	7.66	1.25	3423.3
Amer Sports Corporation	AMEAS	1,124	321	16.48	3.65	690.0
Cargotec Oyj	CGCBV	1,866	388	33.59	5.66	447.0
Elcoteq SE A	ELQAV	577	201	18.45	2.14	493.4
Elisa Oyj	ELI1V	2,596	848	16.35	14.16	1012.5
F-Secure Oyj	FSC1V	455	167	2.70	1.24	2376.3
Finnair Oyj	FIA1S	643	135	12.56	1.76	887.6
Finnlines Oyj	FLG1S	543	59	14.99	1.65	1880.4
Fortum Oyj	FUM1V	13,864	1,638	19.17	52.09	1610.6
Huhtamäki Oyj	HUH1V	1,374	309	14.85	3.64	814.4
KONE Oyj	KNEBV	4,261	653	34.05	9.65	426.4
Kemira GrowHow Oyj	KGH1V	899	195	5.52	1.40	1103.9
Kemira Oyj	KRA1V	797	270	13.88	3.50	919.2
Kesko Oyj B	KESBV	2,310	512	26.23	5.76	410.9
M-real Oyj B	MRLBV	1,384	536	4.57	8.41	3162.0
Metso Oyj	MEO1V	3,274	1,124	28.62	24.44	728.2
Neste Oil Oyj	NES1V	6,122	1,252	26.22	29.99	902.7
Nokia Oyj	NOK1V	64,463	6,345	16.35	515.47	4711.7
Nokian Renkaat Oyj	NRE1V	1,289	912	13.38	12.43	969.9
Nordea Bank AB (publ) FDR	NDA1V	22,729	346	9.45	13.34	3952.3
OKO Pankki Oyj	OKOAS	2,385	406	12.85	5.25	1018.4
Outokumpu Oyj	OUT1V	2,272	830	15.85	14.82	1134.1
Perlos Oyj	POS1V	565	355	7.67	2.88	972.0
Ramirent Oyj	RMR1V	378	115	26.73	1.91	669.3
Rautaruukki Oyj K	RTRKS	2,801	1,035	26.09	16.32	573.3
Sampo Oyj A	SAMAS	8,307	1,346	16.23	39.73	1737.0
Sponda Oyj	SDA1V	344	96	8.44	1.15	1405.7
Stockmann Oyj Abp B	STCBV	880	202	32.77	2.37	368.7
Stora Enso Oyj R	STERV	9,021	1,368	11.80	45.15	2573.7
TeliaSonera AB	TLS1V	20,364	316	4.72	8.02	5254.9
TietoEnator Oyj	TIE1V	1,847	870	28.77	16.63	604.5
UPM-Kymmene Oyj	UPM1V	8,662	1,675	17.76	60.66	1849.7
Uponor Oyj	UNR1V	988	204	21.37	2.44	533.1
Wärtsilä Oyj Abp	WRTBV	2,353	599	29.57	9.40	519.7
YIT Oyj	YTY1V	2,254	635	31.83	10.13	540.5
Mean		5568	752	17.93	27.10	1447.9
Median		1866	406	16.35	8.02	969.9
Standard deviation		11564	1077	9.16	86.33	1261.2
Minimum		344	56	2.70	1.15	368.7
Maximum		64463	6345	34.05	515.47	5254.9
Skewness		4.26	4.35	0.24	5.64	1.70
Kurtosis		20.62	22.32	-1.00	32.66	2.26

trading around the event date. Comerton-Forde (2005) and Hendershott and Jones (2005) do not eliminate any days around the event dates.

3 INFORMED TRADING

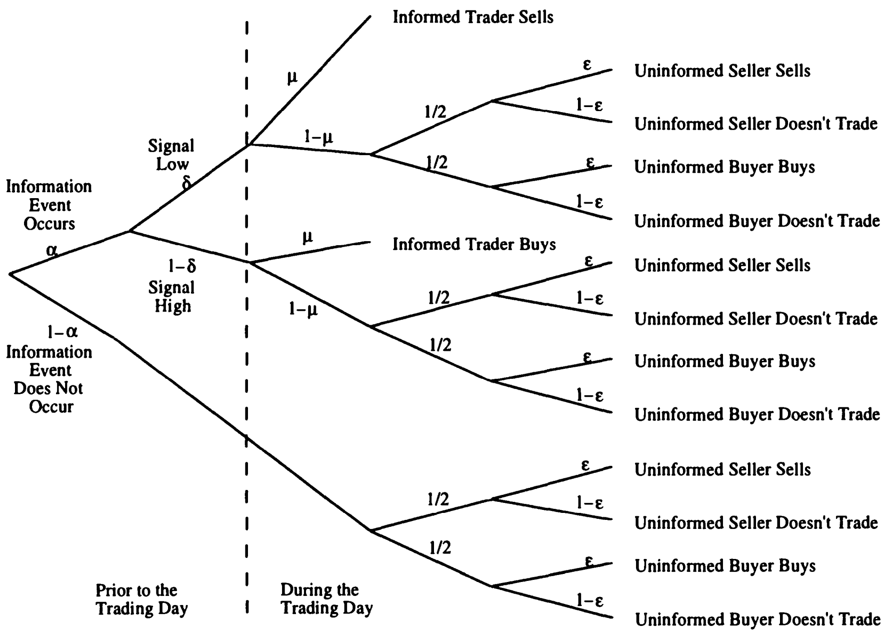
3.1. The PIN (Probability of Informed Trading) model

The Probability of Informed Trading (PIN) measures information asymmetry between investors and traders in a stock. The model was introduced in a series of papers by Easley, Kiefer, O'Hara, and Paperman (1996), and Easley, Kiefer, O'Hara (1996, 1997). The model is based on the Glosten and Milgrom (1985) sequential trading model.

We follow the original PIN model very closely. Figure 1 presents an intuitive overview of the trading process. There are three types of traders in our model: informed traders, uninformed liquidity traders, and risk neutral market makers. At time zero an information event may occur. This event has either a positive ('high'), or negative ('low') effect on the value of the security. An informed trader is risk neutral, takes prices as given, and does not engage in any strategic behavior.

Figure 1 Tree diagram of the trading process, from Easley, Kiefer, and O'Hara (1997).

This diagram is a representation of the trading model discussed in Section III.A. α is the probability of an information event, δ is the probability of a low signal, μ is the probability that the trade comes from an informed trader, ½ is the probability that an uninformed trader is a seller, and ε is the probability that the uninformed trader will actually trade. Nodes to the left of the dotted line occur only at the beginning of the trading day; nodes to the right are possible at each trading interval.



If an information event has occurred in period zero, in period one the informed trader acts upon this information, which only he or she possesses. If an informed trader has seen a high signal, he or she buys the stock if the current price is below the value given the signal; if the signal is low, he or she will sell if the quote is above the value given the signal. If there is no information event, an informed trader does not trade.

The uninformed trader's behavior is more complex. In most microstructure models, the presence of traders with better information dictates that an uninformed trader trading for speculative reasons would always be better off not trading at all. To avoid this no-trade equilibrium, at least some uninformed traders must transact for nonspeculative reasons such as liquidity needs or portfolio considerations. Since the uninformed traders do not have a particular reason to buy or sell, we make the customary assumption that half of them are buyers and the other half are sellers. We assume that when an uninformed trader checks the quote, the probability that he or she will trade is strictly positive.

Before trading starts each day, an information event occurs with probability α. There is thus a probability of $(1-\alpha)$ that there is no information, in which case only the uninformed traders are active in the market. If there is information, it is bad news (“low signal’) with probability δ. The complementary event is that there is good news (‘high signal”) with probability $(1-\delta)$. In the former case, informed traders sell, in the latter, they buy.

During the actual trading process, traders arrive according to a Poisson process. The market maker is always ready to trade during the day, and posts buy and sell quotes accordingly. Orders from informed traders arrive according to a Poisson process, with the intensity parameter μ. Orders from uninformed traders arrive with parameters ε_s and ε_b for sell and buy orders.

Once we have estimated the above parameters, as discussed in Section 3.2.2 below, we are in a position to calculate the Probability of Informed Trading, PIN, as

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_s + \varepsilon_b}, \tag{1}$$

where $\alpha\mu + \varepsilon_s + \varepsilon_b$ is the arrival rate of all orders, and $\alpha\mu$ is the arrival rate of informed orders. The probability of informed trading is thus the ratio of orders from informed traders to the total number of orders.

This model has been adapted to numerous uses, including a study of international analyst coverage in Easley, O’Hara, and Paperman (1998), stock splits in Easley, O’Hara, and Saar (2001), information risk and expected returns in Easley, Hvidkjaer, and O’Hara (2002), and the anonymity of the trading process in Grammig, Schiereck, and Theissen (2001).

3.2. Maximum likelihood estimation of the PIN model

In this section, we present the procedure for estimating the PIN model. We first need to calculate the input data for the estimation. These data are the numbers of buyer and seller initiated trades per trading day. We then use a numerical estimation procedure for the maximum likelihood estimation of the Probability of Informed Trading.

3.2.1. The Lee & Ready (1991) algorithm

Our starting point is ultra-high frequency transactions data. The first step is to determine the number of buyer initiated trades ('buys') and seller initiated trades ('sells') during each trading day. Since this information is not available in the data, it is necessary to employ one of several available methods of inferring the initiator of a trade. The three most common methods are the tick test, which uses changes in trade prices, the quote method, in which trade prices are compared to the prevailing bid and ask prices, and the Lee and Ready (1991) algorithm, which combines the tick and quote methods.[20]

The Lee and Ready (1991) algorithm has two steps in classifying trades. First, if the trading price is closer to the prevailing bid than the offer at the time of the trade, the trade is classified as a sell, and vice versa. Second, for trades that occur at the midpoint of the quote, a tick rule is used. According to this rule, if the last price change was positive (negative), the trade is a buy (sell). Following the findings of Bessembinder (2003), we do not use a time lag when matching quotes with trades. This means that for each trade we use the best bid and ask prices available at the time of the trade.[21]

Boehmer, Grammig, and Theissen (2007) analyze the effects of inaccurately classifying trades as buys and sells. In general, the results of the Lee and Ready algorithm are known to be somewhat inaccurate. Based on both simulation results and empirical evidence, the authors argue that PIN estimates are downward biased, when trade classification is inaccurate. However, Boehmer et al. (2007) also point out that this bias is most pronounced for illiquid stocks. Since our sample consists of the most liquid stocks in the French and German markets, we have no reason to believe that our estimation results are no more prone to suffer from any bias than existing literature in the field. In any case, the evidence presented by Boehmer et al. (2007) calls for some caution when interpreting our results.

3.2.2. The maximum likelihood model

The model is based on the assumption that order arrivals follow independent Poisson processes, as discussed in Section 3.1 above. These arrival intensities induce the following model for the total number of buy and sell trades during a single trading day:

---

[20] See Finucane (2000) and Bessembinder (2003) for a general discussion of the merits of the different methods, and a comprehensive discussion of their use in the literature; Ellis, Michaely, and O'Hara (2000) for tests on the three methods in the NASDAQ market; Chakrabarty et al. (2007) for a discussion on the merits of the methods in modern ECN trading systems; Savickas and Wilson (2003) discuss applications in options markets.

[21] See Bessembinder (2003) for a discussion of the appropriate lag length and alternative algorithms for signing trades. In their original paper, Lee and Ready (1991) recommend using quotes lagged by five seconds when assessing trades. This is done to allow for delays in trade reporting. Bessembinder (2003) also points out, that there are two separate issues at play: classifying trades into buys and sells, and assessing effective trade costs. In the former case, it is best to use a zero lag. In the latter case, it can be argued that a lag should be used, since traders are concerned about the possibility of adverse price movements between the time of the trade decision and trade execution.

$$L\left(\left(B,S\right)\middle|\theta\right)=\alpha\left(1-\delta\right)e^{-(\mu+\varepsilon_b+\varepsilon_b)T}\,\frac{\left(\mu+\varepsilon_b\right)^{B_i}\left(\varepsilon_s\right)^{S_i}}{B_i!S_i!}+$$
$$\alpha\delta e^{-(\mu+\varepsilon_b+\varepsilon_b)}\,\frac{\left(\mu+\varepsilon_s\right)^{S_i}\left(\varepsilon_b\right)^{B_i}}{B_i!S_i!}+(1-\alpha)e^{-(\varepsilon_b+\varepsilon_b)}\,\frac{\left(\varepsilon_b\right)^{B_i}\left(\varepsilon_s\right)^{S_i}}{B_i!S_i!}, \tag{2}$$

where B_i and S_i are the total number of buy and sell trades on the day in question, and $\theta=\left(\mu,\varepsilon_b,\varepsilon_s,\alpha,\delta\right)$ is the parameter vector. This likelihood function emanates from the distributions of the trade outcomes, weighted by the probabilities of the three different types of trading days. The trading day can contain positive news, with a probability of $\alpha(1 - \delta)$, negative news, with a probability of $\alpha\delta$, or no news at all, with a probability of $(1 - \alpha)$.

The likelihood function over multiple trading days is the product of the likelihood functions of a single day (Equation (2) above):

$$L\left(\theta\,|\,M\right)=\prod_{i=1}^{I}L\left(\theta\,|\,B_i,S_i\right), \tag{3}$$

Where B_i, and S_i are the numbers of buy and sell trades for day $i=1,\ldots,I$, and $M=((B_1,S_1),\ldots,(B_I,S_I))$ is the data set. Easley, Kiefer, and O'Hara (1997) test the assumption of independence among the trading days. They are not able to reject the independence assumption.

The direct computation of the maximum likelihood function may result in numerical overflow, since the values of *B!* and *S!* often become very large. We therefore perform the actual calculations using the following approximation, following Easley, Hvidkjaer, and O'Hara (2005). This approximation follows from the above models, after dropping a constant and rearranging terms.

$$L\left(\left(B_t,S_t\right)_{t=1}^{T}\middle|\theta\right)=\sum_{t=1}^{T}\left[-\varepsilon_b-\varepsilon_s+M_t\left(\ln x_b+\ln x_s\right)+B_t\ln\left(\mu+\varepsilon_b\right)+S_t\ln\left(\mu+\varepsilon_s\right)\right]$$
$$+\sum_{t=1}^{T}\ln\left[\alpha\left(1-\delta\right)e^{-\mu}x_x^{S_t-M_t}x_b^{-M_t}+\alpha\delta e^{-\mu}x_b^{B_t-M_t}x_s^{-M_t}+\left(1-\alpha\right)x_s^{S_t-M_t}x_b^{B_t-M_t}\right] \tag{4}$$

where $M_t=\min\left(B_t,S_t\right)+\max\left(B_t,S_t\right)/2$, $x_s=\dfrac{\varepsilon_s}{\mu+\varepsilon_s}$, and $x_b=\dfrac{\varepsilon_b}{\mu+\varepsilon_b}$. The advantages of this factorization are an increase in computing efficiency, and an avoidance of potential overflow problems, when the numbers of buys and sells are large[22].

[22] In the estimation of this likelihood model, we employ the *fminsearch* function in Matlab, which is an implementation of the simplex algorithm. Also, the choice of the initial values is important. We therefore randomize our initial values. Out of the estimated maximum likelihood values we then pick the ones that give the greatest value for the likelihood function.

4 EMPIRICAL ANALYSIS

In this section we analyze the effects of the change to an anonymous market. First we perform a univariate analysis of the changes in market quality, measured in bid-ask spreads, trading volume and volatility. We then seek to explain these findings in a multivariate framework. Subsequently, we examine informed trading in the form of the PIN measure. We study changes in informed trading and use our estimates of PIN as an explanatory variable for changes in market quality. We also study the upstairs market, in comparison to the continuous market, and conclude that no change has occurred there, as expected.

4.1. Univariate analysis

We start this paper by asking the normative question of whether an exchange should be transparent. Even if a categorical answer may not exist, we can still examine the effects of a change in transparency and anonymity, and see whether the market improves in quality. By a high market quality we mean a liquid market with a small cost of trading. In other words, we wish to observe a high trading volume, a low bid-ask spread, and low volatility. We measure the bid-ask spread in three ways: in euros, as a percentage spread, and as an effective spread. The bid-ask spread in euros is the difference between the ask price and the bid price. The percentage spread is defined as $\frac{ask - bid}{(ask + bid)/2} \times 100\%$, i.e. as the ratio of the bid-ask spread and the midquote.

We calculate the effective spread as follows:

$$Spread_{effective} = 2 * \left| P - m_{lag} \right|, \tag{5}$$

where P is the trade price, m_{lag} is the lagged midpoint of the best bid and ask prices. We use a lag of 0, 1, and 5 seconds. The motivation for estimating an effective spread is that it gives a better estimate of the real cost of trading than the quoted bid-ask spread. Originally proposed by Lee and Ready (1991), there is an extensive literature concerning the estimation and use of effective spreads. However, since a modern day electronic limit order market is very different from the NYSE in the early 1990s, the effective spread in our data differs from the quoted spread typically only when a large order executes at multiple prices, i.e. when the order depth at the best bid or ask price is insufficient.

Table 2 presents our univariate results of the measures of market quality, before and after the change to less transparent trading. The results show that there is no significant change in mean bid-ask spreads, whether they are measured as percentage spreads, or in euros. A paired t-test fails to reject the null hypothesis of equal means. However, the median spreads are significantly smaller, according to the Wilcoxon measure. The economic significance of the change, 2.8% in the case of the spread measured in euros, and 3.3% in the percentage spread, is arguably quite small.

Table 2 Univariate results, pre-change (non-anonymous) and post-change (anonymous) periods

This table presents descriptive statistics for pre and post-change periods (transparent and anonymous trading systems, respectively). *Bid-ask spreads* (in euros and in percent) are time-weighted bid-ask spreads during continuous trading hours. *Trading volume* is the average daily trading volume in thousands of euros, and in thousands of shares. *Volatility* is the standard deviation of asset returns on a 30-minute time interval. We report the values calculated for two periods: "pre" and "post", where the former refers to the period of transparent trading before the changes made in March 2006, and the latter to the period of anonymous trading after the changes. We have two sample periods of 62 days each.
We report test statistics (respectively the Wilcoxon z test and the paired Student t test) for the null hypothesis of equal median and mean values, respectively, for the pre and post periods. *, **, and *** indicate significance on a 10%, 5%, and 1% level, respectively. P-values are in parentheses.

	Mean			Median		
	Pre	Post	t-test	Pre	Post	Wilcoxon
Bid-ask spread, euros	0.0410	0.0408	0.545	0.0358	0.0348	0.0475**
Bid-ask spread, percent	0.284%	0.281%	0.45136	0.214%	0.207%	0.0475**
Number of trades	673.14	831.29	0.0005***	423.58	467.16	0.0002***
Trade price	17.65	18.19	0.3347	16.03	16.65	0.0168**
Trading volume, #shares	1,596.80	1,776.00	0.0787	338.57	443.48	0.0259**
Trading volume in euros	24,390	29,808	0.0860	6,803	9,177	0.0037***
Trade size	1,531	1,367	0.0196**	1,003	897	0.0054***
Volatility	0.322%	0.367%	2.59E-05***	0.282%	0.325%	4.53E-05***

N=35.

There are other significant changes in the market, however. There is no significant change in the average daily trading volume, measured in euros, or the number of shares traded. Traders seem to have shifted their behavior in that they have moved to trading in smaller lots. This can be seen in that the average daily number of trades increases, and the average trade size decreases. Both effects are significant, measured both in mean and median values. Last but not least, there is a significant rise in intraday volatility, measured as 30-minute price returns.

For market quality to have improved, bid-ask spreads would have to be lower, trading volume should be higher, and the volatility lower. The evidence is therefore ambiguous as to whether market quality has improved with this change to anonymous trading.

4.2. Multivariate results

Section 4.1 above presents the changes in several variables that describe market quality. The effects described could of course be the result of other factors than the change in the trading system itself. In order to be able to attribute changes in these variables to this transition to an anonymous market in March 2006, we estimate the following regression, using a slightly modified version of the regression used by Foucault, Moinas, and Theissen (2007):

$$s_{i,t} = \gamma_0 + \gamma_1 \log\left(V_{i,t}\right) + \gamma_2 P_{i,t} + \gamma_3 \sigma_{i,t} + \gamma_4 D_t^{post} + \varepsilon_{i,t}\,, \tag{6}$$

where s is the spread of stock i at time t, V is the trading volume of the stock, P the stock price, σ is midquote return volatility. D^{post} is a dummy variable that takes on a value of 0 for the period before the changes (the transparent market), and a value of 1 after the changes (the opaque market)[23].

We first calculate all variables for each stock and each day. We then aggregate the variables over the duration of the two sample periods, producing two observations per stock, one for each sample period. We correct for potential autocorrelation by using Newey-West standard errors in calculating t-statistics[24]. We perform the regression separately for our two sample periods, the pre and post change periods.

Our dependent variable is the quoted bid-ask spread, measured in two ways, in euros and as a percentage. The percentage spread is the difference of the bid and ask prices as a percentage of the stock price; the spread in euros is simply the difference between the two prices.

Table 3 presents our results from the regression model given in Equation (6). The expected result, based on Foucault et al. (2007), and Comerton-Forde et al. (2005) would be a negative sign for the trading period dummy. However, even if a t-test attributes some amount of significance to the trading period dummy, in economic terms it is very close to zero.

