STOCK DATA, TRADE DURATIONS, AND LIMIT ORDER BOOK INFORMATION

Ola Simonsen

Abstract

This thesis comprises four papers concerning trade durations and limit order book information. Paper [1], [2] and [4] study trader durations, e.g., the time between stock transactions in intra-day data. Paper [3] focus on the information content in the limit order book concerning future price movements in stock transaction data.

Paper [1] considers conditional duration models in which durations are in continuous time but measured in grouped or discretized form. This feature of recorded durations in combination with a frequently traded stock is expected to negatively influence the performance of conventional estimators for intraday duration models. A few estimators that account for the discreteness are discussed and compared in a Monte Carlo experiment. An EM-algorithm accounting for the discrete data performs better than those which do not. Empirically, the incorporation of level variables for past trading is rejected in favour of change variables. This enables an interpretation in terms of news effects. No evidence of asymmetric responses to news about prices and spreads is found.

Paper [2] considers an extension of the univariate autoregressive conditional duration model to which durations from a second stock are added. The model is empirically used to study duration dependence in four traded stocks, Nordea, Föreningssparbanken, Handelsbanken and SEB A on the Stockholm Stock Exchange. The stocks are all active in the banking sector. It is found that including durations from a second stock may add explanatory power to the univariate model. We also find that spread changes have significant effect for all series.

Paper [3] empirically tests whether an open limit order book contains information about future short-run stock price movements. To account for the discrete nature of price changes, the integer-valued autoregressive model of order one is utilized. A model transformation has an advantage over conventional count data approaches since it handles negative integer-valued price changes. The empirical results reveal that measures capturing offered quantities of a share at the best bid- and ask-price reveal more information about future short-run price movements than measures capturing the quantities offered at prices below and above. Imbalance and changes in offered quantities at prices below and above the best bid- and ask-price do, however, have a small and significant effect on future price changes. The results also indicate that the value of order book information is short-term.

Paper [4] This paper studies the impact of news announcements on trade durations in stocks on the Stockholm Stock Exchange. The news are categorized into four groups and the impact on the time between transactions is studied. Times before, during and after the news release are considered. Econometrically, the impact is studied within an autoregressive conditional duration model using intradaily data for six stocks. The empirical results reveal that news reduces the duration lengths before, during and after news releases as expected by the theoretical litterature on durations and information flow.

Key words: Finance, Maximum likelihood, Estimation, ACD, News, Multivariate, Intraday, Market microstructure, Granger causality, Time series, INAR, Stock price, Open limit order book.

The following four papers and a summary are included in this thesis:

[1] Brännäs, K. and Simonsen, O. (2005). Discretized Time and Conditional Duration Modelling for Stock Transaction Data. To appear in *Applied Financial Economics.*

[2] Simonsen, O. (2006). An Empirical Model for Durations in Stocks. To appear in *Annals of Finance.*

[3] Hellström, J. and Simonsen, O. (2006). Does the Open Limit Order Book Reveal Information About Short-run Stock Price Movements? *Umeå Economic Studies* 687.

[4] Simonsen, O. (2006). The Impact of News Releases on Trade Durations in Stocks -Empirical Evidence from Sweden. *Umeå Economic Studies* 688.

Introduction

The role for markets is to determine the prices of goods. The question is then, how is this done? The price is determined through an agreement between a buyer and a seller of a quantity of a good. In daily life this may be when a buyer and a seller contact each other directly, for example, when buying a car or a household good. Often there are a number of persons involved in buying or selling goods. The price is then the price at which the quantity for sale is equal to the quantity individuals want to buy. In the economic literature this is the intersection of the demand and supply curves of a specific good. Simply, if more individuals are interested in buying than selling a good the price will increase and if more people are interested in selling than buying a good the price will decrease.