All other explanatory variables have the expected signs. Volatility has a positive estimated coefficient, as expected according to the Foucault, Moinas, and Theissen (2007) model. The reasoning is that uninformed limit order traders are afraid of being picked off by informed traders, and protect themselves by a greater bid-ask spread in the case of greater volatility. Trading volume has a negative sign, meaning that the greater the trading volume, the tighter is the bid-ask spread. This result is expected and in line with previous research. Price has a negative sign when using the percentage spread as the dependent variable, and a positive sign for the bid-ask spread in euros. This is also to be expected, since the greater the stock price, the greater the bid-ask spread, in euros, for the same relative (percentage) spread. Also, percentage spread can be expected to be a negative function of stock price, since the minimum tick size of 1 cent poses a natural lower limit for the spread.

As a robustness check, we perform all the above analyses using effective spreads as well as quoted spreads. As a further robustness check, we perform a company fixed-effects analysis. A fixed stock effect may be a source of correlation in the analysis. The results are similar to the results presented in Table 3, although the significance of the dummy variable is reduced. Both these robustness checks provide largely similar results to the base case presented in the paper. We therefore omit these results; they are available on request.

[23] The only difference to the Foucault, Moinas, and Theissen (2007) regression is that they also include tick size as a variable. However, since the Helsinki market has a uniform tick size of EUR 0.01 for all stocks and all prices, we exclude this variable.
[24] Autocorrelation does not pose problems in practice, since we only use two observations per stock in the regression.

Table 3 Multivariate regression results

This table presents the coefficient estimates of the regression described in Section 4.2. The regression model is

$$s_{i,t} = \gamma_0 + \gamma_1 \log\left(V_{i,t}\right) + \gamma_2 P_{i,t} + \gamma_3 \sigma_{i,t} + \gamma_4 D_t^{post} + \varepsilon_{i,t}\ .$$

The dependent variable is the spread, measured either in euros in or as a percentage. The regression is a pooled regression over all stocks. The explanatory variables are *log(V),* the logarithm of average trading volume in euros, *P* is the average price, σ is the standard deviation of 30-minute logarithmic returns, and D^{post} takes the value 0 for the pre-change (transparent) market, and the value 1 for the post-change (anonymous) market. All variables are first calculated per stock and per trading day, and then averaged over the pre and post periods for each stock, resulting in two observations per stock. The fixed effects regressions include stock specific dummy variables (omitted in this table). *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. Figures in parenthesis are t-statistics. The number of observations is 92.

	Pooled regression		Pooled regression, fixed effects	
	Percentage spread	Spread in euros	Percentage spread	Spread in euros
Constant	0.008***	0.085***	0.003***	0.017***
	(56.66)	(32.89)	(13.98)	(3.14)
Volatility	0.091***	1.506***	0.064***	1.436***
	(13.73)	(12.49)	(10.62)	(9.98)
Log(volume)	-0.000***	-0.005***	-0.000***	-0.002***
	(-43.08)	(-34.25)	(-8.24)	(-8.19)
Price	-0.000***	0.002***	-0.000	0.002***
	(-22.06)	(65.92)	(-0.68)	(13.50)
Pre/post dummy	0.000***	0.001**	0.000*	-0.000
	(3.813)	(2.054)	(1.759)	(-0.256)
Adj. R^2	0.550	0.780	0.854	0.877

4.3. The upstairs market

The upstairs market[25] consists of prearranged trades, which are negotiated in person between brokers. They can occur "in-house", where the buyer and seller are both represented by the same brokerage, or between brokerages. The two main types of upstairs trades on the OMX Helsinki exchange are block trades and contract trades. A block trade is a large trade, defined as a certain minimum percentage of the free float of a stock, and a minimum size in value. These trades can be executed even outside the prevailing bid-ask spread, as long as they are within the limits of the lowest and highest prices for the day. Contract trades are trades that are not automatically matched in the limit order book, but previously agreed upon, either within the same brokerage, or between brokerages.

[25] There is a widely used distinction between the upstairs market, described above, and the downstairs market, which is the automatically matched continuous trading in the electronic limit order book. Smith et al. (2001) find that the upstairs market of the Toronto Stock Exchange enables large non-information based trades to execute at a lower cost. Bessembinder and Venkataraman (2003) come to a similar conclusion for the Paris Bourse.

Studying the upstairs market in the context of a change in transparency in the downstairs market is interesting, since the changes in pre-trade transparency do not affect the upstairs market. One hypothesis is that if there are traders who prefer a transparent trading system, these traders would then switch to the upstairs market when the downstairs market becomes less transparent[26].

The nature of the trading process is in this way the almost complete opposite to the recent proliferation of the so-called dark pools, where buyers and sellers are matched with no knowledge of the counterparties, and with no price discovery. For this reason it is interesting to find that there are no changes in the upstairs market. The internalization rates, i.e. the percentage of trades where a brokerage is able to find counterparty for a trade among its own clients, do not change. The scale of internalization for upstairs trades is similar to the findings of Booth et al. (2000), at around 97.5%. Table 4 presents our results. We find that the internalization rate is consistently very high, both for block trades and for contract trades. A t-test of the internalization rates does not reject the null hypothesis of no change in the internalization rates.

Since there is no evidence of a change in upstairs trading, several different conclusions may be drawn. Either the population of potential switchers is small, or they are unable to change. The first of these hypotheses means that if there are traders who prefer a transparent trading system, their number is too small to be detected in the present analysis. Another possibility is that there are such traders, but they are unable to switch to the upstairs market. Small individual investors could be regarded as such investors, since they are usually only offered the possibility to submit limit orders to the downstairs market.

[26] I wish to thank Christophe Majois for formulating this hypothesis.

Table 4 The upstairs market and internalization rates

This table presents internalization rates for the main non-continuous trade types, block trades and contract trades. Internal trades have the same buyer and seller. Internalization rate is the percentage of internal trades out of all trades. The numbers for block and contract trades are total numbers of trades in the pre- and post – change periods.
The t-test is for the equality of the mean for the internalization rate of the pre-change period and the post-change period. We do no reject the null hypothesis of equal means.

	Pre change (transparent market)						Post change (anonymous market)					
	Block trades		Contract trades		Internalization rate		Block trades		Contract trades		Internalization	
Stock	internal	external	internal	external	Block trades	Contract trades	internal	external	internal	external	Block trades	Contract trades
Amer Sports Oyj	12	0	272	3	100.0%	98.9%	2	0	114	2	100.0%	98.3%
Cargotec Oyj	6	0	163	0	100.0%	100.0%	13	0	214	1	100.0%	99.5%
Elisa Oyj	26	1	425	5	96.3%	98.8%	13	0	400	3	100.0%	99.3%
Fortum Oyj	81	0	730	3	100.0%	99.6%	139	1	1647	3	99.3%	99.8%
Huhtamäki Oyj	2	0	171	1	100.0%	99.4%	6	1	148	3	85.7%	98.0%
Kesko Oyj B	5	0	300	1	100.0%	99.7%	11	0	180	3	100.0%	98.4%
Kone Oyj B	14	0	182	2	100.0%	98.9%	27	0	300	0	100.0%	100.0%
Metso Oyj	47	0	404	0	100.0%	100.0%	35	0	496	7	100.0%	98.6%
M-real Oyj B	5	0	199	1	100.0%	99.5%	34	0	303	0	100.0%	100.0%
Neste Oil Oyj	66	0	517	1	100.0%	99.8%	24	1	410	0	96.0%	100.0%
Nokia Oyj	541	16	2953	20	97.1%	99.3%	513	17	2663	9	96.8%	99.7%
Nokian Renkaat Oyj	26	1	705	5	96.3%	99.3%	12	4	380	9	75.0%	97.7%
Nordea Bank AB (publ)FDR	7	0	139	9	100.0%	93.9%	20	0	123	1	100.0%	99.2%
OKO Bank Oyj A	11	0	280	4	100.0%	98.6%	2	0	97	1	100.0%	99.0%
Outokumpu Oyj	24	0	441	7	100.0%	98.4%	22	0	437	5	100.0%	98.9%
Rautaruukki Oyj K	24	1	414	1	96.0%	99.8%	24	2	266	3	92.3%	98.9%
Sampo Oyj A	74	1	781	1	98.7%	99.9%	73	9	699	4	89.0%	99.4%
Stora Enso Oyj R	84	2	627	2	97.7%	99.7%	99	1	587	0	99.0%	100.0%
TeliaSonera AB	11	0	127	1	100.0%	99.2%	20	1	110	0	95.2%	100.0%
Tietoenator Oyj	35	1	341	2	97.2%	99.4%	30	1	423	4	96.8%	99.1%
UPM-Kymmene Oyj	124	2	575	7	98.4%	98.8%	116	5	590	1	95.9%	99.8%
Wärtsilä Oyj Abp B	17	0	322	0	100.0%	100.0%	12	0	284	2	100.0%	99.3%
YIT-Yhtymä Oyj	15	1	241	5	93.8%	98.0%	15	0	497	1	100.0%	99.8%
Average	54.7	1.1	491.7	3.5	98.8%	99.1%	54.9	1.9	494.3	2.7	96.6%	99.2%
t test											1.796	0.531
											(0.086)	(0.601)

4.4. The Probability of Informed Trading

According to Rindi (2008), given the amount of informed traders, a more transparent market is more liquid. This is in line with most earlier theoretical work. However, with endogenous information acquisition, the analysis is more complex. According to Rindi (2008), limit orders accommodate liquidity shocks caused by liquidity traders. Informed limit order traders are of course in a better position to do this, since they do not face adverse selection costs. Uninformed traders are reluctant to accommodate large market orders, in fear of facing an informed trader. Greater transparency reduces the incentive to acquire information, and thus diminishes the number of liquidity traders.

In order to test the effects of changes in market transparency for informed trading, we estimate the Probability of Informed Trading (PIN) model of Equation (2) above. We perform the estimation for both the pre and post-change periods.

Table 5 presents the results[27]. For most stocks, the estimated PIN value falls between approximately 10% and 20%. There is no clear change with the advent of anonymous trading, contrary to our expectations. The average probability of informed trading coefficient is slightly lower after the change, but not significantly so. Also, there are as many stocks for which PIN increases, as there are stocks with a decrease in PIN.

[27] The maximum likelihood model converges for all stocks and both sample periods, with one exception, Nokian Renkaat Oyj, NRE1V, in the post-change period.

Table 5 The Probability of Informed Trading (PIN)

This table presents the results of the PIN (Probability of Informed Trading) analysis. PIN is defined as

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_s + \varepsilon_b},$$

where α is the probability of an information event, μ is the Poisson parameter for the arrival of informed trades, and ε_s and ε_b are the Poisson densities for the arrival of uninformed sell and buy trades, respectively. *Std errors* (in parenthesis) are the standard error statistics of the PIN estimates. In the table below, δ is the probability of negative news. There is only one value for the ε variable since we make the assumption that uninformed traders are equally likely to be buyers and sellers.
The maximum likelihood model does not converge in a few cases. These values are excluded from our analysis of changes in PIN. The last column indicates whether the estimated PIN has increased or decreased from the first period. We report the average, median, minimum, and maximum values, as well as standard deviation statistics (in parenthesis) for all variables.
The last row presents a paired two-sided t-test for the PIN estimates of the pre-change period and the post-change period. The value in parenthesis is the t-probability value of the t-test.

	Pre-change (transparent) market					Post-change (anonymous) market					
Stock ticker	Alpha	delta	mu	epsilon	PIN	alpha	delta	mu	epsilon	PIN	Change in PIN
ALN1V	0.327	0.867	45.48	27.15	0.215	0.486	0.333	20.60	20.33	0.198	-
	(0.062)	(0.080)	(2.23)	(0.55)	(0.032)	(0.089)	(0.097)	(1.59)	(0.62)	(0.029)	
AMEAS	0.434	0.485	136.53	142.33	0.172	0.450	0.212	142.96	136.55	0.191	+
	(0.065)	(0.098)	(4.03)	(1.36)	(0.021)	(0.063)	(0.078)	(3.45)	(1.22)	(0.022)	
CGCBV	0.145	0.778	242.56	147.59	0.107	0.387	0.458	206.17	204.62	0.163	+
	(0.045)	(0.139)	(6.76)	(1.13)	(0.029)	(0.062)	(0.102)	(4.38)	(1.43)	(0.022)	
ELQAV	0.194	0.586	217.23	84.04	0.201	0.208	0.500	206.05	92.35	0.188	-
	(0.061)	(0.191)	(6.66)	(0.85)	(0.051)	(0.052)	(0.141)	(6.51)	(0.99)	(0.038)	
ELI1V	0.501	0.386	254.28	369.70	0.147	0.564	0.398	244.20	377.75	0.154	+
	(0.064)	(0.088)	(4.95)	(2.01)	(0.016)	(0.063)	(0.083)	(4.71)	(2.07)	(0.015)	
FSC1V	0.212	0.718	126.67	77.24	0.148	0.250	0.585	121.02	65.48	0.188	+
	(0.048)	(0.109)	(4.87)	(1.00)	(0.028)	(0.060)	(0.143)	(4.14)	(0.77)	(0.037)	
FIA1S	0.422	0.343	82.60	60.41	0.224	0.403	0.360	86.03	46.27	0.273	+
	(0.063)	(0.094)	(2.57)	(0.80)	(0.027)	(0.062)	(0.096)	(2.41)	(0.69)	(0.031)	
FLG1S	0.369	0.557	32.63	19.40	0.237	0.444	0.277	42.63	27.78	0.254	+
	(0.066)	(0.108)	(1.85)	(0.48)	(0.033)	(0.067)	(0.087)	(1.92)	(0.60)	(0.029)	
FUM1V	0.569	0.419	382.15	639.39	0.145	1.000	0.002	213.07	246.62	0.302	+
	(0.064)	(0.085)	(6.42)	(2.94)	(0.014)	.	.	.	.	.	

Stock ticker	Pre-change (transparent) market					Post-change (anonymous) market					Change in PIN
	Alpha	delta	mu	epsilon	PIN	alpha	delta	mu	epsilon	PIN	
HUH1V	0.441	0.595	100.77	111.39	0.166	0.440	0.184	140.40	159.55	0.162	-
	(0.064)	(0.097)	(3.13)	(1.11)	(0.021)	(0.064)	(0.074)	(3.84)	(1.36)	(0.020)	
KNEBV	0.416	0.457	181.63	259.91	0.127	0.306	0.527	258.79	344.04	0.103	-
	(0.063)	(0.099)	(4.66)	(1.70)	(0.017)	(0.059)	(0.115)	(5.98)	(1.82)	(0.018)	
KGH1V	0.315	0.564	100.38	91.61	0.147	0.500	0.565	108.66	74.34	0.268	+
	(0.066)	(0.117)	(5.36)	(1.23)	(0.025)	(0.064)	(0.134)	(4.12	(0.84)	(0.026)	
KRA1V	0.211	0.125	168.07	114.21	0.135	0.464	0.429	115.91	122.08	0.180	+
	(0.052)	(0.044)	(5.48)	(1.05)	(0.029)	(0.067)	(0.093)	(3.75)	(1.34)	(0.021)	
KESBV	0.607	0.709	157.39	177.38	0.212	0.372	0.482	247.22	268.66	0.146	-
	(0.063)	(0.074)	(3.41)	(1.52)	(0.018)	(0.062)	(0.105)	(5.22)	(1.66)	(0.021)	
MRLBV	0.119	0.000	242.92	168.37	0.079	0.258	0.188	296.95	279.24	0.121	+
	(0.042)	.	(7.76)	(1.23)	(0.026)	(0.056)	(0.098)	(6.34)	(1.62)	(0.023)	
MEO1V	0.274	0.823	436.15	415.85	0.126	0.419	0.308	414.45	590.20	0.128	+
	(0.057)	(0.093)	(7.39)	(1.97)	(0.023)	(0.063)	(0.091)	(6.74)	(2.47)	(0.017)	
NES1V	0.424	0.600	378.26	482.69	0.142	0.274	0.710	470.48	577.71	0.100	-
	(0.064)	(0.098)	(6.34)	(2.29)	(0.019)	(0.054)	(0.077)	(7.63)	(2.36)	(0.018)	
NOK1V	0.431	0.012	1453.06	2545.42	0.110	0.309	0.021	1672.43	2904.86	0.082	-
	(0.065)	(0.009)	(13.71)	(5.29)	(0.015)	(0.062)	(0.015)	(17.33)	(5.59)	(0.015)	
NRE1V	0.328	0.251	326.09	425.27	0.112						-
	(0.060)	(0.097)	(6.59)	(2.06)	(0.018)						
NDA1V	0.210	0.500	182.30	133.22	0.125	0.551	0.119	155.77	150.91	0.221	+
	(0.052)	(0.139)	(5.27)	(1.11)	(0.027)	(0.064)	(0.056)	(3.46)	(1.39)	(0.021)	
OKOAS	0.306	0.684	230.40	196.90	0.152	0.353	0.518	148.63	163.94	0.138	-
	(0.059)	(0.107)	(4.99)	(1.38)	(0.025)	(0.061)	(0.110)	(4.18)	(1.30)	(0.021)	
OUT1V	0.377	0.435	260.07	274.19	0.152	0.355	0.409	391.03	451.45	0.133	-
	(0.062)	(0.103)	(5.10)	(1.66)	(0.021)	(0.061)	(0.105)	(6.65)	(2.12)	(0.020)	
POS1V	0.413	0.447	143.58	131.47	0.184	0.295	0.278	179.73	122.90	0.178	-
	(0.065)	(0.100)	(3.92)	(1.26)	(0.024)	(0.058)	(0.106)	(4.29)	(1.09)	(0.029)	
RAIVV	0.145	0.332	127.37	47.78	0.162	0.496	0.125	48.00	38.97	0.234	+
	(0.045)	(0.158)	(4.49)	(0.65)	(0.042)	(0.067)	(0.046)	(2.01)	(0.72)	(0.025)	
RMR1V	0.334	0.654	61.66	40.48	0.203	0.350	0.461	82.20	55.74	0.205	+

	(0.062)	(0.107)	(2.77)	(0.69)	(0.030)	(0.061)	(0.108)	(2.83)	(0.78)	(0.029)	
RTRKS	0.371	0.477	269.73	376.12	0.118	0.492	0.200	343.61	541.91	0.135	+
	(0.061)	(0.105)	(5.70)	(1.94)	(0.017)	(0.064)	(0.073)	(5.96)	(2.43)	(0.015)	
SAMAS	0.339	0.237	410.97	524.32	0.117	0.516	0.313	415.55	684.45	0.136	+
	(0.060)	(0.082)	(7.07)	(2.26)	(0.018)	(0.063)	(0.082)	(6.47)	(2.73)	(0.015)	
SDA1V	0.505	0.565	44.10	35.67	0.238	0.509	0.500	46.91	45.30	0.208	-
	(0.067)	(0.092)	(1.87)	(0.69)	(0.025)	(0.067)	(0.092)	(2.00)	(0.77)	(0.023)	
STCBV	0.455	0.649	84.34	82.65	0.189	0.502	0.389	78.07	90.23	0.179	-
	(0.065)	(0.090)	(2.83)	(0.99)	(0.022)	(0.077)	(0.092)	(3.68)	(1.43)	(0.022)	
STERV	0.356	0.009	443.80	473.81	0.143	0.412	0.404	389.63	697.19	0.103	-
	(0.062)	(0.004)	(7.12)	(2.23)	(0.022)	(0.063)	(0.099)	(7.52)	(2.77)	(0.014)	
TLS1V	0.695	0.063	109.37	110.21	0.257	0.258	0.250	175.30	156.33	0.127	-
	(0.063)	(0.039)	(2.84)	(1.42)	(0.018)	(0.056)	(0.153)	(4.80)	(1.21)	(0.024)	
TIE1V	0.484	0.338	257.19	111.24	0.359	0.450	0.305	327.39	513.13	0.126	-
	(0.063)	(0.080)	(3.70)	(1.09)	(0.031)	(0.070)	(0.096)	(6.72)	(2.60)	(0.017)	
UPM1V	0.271	0.001	500.18	619.45	0.099	0.532	0.075	451.20	839.44	0.125	+
	(0.058)	(0.001)	(8.78)	(2.47)	(0.019)	(0.063)	(0.030)	(6.95)	(3.04)	(0.013)	
UNR1V	0.250	0.047	72.68	70.84	0.114	0.025	0.675	166.50	109.64	0.019	-
	(0.049)	(0.017)	(2.76)	(0.86)	(0.020)	(0.006)	(0.113)	(4.81)	(1.10)	(0.004)	
WRTBV	0.507	0.692	197.60	217.48	0.187	0.451	0.250	223.66	297.55	0.145	-
	(0.064)	(0.085)	(4.04)	(1.58)	(0.020)	(0.063)	(0.079)	(4.74)	(1.79)	(0.018)	
YTY1V	0.499	0.711	164.18	199.23	0.171	0.373	0.393	271.71	373.84	0.119	-
	(0.064)	(0.082)	(3.75)	(1.48)	(0.018)	(0.062)	(0.103)	(6.20)	(2.02)	(0.018)	
Mean					0.166					0.162	
Median					0.150					0.150	
Minimum					0.079					0.019	
Maximum					0.359					0.302	
Standard deviation					0.056					0.059	
t-test										0.481	
										(0.713)	

4.5. Informed trading as an explanation for changes in bid-ask spreads

This regression analysis is similar in many ways to the analysis of Section 4.2 above. The dependent variable is the bid-ask spread, measured both in euros and as a percentage. The explanatory variables are the same as above, with the addition of the Probability of Informed Trading, PIN. In other words, our explanatory variables are the logarithmic trading volume, the average trade price, 30-minute return volatility, a pre-post change dummy, and the PIN.

There are also some notable differences with the analysis presented in Section 4.B above. Instead of a pooled regression with values for all dates and all stocks in the sample, this is a cross-sectional regression over all stocks. We calculate the average value of each variable, separately for the pre- and post-change periods (transparent and non-transparent trading systems, respectively). This results in two estimated values for each stock. The reason for using this smaller number of observations is the impossibility of obtaining daily estimates of the Probability of Informed Trading. As we pointed out in Section 2 above, a minimum of 60 trading days is generally regarded as necessary for a reliable estimation of the PIN measure. We therefore have two values for the PIN for each stock, one for the pre-change period, and another for the post-change period.

We estimate the following model:

$$s_{i,t} = \gamma_0 + \gamma_1 \log\left(V_{i,t}\right) + \gamma_2 P_{i,t} + \gamma_3 \sigma_{i,t} + \gamma_4 D_t^{post} + \gamma_5 PIN_t + \varepsilon_{i,t}\,, \qquad (7)$$

where s is the spread of stock i at time t, V is the trading volume of the stock, P the average stock price, σ is midquote return volatility; D^{post} is a dummy variable that takes on a value of 0 for the period before the changes (the transparent market), and a value of 1 after the changes (the opaque market); *PIN* is the probability of informed trading. There are only two time periods: the pre-change and the post-change periods.

Table 6 presents the results of the regression of Equation (7). Trading volume and stock price are significant explanatory variables in all regressions. Both also have the expected sign, similarly to the earlier analysis of Section 4.B: the greater the trading volume and the greater the price, the smaller is the bid-ask spread. For the regression in Panel B, where the bid-ask spread is measured in euros, the stock price has the opposite sign, which is also expected. The PIN variable is not significant in any of these regressions, albeit it does have the expected sign, being positive. However, the column (5) presents results from a regression with the PIN variable and the dummy variable only. Here the PIN is both significant and positive. However, since this regression does not control for any of the other variables, we do wish to place very much emphasis on this result. In the last column in Table 6 we use an interaction term PIN*dummy. This does not change the results significantly.[28]

[28] We also run the regression using first differences. In this case the period dummy is of course excluded, since its effect is captured by the intercept. These results snow no significant explanatory power for the regression, exhibiting a negative R^2.

Table 6 Cross-sectional regression, bid-ask spread as the dependent variable

This table presents the results of our analysis of informed trading as an explanatory variable of market quality. *Percentage spread* is the average percentage quoted bid-ask spread, and the *Spread in euros* is the average difference between the bid and the ask prices.
The explanatory variables are the following. *Log volume* is the logarithm of average trading volume in euros. *Price* is the average price. *Volatility* is the standard deviation of 30-minute logarithmic returns. *PIN* is the Probability of Informed Trading. *Dummy* takes the value 0 for the transparent market, and 1 for the anonymous market. All variables are stock-specific averages for the pre-change and the post-change periods. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Panel A: Percentage spread as the dependent variable					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.016***	0.018***	0.018***	0.017***	0.0007	0.018***
	9.321	13.256	13.113	9.869	1.029	(11.946)
Log volume	-0.001***	-0.001***	-0.001***	-0.001***		-0.001***
	-8.921	-10.461	-10.386	-9.037		(-10.185)
Price	-4.5E-05***	-4.0E-05***	-4.0E-05***	-3.0E-05***		-0.000***
	-3.277	-2.931	-2.914	-2.904		(-3.327)
Volatility	0.181*					0.191*
	1.752					(1.854)
PIN	0.003			0.003	0.013***	
	1.101			1.253	3.820	
Dummy	0.000		0.000		3.4E-05	
	-0.086		0.216		0.086	
PIN*Dummy						-0.000
						(-0.115)
R^2	0.681	0.672	0.667	0.675	0.158	0.680

	Panel B: Spread in euros as the dependent variable					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.198***	0.210***	0.211***	0.191***	0.019*	0.217***
	6.544	9.015	8.936	6.543	1.808	(8.614)
Log volume	-0.012***	-0.013***	-0.013***	-0.012***		-0.013***
	-7.274	-8.482	-8.415	-7.271		(-8.383)
Price	0.002***	0.002***	0.002***	0.002***		0.002***
	8.080	8.108	8.048	8.150		(7.996)
Volatility	-1.743					-1.601
	-0.988					(-0.908)
PIN	0.048			0.044	0.135**	
	1.174			1.098	2.30	
Dummy	0.000		-0.001		0.0003	
	0.085		-0.146		0.047	
						0.004
						(0.182)
R^2	0.617	0.622	0.616	0.623	0.049	0.615

N=68.

The period dummy is not significant in any of the regressions. We conclude that our analysis does not find evidence of the significance of the change into a less pre-trade transparent trading system. This result is at odds with e.g. Foucault, Moinas and Theissen (2007). Two explanations immediately suggest themselves. Either the change in transparency has not had an impact on market quality, measured by bid-ask spreads, or our regression specification and the other explanatory variables fail to capture other changes occurring in the market at the same time.

As a robustness test, we also use a smaller sample of the largest and most liquid stocks. We select the 20 most liquid companies, excluding Nokia, which is in a class of its own, measured in trading volume and market capitalization. We run the same regression model as above. The results are substantially the same, and are available upon request.

5 CONCLUSIONS

In this paper, we study a change from a fully pre-trade transparent limit order book to a less transparent one, at the OMX Helsinki Stock Exchange (the former HEX). The exchange changed their trading system in March 2006, by eliminating the display of buyer and seller identities in the electronic limit order book. However, the market remained fully post-trade transparent, in that the identities of the buying and selling counterparties are known to all market participants immediately after a trade takes place.

Our main results are the following. We use the Probability of Informed Trading (PIN) as a proxy for the participation rate of informed traders. We use several trading related variables, such as trading volume, stock price, intraday return volatility, the estimated values of PIN, and a period dummy variable. In a regression with stock-specific averages for these explanatory variables, for the pre-change and post-change periods, we do not find evidence of any significant explanatory power for the PIN variable.

We also study the broader question of the effects of a change from a pre-trade transparent to an anonymous electronic limit order book market. The unresolved question, posed by many market regulators and exchange officials the world over, is whether transparency improves market quality. It is commonly agreed that market quality improves with a decrease in the bid-ask spread, an increase in trading volume, and a decrease in return volatility.

We find that the effects of a switch to a less pre-trade transparent limit order book are not unambiguously positive. Our analysis shows no significant change in average bid-ask spreads, and a significant increase in intraday volatility. However, trading volume increases after the changes, mitigating the two negative effects. We perform a cross-sectional analysis, using trading volume, stock price, and intraday volatility in addition to a period dummy variable as explanatory variables for the bid-ask spread. This analysis fails to provide evidence of the change in pre-trade transparency as a significant factor in explaining changes in market quality after the switch.

As a comparison with the changes in transparency in the electronic limit order book market, we also study trading in the upstairs market. The changes in the trading system do not affect this segment, where pre-trade anonymity does not exist. We find that there indeed is no change in the number of block trades and contract trades. Neither does the internalization rate of trades change.

REFERENCES

Admati, Anat R., and Paul Pfleiderer, 1991, Sunshine Trading and Financial Market Equilibrium, *Review of Financial Studies* 4, 443–481.

Baruch, Shmuel, 2005, Who Benefits from an Open Limit-Order Book?, *Journal of Business* 78, 1267–1306.

Bessembinder, Hendrik, 2003, Issues in assessing trade execution costs, *Journal of Financial Markets* 6, 233–257.

Bessembinder, Hendrik, and Kumar Venkataraman, 2004, Does an electronic stock exchange need an upstairs market?, *Journal of Financial Economics* 73, 3–36.

Bloomfield, Robert, and Maureen O'Hara, 1999, Market transparency: Who wins and who loses?, *Review of Financial Studies* 12, 5–35.

Bloomfield, Robert, and Maureen O'Hara, 2000, Can transparent markets survive?, *Journal of Financial Economics* 55, 425–459.

Bloomfield, Robert, Maureen O'Hara, and Gideon Saar, 2005, The 'Make' or 'Take' Decision in an Electronic Market: Evidence on the Evolution of Liquidity, *Journal of Financial Economics* 75, 165–200.

Boehmer, Ekkehart, Joachim Grammig, and Erik Theissen, 2007, Estimating the probability of informed trading – does trade misclassification matter?, *Journal of Financial Markets* 10, 26–47.

Boehmer, Ekkehart, Gideon Saar, and Lei Yu, 2005, Lifting the Veil: An Analysis of Pre-trade Transparency at the NYSE, *Journal of Finance* 60, 783–815.

Booth, G. Geoffrey, Juha-Pekka Kallunki, Ji-Chai Lin, and Teppo Martikainen, Internalization and stock price clustering: Finnish evidence, *Journal of International Money and Finance* 19, 737–751.

Chakrabarty, Bidisha, Bingguang Li, Vanthuan Nguyen, and Robert A. Van Ness, 2007, Trade classification algorithms for electronic communications network trades, *Journal of Banking and Finance* 31, 3806–3821.

Chan, Louis K.C., and Josef Lakonishok, 1995, The behavior of stock prices around institutional trades, *Journal of Finance* 50, 1147–1174.

Chowdhry, Bhagwan, and Vikram Nanda, 1991, Multimarket trading and market liquidity, *Review of Financial Studies* 4, 483–511.

Comerton-Forde, Carole, Alex Frino, and Vito Mollica, 2005, The impact of limit order anonymity on liquidity: Evidence from Paris, Tokyo and Korea, *Journal of Economics and Business* 57, 528–540.

Copeland, Thomas E., and Dan Galai, 1983, Information Effects and the Bid-Ask Spreads, *Journal of Finance* 38, 1457–69.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2002, Is information risk a determinant of asset returns?, *Journal of Finance* 57, 2185–2221.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2005, Factoring information into returns, (July 27, 2005), EFA 2004 Maastricht Meetings Paper No. 4118, http://ssrn.com/abstract=556079.

Easley, David, Nicholas M. Kiefer, Maureen O'Hara and Joseph B. Paperman, 1996. Liquidity, Information, and Infrequently Traded Stocks, *Journal of Finance* 51, 1405–1436.

Easley, David, Nicholas M. Kiefer, and Maureen O'Hara, 1996, Cream-Skimming or Profit-Sharing? The Curious Role of Purchased Order Flow, *Journal of Finance* 51, 811–833.

Easley, David, Nicholas M. Kiefer, and Maureen O'Hara, 1997, One Day in the Life of a very Common Stock, *Review of Financial Studies* 10, 805–835.

Easley, David and Maureen O'Hara, 1992, Time and the Process of Security Price Adjustment, *Journal of Finance* 57, 577–605.

Easley, David, Maureen O'Hara, and Joseph Paperman, 1998, Financial analysts and information-based trade, *Journal of Financial Markets* 1, 175–201.

Easley, David, Maureen O'Hara, and Gideon Saar, 2001, How Stock Splits Affect Trading: A Microstructure Approach, *Journal of Financial and Quantitative Analysis* 36, 25–51.

Ellis, Katrina, Roni Michaely, and Maureen O'Hara, 2000, The accuracy of trade classification rules: Evidence from Nasdaq, *Journal of Financial and Quantitative Analysis* 35, 529–551.

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

Fishman, Michael J. and Kathleen M. Hagerty, 1995, The Mandatory Disclosure of Trades and Market Liquidity, *Review of Financial Studies* 8, 637–676

Flood, Mark D., Ronald Huisman, Kees G. Koedijk, Ronald J. Mahieu, 1999, Quote disclosure and price discovery in multiple-dealer financial markets, *Review of Financial Studies* 12, 37–59.

Foucault, Thierry, Sophie Moinas, and Erik Theissen, 2007, Does anonymity matter in electronic limit order markets?, *Review of Financial Studies* 20, 1707–1747.

Garfinkel, Jon A., and M. Nimalendran, 2003, Market Structure and Trader Anonymity: An Analysis of Insider Trading, *Journal of Financial and Quantitative Analysis* 8, 591–610.

Glosten, Lawrence R., and Paul R. Milgrom, 1985, *Journal of Financial Economics* 14, 71–100.

Grammig, Joachim, Dirk Schiereck, Erik Theissen, 2001, Knowing me, knowing you: Trader anonymity and informed trading in parallel markets, *Journal of Financial Markets* 4, 385–412.

Harris, Jeffrey H. and Paul H. Schultz, 1997, The importance of firm quotes and rapid executions: Evidence from the January 1994 SOES rules change, *Journal of Financial Economics* 45, 135–166.

Hendershott, Terrence, and Charles M. Jones, 2005, Island goes dark: Transparency, fragmentation, and regulation, *Review of Financial Studies* 18, 743–793.

Huang, Roger D. and Hans R. Stoll, 1996, Dealer versus auction markets: A paired comparison of execution costs on NASDAQ and the NYSE, *Journal of Financial Economics* 41, 313–357.

de Jong, Frank, Theo Nijman and Ailsa Roell, 1996, Price Effects of Trading and Components of the Bid-ask Spread on the Paris Bourse, *Journal of Empirical Finance* 3, 193–213.

Jones, Charles M., Gautam Kaul, and Marc L. Lipson, 1994, Transactions, Volume and Volatility, *Review of Financial Studies* 7, 631–51.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1336.

Lee, Charles M. C., and Mark J. Ready, 1991, Inferring Trade Direction from Intraday Data, *Journal of Finance* 46, 733–746.

Madhavan, Ananth, 1995, Consolidation, Fragmentation, and the Disclosure of Trading Information, *Review of Financial Studies* 8, 579–603.

Madhavan, Ananth, 1996, Security Prices and Market Transparency, *Journal of Financial Intermediation* 5, 255–283.

Madhavan, Ananth, and Minder Cheng, 1997, In Search of Liquidity: Block Trades in the Upstairs and Downstairs Markets, *Review of Financial Studies* 10,175–203.

Madhavan, Ananth, David Porter, and Daniel Weaver, 2005, Should securities markets be transparent?, *Journal of Financial Markets* 8, 266–288.

Maher, Owen, Peter L. Swan, and P. Joakim Westerholm, 2008, Twilight Falls on the Limit Order Book: Endogeneity and the Demise of Broker Identity, http://ssrn.com/abstract=1090001

Pagano, Marco and Ailsa Roell, 1996, Transparency and Liquidity: A Comparison of Auction and Dealer Markets with Informed Trading, *Journal of Finance* 51, 579–611

Rindi, Barbara, 2008, Informed Traders as Liquidity Providers: Anonymity, Liquidity and Price Formation, *Review of Finance* 12, 497–532.

Savickas, Robert, and Arthur J. Wilson, 2003, On Inferring the Direction of Option Trades, *Journal of Financial and Quantitative Analysis* 38, 881–902.

Seppi, Duane J., 1990, Equilibrium Block Trading and Asymmetric Information, *Journal of Finance* 45, 73–94.

Simaan, Yusif, Daniel G. Weaver, and David K. Whitcomb, 2003, Market maker quotation behavior and pretrade transparency, *Journal of Finance* 58, 1247–1267.

Smith, Brian F., Alasdair S. Turnbull, and Robert W. White, 2001, Upstairs Market for Principal and Agency Trades: Analysis of Adverse Information and Price Effects, *Journal of Finance* 56, 1723–1746.

Theissen, Erik, 2003, Trader Anonymity, Price Formation and Liquidity, *European Finance Review* 7, 1–26.

Essay 3

MACROECONOMIC ANNOUNCEMENTS AND CROSS-LISTED STOCKS

Abstract

We study the reactions of large European stocks cross-listed in the United States to scheduled U.S. macroeconomic announcements. Our main result is that the U.S. share of price discovery is lower on announcement days than on days with no announcements. We attribute this result to the information content of the macroeconomic releases. When comparing days with one or more macroeconomic announcements with days with no announcements, we observe an increase in average daily returns, return correlations, and intraday return volatility. The number of years listed as an ADR and the size of the largest shareholding are the most significant variables in explaining differences in shares of price discovery.

Keywords: macroeconomic announcements, price discovery, multiple listings, international markets.

JEL Classification: G12, G14, G15.

1 INTRODUCTION

In this study, we combine two areas of research. On the one hand, there are a number of papers on the reactions of different asset classes in different countries to U.S. macroeconomic releases. The most comprehensive study in this research area is Andersen et al. (2007), who study price responses in equities, fixed-income, and foreign exchange markets, in the U.S., the U.K., and Germany. On the other hand, we study price discovery in a multiple markets setting, i.e. where a cross-listed stock is traded simultaneously in two or more markets. Examples of these studies include Eun and Sabherwal (2003), Grammig, Melvin, and Schlag (2005), and Pascual, Pascual-Fuster, and Climent (2006), who study Canadian, German, and Spanish stocks, respectively, cross-listed in the U.S.

We use a sample of scheduled releases of macroeconomic data in the United States. We study price discovery, return volatility, and return correlations for large French and German companies that are cross-listed in the U.S. market. In particular, we focus on the differences between days with one or more releases of macroeconomic data, and days with no new macroeconomic data.

Theoretical literature on the effects of public news releases falls into three broad categories: information based models, rational expectations models, and behavioral models; see Nofsinger and Prucyk (2003) for a discussion. In the first group, Kim and Verrecchia (1991a) assume that traders form their own expectations prior to news releases, based on private information. Traders update their beliefs after the data is released. Trading is subsequently caused by the unexpected element, if any, of the data. Greater than usual volume is positively correlated with the absolute price change after an announcement. Kim and Verrecchia (1991b) is an extension of the previous model.

By using a reversal of the assumptions made in their previous models, Kim and Verrecchia (1994) assume that traders are unable to obtain any private information prior to a news event. This creates a temporary asymmetry, with high trading volume, as traders seek to understand the significance of the news in question.

Within the rational expectations framework, Veronesi (1999) studies the significance of the state of the market, a bull or a bear market, to the reaction to news. According to the model, investors are most anxious about signs pointing at a reversal of the current regime. This means that they overreact to bad news during economic expansion and to good news during a contraction.

Finally, in the behavioral group of explanations, Barberis, Shleifer, and Vishny (1998) create a model where investors suffer from a cognitive error when studying news releases, a so-called representativeness bias. Investors take whatever news there is as a representation of things to come. In this way, positive surprises are taken as a sign of good future news, and more bad news are expected after a negative surprise.

Correlation, in addition to stock returns, is another important variable in studies of international markets and diversification benefits. Bracker and Koch (1999) study changes in correlation over time, using a theoretical model to explain these changes. Martens and Poon (2001) focus on the problems caused by non-synchronous closing times in studying international market correlations. They also find significantly higher correlations when market returns are extremely negative. Bollerslev, Cai, and Song (2000) find that many macroeconomic releases have a significant impact on intraday volatility in the U.S. bond market. For the stock market, Graham, Nikkinen, and Sahlström (2003) find that some announcements have a significant effect on stock valuations, whereas the effect of other announcements is negligible. Becker, Finnerty, and Friedman (1995) study the effects for the U.K. market, Kim (2003) for Asia-Pacific markets, Nikkinen and Sahlström (2004) for the Finnish and German markets.

Theory does not give clear predictions about the effects of news releases on stock market returns or price discovery in a multiple markets setting. In this paper, we test the ways in which a release of macroeconomic news affects the stock market. If the most important effect is the general significance of the macroeconomic news, then the U.S. market is expected to be better and faster at interpreting the data. This should show up as a greater U.S. share of price discovery during days of data releases. On the other hand, if the importance of macroeconomic data is mostly in how it affects individual companies, then the local market is expected to gain in price discovery, in line with studies such as Hau (2001). This finding would highlight the significance of the local market, in line with earlier studies on price discovery in general (see e.g. Eun and Sabherwal (2003), Grammig, Melvin, and Schlag (2005))

Our sample includes twenty of the most liquid French and German stocks, cross-listed in the United States. We study their return correlations and shares of price discovery during the second half of 2004 (July 1st, 2004 to December 31st, 2004), using high-frequency data. We use a set of the most important U.S. macroeconomic announcements, and divide the sample into two groups: days with one or more macroeconomic announcements and days with no announcement. Our main results are that correlations between the stocks themselves, as well as between stocks and stock market indices, are higher on days when macroeconomic announcements are made. Also, the U.S. share of price discovery is lower on dates with new macroeconomic data. Another way to state this finding is that most of the adjustment after an external price innovation occurs in the local market. Our proposed explanation for this is that as the local market is closest to information pertaining to the company, it dominates price discovery even when the information originates from abroad. Using a cross-sectional regression, we find that the number of years listed as an ADR in the NYSE, and the size of the largest shareholding are the most significant explanatory variables in explaining differences in changes in shares of price discovery between days with macroeconomic news and days with no news.