When buying and selling financial assets the buyers and sellers, denoted traders in the finance literature, may either contact each other directly through a specialist or trade through a computerized system. A common wish among traders is to trade at a market that is liquid. This may be defined as the ability to quickly trade a quantity of shares with low price impact. To accomplish this efficiently several trading mechanisms have emerged. The three main solutions are order driven markets, Walrasian auction and price driven markets. The Stockholm Stock Exchange (SSE) is an order driven market. At the SSE traders enter their orders directly into a computerized system and orders are executed when they can be matched. Since the market is computerized small as well as large investors have direct access to the market. The access may be, e.g., through trading platforms on the internet. A second market mechanism used at financial markets is the Walrasian auction. It is one of the oldest market types and is used, e.g., at the London gold market. Finally, a third alternative market mechanism is the price driving market. In that market traders trade through a market maker. The market maker has to provide liquidity to the market, i.e. always standing ready to buy and sell assets. A leading example of such a market is the New York Stock Exchange (NYSE).

The common feature of the different trading mechanisms is that they are governed by rules. For example, the order driven market at the SSE is a continuous market. Traders may enter their orders at any time during the opening hours and, hence, orders are matched and executed continuously over time. The implication of this is that trades occur irregularly and may be clustered over time. In the finance literature the irregularly spaced time between transactions has received attention as it is thought to reflect the information flow to the market. In order driven markets traders enter their orders to an order book. The order book is monitored by a computer and the content is visible to traders. Accordingly, traders' order placements are observed and may influence other market participants. Another common feature of the SSE is the existence of automated trading or algorithm trading among institutional and private traders. Automated trading is performed by a computer, i.e. the trading decisions is transferred from a human trader to a computer. Strategies based on variables such as liquidity, time and price may trigger buy and sell orders. The purpose of the automated trading may be, for example, to lower transaction costs or to make profit out of trading strategies. Institutional investors may have research divisions devoted to automated trading or use software from companies providing sophisticated trading solutions. Also, private investors may have access to automated trading through online brokerage firms. The use of automated trading may have implications for the market characteristics at the SSE. One possible complication may be that reactions after events are faster and more powerful than without automated trading. The reason for this may be that the automated trading triggers a chain reaction. For example, an event resulting in a price movement may initiate a number of trades performed by computers.

In this thesis high frequency data are utilized. High frequency data or intradaily data is financial data where every transaction or event is recorded. The recording frequency is often on a second scale in contrary to, e.g., daily or weekly. Such high frequency data is characterized by a lack of synchronization, i.e. event occurs irregularly over time. Recorded events in high frequency data may be, e.g., how market participants place their orders or every transaction with associated price, spread and volume. Consequently, even short time series of intradaily data contain huge amounts of data. They give much finer information of events at the market and how market participants are acting. The availability of high frequency data has had an important impact on research of how the market mechanisms are working and may be modelled. For example, economic questions like the role of information to stock market participants may be studied and modelled. The availability of high frequency data has also led to the development of new econometric tools. For example, the feature of irregularly spaced times between events has led to new econometric tools, e.g., for the modelling of the time between transactions. For an overview of high frequency data, market microstructure and econometrics, see, e.g., Bauwens and Giot (2001) and Tsay (2002, ch. 5).

Paper [1], [2] and [4] in this thesis study trade durations. Trade durations are the time between two consecutive transactions and a transaction is the agreement between a buyer and seller of a volume of stocks at a given price. Paper [3] studies the information content in the order book on future price changes using data from the Stockholm Stock Exchange.

Trade durations

Trade durations correspond to the time between two consecutive transactions and a transaction refers to a trade between a buyer and a seller of a volume of stocks at a given price. Trade durations have played a central roll in market microstructure research in recent decades. The time elapsed between transactions is thought to carry information about the information flow to market participants. Relevant information may be, e.g., related to the valuation of the stock. Consequently, trade durations may be important for understanding the price process. The idea to study durations originates from the information based model of Glosten and Milgrom (1985). In this model traders are either informed or uninformed, with information varying with regard to the value of the asset they trade. Uninformed traders mainly trade for liquidity reasons, while informed traders act on their superior information on the value of the asset. Diamond and Verrecchia (1987) show that short durations correspond to the presence of bad news. Easley and O'Hara (1992) extend the model of Glosten and Milgrom (1985) by highlighting the importance of time to distinguish between informed and uninformed traders. They stress that new information to market participants leads to increased trade intensity, i.e. shorter durations between transactions. This corresponds to an increased number of informed agents trying to exploit their new information. The information may either be public or private. The public information may, e.g., be news announcements from news agencies or press releases. The part denoted private information may be not publicly released information, e.g., analysis performed by influential investment banks.