The rest of this paper is organized as follows. We discuss our data set in Chapter 2. We then describe our methodology in Chapter 3, including a discussion of the error correction method used in analysis of price discovery. We present our results on returns, correlation, volatility, and price discovery in Chapter 4. And finally, in Chapter 5 we offer some concluding remarks.

2 DATA

2.1. Equities and foreign exchange data

The data sample consists of high-frequency trade and quote data for the U.S., French, and German markets. Our selection criterion is trading volume, i.e. we pick the ten most liquid companies for both countries, as measured by the trading volume in the NYSE market during our sample period. We use the TAQ (Trade and Quote) database for the U.S. markets. This is a complete intraday data set which contains all quotes and trades for all U.S. stock exchanges in 2004. Data for the European markets, Euronext Paris and Deutsche Börse Frankfurt, are from the exchanges themselves. These data sets also consist of a complete set of trades and quotes for the markets in question.

Foreign exchange data for the USD/EUR exchange rate are provided by Olsen Associates. This is a data set of indicative bid and ask quotes with 1-minute intervals. According to Olsen Data, their data collection software interfaces with several data feeds, such as Reuters, Tenfore, and Bloomberg.

The use of transactions data is generally known to suffer from autocorrelation bias. This was first analyzed by Roll (1984), when he introduced the concept of the bid-ask bounce. Quotes, however, can be updated even in the absence of trading; they do not suffer from autocorrelation bias in the same way. This leads us to the use of regularly spaced quotes in our analysis of price discovery. We take the bid and the ask quote with regular intervals and calculate the midpoint. Following accepted practice, we use logarithmic midquotes, i.e. logarithms of the average of the bid and offer prices.

The selection criteria for the stock sample are that the stocks be cross-listed and liquid. By liquidity we mean an active trading process, enabling us to perform an analysis of price discovery using high frequency data. Thus, out of a sample of all French and German companies listed at the NYSE, we take the stocks with the greatest turnover in the sample year (2004). We find that the ten most liquid stocks in each market had a sufficient number of trades for our analysis. Our sample of stocks, including some descriptive statistics, is listed in Table 1.

We filter our price series for errors, requiring among other things that the price be a multiple of the tick size for the exchange in question. The number of price points excluded in this way is negligible, however. We also exclude quotes flagged as odd lots, pre-opening, opening, closing, or halted.

All European stocks in our sample are traded as ADRs (American Depositary Receipts) on the NYSE. These are claims against home-market common shares issued by a U.S. depositary bank. They are quoted, traded and settled in U.S. dollars, and dividends are paid in dollars. They are directly exchangeable with common stock; issuances and cancellations are possible every day through the custodian bank for a small fee.

The dataset is from the period July 2004 – December 2004. It includes quote and trade data for 10 French and 10 German stocks, cross-listed in the U.S. We study the period of simultaneous trading in the U.S. and European markets. This means that all price data are only for this period, which starts when the U.S. market opens, at 9.30 a.m. EST (3.30 p.m. CET), and ends at 11.25 a.m. EST (5.25 p.m. CET), with the closing of the Paris market.

Table 1 Descriptive statistics of sample companies

This table presents descriptive statistics for our sample companies. *Industry classification* is the Industry Classification Benchmark (ICB) by FTSE and Dow Jones. *Market cap* is market capitalization in millions of USD, as of 31st December, 2004, as reported by Reuters. *Exchange* is either Euronext Paris, or Deutsche Börse Frankfurt am Main. *Last price* is the last traded price of the year 2004. *Trading volume U.S.*, and *Trading volume, home* are the average daily number of shares traded in the U.S. and home markets, respectively, in thousands of shares. *Number of trades, U.S.*, and *Number of trades, home,* are the average daily number of trades for the U.S., and home markets, respectively.

Company name	Industry classification	Market cap MUSD	Exchange	Last price, EUR	Trading volume, U.S.	Trading volume, home	Number of trades, U.S.	Number of trades, home
Alcatel	Telecommunications Equipment	21,02	Paris	11.41	1,686	1,302	823.0	6268.1
AXA	Full Line Insurance	45,69	Paris	18.17	295	828	580.2	5378.8
Danone	Food Products	23,112	Paris	67.7	35	121	179.3	3416.9
France Telecom	Fixed Line Telecommunications	80,849	Paris	24.48	133	917	252.3	10377.6
Lafarge	Building Materials & Fixtures	16,253	Paris	71.25	36	89	113.9	3117.7
Sanofi-Aventis	Pharmaceuticals	69,022	Paris	59.05	535	424	1114.1	4970.9
Suez	Multiutilities	26,619	Paris	19.43	28	485	83.6	4843.7
Thomson	Broadcasting & Entertainment	7,218	Paris	19.29	47	216	89.3	3195.1
Total	Integrated Oil & Gas	133,610	Paris	161.7	629	1,159	1508.2	6806.8
Vivendi	Broadcasting & Entertainment	34,389	Paris	23.54	331	651	433.3	5230.0
Allianz	Life Insurance	48,295	Frankfurt	97.13	122	32	590.2	7355.2
BASF	Specialty Chemicals	39,518	Frankfurt	52.81	36	25	219.8	6702.1
Bayer	Specialty Chemicals	24,817	Frankfurt	24.9	67	44	542.4	856.2
Daimler Chrysler	Automobiles	48,665	Frankfurt	35.31	274	65	303.6	4128.4
Deutsche Bank	Banks	40,189	Frankfurt	65.29	51	412	641.5	10656
Deutsche Telekom	Fixed Line Telecommunications	95,143	Frankfurt	1658	298	175	300.3	4501.2
E.ON	Multiutilities	59,840	Frankfurt	66.89	22	27	1240.2	6432.3
Infineon	Semiconductors	7,645	Frankfurt	7.94	384	1,684	942	5450.5
SAP	Software	54,933	Frankfurt	131.83	822	960	226.5	4409.4
Siemens	Electronic Equipment	65,671	Frankfurt	62.16	111	603	205	2779.5

2.2. Macroeconomic announcements

We collect data for the most important regularly released U.S. macroeconomic variables during July – December 2004. Our choice of variables is based on the Bureau of Labor Statistics classification. Previous studies provide support for our choice of variables; see e.g. Bollerslev et al. (2000). They include the following monthly reports: Employment report, Durable goods (Advance report), Retail Sales, Leading indicators, Housing starts, Factory orders, Consumer confidence, NAPM (National Association of Purchasing Management) manufacturing, and NAPM non-manufacturing.

Macroeconomic news are published at either 8.30 a.m., or 10.00 a.m., Eastern Standard Time, which corresponds to 2.30 p.m., or 4.00 p.m. Central European Time, respectively. In 2004 both European countries and the U.S. switched from Daylight Saving Time to regular time on October 31st, so no adjustment for different times of changing to regular time needs to be made. In the case of the earlier release time the U.S. market is not yet open. The closing time of the French market is 5.25 p.m., which means that markets have at least one hour and 25 minutes of simultaneous trading time to adapt to the news release before the European closing.

Table 2 lists the macroeconomic news releases used in this study, including their release dates. We divide the sample of 118 trading days into two parts: days with one or more macroeconomic releases, shown in Table 2, and days with no releases. There are 45 days with one or more macroeconomic release, and 73 days with no releases. We then calculate the mean of both regular returns and absolute returns, and average the results over the two groups.

Table 2 Macroeconomic announcements and their release dates

This table lists the macroeconomic indicators released by U.S agencies, and their release dates used in this sample.

Report name	Time of release	Dates of release					
Employment report	8:30AM	2-Jul	6-Aug	3-Sep	8-Oct	5-Nov	3-Dec
Durable goods (advance report)	8:30AM	8-Jul	25-Aug	24-Sep	27-Oct	24-Nov	23-Dec
Retail Sales	8:30AM	14-Jul	12-Aug	4-Sep	5-Oct	12-Nov	13-Dec
Leading indicators	10:00AM	2-Jul	19-Aug	23-Sep	21-Oct	18-Nov	20-Dec
Housing starts	8:30AM	20-Jul	17-Aug	21-Sep	19-Oct	17-Nov	16-Dec
Factory orders	10:00AM	2-Jul	4-Aug	2-Sep	4-Oct	3-Nov	2-Dec
Consumer confidence	10:00AM	27-Jul	31-Aug	28-Sep	26-Oct	30-Nov	28-Dec
NAPM manufacturing	10:00AM	1-Jul	2-Aug	1-Sep	1-Oct	1-Nov	1-Dec
NAPM non-manufacturing	10:00AM	6-Jul	4-Aug	3-Sep	5-Oct	3-Nov	3-Dec

3 PRICE DISCOVERY

We study price discovery after a public information release for European stocks cross-listed in the U.S. market. We follow Harris et al. (1995) in estimating an error correction model. We perform the analysis for each trading day separately, using 1-minute interval data. This results in 115 or 120 periods per trading day (French and German markets, respectively). We then compare the shares of price discovery during days of macroeconomic announcements and days without announcements.

3.1. Unit root and cointegration tests

A requirement of the price discovery analysis is that the two price series, the home market price and the cross-listed price, be cointegrated. Intuitively this makes sense as long as a no-arbitrage criterion (the law of one price) applies to the two prices. A necessary condition for two price series to be cointegrated is that they individually contain a unit root.

We begin by determining the appropriate lag length for each stock, using the Schwarz Information Criterion. We also employ the rule of thumb of a maximum lag length of $T^{1/3}$ (where T is the number of observations), resulting in a maximum lag length of 20. The optimum lags are 1 or 2 for most stocks, with a maximum value of 3.

We perform unit root tests for all equity data series, using the augmented Dickey-Fuller (1981) test, following standard procedure in the literature. This test checks for the presence of a unit root by including lagged first differences of the price series. We consider the following three standard formulations of the model: a random walk, a random walk with a drift term, and a random walk with a drift term and a time trend.

$$\Delta x_t = \delta x_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta x_{i-1} + e_t \tag{1}$$

$$\Delta x_t = c_0 + \delta x_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta x_{i-1} + e_t \tag{2}$$

$$\Delta x_t = c_0 + \delta x_{t-1} + c_1 t + \sum_{i=1}^{p} \varphi_i \Delta x_{i-1} + e_t \tag{3}$$

In each case, the null hypothesis is that δ equals zero, and that the price series thus contains a unit root. This means that the price series is nonstationary, but that the changes are stationary.

Since the augmented Dickey-Fuller tests have a low power to reject the null of *I(1)*, we complement the tests with Phillips-Perron tests.

We perform the tests for price series from the home exchange, i.e. Paris, and the price of the ADR traded in New York, converted into euros at the prevailing midquote of EUR/USD. Using a significance level of 5%, we accept the null hypothesis of *I(1)* for all stocks. These results are available on request.

Simple arbitrage bounds for a cross-listed stock imply that the quotes in the home and cross-listed markets cannot be expected to diverge significantly from each other. In other words, we can expect the two price series to be cointegrated. If the series are indeed cointegrated, then the following is also true: series are *I(1)*, i.e. they contain a unit root, and there exists a so-called cointegration vector $\beta = \left(\beta^{EU}, \beta^{US}\right)$, such that $\beta^{EU} P_t^{EU} + \beta^{US} P_t^{US}$ is *I(0)*.

To test for cointegration of the price series we use the Johansen (1988) method and the number of cointegration vectors. We rewrite a *p*th order autoregressive process,

$$x_t = A_1 x_{t-1} + A_2 x_{t-2} + \ldots + A_p x_{t-p} + \varepsilon_t \tag{4}$$

as

$$\Delta x_t = \sum_{i=1}^{p-1} \Pi_i \Delta x_{t-i} + \Pi x_{t-p} + \varepsilon_t , \tag{5}$$

where Δ is the first-difference lag operator, x_t is a vector of *I(1)* time series, ε_t is a zero-mean n-dimensional white noise vector, Π_i are *(n x n)* matrices of parameters, and Π is a *(n x n)* matrix of parameter vectors.

The tested hypothesis is that the number of cointegrating vectors is equal to or less than *r*. We conduct this test using one of two statistics: $\lambda_{trace}(r)$ or $\lambda_{\max}(r,r+1)$. These statistics are given by

$$\lambda_{trace}(r) = -T \sum_{i=t+1}^{n} \ln(1 - \hat{\lambda}_t) \tag{6}$$

and

$$\lambda_{\max}(r, r+1) = -T \ln(1 - \hat{\lambda}_{r+1}) , \tag{7}$$

where T is the number of observations, and $\hat{\lambda}$ is the estimated value of the characteristic root, obtained from the estimated *Π* matrix.

The two statistics test different hypotheses. The null hypothesis for $\lambda_{trace}(r)$ is that the number of distinct cointegrating vectors is less than or equal to *r*. The statistic $\lambda_{\max}(r,r+1)$ tests for the null hypothesis of *r* cointegrating vectors against an alternative of *r+1* vectors. Johansen and Juselius (1990) provide critical values for both statistics, which do not follow any standard distributions.

We reject the null hypothesis of no cointegrating vector for all stocks in the sample. These results are available on request[29].

---

[29] We also employ a direct test for the stationarity of the logarithmic price difference between the prices in the two markets. These results lend support to the above conclusion, in that the price differential is indeed stationary, and very close to zero at all times.
This is indeed the expected results, since all stocks are very liquid, and there is no reason for the two prices to deviate. An analysis of deviations from the law of one price indicates that any deviation above the

3.2. The Error Correction Model

Cointegration between the European and U.S. prices of a stock implies the time series of both prices are influenced by any deviation from equilibrium. If there is a price innovation in one of the markets, one or both prices have to adjust.

We estimate the following equations:

$$\Delta P_t^{EU} = \alpha_0^{EU} + \alpha^{EU}\left(P_{t-1}^{US} - P_{t-1}^{EU}\right) + \sum_{i=1}^{p} \gamma_i \Delta P_{t-i}^{EU} + \sum_{i=1}^{p} \delta_i \Delta P_{t-i}^{US} + \varepsilon_t^{EU} \qquad (8)$$

$$\Delta P_t^{US} = \alpha_0^{US} + \alpha^{US}\left(P_{t-1}^{US} - P_{t-1}^{EU}\right) + \sum_{j=1}^{p} \gamma_j \Delta P_{t-j}^{EU} + \sum_{j=1}^{p} \delta_j \Delta P_{t-j}^{US} + \varepsilon_t^{US} \qquad (9)$$

where P^{EU} and P^{US} are the prices of the stock in the European market and the U.S. market, respectively; $\gamma_{i,j}$ are error correction parameters which reflect idiosyncratic adjustment of P_t to disparities in shocks to private value that cause the cointegrated series to diverge; p is the lag length determined earlier using the Schwarz Information Criterion; and ε_t is a zero-mean, covariance-stationary random disturbance, which is assumed to be identically distributed but may be autocorrelated.[30]

We estimate the coefficients α^{EU} and α^{US} above in the error correction model. These coefficients give us the average adjustment of each price series in the case of a deviation from the law of one price. In order to be able to study the amount of adjustment occurring in one exchange as a proportion of total adjustment, we construct the variable USPD, for the share of price discovery of the U.S. market.

We measure of the share of total adjustment that occurs in the cross-listed market, NYSE, as first proposed by Schwarz and Szakmary (1994) and used by, among others, Theissen (2002), and Eun and Sabherwal (2003). This is simply the European adjustment coefficient as a share of the summed adjustment coefficients:

$$USPD = \frac{\left|\alpha^{EU}\right|}{\left|\alpha^{EU}\right| + \left|\alpha^{US}\right|}. \qquad (10)$$

A high value of the variable *USPD* means that the coefficient α^{EU} is high. The interpretation of this is that the U.S. market does most of the adjusting to price innovations. In the hypothetical case of no feedback from the U.S. to the European market, α^{EU} is zero, and thus also *USPD* is zero.

---

arbitrage bounds imposed by the conversion cost of USD 0.05 per share is very short-lived. These results are available on request.

[30] We also use a trivariate specification, where the third error correction equation is as follows:

$$\Delta P_t^{FX} = \alpha_0^{FX} + \alpha^{FX}\left(P_{t-1}^{US} - P_{t-1}^{EU}\right) + \sum_{j=1}^{p} \gamma_j \Delta P_{t-j}^{EU} + \sum_{j=1}^{p} \delta_j \Delta P_{t-j}^{US} + \varepsilon_t^{FX},$$

where P^{FX} is the USD/EUR exchange rate; other parameters are as per the above discussion.

4 RESULTS

4.1. Stock price returns

We calculate average daily logarithmic returns for all stocks in the sample. Table 3 reports returns averaged over all stocks, and for three kinds of subsamples: days with macroeconomic announcements, days with no announcements, and the entire sample. We also report the same averages using absolute values of daily returns. We see that in both regular and absolute returns, days with macroeconomic announcements have a higher average return. We also test for the equality of returns on “Macro” and “No macro” days. When comparing the returns for each stock, averaged over the days of macroeconomic announcements with the average returns of each stock during days with no announcements, we reject the hypothesis of equal averages.

We perform the same analysis using foreign exchange data for comparison purposes. Days with macroeconomic releases exhibit greater price movement even in foreign exchange. As the exchange rate is defined as the dollar price of euros (USD/EUR), the fact that the average return is greater during macroeconomic release days means that the euro has appreciated and the dollar depreciated on average on those days.

We test for the equality of the means of the two return series, i.e. the average stock-specific returns, for Macro and Non-macro days. In all four cases, using both regular returns and absolute returns, and for both stock data and foreign exchange data, the null hypothesis of equal means is rejected with a high significance.

Table 3 Descriptive statistics: daily returns

This table presents average daily returns, for days with macroeconomic announcements, and for days with no announcements. Returns are measured as logarithmic returns; absolute returns take absolute values of all returns, thus eliminating negative returns from the sample. *Macro* refers to days with macroeconomic announcements; *No macro* to days with no announcements; *All data* contains all observations. All statistics are calculated over all days and all stocks in the sample. Panel A presents data for daily stock returns, and Panel B the equivalent data for the EUR/USD exchange rate.
The t-test tests for the difference between returns on "Macro" days and "No macro" days. Values in parenthesis are p-values. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A: stock data

	Returns			Absolute returns		
	Macro	No macro	All data	Macro	No macro	All data
Average	0.19%	-0.02%	0.06%	1.09%	0.92%	0.98%
Median	0.24%	-0.06%	0.06%	0.87%	0.73%	0.77%
Standard deviation	1.42%	1.20%	1.29%	0.93%	0.77%	0.84%
Max	7.54%	5.36%	7.54%	7.54%	5.36%	7.54%
Min	-4.84%	-4.78%	-4.84%	0.00%	0.00%	0.00%
Skewness	-0.13	0.14	1.53	1.54	0.05	1.58
Excess kurtosis	1.40	1.16	3.75	3.40	1.35	3.90
T test	4.602***			6.164***		
	(0.00019)			(0.00001)		

Panel B: foreign exchange data

	Returns			Absolute returns		
	Macro	No macro	All data	Macro	No macro	All data
Average	0.16%	0.02%	0.08%	0.46%	0.40%	0.43%
Median	0.10%	-0.01%	0.08%	0.35%	0.26%	0.30%
Standard deviation	0.57%	0.55%	0.56%	0.38%	0.37%	0.37%
Max	1.74%	1.22%	1.74%	1.74%	1.33%	1.74%
Min	-1.06%	-1.33%	-1.33%	0.00%	0.00%	0.00%
Skewness	0.14	-0.12	1.09	0.99	-0.01	1.01
Excess kurtosis	0.21	0.28	0.97	-0.35	0.29	0.17
	2.688**			2.926***		
	0.015			0.009		

4.2. Volatility

We calculate volatility as standard deviations of daily logarithmic return data.[31] For each stock we report three volatility numbers in Table 4: one for "macro" days, one for "Non-macro" days, and one using the entire sample. We find a significant increase in average volatilities for the "Macro" days over the "Non-macro" days. The average of "Macro" volatilities is 18% greater than the average "Non-macro" volatility.

This result is also expected, and in line with previous research. The general conclusion in volatility spillover literature is that asset prices and volatilities are affected by macroeconomic news announcements; see e.g. Andersen and Bollerslev (1998) for a discussion.

Table 4 Volatility

Volatility is as the standard deviation of daily logarithmic midquote returns. We report average volatilities over all days with macroeconomic announcements ("Macro"), and over all days with no announcements ("No macro"). The last column reports the average volatility over the entire sample.
We test for the hypothesis of equal volatilities for macro and no macro days. *** denotes significance at the 1% level.

Company name	Exchange	Macro	No Macro	All data
Alcatel	Paris	0.022	0.018	0.020
AXA	Paris	0.015	0.012	0.014
Danone	Paris	0.011	0.010	0.010
France Telecom	Paris	0.015	0.013	0.013
Lafarge	Paris	0.012	0.010	0.011
Sanofi-Aventis	Paris	0.012	0.012	0.012
Suez	Paris	0.014	0.009	0.011
Thomson	Paris	0.015	0.015	0.015
Total	Paris	0.009	0.008	0.008
Vivendi	Paris	0.015	0.012	0.014
SAP	Frankfurt	0.017	0.013	0.014
Daimler Chrysler	Frankfurt	0.012	0.011	0.012
Siemens	Frankfurt	0.011	0.012	0.011
Infineon	Frankfurt	0.020	0.017	0.019
Deutsche Telekom	Frankfurt	0.012	0.011	0.011
Deutsche Bank	Frankfurt	0.013	0.010	0.012
BASF	Frankfurt	0.011	0.010	0.011
Bayer	Frankfurt	0.012	0.012	0.012
E.ON	Frankfurt	0.012	0.008	0.010
Allianz	Frankfurt	0.016	0.011	0.013
Average		0.014	0.012	0.013
Median		0.013	0.012	0.012
T test			5.332***	
			(0.00004)	

[31] We also calculate volatility using 30-minute interval data. The results are very similar to the results obtained using daily data.

4.3. Correlation

We analyze correlation in two main ways. First, we calculate cross-correlations for our sample of 20 stocks. Second, we analyze correlations between individual stocks and market indices. Correlations are fundamental in international diversification. They are also a proxy for market integration. For a long-term view on international correlations and a survey of earlier literature, see Goetzmann, Li, and Rouwenhorst (2005).

Table 5 presents our results. We divide the sample into two parts, days with macro announcements, and days with no announcements. We choose one broad-based index as a representative benchmark for both the European and U.S. market, the STOXX 600 index for Europe, and the S&P 500 for the U.S.

As expected, average correlations are higher on days with macroeconomic announcements. This result applies to both cross-correlations within the sample group of stocks, and to the correlations between the stocks and market indices. T tests give a significant result only in the case of the S&P 500 index, when testing for the equality of means of "Macro" and "Non-macro" days. Apart from testing for the means of average correlations, we can test for the equality of an entire correlation matrix. Using a Jennrich (1970)[32] χ^2 test for the equality of two correlation matrices, we conclude that the cross-correlation matrix is not equal for "Macro" and "Non-macro" data[33].

---

[32] See Goetzmann, Li, and Rouwenhorst (2005) for a detailed discussion.

[33] Note that we are only able to perform the Jennrich (1970) for square matrices. The cross-correlation matrix is a 20 x 20 matrix, whereas the correlation matrixes between stock returns and index returns are of size 20 x 2.

Table 5 Return correlations

This table presents correlation statistics. In all calculations, the sample period is divided into two parts: days with macroeconomic announcements ("Macro"), and days with no announcements ("No macro"). "Cross-correlations" are the stock-specific average cross-correlations, which are calculated with all other 19 stocks in the sample. We also calculate correlations between each stock and two broad-based stock market indices: S&P 500 and STOXX. In all cases, days with macroeconomic announcements have greater average correlations.
T-statistics are for testing the equality of the correlation matrices; p values are in parenthesis. Wilcoxon's z test measures the equality of two square matrices. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Cross-correlations		S&P 500 index		STOXX index	
Company name	macro	no macro	macro	no macro	macro	no macro
Alcatel	0.557	0.512	0.277	0.112	0.669	0.660
AXA	0.660	0.625	0.497	0.112	0.873	0.814
Danone	0.562	0.444	0.384	0.062	0.731	0.557
France Telecom	0.615	0.589	0.471	0.256	0.751	0.731
Lafarge	0.567	0.531	0.381	0.087	0.770	0.722
Sanofi-Aventis	0.362	0.438	0.150	0.106	0.453	0.583
Suez	0.522	0.496	0.282	0.205	0.678	0.662
Thomson	0.318	0.507	0.230	0.095	0.350	0.691
Total	0.455	0.372	0.352	0.234	0.613	0.480
Vivendi	0.584	0.483	0.298	0.252	0.734	0.582
SAP	0.634	0.577	0.421	0.276	0.832	0.772
Daimler Chrysler	0.629	0.543	0.537	0.324	0.841	0.742
Siemens	0.651	0.611	0.488	0.232	0.874	0.804
Infineon	0.539	0.552	0.429	0.239	0.671	0.682
Deutsche Telekom	0.585	0.555	0.444	0.282	0.746	0.704
Deutsche Bank	0.555	0.550	0.454	0.287	0.691	0.775
BASF	0.599	0.474	0.381	0.077	0.812	0.588
Bayer	0.653	0.598	0.421	0.358	0.849	0.775
E.ON	0.390	0.434	0.142	0.145	0.543	0.474
Allianz	0.580	0.636	0.349	0.292	0.782	0.830
Average	0.551	0.526	0.369	0.202	0.713	0.681
t test		1.485		7.086***		1.175
		(0.154)		(0.000)		(0.255)
Jennrich χ^2		477.58***				
		(0.000)				

4.4. Price discovery

Our main contribution is the analysis of how a cross-listed stock reacts to news releases that are global, rather than stock-specific in nature. Macroeconomic data is obviously global in nature. Due to the significant size of both the real economy and the market capitalization of stock markets, it can be expected that releases of U.S. data are significant indicators of the health of the global economy. Earlier results have established that most price discovery occurs on the home exchange. Our main hypothesis is that if macroeconomic data releases are general in nature, in that the economic news is the most important, then the U.S. market can be expected to gain in dominance in price discovery. If, on the other hand, macroeconomic news releases are most important because of how the news affects individual companies, the local exchange may still be better at interpreting this information.

We estimate the error correction model of Equations (8) and (9) using three sets of data. First, we use a combined sample of all “Macro” days taken together. Second, we combine all “Non-macro” days. Third, we use all available data for the entire sample period. The average U.S. share of price discovery is smaller when macroeconomic data are released, indicating that the company-specific interpretation of data is of paramount importance. We reject the hypothesis of equal means at the 10% significance level. The U.S. share of price discovery on “Non-macro” days is on average 6% greater than on “Macro” days.

Table 6 U.S. shares of price discovery

This table presents the average shares of price discovery, as measured by the error correction model of Equations (8) and (9). Shares of price discovery are calculated separately for days with macroeconomic announcements ("Macro") and for days with no announcements ("No macro").
T-statistics are for testing the equality of the matrices of shares of price discovery. * indicates significance at the 10% level.

		U.S. share of price discovery	
Stock name	Exchange	Macro	No macro
Alcatel	Paris	0.227	0.205
AXA	Paris	0.175	0.178
Danone	Paris	0.385	0.409
France Telecom	Paris	0.256	0.216
Lafarge	Paris	0.297	0.376
Sanofi-Aventis	Paris	0.223	0.274
Suez	Paris	0.267	0.291
Thomson	Paris	0.363	0.432
Total	Paris	0.279	0.301
Vivendi	Paris	0.252	0.222
SAP	Frankfurt	0.311	0.336
Daimler Chrysler	Frankfurt	0.264	0.297
Siemens	Frankfurt	0.297	0.325
Infineon	Frankfurt	0.366	0.338
Deutsche Telekom	Frankfurt	0.325	0.263
Deutsche Bank	Frankfurt	0.242	0.254
BASF	Frankfurt	0.273	0.306
Bayer	Frankfurt	0.227	0.334
E.ON	Frankfurt	0.336	0.317
Allianz	Frankfurt	0.363	0.386
Average		0.286	0.303
t stat			1.748*
			(0.097)

4.5. Cross-sectional analysis of differences in price discovery

In the previous chapter we have estimated what the effects of macroeconomic releases are for the shares of price discovery. We are now in a position to analyze these differences, using a set of descriptive variables

4.5.1. Explanatory variables

We use a broad range of explanatory variables, described in Table 7. They can be grouped into three categories: trading related, ownership variables, and company descriptive variables. Below we describe all variables in depth.

Table 7 Cross-sectional variables

Market cap is market capitalization, in U.S. dollars, the logarithm of which is used in the regression. *Home volume* is the share of trading volume of the home market, as a share of total trading value. *Spread ratio* is the ratio of the average U.S. bid-ask spread and the average home market bid-ask spread, where spread is (Ask-Bid)/((Ask+Bid)/2).*Years listed* is the number of years the company has been listed as ADRs in the U.S. *Int'l sales* is the share of the sales taking place outside the company's home market. *Largest shareholder* is the share of the largest shareholder as a percentage of total shares, and other ownership variable, *U.S. holdings,* is the share of U.S. based investors as a percentage of total shares.

Company name	Market cap, MUSD	Home volume	Spread ratio	Years listed	Int'l sales	Largest shareholder	U.S. holdings
Alcatel	21,020	54%	1.15	12	36%	9	10.3
AXA	45,669	19%	1.57	8	23%	17	13
Danone	23,112	84%	2.77	7	68%	4	1.1
France Telecom	80,849	81%	1.98	7	59%	23	2
Lafarge	16,253	37%	2.50	3	16%	9	2.3
Sanofi-Aventis	69,022	89%	0.91	2	49%	12	20
Suez	26,619	41%	3.15	3	23%	7	9
Thomson	7,218	40%	2.64	5	23%	6	1
Total	133,610	55%	0.59	13	19%	5	23
Vivendi	34,389	21%	1.48	9	56%	10	2
SAP	54,933	44%	1.17	6	27%	10	8.1
DaimlerChrysler	48,665	74%	1.19	6	16%	12	17
Siemens	65,671	78%	1.77	3	32%	9	12
Infineon	7,645	87%	1.17	4	23%	28	3
Deutsche Telekom	95,143	71%	1.31	8	57%	26	29
Deutsche Bank	40,189	44%	1.79	3	21%	3	8.5
BASF	39,518	94%	2.13	4	41%	0	14
Bayer	24,817	82%	1.68	2	49%	6	8
E.ON	59,840	65%	2.20	7	64%	6	19.7
Allianz	48,295	66%	4.16	4	47%	12	7.2

Our trading related variables are a direct measure of the ease and expense of trading the stock. We define home market trading volume as the number of shares traded in the home market, Frankfurt or Paris, as a proportion of total shares traded in the home market and the U.S. market during simultaneous trading hours. We of course adjust, when needed, for the ratio of ADRs to common shares. Hasbrouck (1995) studies the relative contributions of the NYSE and regional exchanges to price discovery of the 30 stocks in the Dow Jones Industrial Average. He finds a positive and statistically significant relation between shares of trading volume and the shares of price discovery between the NYSE and U.S. regional exchanges. Since the bid-ask spread is a major component of trading costs, we can expect it to influence price discovery, see e.g. Fleming, Ostdiek and Whaley (1996), Harris, et al. (2002), and Eun and Sabherwal (2003). Fleming, Ostdiek, and Whaley (1996) suggest that the market with a smaller bid-ask spread will have a greater share of informed trading. Harris et al. (2002) find evidence that the NYSE share of price discovery is negatively related to the size of its bid-ask spread.

We also include two variables to reflect the different ownership structures of the companies. The first variable is the share of the largest shareholder, as a percentage of total shares outstanding. The second ownership variable is the total share of U.S. based owners, as a percentage of the total number of shares.

The rest of our variables describe the company itself in greater detail. Exchange, years listed, U.S. share of sales. We include a dummy variable for the home exchange of the company, 1 for Paris and 0 for Frankfurt. This variable of course also serves as an indicator of the home country of the company. We also include the number of years the company's stock has been listed on the New York Stock Exchange as an ADR. "International sales" is the percentage share of the company's total sales that take place outside the home market. Logarithmic market capitalization is a measure of the size of the company and its importance in the capital markets.

As an additional robustness test, we also include a measure of the difference in trading volume between days with a macroeconomic release, and days without data releases. We calculate this measure ("volume ratio") in the following way. For each stock and each trading day, we first calculate the ratio of the home market trading volume and the U.S. trading volume. We then separate the sample into days with macroeconomic data ("Macro days") and days with no data ("Non-macro days"). We then obtain the average ratios for Macro days and Non-macro days. The final variable is then the ratio of the Macro average ratios and the Non-macro average ratios. This variable reflects the increase in the U.S. relative trading volume of each stock, when there is macroeconomic information, compared with days with no new information.

4.5.2. Cross-sectional results

We wish to study the changes in price discovery, on days when there are macroeconomic announcements, compared to days with no announcements. For this purpose, we construct two measures of the change in shares of price discovery. Our first dependent variable, PD_{diff} is the difference between the average foreign share of price discovery during days with announcements and the average home market share of price discovery during days with no announcements.

$$PD_{diff} = PD_{macro} - PD_{no\_macro} \tag{11}$$

Our second dependent variable, PD_{ratio}, is the ratio of the above mentioned average shares of price discovery:

$$PD_{ratio} = \frac{PD_{macro}}{PD_{no\_macro}} \tag{12}$$

We calculate both of these variables for each stock, which gives us 20 variables to work with the cross-sectional regression.

Our results are shown in Table 8. We are able to explain between 55% and 72% of the variation in PD_{ratio}, and between 48% and 57% of the variation in PD_{diff}. None of our trading related variables, such as the bid-ask spread, or trading volume, is significant. Rather, the number of years listed in the U.S., and the size of the holding of the largest shareholder are the most significant in all regressions. We attribute this finding to information asymmetries and the transmission of information. The estimated coefficient for both variables is higher. This means that the U.S. share of price discovery increases more when there are macroeconomic announcements for companies with an established presence in the U.S. market. Also, the larger the holding of the largest shareholder, the more significant the U.S. price discovery is during announcement days. It is to be noted that our ownership data do not identify the largest owner. Thus, we do not know whether this owner is local, U.S., or perhaps based in a third country. The relative increase of U.S. trading volume on days with macroeconomic information is not a significant explanatory variable.

These results are highly tentative, due to the small number of companies in the regression. We attribute these results to the different ways in which investors react to news.

Table 8 Cross-sectional results: Differences in shares of price discovery

This table presents regression results, where the dependent variable is the difference or ratio between the average U.S. share of price discovery on days of macroeconomic releases, and the average U.S. share of price discovery on days with no macroeconomic announcements.
The explanatory variables are the following. *Market cap* is the logarithmic market capitalization as of December 31, 2004. *Volume* is the share of trading volume of the home market, as a percentage of total trading. *Spread ratio* is the ratio of U.S. bid-ask spread and the home market bid-ask spread. *Exchange* is a dummy variable, 1 for Paris and 0 for Frankfurt. *Years listed* is the number of years the company has been listed as ADRs in the U.S. *Int'l sales* is the share of the sales taking place outside the company's home market. *Largest shareholder* is the share of the largest shareholder as a percentage of total shares. *U.S. holdings*, is the share of U.S. based investors as a percentage of total shares. Numbers in parenthesis are adjusted t-statistics.
Panel A presents OLS results, where the dependent variable is the difference between the average shares of price discovery:

$$PD_{diff} = PD_{macro} - PD_{no\_macro}$$

Panel B presents OLS results, where the dependent variable is the ratio of the average shares of price discovery:

$$PD_{ratio} = \frac{PD_{macro}}{PD_{no\_macro}}$$

Panel A PD$_{diff}$ as the dependent variable

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-0.170	-0.093***	-0.108***	-0.106***	-0.140	0.004
	(-1.310)	(-4.283	(-5.178)	(-5.419)	(-0.905)	(0.026)
Market cap	0.008				0.015	0.016
	(0.560)				(1.056)	(1.189)
Volume	-0.135	-0.128			-0.184	-0.249
	(-0.852)	(-0.954)			(-0.839)	(-1.576)
Spread ratio	0.003				-0.010	-0.003
	(0.251)				(-0.799)	(-0.328)
Exchange	-0.019	-0.018			-0.002	0.001
	(-1.181)	(-1.342)			(-0.104)	(0.033)
Years listed	0.008**	0.008***	0.006**	0.006***		0.006
	(2.791)	(3.432)	(2.824)	(3.078)		(2.105)*
Int'l sales	0.000	0.000	0.001	0.001		0.000
	(0.833)	(1.303)	(1.556)	(1.608)		1.167
Largest shareholder	0.003**	0.003***	0.003***	0.003***		0.004***
	(3.033)	(3.350)	(3.297)	(3.418)		(3.753)
US holdings	-0.001		0.000			-0.001
	(-0.383)		(0.289)			(-0.492)
Volume ratio						-0.250
						(-1.786)
Adjusted R^2	0.480	0.575	0.541	0.567	-0.086	0.566
Durbin-Watson	2.355	2.424	2.324	2.272	1.769	2.747
Number of observations	20	20	20	20	20	20
Number of variables	9	6	5	4	5	10

Panel B: PD_{ratio} as the dependent variable

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.485	0.691***	0.653***	0.648***	0.616	0.951*
	(1.236)	(10.303)	(10.266)	(10.794)	(1.181)	(1.986)
Market cap	0.020				0.041	0.042
	(0.490)				(0.878)	(1.014)
Volume	-0.476	-0.486			-0.642	-0.781
	(-0.997)	(-1.180)			(-0.870)	(-1.587)
Spread ratio	0.012				-0.029	-0.004
	(0.392)				(-0.675)	(-0.072)
Exchange	-0.057	-0.044			0.007	-0.004
	(-1.167)	(-1.067)			(0.109)	(-0.072)
Years listed	0.028***	0.025***	0.022***	0.022***		0.022**
	(3.296)	(3.699)	(3.357)	(3.477)		(2.605)
Int'l sales	0.001	0.002	0.002**	0.002**		0.002
	(1.097)	(1.595)	(1.906)	(1.955)		(1.378)
Largest shareholder	0.010***	0.010***	0.010***	0.010***		0.012***
	(3.421)	(3.680)	(3.632)	(3.723)		(3.937)
US holdings	-0.003		-0.001			-0.003
	(-0.673)		(-0.336)			(-0.775)
						-0.669
						(-1.535)
Adjusted R^2	0.565	0.728	0.606	0.628	-0.124	0.613
Durbin-Watson	2.248	2.385	2.264	2.327	1.752	2.614
Number of observations	20	20	20	20	20	20
Number of variables	9	6	5	4	5	10

5 CONCLUSION

In this paper, we contribute to two areas of literature. First, the scheduled releases of macroeconomic data in the U.S. have well documented effects on asset prices and volatility, both in the U.S., and abroad. We study the effects of these data releases for the period of simultaneous trading in cross-listed stocks. Second, we contribute to the existing literature on cross-listings and price discovery in a multiple markets setting.

Our sample consists of 20 large French and German companies cross-listed in the NYSE. Our main result is that the U.S. share of price discovery declines on days with macroeconomic announcements, compared to days with no announcements. Shares of price discovery are estimated as a proxy for the adjustment process of the price of a cross-listed stock to price innovations, meaning that the more the price of the U.S. market adjusts to changes in the local market, the greater the price discovery share of the local market. This finding supports our hypothesis of the importance of the stock-specific component of macroeconomic releases. Also, it highlights the importance of the local market in price discovery, in line with earlier literature in the field.

We also find that daily returns are higher on announcement days, both in absolute terms and in logarithmic returns. This result applies to both stock prices and the EUR/USD exchange rate. As expected on the basis of a large literature on international stock market correlations, both cross-correlations within the sample of 20 stocks, and correlations between the stocks and two broad market indices (STOXX 600, and S&P 500) are higher when macroeconomic data are released. We also analyze volatility, which is significantly higher on days with macroeconomic releases.

A cross-sectional analysis of the changes in shares of price discovery, when comparing days with macroeconomic announcements with days without announcements, we find that the two most significant explanatory variables are the number of years listed as an ADR in the U.S. market, and the size of the largest shareholding in the company. We attribute these findings to differences in the dissemination of information in the stocks in question.

REFERENCES

Andersen, Torben, Tim Bollerslev, Francis X. Diebold, and Clara Vega, 2007, Real-time price discovery in global stock, bond and foreign exchange markets, Journal of International Economics 73, 251–277.

Andersen, Torben, and Tim Bollerslev, 1998, Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies, *Journal of Finance* 53, 219–265.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307–343.

Becker, Kent G., Joseph E. Finnerty, and Joseph Friedman, 1995, Economic news and equity market linkages between the U.S. and U.K., *Journal of Banking and Finance* 19, 1191–1210

Bollerslev, Tim, Jun Cai, and Frank M. Song, 2000, Intraday periodicity, long memory volatility, and macroeconomic announcement effects in the US Treasury bond market, *Journal of Empirical Finance* 7, 37–55.

Bracker, Kevin, and Paul D. Koch, 1999, Economic determinants of the correlation structure across international equity markets, *Journal of Economics and Business* 51, 443– 471.

Eun, Cheol S., and Sanjiv Sabherwal, 2003, Cross-Border Listings and Price Discovery: Evidence from U.S.-Listed Canadian Stocks, *Journal of Finance* 58, 549–575.

Fleming, Jeff, Barbara Ostdiek, and Robert E. Whaley, 1996, Trading costs and the relative rates of price discovery in stock, futures, and option markets, *Journal of Futures Markets* 16, 353–387.

Grammig, Joachim, Michael Melvin, and Christian Schlag, 2005, Internationally Cross-Listed Stock Prices during Overlapping Trading Hours: Price Discovery and Exchange Rate Effects, *Journal of Empirical Finance* 12, 139–164.

Goetzmann, William, Lingfeng Li, and K. Geert Rouwenhorst, 2005, Long-Term Global Market Correlations, *Journal of Business* 78, 1–38.

Graham, Michael, Jussi Nikkinen, and Petri Sahlström, 2003, Relative Importance of Scheduled Macroeconomic News for Stock Market Investors, *Journal of Economics and Finance* 27, 153–165.