News may not only be important for the particular stock it concerns it may also be important for related stocks. For example, private information, discovered by increased trade intensity in a stock may influence trading in related stocks, e.g., in the same industry. If this is true we may find duration dependence between stocks. This feature has rendered attention in the finance literature. Recent studies of dependence between stock transaction series are, e.g., Bauwens and Hautsch (2004) and Spierdijk et al. (2002) who found dependence between transaction series.

Paper [2] in this thesis examines the dependence between stocks in the banking sector at the Stockholm Stock Exchange. The result shows that there is Granger causality (Granger, 1969) between the stocks in the banking sector at the Stockholm Stock Exchange, i.e. increased trade intensity in one of the stocks from the banking sector influence trading in other stocks in the same sector.

News agencies distribute price driving information from stock market companies. Companies are forced to release their price driving information publicly as the use of not publicly released price driving information is prohibited in most countries, including Sweden. Hence, news releases from news agencies may shorten durations. Several studies have focused on the impact of news releases on price and volatility rather than durations, e.g., Berry and Howe (1994), Ederington and Lee (1993), Mitchell and Mulherin (1994), Bollerslev et al. (2000), Bauwens et al. (2005) and Kalev et al. (2004). Their results show that news influence price and volatility. The potential link between price and trade durations has also been studied by, e.g., Grammig and Wellner (2002), Dufour and Engle (2000) and Engle (2000) who study the interdependence between intradaily prices, price volatility and trade durations. Dufour and Engle (2000) find that as the time between transactions become shorter the speed of price adjustment increases suggesting that an active market with short durations demonstrates presence of informed traders. Accordingly, publicly released news announcements that contains price driving information may not only reduce duration lengths but may also affect prices. Paper [4] in this thesis deals with the impact of public news releases on durations, i.e. how trade durations are affected by news releases from news agencies.

Autoregressive conditional duration model

For econometric modelling of the time between transactions standard econometric tools may not be appropriate as transactions are irregularly spaced over time. An influential suggestion of how to model the irregularly spaced times is due to Engle and Russell (1998). Their autoregressive conditional duration (ACD) model explains the length of the next duration by conditioning on the length of past durations and explanatory variables. Several extensions and applications of the original ACD model have been presented by, e.g., Engle and Lunde (2003) and Bauwens and Giot (2001).

A duration arises as the time between two consecutive transactions at t_{i-1} and t_i, i.e. $d_i = t_i - t_{i-1}$. The conditional expectation of a duration d_i is specified as $E(d_i \,| d_{i-1}, ..., d_1; x) = \theta_i$. Here d_i, is conditioned on past durations and other explanatory variables x, and θ_i is specified in such a way that $\epsilon_i = d_i/\theta_i$ is independent and identically distributed. The θ_i may be parameterized as

$$\theta_i = \omega + \sum_{j=1}^{p} \alpha_j d_{i-j} + \sum_{j=1}^{q} \beta_j \theta_{i-j} + \pi' x_{i-1}.$$

Here, θ_i is a function of p lagged durations and q lags of conditional durations. This is called an ACD(p,q,x), where p and q are the orders of the lags in the mean function and x_i is a vector of explanatory variables such as volume, spread and price.

Engle and Russell (1998) popularized the quasi maximum likelihood (QML) estimator building on the exponential distribution for the estimation of the unknown parameters of the ACD model. The estimator maximizes the log-likelihood function

$$\ln \ell = -\sum \left[\ln \theta_i + \frac{d_i}{\theta_i} \right].$$

The estimator is widely used for estimating these types of models and, hence, an increasing amount of empirical results builds on estimates from this estimator. The estimator is consistent if the model for the conditional mean is correctly specified and the duration is in the exponential family of distributions. A potential problem with the estimator is with the empirical duration data often used in the estimation. The data are discretized with a large fraction of short and zero durations, e.g., more than one transaction occurring during one second. The discreteness of the recorded transaction data is due to the second scale and data are hence integer-valued. As the distribution underlying the QML estimator is continuous the discretness may negatively affect the performance of the estimator. Kulldorff (1961, ch. 2) shows that inconsistency of the maximum likelihood (ML) estimator may arise from, e.g., using mid-interval values to represent the interval when data are discretized or grouped. In paper [1] the problem is highlighted and estimators are proposed to account for the problem.