Harris, Frederick H. DeB, Thomas H. McInish, Gary L. Shoesmith, and Robert A. Wood, 1995, Cointegration, Error Correction, and Price Discovery on Informationally Linked Security Markets, *Journal of Financial and Quantitative Analysis* 30, 563–579.

Harris, Frederick H. deB., Thomas McInish, and Robert A. Wood, 2002, Security price adjustment across exchanges: an investigation of common factor components for Dow stocks, *Journal of Financial Markets* 5, 277–308.

Hasbrouck, Joel, 1995, One security, many markets: determining the contributions to price discovery, *Journal of Finance* 50, 1175–1199.

Hau, Harald, 2001, Location matters: An examination of trading profits, *Journal of Finance* 56, 1959–1983.

Jennrich, Robert I, 1970, An asymptotic chi-square test for the equity of two correlation matrices. *Journal of the American Statistical Association* 65, 904–912.

Johansen, Soren, 1988, Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* 12, 231–254.

Kim, Oliver, and Robert E. Verrecchia, 1991a, Trading volume and price reactions to public announcements, *Journal of Accounting Research* 29, 302–321.

Kim, Oliver, and Robert E. Verrecchia, 1991b, Market reaction to anticipated announcements, *Journal of Financial Economics* 30, 273–309.

Kim, Oliver, and Robert E. Verrecchia, 1994, Market liquidity and volume around earnings announcements, *Journal of Accounting and Economics* 17, 41–67.

Kim, Suk-Joong, 2003, The spillover effects of US and Japanese public information news in advanced Asia-Pacific stock markets, *Pacific-Basin Finance Journal* 11, 611–630.

Martens, Martin, and Ser-Huang Poon, 2001, Returns synchronization and correlation dynamics between international stock markets, *Journal of Banking and Finance* 25, 1805–1827.

Nikkinen, Jussi, and Petri Sahlström, 2004, Scheduled domestic and US macroeconomic news and stock valuation in Europe, *Journal of Multinational Financial Management* 14, 201–215.

Nikkinen, Jussi, Mohammed Omran, Petri Sahlström, and Janne Äijö, Global stock market reactions to scheduled U.S. macroeconomic news announcements, *Global Finance Journal* 17, 92–104.

Nofsinger, John R., and Brian Prucyk, 2003, Option volume and volatility response to scheduled economic news releases, *Journal of Futures Markets* 23, 315–345.

Schwarz, Thomas V., and Andrew C. Szakmary, 1994, Price discovery in petroleum markets: Arbitrage, cointegration, and the time interval of analysis, *Journal of Futures Markets* 14, 147–167.

Theissen, Erik, 2002, Price discovery in floor and screen trading systems, *Journal of Empirical Finance* 9, 455–474.

Veronesi, Pietro, 1999, Stock market overreaction to bad news in good times: A rational expectations equilibrium model, *Review of Financial Studies* 12, 975–1007.

Essay 4

INTRA-MARKET PRICE DISCOVERY:

ARE LOCAL OR FOREIGN INVESTORS MORE INFORMED?

Abstract

There is no agreement in existing literature on whether local or foreign investors are more informed. We study price discovery between local and foreign members. The results of an error correction model indicate that local brokers dominate the price discovery process for most stocks. Market capitalization and the existence of a cross-listing abroad are directly related to a higher foreign share of price discovery, as expected. However, trading volume, measured both as turnover and the number of trades, is inversely related to the foreign share of price discovery.

Keywords: price discovery, information asymmetry, home bias, equities.

JEL classification: G10, G14.

1 INTRODUCTION

Are local investors better informed than foreign investors? This is a controversial question, on which there is no general agreement in the existing literature. There are two opposing theoretical arguments for whether local or foreign investors are more informed. Factors favoring local investors are the physical proximity to domestic companies, and the relatively fewer trade and language barriers that they face. Hau (2001) provides empirical support for this view. Using data for the electronic Xetra market in Germany, he finds that traders located outside Germany, as well as outside German speaking countries, exhibit lower proprietary trading profits. Choe, Kho, and Stulz (2005) reach a similar conclusion for the Korean market. Other studies supporting a leading role for local investors include Brennan and Cao (1997), and Dvorak (2005). The opposite theoretical argument is that foreign investors possess superior analytical resources, leading to more informed investment decisions. Using a unique database of the identities of stock market investors, Grinblatt and Keloharju (2000) find that foreign investors are typically momentum traders, whereas local investors are contrarian investors. They also find that foreign investors outperform local households in the Finnish market. They interpret this finding as evidence of the higher degree of sophistication of foreign investors. Froot and Ramadorai (2008) reach a similar conclusion using a cross-section of 25 countries.

In this paper, we bring a novel point of view to the analysis of local and foreign trading in a stock market in that we use the tools of price discovery analysis in a single market setting. In particular, we use the Error Correction Method of Harris et al. (1995) to analyze shares of price discovery between the local and foreign traders in a single market, the OMX Helsinki Stock Exchange.

All major European stock exchanges use an electronic limit order book as their trading platform. This enables traders to connect to the system from anywhere, rather than being bound by the physical location of the exchange itself. This has led to a proliferation of remote members in many exchanges, including the Nordic markets. The OMX Helsinki Stock Exchange classifies its members according to the physical location of their trading floor. For the purposes of this study, we group the exchange members into two groups, those with a local office in Helsinki, and those with remote access only, from an office outside Helsinki. We then study the relative contributions of the local and the foreign traders to price discovery in large-cap stocks.

The existing literature on price discovery studies prices of the same security in two or more markets[34], either internationally or within one country with several competing exchanges, such as the United States. One of the first studies on price discovery of cross-listed stocks using high frequency data is Hasbrouck (1995), who studies price discovery for the 30 stocks included in the Dow Jones Industrial Average. Eun and Sabherwal (2003) study Canadian companies cross-listed in the United States. They find large differences in shares of price discovery. Three more recent studies are Grammig, Melvin, and Schlag (2005), Pascual, Pascual-Fuster, and Climent (2006), and Phylaktis and Korczak (2007), who study German, Spanish, and U.K. cross-listed stocks, respectively.

[34] See e.g. Karolyi (2006) for a comprehensive overview of the literature on cross-listings

Our study differs from most earlier papers on price discovery in that we analyze only one market, consisting of two types of exchange members, domestic and foreign[35]. The domestic segment consists of all trades where the buyer and the seller are a local broker, and the foreign segment of all trades with a foreign buyer and a foreign seller. Compared with a multiple markets setting, the main difference is that the law of one price is enforced by the nature of the trading process. Due to the nature of the trading process, no continuously matched trades can occur outside the boundaries set by the existing best bid and best ask in the limit order book. There is a related study, Sapp (2002), who studies price discovery for Deutschmark – U.S. dollar spot market. He studies the relative contributions of five large dealers in this over-the-counter market, and finds differences between the dealers, as well as between different market conditions.

Our main results are the following. First, applying an error correction model to 10-minute interval trade data of the OMXH25 index, we find that 28% of price discovery is attributed to the foreign segment of the market, and correspondingly 72% to the domestic brokers. This result is not a reflection of the market shares of the domestic and foreign segments of the market, as approximately 70% of all trading is done by foreigners. As a robustness check, we also perform the same analysis using 5 and 30 minute interval data; the results are largely similar. Second, estimating a cross-sectional regression of the foreign share of price discovery, we find that market capitalization is directly related to the foreign share of price discovery. Also, seven companies are cross-listed in either Stockholm and/or New York. A dummy variable, which takes the value 1 for cross-listed stocks, and the value 0 for stock with no cross-listing, is significant and positive, meaning that a cross-listing is directly related to a higher foreign share of price discovery. Both these results are expected. However, liquidity, measured as either turnover in euros, or the number of shares traded, is inversely related to the foreign share of price discovery. This result is contrary to our expectations, since foreign investors are usually seen as keen to own liquid securities in a relatively remote market, such as the Finnish one.

The rest of this paper is organized as follows. Chapter 2 discusses our data sample. Chapter 3 presents our method of dividing trades into local and foreign trades. Chapter 4 presents the error correction analysis. Chapter 5 contains a cross-sectional analysis of the results of the previous chapter. Chapter 6 concludes.

[35] Two related studies are Anand and Subrahmanyam (2008), and Kurov (2008). The former paper uses trader classifications in the Toronto Stock Exchange to study differences in price discovery between market intermediaries, such as specialists and proprietary traders, and their clients. Kurov (2008) studies differences between different types of traders in the E-mini futures markets, and their different contributions to both price discovery and volatility.

2 DATA SAMPLE

We obtain our dataset directly from the OMX Helsinki Stock Exchange. Our dataset contains all trades in all stocks traded in the Helsinki market during the sample period, the calendar year 2007 (January 1, 2007 to December 31, 2007).

Our sample contains the stocks that have at some point during 2007 been included in the OMXH25 index[36]. This index is the leading market index. It contains the 25 most traded share series in terms of their daily median turnover for the preceding calendar half-year on the Main list of the Helsinki Stock Exchange. The number of companies selected in this way is 28. Table 1 presents descriptive statistics for our sample companies.

Even if all companies in our sample are large measured by market capitalization, there are considerable differences among them. Nokia is by far the largest company, both by market capitalization and measured by turnover. The average turnover of Nokia is about ten times greater than that of the next most liquid stock, Fortum. There is a second tier of companies, which could be said to include Fortum, Metso, Neste Oil, Outokumpu, Sampo, Stora Enso, and UPM-Kymmene. In order to be able to measure the price discovery process using high frequency data, it is important that the stock be liquid enough. For most stocks, the average number of trades per trading day is in the thousands. The smallest average daily number of trades is for Nordea Bank, 574, which is approximately one trade per minute. The last column of Table 1 lists the cross-listings of the sample stocks. Four stocks, Metso, Nokia, Stora Enso, and UPM-Kymmene, are cross-listed in the NYSE in the form of ADRs. Nordea Bank and TeliaSonera have their main listing in Stockholm in the sense that most trading occurs in that exchange. In addition, Tietoenator and Stora Enso are also listed in Stockholm.

---

[36] The selection process for the OMXH25 index is as follows. Based on the largest median daily turnover, the index components are selected every six months, at the end of January and at the end of July. This results in three periods with different compositions of the index during our sample period: first, January 2007, second, February – July 2007, and third, August – December 2007.
The changes in index composition during our sample period are the following. Starting February 1, 2007, Orion Corporation B (ORNBV) and Outotec Oyj (OTE1V) were included, and Amer Sports Corporation (AMEAS) and OKO Bank Oyj A (OKOAS) were excluded. Starting August 1, 2007, Amer Sports Corp. (AMEAS), and Uponor Oyj (UNR1V) were included; Huhtamäki Oyj (HUH1V), and Orion B (ORNBV) were excluded.

Table 1 Descriptive statistics

This table presents descriptive statistics for our sample. *Ticker* is the official ticker code of each stock. *Company name* is the official name of the company. *DJ Industry classification* is the Dow Jones classification of the company's most important line of business. *Market cap* is the market capitalization as of 31st December 2007, in millions of euros. *Turnover* is the average daily turnover in millions of euros. *#trades* is the average daily number of trades. The last column, *CL*, indicates whether the company is cross-listed at the OMX Stockholm ('S'), or the NYSE ('N').
Our data source for the industry classification and the market capitalization is Thomson Financial; the figures for the daily turnover and the number of trades are from the exchange.
We report the mean, median standard deviation, minimum, maximum values, skewness, and excess kurtosis for the entire sample.

Ticker	Company name	DJ industry classification	Market cap	Turnover	#trades	CL
AMEAS	Amer Sports Corporation	Other Recreational Products & Services	1,329.1	11.27	697.5	
CGCBV	Cargotec 'B' OYJ	Machinery Makers	1,972.3	11.52	933.0	
ELI1V	Elisa OYJ	Fixed Line Communications	3,323.4	26.95	1,723.2	
FUM1V	Fortum OYJ	Electric Utilities	27,318.7	74.25	2,880.2	
HUH1V	Huhtamaki OYJ	Containers & Packaging	815.5	5.93	582.5	
KCR1V	Konecranes OYJ	Machinery Makers	1,379.6	14.06	1,132.7	
KESBV	Kesko Oyj	Food Retailers & Wholesalers	3,687.8	21.82	1,585.0	
KNEBV	Kone OYJ	Factory Equipment	6,026.6	17.95	1,267.6	
MEO1V	Metso Oyj	Factory Equipment	5,281.7	58.03	2,642.0	N
MRLBV	M-Real Oyj	Paper Products	1,066.5	9.45	825.9	
NDA1V	Nordea Bank Ab	Banks, All	29,631.8	21.00	573.6	S
NES1V	Neste Oil Oyj	Oil, Integrated majors	6,175.0	47.90	2,350.7	
NOK1V	Nokia Corporation	Communications Technology	101,994.6	740.52	8,700.5	N
NRE1V	Nokian Renkaat Oyj	Tires & Rubber	2,974.9	21.19	1,517.0	
OKOAS	OKO Bank Oyj	Banks, All	2,657.8	8.10	988.7	
ORNBV	Orion Corporation	Pharmaceutical Companies	2,264.4	7.06	785.0	
OTE1V	Outotec OYJ	Factory Equipment	1,579.2	20.01	1,010.6	
OUT1V	Outokumpu Oyj	Steel	3,820.0	51.53	2,776.1	
RTRKS	Rautaruukki Oyj	Steel	4,113.0	33.77	2,078.6	
SAMAS	Sampo Oyj	Insurance, Property & Casualty	10,381.7	64.27	2,416.1	
STERV	Stora Enso OYJ	Paper products	8,075.3	64.64	2,595.4	N, S
SWS1V	Sanoma Corporation	Publishing	3,196.2	8.06	766.8	
TIE1V	Tietoenator OYJ	Diversified Technology Services	1,100.7	24.77	1,536.6	S
TLS1V	TeliaSonera Ab	Wireless Communications	28,751.1	15.32	564.9	S
UNR1V	Uponor Oyj	Building Materials	1,260.6	9.45	998.9	
UPM1V	UPM-Kymmene OYJ	Paper Products	7,083.7	65.89	2,617.2	N
WRTBV	Wärtsilä Oyj	Machinery Makers	4,999.1	26.74	1,533.1	
YTY1V	YIT Oyj	Heavy Construction	1,907.0	21.79	1,567.7	
Mean			9791.7	53.69	1773.1	
Median			3505.6	21.49	1525.1	
Std. dev.			19837.8	136.22	1551.5	
Minimum			815.5	5.93	564.9	
Maximum			101994.6	740.52	8700.5	
Skewness			4.1	5.10	3.5	
Kurtosis			18.4	26.55	15.2	

We only analyze automatically matched trades during the continuous trading period. The reason for this is two-fold. First, by definition, continuous trades occur in real time, when a buyer and a seller are automatically matched in the electronic limit order book. Block trades and contract trades are, by contrast, agreed in advance, and only printed at some later point in time. They are thus not conducive to a study of price discovery using high frequency data. Second, the number of block trades and contract trades is relatively small compared to the number of automatically matched continuous trades. We also exclude the opening and closing auctions, as well as any other types of trades than continuous trades. Specifically, we exclude pre-arranged contract trades and block trades, as well as all trades in the pre- and post-market trading periods.

For our analysis of price discovery, we construct 10-minute interval data, with 5-minute and 30-minute intervals as a robustness check. This results in 50 periods per trading day (100, and 16 periods, for the 5 and 30 minute intervals, respectively)[37]. For each interval we take the most recent trading price within the same day. The start of a trading day is a special case, however. If there are no trades available in the first one or more intervals of the day, we exclude these intervals.

[37] OMX Helsinki observes the following trading times. After an opening auction, continuous trading begins at 10 a.m. local time (9 a.m. CET). Continuous trading ends at 6.20 p.m. local time (5.20 p.m. CET), after which there is a closing auction. Since we use prices, not returns, we do not lose the first interval. The exchange's home page at http://www.nasdaqomxnordic.com/ provides more information.

3 LOCAL AND FOREIGN TRADING

In this paper, we use a novel approach to construct two market segments within one market. We construct a domestic and a foreign market segment in three different ways, based on the identities of the buying and selling broker.

We use the geographical location of the trading desk of the broker as the defining criterion in grouping brokers. If the broker has an office in Helsinki, we classify this broker as a local broker. These brokers are typically either Finnish investment banks, or branch offices of Scandinavian banks and investment banks. The foreign group consists of the so-called remote members. These are typically large investment banks with trading desks in London[38]. We also use a second classification, in which the definition of the first group is narrower, in that we exclude foreign firms from the group. This means that our group of local traders consists of local firms with a local office in Helsinki, and our group of foreign traders is all remote brokers with no physical presence in Helsinki. Foreign banks with a Helsinki office are excluded in this second classification.

Table 2 lists all brokers, their geographical classifications, and their market shares. It is clear from Table 2, that foreign brokers dominate the market in trading volume. Measured both as the number of trades, and as total turnover, these foreign (remote) brokers control approximately 70% of the market. The top 10 members by turnover handle just over half of all trading, measured by turnover in euros (51%). There is only one local broker (Evli) among these ten largest members by market share. It is notable that all the other large brokers are remote members.

---

[38] The categories, according to OMX Helsinki are the following (source: the exchange's home page http://www.nasdaqomxnordic.com/):

1. Domestic banks and securities companies
2. Foreign branches domestically
3. Remote Members

Table 2 Brokers and their market shares

This table presents all brokers active on the OMX Helsinki market in 2007. *ID* is the broker ID. *Category* is the official grouping by the exchange:

A Domestic banks and securities companies
B Foreign branches domestically
C Remote Members

For the purposes of this study, groups A and B (with a local office) are classified as *local*, and C (typically large London-based investment banks) as *remote*. *Turnover* and *Trades* are the total turnover over the entire year 2007 in millions of euros, and the total number of trades, in thousands, respectively. The data are provided by the OMX Helsinki Stock Exchanges.

Rank					Turnover		Trades	
Turnover	# of trades	Member name	ID	Category	MEUR	%	000's	%
1	1	Morgan Stanley & Co. International Ltd.	MSI	C	58,310.9	7.33	2,566.1	8.80
2	13	Evli Bank Abp	EVL	A	45,996.6	5.78	920.7	3.16
3	4	Glitnir banki hf.	GLB	C	45,982.1	5.78	1,621.4	5.56
4	7	Goldman Sachs International	GSI	C	43,538.7	5.47	1,270.6	4.36
5	5	Deutsche Bank AG	DBL	C	39,381.3	4.95	1,548.4	5.31
6	3	Credit Suisse Securities (Europe) Ltd	CSB	C	39,332.9	4.94	1,877.7	6.44
7	10	Merrill Lynch International	MLI	C	37,373.9	4.70	1,156.5	3.96
8	2	Lehman Brothers International	LBI	C	32,678.9	4.11	2,039.6	6.99
9	14	Citigroup Global Markets Limited	SAB	C	32,131.3	4.04	912.7	3.13
10	9	UBS Limited	UBS	C	31,867.4	4.01	1,174.7	4.03
11	12	Société Générale S.A.	SGP	C	29,738.8	3.74	996.8	3.42
12	6	Nordea Bank Finland Plc	NRD	A	28,278.2	3.55	1,527.9	5.24
13	20	Carnegie Investment Bank AB	CAR	B	27,817.7	3.50	609.8	2.09
14	8	eQ Pankki Oyj	EQB	A	25,792.9	3.24	1,215.9	4.17
15	16	Svenska Handelsbanken AB	SHB	B	25,568.7	3.21	741.0	2.54
16	21	Skandinaviska Enskilda Banken AB	ENS	B	25,097.7	3.16	570.4	1.96
17	11	OKO Pankki Oyj	OPS	A	21,101.5	2.65	1,143.4	3.92
18	15	Danske Bank A/S Helsinki Branch	MDT	B	19,965.0	2.51	766.6	2.63
19	17	Crédit Agricole Cheuvreux Nordic AB	CDV	C	19,500.0	2.45	678.3	2.33
20	23	BNP Paribas Arbitrage SNC	BPP	C	15,715.4	1.98	513.8	1.76
21	26	Kaupthing Bank hf	KFI	A	15,654.5	1.97	294.8	1.01
22	22	NeoNet Securities AB	NEO	C	13,118.3	1.65	566.3	1.94
23	25	JP Morgan Securities Ltd	JPM	C	10,423.1	1.31	308.3	1.06
24	18	Barclays Capital Securities Limited Plc	BRC	C	10,044.9	1.26	658.2	2.26
25	24	E*Trade Sverige AB	ETS	C	9,029.0	1.14	386.2	1.32
26	37	Dresdner Kleinwort Securities Limited	DRS	C	8,271.7	1.04	103.3	0.35
27	29	ABN AMRO Bank N.V., London Branch	ABN	C	7,411.9	0.93	188.3	0.65
28	28	ABG Sundal Collier AB	ABC	C	7,172.4	0.90	203.0	0.70
29	19	Nordnet Bank AB	NON	C	6,793.9	0.85	650.2	2.23