One of the prominent features of the results from the ACD model, e.g., from papers [1],[2] and [4], is the quick response of, for example, price changes and changes of traded volumes. The response takes place within seconds after events occur and fades out quickly thereafter. The presence of automated trading at the Stockholm Stock Exchange may be one explanation for this pattern. Trading rules programmed to react to, e.g., price moves of more than a predetermined number of ticks, may influence the characteristics of the trading behavior.

Limit order book

In order driven markets no market maker is involved in the trading process. Traders directly enter their orders in an order book monitored by a computer. The order book is visible to traders, private as well as institutional traders. Traders may enter two types of orders, either a market order or a limit order. Limit orders is placed in queue in the order book and is executed when they can be matched. The priority of limit order execution is determined by price and time of order placement, i.e. the stocks with the lowest limit sell price are sold first and the stocks with the highest limit buy price are bought first. Market orders are, on the other hand, executed immediately to the best bid or ask price. For example, the publicly visible limit order book for a stock listed on the SSE shows the first five levels on the bid and ask side, respectively. This is illustrated in Figure 1 where P_i^d and P_i^s are the prices on the bid- (demand) and ask- (supply) side of an arbitrary order book for the levels $i = 1, 2...5$. The bid- and ask-volumes contained at level i are denoted Q_i^d and Q_i^s.

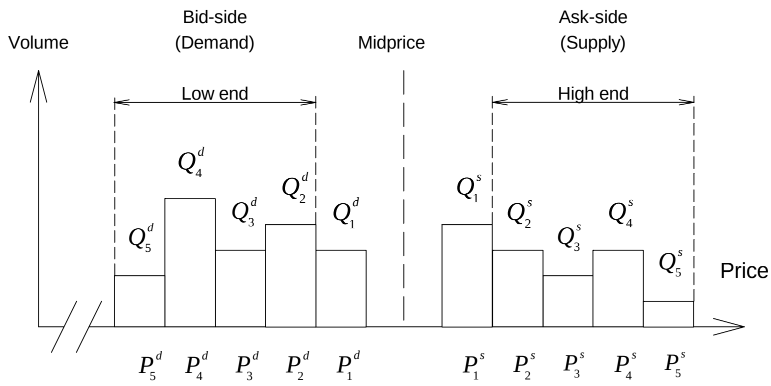

Figure 1: Illustration of the limit orderbook.

In the theoretical literature of limit order books the assumption is that informed traders always use market orders instead of using limit orders, e.g., Glosten (1994), Rock (1996) and Seppi (1997). Accordingly, there should be limited information of observing offered quantities of a share at other prices than the best bid- or ask-price, i.e. at the high and low end of the order book. However, recent papers of Cao et al. (2004) and Bloomfield et al. (2005) find evidence of information also in the order book beyond the first levels. Cao et al. (2004) introduce measures capturing the shape of the order book, e.g., if the order book is asymmetric, i.e. if there is more value on the bid- (ask-) side relative to the ask- (bid-) side. The measure is used to study if the asymmetry of the order book is informative regarding future price movements. The findings are that the order book contains information regarding future price movements using intradaily data with 5 minutes aggregation level. Further, Foucault (1999) argues that an increase in asset volatility increases the proportion of limit order traders and the limit order traders have to post higher ask prices and lower bid prices, i.e. market depth increase.

In paper [3] of this thesis the purpose is to empirically study the information contained in the open limit order book about future short-run stock price movements. To assess the information contained in the order book the paper presents new measures and extensions of existent measures. A new measure capturing the market depth discussed in Foucault (1999) is presented. We find the appropriate aggregation level to be 1 minute in contrary to Cao et al. (2004), who use 5 and 10 minutes aggregation levels.

Summary of the papers

Paper [I] Discretized Time and Conditional Duration Modelling for Stock Transaction Data

The paper considers conditional duration models in which durations are in continuous time but measured in grouped or discretized form. This feature of recorded durations in combination with a frequently traded stock is expected to negatively influence the performance of conventional estimators for intraday duration models. Estimators that account for the discreteness are discussed and compared in a Monte Carlo experiment, e.g., grouped maximum likelihood and EM-algorithm. In the small Monte Carlo study the EM-algorithm that accounts for the discrete nature of the data both in the outcome and the lagged explanatory variables comes out as the best estimator of the compared ones.

In the empirical part of the paper the differences between estimators are generally quite small and the EM-algorithm and ML estimators based on discrete data are not too different from ML based on grouped data and Weibull and Burr models. When it comes to the effects of explanatory variables the study provides support for the use of changes rather than levels to reflect news. There is throughout a significant and positive effect of news about prices and a negative effect of a change in the spread. The spread effect is, however, not significant. A higher volume has an insignificant but prolonging effect in most cases. We could not find statistically significant support for an asymmetric response to news about spreads nor about prices. A contributory cause of the quick response after changes in the explanatory variables may be the computerized trading at the stock market. The log-likelihood function value of the Burr is larger than for other models but the models are not nested. In addition, the serial correlation properties of the exponential and Weibull models speak in favor of these two models. A generalized gamma was also employed and provided a better fit to the data than both the exponential and Weibull models.

Paper [2] An Empirical Model for Durations in Stocks

This paper considers an extension of the univariate autoregressive conditional duration model to which durations from a second stock are added. By including durations from a second stock dependence between duration series is captured in the model. For example, Figure 2 illustrates the dependence from stock 2 to stock 1. Completed durations are added from stock 2, $d^2_{N^2(t)}$, to stock 1, and the completed durations from stock 2 are given weights dependent on the size of τ_1, i.e. depending on how far away in time the completed durations are.

The model is empirically used to study duration dependence in four traded stocks, Nordea, Föreningssparbanken, Handelsbanken and SEB A on the Stockholm Stock Exchange. The stocks are all in the banking sector. In the empirical part we find Granger causality between all the stocks. This result indicates that there may be duration dependence between stocks active in the banking sector on the Stockholm Stock Exchange. The dependence may be caused by new information revealed by the trade intensity. In view of this empirical result the suggested model extension is able to capture dependence between duration series and to provide an improvement of the econometric specification of the model.

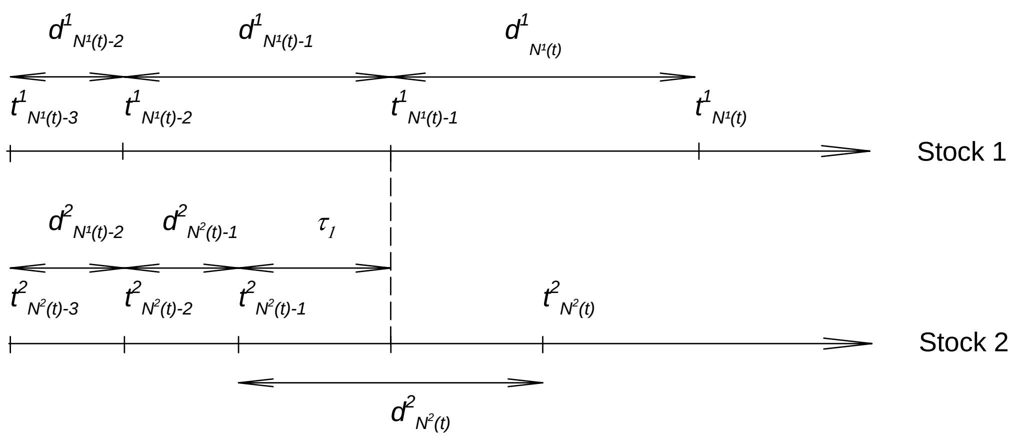

Figure 2: Illustration of two stocks, stock 1 and stock 2, with transaction times t and durations d. τ_1 is the observed length of the most recent duration in stock 2 at time $t^1_{N^1(t)-1}$.

Paper [3] Does the Open Limit Order Book Reveal Information About Short-run Stock Price Movements?