30	36	Optiver VOF	OPV	C	6,534.7	0.82	116.3	0.40
31	38	Pan Capital AB	PAN	C	5,349.0	0.67	99.8	0.34
32	40	Öhman J:or Fondkommission AB, E.	OHM	C	5,248.0	0.66	78.3	0.27
33	33	Bear, Stearns International Ltd	BSI	C	5,199.0	0.65	148.0	0.51
34	45	HSBC Bank plc	HBC	C	5,084.5	0.64	36.7	0.13
35	30	Danske Bank A/S	DDB	C	4,836.7	0.61	168.0	0.58
36	27	Instinet Europe Limited	INT	C	4,783.8	0.60	275.2	0.94
37	34	ABN AMRO Bank N.V., Helsinki Branch	ABF	A	4,697.2	0.59	129.7	0.44
38	42	Van der Moolen Effecten Specialist B.V	VDM	C	4,311.5	0.54	61.2	0.21
39	31	Ålandsbanken Abp	AAL	A	2,433.4	0.31	155.6	0.53
40	39	IMC Securities B.V	IMC	C	2,154.0	0.27	81.6	0.28
41	47	Swedbank AB	SWB	C	2,097.8	0.26	26.8	0.09
42	32	Timber Hill Europe AG	TMB	C	1,420.0	0.18	151.6	0.52
43	35	Commerzbank AG	CBK	C	1,332.1	0.17	122.2	0.42
44	44	HQ Bankaktiebolag	HQB	C	1,251.3	0.16	43.7	0.15
45	43	United Bankers Securities Limited	UB	A	1,142.0	0.14	58.3	0.20
46	49	Cazenove	CAZ	C	1,017.3	0.13	22.2	0.08
47	48	Knight Equity Markets International Ltd	KEM	C	840.9	0.11	23.1	0.08
48	51	Landsbanki Íslands hf.	LAI	C	778.3	0.10	12.6	0.04
49	41	Aktia Säästöpankki Oyj	AKT	A	486.4	0.06	66.8	0.23
50	50	Erik Penser Fondkommission AB	EPF	C	464.5	0.06	20.4	0.07
51	46	Bankaktiebolaget Avanza	AVA	C	358.7	0.05	35.9	0.12
52	52	Kaupthing Bank Sverige AB	KAB	C	154.8	0.02	6.4	0.02
53	58	All Options International B.V.	AOI	C	95.8	0.01	1.4	0.00
54	56	DnB NOR Bank ASA	DNM	C	57.6	0.01	2.3	0.01
55	55	Saga Capital fjárfestingarbanki hf.	SGA	C	47.1	0.01	2.7	0.01
56	53	Nyenburgh Beheer B:V	NYE	C	44.5	0.01	3.3	0.01
57	54	Remium AB	REM	C	44.5	0.01	3.2	0.01
58	57	Sydbank A/S	SYD	C	37.3	0.00	1.9	0.01
59	61	Arctic Securities ASA	ARC	C	32.8	0.00	0.8	0.00
60	59	MP Fjárfestingarbanki hf.	MPB	C	31.7	0.00	1.2	0.00
61	62	Parex banka	PRX	C	3.5	0.00	0.1	0.00
62	60	Straumur-Burðarás Fjárfestingabanki hf.	STR	C	1.8	0.00	1.2	0.00
63	63	NordVest verðbréf hf.	NOV	C	0.8	0.00	0.0	0.00
		Total			795,462.8	100	29,169.3	100
		Total, local members			244,031.7	30.68	8,200.9	28.11
		Total, remote members			551,431.1	69.32	20,968.3	71.89
		Total, top 10			406593.9	51.1	15998.7	54.8

In order to classify trades as domestic and foreign, our primary classification is as follows. We infer trade direction for each trade using the tick test.[39] We then use the identity of the trade initiator to classify trades as local or foreign.

As a robustness test, we also use two other methods for classifying trades. The first additional method is as follows. Any trade with a local buyer and a local seller is a local trade. Similarly, any trade with a foreign member as both the buyer and the seller is a remote trade. We exclude all trades where the buyer is local and the seller is foreign, or vice versa. This results in excluding approximately half of all trades from the sample. Also, we do not make a distinction between internal trades, i.e. trades with the same buyer and seller, and trades made between two different brokers. Arguably, this would be necessary if our sample included block and contract trades. Since only automatically matched continuous trades are included, internalization of trades should not influence our results.

We also use a second additional trade classification algorithm. In this algorithm, no trades are discarded. Rather, we classify trades in two additional ways: according to the identity of the buyer and according to the identity of the seller. This means classifying all trades where the buyer is local as local trades, irrespective of the identity of the seller. Similarly, all trades with a foreign buyer are then classified as foreign trades. These results for both additional classification algorithms are largely similar, and are available upon request[40].

---

[39] We are unable to employ the most widely used methodology for inferring trade direction, i.e. the Lee and Ready (1991) algorithm, since we lack quote data. See e.g. Finucane (2000) for a detailed discussion of different methodologies for determining trade direction.

[40] The most widely used method to sign trades is the Lee-Ready (1991) algorithm. However, the use of this method is not possible using the present dataset, however, since we only have trade data. The Lee-Ready algorithm requires the use of both trade and quote data.

4 ANALYSIS OF PRICE DISCOVERY

Price discovery is a central contribution of an exchange. Hasbrouck (1995) describes price discovery as the process by which markets impound new information and find the equilibrium price. We study price discovery within one electronic limit order book market, OMX Helsinki. This is in contrast to earlier work in price discovery, where two separate markets are studied[41]. Instead, we analyze the OMX Helsinki market as consisting of two separate sets of traders, the locals and the foreigners. As a proxy for these two sets of traders we use the identities of the brokers conducting the trading.

There are two main methodological approaches to measuring price discovery for cross-listed stocks: the error correction approach, also known as the "Permanent-Temporary" decomposition (PT), of Harris et al. (1995) and the so-called "Information Shares" (IS) method of Hasbrouck (1995). The models are directly related and provide similar results in most cases, see e.g. de Jong (2002), Baillie, Booth, Tse, and Zabotina (2002). Fundamentally, both methods decompose the impact of a price innovation into permanent and temporary components. The Hasbrouck method employs Choleski factorization of the covariance matrix of price innovations in the two exchanges, which requires that the prices be ordered. Unfortunately, this also means that the information shares obtained by this method are not unique. Rather, the methodology results in upper and lower bounds for the information shares, as they depend on the ordering of prices (see e.g. Hasbrouck (1995), Eun and Sabherwal (2003), and Lehmann (2002)). The difference between the upper and lower bounds is small when using extremely high frequency data, such as the one-second interval data in Hasbrouck (1995). On the other hand, Booth, Lin, Martikainen, and Tse (2002) and Huang (2002) find that using intervals in the range of a few minutes, the difference between the upper and lower bounds obtained for the information shares can be significant. In our study we are unable to use very high frequency data, since not all stocks in our sample are extremely liquid in the U.S. market. For this reason we conclude that the Error Correction Methodology of Harris et al. (1995) serves the purposes of our study.

Both methods are widely used. Eun and Sabherwal (2003), and Phylaktis and Korczak (2007) use the Harris (1995) error correction methodology; Hasbrouck (2002), Grammig, Melvin, and Schlag (2005), and Pascual, Pascual-Fuster, and Climent (2006) use the Hasbrouck (1995) information shares method.

4.1. The Error Correction Model

In their seminal paper, Engle and Granger (1987) provide the link from two or more price series to the Error Correction Model. It is necessary for the price series to be cointegrated[42] for the Error Correction Model to be applicable (see, e.g. Harris et al. (1995) and Eun and Sabherwal (2003) for a discussion). In the case of a stock that is

[41] Studies on price discovery in multiple markets setting using low frequency (daily) data include Lau and Diltz 1994, Lieberman, Ben-Zion, and Hauser (1999), Kim, Szakmary, and Mathur (2000), and Wang, Rui, and Firth (2002). Studies using high-frequency (intraday) data include Hasbrouck (2002), Eun and Sabherwal (2003), Grammig, Melvin, and Schlag (2005), and Pascual, Pascual-Fuster, and Climent (2006), and Phylaktis and Korczak (2007).

[42] A random walk, or more formally, an integrated process *I(1)*, is nonstationary. However, it has the property that the differences of that process are stationary. Two random walk processes are said to be cointegrated, when a linear combination of them is stationary.

traded simultaneously on two exchanges, the law of one price keeps the two prices of the stock within certain arbitrage limits. These limits are imposed by factors such as the cost of converting common stocks to ADRs and the potential exchange rate risk. In this paper, we study price discovery for two segments of one market. It is therefore clear that the local and foreign prices cannot deviate from each other in a significant way.

We estimate the following equations:

$$\Delta P_t^R = \alpha_0^R + \alpha^R \left(P_{t-1}^L - P_{t-1}^R \right) + \sum_{i=1}^{p} \gamma_i \Delta P_{t-i}^R + \sum_{i=1}^{p} \delta_i \Delta P_{t-i}^L + \varepsilon_t^R \tag{1}$$

$$\Delta P_t^L = \alpha_0^L + \alpha^L \left(P_{t-1}^L - P_{t-1}^R \right) + \sum_{j=1}^{p} \gamma_j \Delta P_{t-j}^R + \sum_{j=1}^{p} \delta_j \Delta P_{t-j}^L + \varepsilon_t^L \tag{2}$$

where i and P^R are transaction prices of the local and the remote brokers, respectively; α_0 are constant terms, α^R and α^L are the main error correction parameters, since they reflect the extent to which the price series responds to a price innovation; the $\gamma_{i,j}$ parameters reflect idiosyncratic adjustment of P_t to disparities in shocks to private value that cause the cointegrated series to diverge; p is the lag length determined using the Schwarz Information Criterion[43]; and ε_t is a zero-mean, covariance-stationary random disturbance, which is assumed to be identically distributed but may be autocorrelated.

We measure the share of total adjustment that can be attributed to the remote brokers using a measure first proposed by Schwarz and Szakmary (1994), and used by, among others, Theissen (2002b), and Eun and Sabherwal (2003). This is simply the local adjustment coefficient as a share of the summed adjustment coefficients:

$$PD = \frac{\left|\alpha^L\right|}{\left|\alpha^L\right| + \left|\alpha^R\right|}. \tag{3}$$

A high value of the variable *PD* means that the coefficient α^L is high. The interpretation of this is that the local brokers do most of the adjusting to price innovations. In the hypothetical case of no feedback from the remote brokers to the local brokers, local brokers perform no price adjustment at all, α^L is zero, and thus *PD* is zero.

4.2. Results

We estimate the shares of price discovery using 10 minute interval data. The estimates of α^L and α^R above are of the greatest interest to us. These coefficients give us the average adjustment of each price series to an external shock or price innovation.

[43] We determine the appropriate lag length for each stock using the Schwarz Information Criterion. We employ the rule of thumb of a maximum lag length of (where T is the number of observations), resulting in a maximum lag length of 20. The optimum lags are 1 or 2 for most stocks, with a maximum value of 3. We do not report these results.

Table 3 presents our estimates of the α coefficients. We see that all values for the local price adjustment, α^L, are positive, and all values for the remote price adjustment, α^R, are negative. This is as expected, because of the way the Error Correction Model of Equations (1) and (2) is constructed: the local price less the remote price. This means that when the local price is lower than the remote price, the local price needs to rise, and/or the remote price to decline, for equilibrium to be restored. Similarly, when the local price is higher than the remote price, the reverse needs to happen for prices to move towards restoring the equilibrium prices in the two marketplaces.

Table 3 Shares of price discovery

This table presents shares of price discovery within the Helsinki Stock Exchange. Under the heading *Price adjustment coefficients* we report estimated values for α^L and α^R. These are the local and remote adjustment coefficients in the Error Correction Model of Equations (1) and (2). *Remote Share PD* is the share of price discovery of remote members, calculated as $PD = \frac{|\alpha^L|}{|\alpha^L| + |\alpha^R|}$. We report estimation results using three different sets of data: 5, 10 and 30 minute interval data. The 10-minute interval results are our main result; 5 and 30 minute intervals are for robustness test purposes. T-statistics for the hypothesis of $\alpha = 0$ are in parenthesis.
We report shares of price discovery for two different methods of categorizing trades:

- Panel A: members groups 1 and 2, i.e. domestic companies and foreign banks with an office in Helsinki as locals, and group 3, remote members as foreign traders.
- Panel B: group 1, domestic companies, form the local segment of the market; group 2, foreign banks with an office in Helsinki, are excluded, and group 3, remote members, are grouped as foreign traders.

Panel A: Domestic and foreign with local presence grouped as locals

	Price adjustment coefficients												Remote share PD		
	5 minute intervals				10-minute interval data				30-minute interval data				5 min	10 min	30 min
Stock	α^L	t-stat	α^R	t-stat	α^L	t-stat	α^R	t-stat	α^L	t-stat	α^R	t-stat	PD	PD	PD
AMEAS	0.040	11.03	-0.148	-40.28	0.084	12.29	-0.250	-37.19	0.165	8.75	-0.443	-24.19	21.1%	25.2%	27.2%
CGCBV	0.074	17.18	-0.190	-43.26	0.077	8.97	-0.339	-39.33	0.208	7.57	-0.511	-19.21	28.2%	18.5%	28.9%
ELI1V	0.074	14.35	-0.302	-59.65	0.115	11.47	-0.458	-47.72	0.219	7.19	-0.624	-21.35	19.7%	20.1%	26.0%
FUM1V	0.089	16.06	-0.357	-66.13	0.157	14.69	-0.475	-47.27	0.220	7.06	-0.616	-20.42	19.9%	24.8%	26.3%
HUH1V	0.067	16.68	-0.138	-33.57	0.117	14.70	-0.229	-28.92	0.236	10.27	-0.407	-18.32	32.8%	33.8%	36.7%
KCR1V	0.064	14.27	-0.220	-48.76	0.223	8.76	-0.499	-20.63	0.223	8.76	-0.499	-20.63	22.6%	30.9%	30.9%
KESBV	0.060	14.30	-0.201	-48.47	0.115	13.49	-0.316	-39.46	0.248	9.92	-0.478	-20.52	23.0%	26.6%	34.1%
KNEBV	0.043	10.24	-0.232	-54.90	0.095	11.52	-0.366	-46.78	0.204	8.84	-0.598	-28.06	15.5%	20.6%	25.5%
MEO1V	0.069	12.03	-0.313	-55.22	0.131	11.37	-0.417	-37.48	0.173	5.16	-0.627	-19.33	18.0%	23.9%	21.6%
MRLBV	0.039	11.92	-0.149	-43.97	0.066	10.25	-0.240	-37.65	0.188	9.61	-0.405	-22.17	20.9%	21.6%	31.7%
NDA1V	0.290	45.30	-0.151	-24.20	0.473	38.65	-0.196	-16.19	0.707	19.15	-0.202	-5.46	65.8%	70.7%	77.8%
NES1V	0.118	19.14	-0.377	-64.04	0.179	14.70	-0.499	-43.40	0.287	8.50	-0.604	-18.68	23.8%	26.4%	32.2%
NOK1V	0.271	15.34	-0.577	-33.44	0.408	11.40	-0.539	-15.24	0.579	5.14	-0.439	-3.91	32.0%	43.1%	56.8%

NRE1V	0.079	13.79	-0.317	-56.76	0.142	12.92	-0.470	-44.76	0.205	6.04	-0.654	-20.10	20.0%	23.3%	23.8%
OKOAS	0.058	14.34	-0.211	-50.16	0.100	12.64	-0.337	-42.96	0.232	10.00	-0.499	-22.63	21.7%	23.0%	31.8%
ORNBV	0.075	20.68	-0.136	-40.30	0.156	21.28	-0.229	-34.86	0.258	12.74	-0.458	-24.41	35.4%	40.5%	36.0%
OTE1V	0.071	14.92	-0.231	-49.02	0.118	12.56	-0.369	-40.44	0.239	8.20	-0.511	-18.42	23.5%	24.3%	31.8%
OUT1V	0.107	16.94	-0.366	-60.37	0.146	11.87	-0.515	-43.61	0.221	5.78	-0.655	-17.57	22.6%	22.1%	25.2%
RTRKS	0.065	12.01	-0.282	-52.69	0.112	10.29	-0.417	-40.23	0.239	7.19	-0.584	-18.15	18.9%	21.1%	29.0%
SAMAS	0.091	15.96	-0.346	-62.86	0.152	13.59	-0.480	-45.52	0.301	9.04	-0.580	-18.36	20.8%	24.1%	34.1%
STERV	0.094	15.51	-0.344	-58.58	0.151	12.88	-0.478	-42.83	0.304	9.04	-0.594	-18.23	21.5%	24.0%	33.9%
SWS1V	0.050	13.90	-0.182	-47.44	0.092	12.86	-0.302	-41.92	0.193	9.49	-0.475	-24.18	21.5%	23.4%	28.9%
TIE1V	0.094	14.74	-0.265	-41.67	0.149	11.90	-0.382	-31.13	0.287	7.65	-0.485	-13.35	26.2%	28.1%	37.1%
TLS1V	0.261	45.51	-0.138	-25.93	0.398	36.66	-0.207	-19.99	0.603	20.26	-0.265	-8.94	65.5%	65.8%	69.5%
UNR1V	0.049	11.61	-0.211	-48.47	0.091	10.88	-0.328	-39.61	0.141	5.41	-0.588	-23.50	18.8%	21.7%	19.3%
UPM1V	0.091	14.98	-0.393	-65.47	0.116	9.92	-0.550	-48.63	0.268	7.65	-0.635	-18.74	18.8%	17.4%	29.6%
WRTBV	0.042	8.64	-0.254	-51.18	0.064	6.60	-0.397	-41.58	0.136	4.52	-0.626	-21.49	14.2%	13.9%	17.9%
YTY1V	0.066	12.89	-0.282	-54.53	0.102	10.15	-0.418	-42.67	0.228	7.76	-0.551	-19.28	19.1%	19.6%	29.2%
Average													25.4%	27.8%	33.3%
Median													21.5%	23.9%	30.3%
Minimum													14.2%	13.9%	17.9%
Maximum													65.8%	70.7%	77.8%

Panel B: foreign banks with a local presence excluded; groups 1 and 3 only.

	Price adjustment coefficients												Remote share PD		
	5 minute intervals				10-minute interval data				30-minute interval data				5 min	10 min	30 min
Stock	α^L	t-stat	α^R	t-stat	α^L	t-stat	α^R	t-stat	α^L	t-stat	α^R	t-stat	PD	PD	PD
AMEAS	0.079	18.39	-0.211	-48.63	0.092	11.64	-0.345	-42.67	0.226	9.93	-0.492	-22.22	27.2%	21.1%	31.5%
CGCBV	0.132	22.84	-0.290	-51.31	0.169	14.80	-0.445	-40.12	0.367	10.37	-0.531	-15.37	31.3%	27.5%	40.9%
ELI1V	0.114	15.94	-0.425	-61.39	0.156	11.30	-0.532	-40.13	0.315	7.61	-0.583	-14.57	21.1%	22.7%	35.1%
FUM1V	0.216	24.03	-0.564	-65.54	0.262	15.48	-0.636	-39.11	0.276	5.60	-0.688	-14.14	27.7%	29.2%	28.6%
HUH1V	0.122	21.71	-0.214	-39.17	0.190	17.76	-0.321	-30.89	0.355	11.45	-0.484	-16.16	36.3%	37.2%	42.3%
KCR1V	0.150	22.15	-0.406	-61.45	0.195	14.88	-0.546	-43.26	0.282	7.39	-0.643	-17.23	27.0%	26.3%	30.5%
KESBV	0.105	16.41	-0.416	-64.81	0.144	11.26	-0.558	-44.54	0.232	6.20	-0.676	-18.28	20.1%	20.6%	25.6%

	Price adjustment coefficients												Remote share PD		
	5 minute intervals				10-minute interval data				30-minute interval data				5 min	10 min	30 min
Stock	α^L	t-stat	α^R	t-stat	α^L	t-stat	α^R	t-stat	α^L	t-stat	α^R	t-stat	PD	PD	PD
KNEBV	0.082	13.93	-0.433	-73.16	0.131	11.38	-0.574	-51.84	0.253	7.48	-0.640	-19.53	16.0%	18.6%	28.3%
MEO1V	0.169	18.27	-0.526	-59.23	0.275	14.96	-0.572	-32.33	0.409	7.24	-0.574	-10.35	24.4%	32.5%	41.6%
MRLBV	0.082	18.14	-0.221	-49.00	0.136	15.67	-0.318	-37.75	0.269	10.77	-0.426	-17.77	27.1%	29.9%	38.7%
NDA1V	0.320	40.84	-0.284	-37.78	0.395	25.98	-0.414	-27.82	0.538	11.81	-0.431	-9.52	53.0%	48.9%	55.5%
NES1V	0.247	25.19	-0.550	-58.55	0.368	19.12	-0.552	-29.62	0.504	8.72	-0.488	-8.54	31.0%	40.0%	50.8%
NOK1V	0.304	14.19	-0.633	-29.90	0.412	9.34	-0.556	-12.68	0.794	5.50	-0.194	-1.34	32.5%	42.6%	80.4%
NRE1V	0.134	16.84	-0.463	-60.27	0.209	13.88	-0.565	-38.80	0.296	6.22	-0.651	-13.92	22.5%	27.0%	31.3%
OKOAS	0.129	22.60	-0.329	-59.00	0.200	18.02	-0.440	-41.48	0.338	10.82	-0.542	-18.08	28.2%	31.3%	38.4%
ORNBV	0.111	24.15	-0.192	-44.46	0.216	23.79	-0.258	-31.17	0.318	12.52	-0.407	-17.20	36.7%	45.6%	43.9%
OTE1V	0.151	22.35	-0.334	-52.27	0.241	18.08	-0.448	-35.57	0.354	9.16	-0.560	-15.04	31.1%	35.0%	38.7%
OUT1V	0.325	30.14	-0.521	-50.71	0.281	13.24	-0.645	-30.91	0.370	5.44	-0.577	-8.49	38.4%	30.3%	39.1%
RTRKS	0.196	20.26	-0.527	-56.24	0.282	14.95	-0.588	-32.05	0.341	5.92	-0.593	-10.43	27.1%	32.4%	36.5%
SAMAS	0.232	24.73	-0.528	-58.77	0.340	18.61	-0.544	-30.82	0.478	8.90	-0.481	-9.10	30.5%	38.4%	49.8%
STERV	0.126	16.27	-0.416	-55.78	0.162	10.99	-0.522	-36.61	0.235	5.49	-0.641	-15.24	23.2%	23.7%	26.8%
SWS1V	0.104	20.91	-0.290	-57.90	0.171	17.83	-0.408	-43.96	0.288	11.20	-0.524	-20.93	26.4%	29.5%	35.5%
TIE1V	0.088	12.96	-0.266	-39.30	0.146	10.99	-0.374	-28.81	0.240	6.13	-0.466	-12.16	24.9%	28.0%	34.0%
TLS1V	0.325	45.30	-0.281	-41.29	0.427	32.38	-0.360	-28.09	0.515	13.99	-0.444	-12.24	53.6%	54.3%	53.7%
UNR1V	0.109	19.60	-0.294	-53.42	0.177	16.13	-0.387	-36.53	0.268	7.60	-0.541	-15.91	27.0%	31.3%	33.1%
UPM1V	0.168	19.16	-0.504	-59.74	0.189	11.32	-0.644	-40.00	0.352	7.10	-0.617	-12.67	25.0%	22.7%	36.3%
WRTBV	0.111	15.98	-0.399	-58.42	0.159	11.87	-0.525	-39.81	0.220	5.04	-0.648	-15.04	21.7%	23.3%	25.4%
YTY1V	0.139	18.55	-0.391	-53.81	0.216	14.75	-0.479	-33.98	0.304	7.19	-0.544	-13.12	26.3%	31.1%	35.9%
Average													29.2%	31.5%	38.9%
Median													27.1%	30.1%	36.4%
Minimum													16.0%	18.6%	25.4%
Maximum													53.6%	54.3%	80.4%

Table 3 also presents our estimates of the foreign share of price discovery, calculated on the basis of our error correction model estimates, using Equation (3). With the exception of two companies cross-listed in Stockholm, Nordea Bank and TeliaSonera, the majority of price discovery occurs among local brokers. These two companies are also the exception in the sense that they are headquartered in Sweden, rather than in Finland. Nokia, by far the most liquid and valuable company among our sample stocks, also exhibits a large foreign share of price discovery, 43%. This is also as expected, since Nokia is one of the most liquid stocks in all of Europe. It is cross-listed in several European markets, as well as the NYSE.

The dominant position of local brokers in price discovery is not a result of a greater market share. On the contrary, based on market share data presented in Table 2, it is clear that the foreign market, represented by the remote brokers, is the dominant one in terms of volume. Eight of the ten largest members by market share are remote members. Also, the total market share of the remote members is 69% of turnover, and 72% in numbers of traded shares.

As a robustness test, we classify traders into the domestic and foreign segments in two different ways. Panels A and B in Table 3 reflect the differences between these two classifications. Panel A presents results from our first classification, where domestic firms and foreign brokers with a local office are defined as domestic market participants, and the foreign segment consists of the remote members. The second classification is similar, except that the foreign brokers with a local office are excluded. The foreign shares of price discovery are consistently higher in this case, with an average of 31.5% compared to 27.8%, when using 10-minute interval data.

These results are interesting as a comparison with Grinblatt and Keloharju (2000). They study the Finnish market in 1995 and 1996, using a dataset containing information about individual investors and their paper contains two main findings: that foreign investors are mainly momentum investors, whereas domestic investors are contrarians. Foreign investors also seem to outperform the local investors, even after controlling for differences in investor behavior.

Unfortunately, it is not possible to perform the price discovery analysis of this paper using the same dataset. The reason is the very low liquidity of the market during those years. The sample used by Grinblatt and Keloharju (2000) contains 16 stocks, most of which only trade once an hour on average. It is clear that performing an analysis of price discovery using even 30-minute interval data would give questionable results in this case.

We are therefore unable to provide a direct comparison with the analysis of Grinblatt and Keloharju (2000). One explanation is that especially during the second sample year, 1996, the market began its rise, which later culminated in the year 2000. This rise was also characterized by the increasing presence of foreign investors in the Finnish market. Thus, if these foreign investors arrived in the market, being in this way net buyers and momentum investors during a bull market, they would make a profit.

5 CROSS-SECTIONAL REGRESSION

In the previous section we estimate shares of price discovery for local and foreign traders. We find that there are large differences between stocks, but in general the local traders are responsible for the majority of price discovery. In this section, we turn to the analysis of the factors affecting the differences between different companies in shares of price discovery. We test for differences in shares of price discovery between cross-listed and non-cross-listed stocks on the one hand, and between large-cap and small-cap firms on the other. Two-sided, non-paired t-tests fail to reject the null of equal means.

Our explanatory variables are logarithmic market capitalization, the average daily turnover, the average daily number of trades, and a dummy variable, which takes the value of 1 when the company is cross-listed in either OMX Stockholm or the NYSE.

A general finding in price discovery literature is that trading volume is an important explanatory variable for differences in shares of price discovery. Hasbrouck (1995) finds that the NYSE contribution to price discovery is positively and statistically significantly related to the share of trading taking place at the NYSE in that particular stock. Eun and Sabherwal (2003) find that the relative trading volumes at the Toronto Stock Exchange and the NYSE are a significant explanatory variable for the NYSE share of price discovery in Canadian cross-listed stocks. In our study the turnover closely follows market capitalization, with a correlation of 72% between the two variables.

High correlations between our explanatory variables may lead to multicollinearity, if highly correlated variables are simultaneously included in regressions. Table 4 presents results for the correlation analysis. As expected, the two measures of turnover, trading volume in euros and the number of daily trades, are highly correlated. Also, a high market capitalization is highly correlated with turnover measured in euros. Multicollinearity may often lead to increased standard errors in the regression coefficients, resulting in less significant or even insignificant estimation coefficients. Since the results of our estimation are significant, multicollinearity does not seem to pose any great problems for us.[44] A test of multicollinearity only yields high readings when both the number of trades and turnover in euros are included as explanatory variables. When we only use turnover as a variable, there is no multicollinearity problem.

[44] We test for multicollinearity using variance inflation factors (VIF). This is a measure given by

$$\frac{1}{\left(1-R_i^2\right)},$$

where R_i^2 are the estimated values of R^2 from a regression of independent variable *i* on all other independent variables. A rule of thumb is that a VIF greater than 10 indicates harmful multicollinearity. The effects of multicollinearity can then be directly measured in the standard errors, which are inflated by the square root of the VIF factor. See e.g. Kennedy (2003) for a discussion of VIF.
We obtain values no greater than 2.3 for the VIF in all regressions where the number of trades and turnover are not simultaneously included. However, when they are both included in the regression, values higher than 20 are obtained, indicating a multicollinearity problem. These results are available on request.

Table 4 Correlation analysis for the cross-sectional variables

This table presents correlations between the explanatory variables used in cross-sectional regression, the results of which are presented in Section 5. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, for the null hypothesis of zero correlation.

	Market cap	Turnover	Number of trades	Cross-listing dummy
Market cap	1.000	0.723***	0.494***	0.522
Turnover		1.000	0.921***	0.472**
Number of trades			1.000	0.251
Cross-listing dummy				1.000

N=27

Table 5 presents the results of our cross-sectional analysis. As expected, the larger the company, as measured by market capitalization, the greater the remote share of price discovery. A potential explanation is that foreign investors are better informed and more willing to trade the stocks of large companies. On the other hand, the number of trades in a stock is inversely related to remote price discovery. This is contrary to our expectations, since foreign investors are usually expected to be willing to hold only liquid stocks, especially in a small market, such as the Finnish one. One possible explanation is that foreign traders only trade when informed. Also, the analysis is mostly focused on larger companies, especially in a small, arguably peripheral, market like Finland.

Table 5 Cross-sectional regression, Remote share of price discovery as an independent variable

This table presents OLS regression results for three different regressions, with the remote members' share of price discovery as the independent variable. The explanatory variables are all logarithmic, with the exception of a dummy variable, which obtained the value 1 when the company is cross-listed either at OMX Stockholm or at the NYSE. Values in parentheses are t-values. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Constant	1.390**	0.705***	1.721***	-0.220
	(2.701)	(4.204)	(3.499)	(-1.640)
Logarithmic market capitalization	0.055**	0.077***	0.053**	0.097***
	(2.540)	(5.012)	(2.343)	(4.679)
Logarithmic average daily turnover in euros	0.090		0.144**	-0.106***
	(1.403)		(2.474)	(-4.607)
Logarithmic average daily number of trades	-0.259***	-0.151***	-0.322***	
	(-3.207)	(-6.024)	(-4.380)	
Cross-listed dummy	0.069	0.097**		0.131***
	(1.670)	(2.666)		(3.075)
Adjusted R^2	0.713	0.701	0.691	0.601

N=27

6 CONCLUSION

We study the unresolved question of who are more informed, domestic or foreign investors. We do this by studying price discovery in a small limit order book market, OMX Helsinki. We use a novel way to study the relative contributions to price discovery of domestic and foreign traders. Using the Error Correction Model of Harris et al. (1995), we estimate relative shares of price discovery for the local and foreign trades in the market. We conclude that for most stocks the local traders determine the majority of price discovery, even if most trading is done by foreign brokers, measured both by the number of trades and turnover. The two stocks with a greater than 50% foreign share of price discovery, Nordea Bank and TeliaSonera, are both cross-listed and heavily traded in the Stockholm market.

We also perform a cross-sectional analysis of the shares of price discovery. We find that both a cross-listing either at OMX Stockholm Stock Exchange or the NYSE, as well as a high market capitalization, are directly related to a higher share of foreign price discovery, as expected. However, contrary to our expectation, liquidity, whether measured by the number of trades, or total turnover in euros, is inversely related to the foreign share of price discovery.

Our results shed new light on the unresolved issue of whether local or foreign investors are better informed. Our results support the findings of Brennan and Cao (1997), Hau (2001), Choe, Kho, and Stulz (2005), and Dvorak (2005), who find that the local investors are more informed. This study also contributes to the literature on price discovery. We use the standard methods of analyzing price discovery in a new setting. Rather than analyzing price discovery in a multiple markets setting, either with a cross-listing, or competition between several markets, we analyze price discovery between the domestic and foreign segments of one market.

REFERENCES

Anand, Amber, and Avanidhar Subrahmanyam, Information and the Intermediary: Are Market Intermediaries Informed Traders in Electronic Markets?, Journal of Financial and Quantitative Analysis 43, 1–28.

Brennan, Michael J, and H. Henry Cao, 1997, International Portfolio Investment Flows, Journal of Finance 52, 1851–1880.

Choe, Hyuk, Bong-Chan Kho, and René M. Stulz, 2005, Do Domestic Investors Have an Edge? The Trading Experience of Foreign Investors in Korea, Review of Financial Studies 18, 795–829.

Dickey, David A., and Wayne A. Fuller, 1981, Likelihood ratio statistics for autoregressive time series with a unit root, Econometrica 49, 1057–1072.

Dvorak, T., 2005, Do domestic investors have an informational advantage? Evidence from Indonesia, Journal of Finance 60, 817–839.

Eun, Cheol S., and Sanjiv Sabherwal, 2003, Cross-Border Listings and Price Discovery: Evidence from U.S.-Listed Canadian Stocks, Journal of Finance 58, 549–575.

Froot, K.A., and T. Ramadorai, 2008, Institutional Portfolio Flows and International Investments, Review of Financial Studies 21, 937–971.

Grammig, Joachim, Michael Melvin, and Christian Schlag, 2005, Internationally Cross-Listed Stock Prices during Overlapping Trading Hours: Price Discovery and Exchange Rate Effects, Journal of Empirical Finance 12, 139–164.

Grinblatt, Mark, and Matti Keloharju, 2000, The investment behavior and performance of various investor types: a study of Finland's unique data set, Journal of Financial Economics 55, 43–67.

Harris, Frederick H. DeB, Thomas H. McInish, Gary L. Shoesmith, and Robert A. Wood, 1995, Cointegration, Error Correction, and Price Discovery on Informationally Linked Security Markets, Journal of Financial and Quantitative Analysis 30, 563–579.

Hasbrouck, Joel, 1995, One security, many markets: determining the contributions to price discovery, Journal of Finance 50, 1175–1199.

Hasbrouck, Joel, 2002, Stalking the “efficient price” in market microstructure specifications: an overview, Journal of Financial Markets 5, 329–339.

Hau, Harald, 2001, Location matters: An examination of trading profits, Journal of Finance 56, 1959–1983.

Johansen, Soren, 1988, Statistical analysis of cointegration vectors, Journal of Economic Dynamics and Control 12, 231–254.

Johansen, Soren, and Katarina Juselius, 1990, Maximum likelihood estimation and inference on cointegration with application to the demand for money, Oxford Bulletin of Economics and Statistics 52, 169–209.

Karolyi, G. Andrew, 2006, The world of cross-listings and cross-listings of the world: Challenging conventional wisdom, Review of Finance 10, 99–152.

Kennedy, Peter, 2003, A Guide to Econometrics, The MIT Press, Cambridge, Mass.

Kim, Minho, Andrew C. Szakmary, and Ike Mathur, 2000, Price transmission dynamics between ADRs and their underlying foreign securities, Journal of Banking & Finance 24, 1359–1382.

Kurov, Alexander, 2008, Information and noise in financial markets: Evidence from the E-Mini index futures, Journal of Financial Research 31, 247–270.

Lau, Sie Ting, and J. David Diltz, 1994, Stock returns and the transfer of information between the New York and Tokyo stock exchanges, Journal of International Money and Finance 13, 211–222.

Lieberman, Offer, Uri Ben-Zion, and Shmuel Hauser, 1999, A characterization of the price behavior of international dual stocks: an error correction approach, Journal of International Money and Finance, 18, 289–304.

Pascual, Roberto, Bartolomé Pascual-Fuster, and Francisco Climent, 2006, Cross-listing, price discovery and the informativeness of the trading process, Journal of Financial Markets 9, 144–161.

Phylaktis, Kate, and Piotr Korczak, 2007, Specialist Trading and the Price Discovery Process of NYSE-Listed Non-US Stocks, Working paper, SSRN: http://ssrn.com/abstract=567104.

Schwarz, Thomas V., and Andrew C. Szakmary, 1994, Price discovery in petroleum markets: Arbitrage, cointegration, and the time interval of analysis, Journal of Futures Markets 14, 147–167.

Theissen, Erik, 2002a, Floor versus Screen Trading: Evidence from the German Stock Market, Journal of Institutional and Theoretical Economics, 32–54.

Theissen, Erik, 2002b, Price discovery in floor and screen trading systems, Journal of Empirical Finance 9, 455–474.

Wang, Steven Shuye, Oliver Meng Rui, and Michael Firth, 2002, Return and volatility behavior of dually-traded stocks: the case of Hong Kong, Journal of International Money and Finance 21, 265–293.

EKONOMI OCH SAMHÄLLE
Skrifter utgivna vid Svenska handelshögskolan

ECONOMICS AND SOCIETY
Publications of the Hanken School of Economics

164. OSSI AURA: Worksite Fitness Policy in an Intellectual Capital Framework. Helsingfors 2006.

165. ANN-MARIE IVARS, MIKAEL REUTER, PIA WESTERBERG och ULLA ÅDAHL-SUNDGREN (Red.): Vårt bästa arv. Festskrift till Marika Tandefelt den 21 december 2006. Helsingfors 2006.

166. GYÖNGYI KOVÁCS: Corporate Environmental Responsibility in Demand Networks. Helsingfors 2006.

167. LEIF RYTTING: Visst gör kunden en stor del av jobbet. Referensramar för kunders medverkan vid tillkomsten av konsumenttjänster. Helsingfors 2006.

168. ANETTE BJÖRKMAN: Towards Explaining the Use of Control Mechanisms in Foreign Subsidiaries of MNCs. Helsingfors 2007.

169. ÅSA HAGBERG-ANDERSSON: Adaptation in a Business Network Cooperation Context. Helsingfors 2007.

170. MIKAEL BERNDTSON: Informell marknadskommunikation. Teoretisk analys jämte en studie av användningsmöjligheter inom banksektorn. Helsingfors 2007.

171. MIKAEL JUSELIUS: A Cointegration Approach to Topics in Empirical Macroeconomics. Helsingfors 2007.

172. KAROLINA WÄGAR: The Nature of Learning about Customers in a Customer Service Setting. A Study of Frontline Contact Persons. Helsingfors 2007.

173. VELI-MATTI LEHTONEN: Henkilöstöjohtamisen tehostaminen valtionhallinnossa henkilöstötilinpäätösinformaation avulla. Empiirinen tutkimus Suomen valtionhallinnossa tuotettavan henkilöstötilinpäätösinformaation arvosta johtamisessa. Strengthening Personnel Management in State Administration with the Support of Information from the Human Resource Report. With an English Summary. Helsinki 2007.

174. KARI PÖLLÄNEN: The Finnish Leadership Style in Transition. A Study of Leadership Criteria in the Insurance Business, 1997-2004. Helsinki 2007.

175. ANNE RINDELL: Image Heritage. The Temporal Dimension in Consumers' Corporate Image Constructions. Helsinki 2007.

176. MINNA PIHLSTRÖM: Perceived Value of Mobile Service Use and Its Consequences. Helsinki 2008.

177. OGAN YIGITBASIOGLU: Determinants and Consequences of Information Sharing with Key Suppliers. Helsinki 2008.

178. OANA VELCU: Drivers of ERP Systems' Business Value. Helsinki 2008.

179. SOFIE KULP-TÅG: Modeling Nonlinearities and Asymmetries in Asset Pricing. Helsinki 2008.

180. NIKOLAS ROKKANEN: Corporate Funding on the European Debt Capital Market. Helsinki 2008.

181. OMAR FAROOQ: Financial Crisis and Performance of Analysts' Recommendations. Evidence from Asian Emerging Markets. Helsinki 2008.

182. GUY AHONEN (Ed.): Inspired by Knowledge in Organisations. Essays in Honor of Professor Karl-Erik Sveiby on his 60th Birthday 29th June 2008. Helsinki 2008.

183. NATAŠA GOLIK KLANAC: Customer Value of Website Communication in Business-to-Business Relationships. Helsinki 2008.

184. ANNIKA SANDSTRÖM: Political Risk in Credit Evaluation. Empirical Studies and Survey Results. Helsinki 2008.

185. MARIA SUOKANNAS: Den anonyma seniorkonsumenten identifieras. Om identitetsskapande processer i en marknadsföringskontext. Helsingfors 2008.

186. RIIKKA SARALA: The Impact of Cultural Factors on Post-Acquisition Integration. Domestic and Foreign Acquisitions of Finnish Companies in 1993-2004. Helsinki 2008.

187. INGMAR BJÖRKMAN et al. (Eds.): Innovation, Leadership, and Entrepreneurship. A Festschrift in Honour of Professor Martin Lindell on his 60th Birthday. Helsinki 2008.

188. JOACIM TÅG: Essays on Platforms. Business Strategies, Regulation and Policy in Telecommunications, Media and Technology Industries. Helsinki 2008.

189. HENRIK TÖTTERMAN: From Creative Ideas to New Emerging Ventures. Entrepreneurial Processes Among Finnish Design Entrepreneurs. Helsinki 2008.

190. ANNIKA RAVALD: Hur uppkommer värde för kunden? Helsingfors 2008.

191. TOM LAHTI: Angel Investing in Finland: An Analysis Based on Agency Theory and the Incomplete Contracting Theory. Helsinki 2008.

192. SYED MUJAHID HUSSAIN: Intraday Dynamics of International Equity Markets. Helsinki 2009.

193. TEEMU TALLBERG: The Gendered Social Organisation of Defence. Two Ethnographic Case Studies in the Finnish Defence Forces. Helsinki 2009.

194. JONAS HOLMQVIST: Language Influence in Services. Perceived Importance of Native Language Use in Service Encounters. Helsinki 2009.

195. ENSIO ERÄ-ESKO: Beskattningsrätt och skattskyldighet för kyrkan i Finland. Steuerrecht und Versteuerung der Kirche in Finnland. Mit einer deutschen Zusammenfassung. Helsingfors 2009.

196. PIA BJÖRKWALL: Nyttighetsmodeller: ett ändamålsenligt innovationsskydd? Helsingfors 2009.