The purpose of this paper is to empirically study the information contained in the open limit order book about future short-run stock price movements. Specifically, attention is paid to whether changes or asymmetries in the order book concerning offered quantities of a share at prices below the best bid price (low end of the order book) and above the best ask price (high end of the order book) are informative. To assess the information contained in the order book the paper presents new measures as well as extensions to existing measures summarizing order book movements.

To account for the discrete nature of price changes, the integer-valued autoregressive model (e.g., McKenzie, 1985, 1986, Al-Osh and Alzaid, 1987) is utilized to model the discrete nature of high frequency stock price data.

The empirical results reveal that measures capturing offered quantities of a share at the best bid- and ask-price reveal more information about future short-run price movements than measures capturing the quantities offered at prices below and above.

However, imbalance and changes in offered quantities at prices below and above the best bid- and ask-price do have a small and significant effect on future price changes. The results also indicate that the value of order book information is short-term. This can be compared to Cao et al. (2004) who found informational value of the higher levels of the order book at an aggregation level of 5 and 10 minutes. The effects in this paper are most apparent at a low aggregation level (1 minute) while estimation results for higher aggregation levels (2, 5 and 10 minutes) show mostly insignificantly results.

Paper [4] The Impact of News Releases on Trade Durations in Stocks - Empirical Evidence from Sweden

This paper examines empirically the short-run impact of public news announcements on trade durations in stocks traded on the Stockholm Stock Exchange in Sweden. Econometrically, the impact is studied within an autoregressive conditional duration model using intradaily data for six stocks. The news are categorized into four groups and added as explanatory variables to the model. The news categories are Company/Sector, EU macro, Swedish macro and US macro news. The news categories are added to the model through a dummy variable structure that captures the impact before, during and after the news.

The empirical results reveal that news reduces the duration lengths before, during and after news releases. For example, Company/Sector related news shortens durations by 20 to 40 percent. The cause of reduced durations before the actual news releases may originate from anticipated news releases, while the significant result of the impact during and after the news release may be due to both anticipated and unanticipated news events. The result supports the predictions in the theoretical literature of shorter durations in connection to news.

References

Al-Osh, M.A. and Alzaid, A.A. (1987). First-Order Integer Valued Autoregressive Process. *Journal of Time Series Analysis,* 8, 261-275.

Bauwens, L. and Giot, P. (2001). *Econometric Modelling of Stock Market Intraday Activity.* Kluwer, Dordrecht.

Bauwens, L. and Hautsch, N. (2004). Dynamic Latent Factor Models for Intensity Processes. Core Discussion Paper 2003/103.

Bauwens, L., Omrane, W.B. and Giot P. (2005). News Announcements, Market Activity and Volatility in the Euro/Dollar Foreign Exchange Market. *Journal of International Money and Finance*, 24, 1108-1125.

Berry, T.D. and Howe, K.M. (1994). Public Information Arrival, *Journal of Finance,* 49, 1331-1346.

Bloomfield, R., O'Hara, M. and Saar, G. (2005). The "Make or Take" Decision in an Electronic Market: Evidence on the Evolution of Liquidity. *Journal of Financial Economics*, 75, 165-199.

Bollerslev, T., Cai, J. and Song F.M. (2000). Intraday Periodicity, Long Memory Volatility, and Macroeconomic Announcement Effects in the US Treasury Bond Market. *Journal of Empirical Finance*, 7, 37-55.

Cao, C., Hansch, O. and Wang, X. (2004). The Informational Content of an Open Limit Order Book. Working Paper. Penn state university.

Diamond, D.W. and Verrecchia, R.E. (1987). Constraint on Short-Selling and Asset Price Adjustment to Private Information. *Journal of Financial Economics,* 18, 277-311.

Dufour, A. and Engle, R.F. (2000). Time and the Price Impact of Trade.*The Journal of Finance,* 55, 2467-2498.

Easley, D. and O'Hara, M. (1992). Time and the Process of Security Price Adjustment. *Journal of Finance,* 47, 577-606.

Ederington, L.H. and Lee, J.H. (1993). How Markets Process Information: News Releases and Volatility, *Journal of Finance,* 48, 1161-1191.

Engle, R.F. (2000). The Econometrics of Ultra-High-Frequency Data. *Econometrica,* 68, 1-22.

Engle, R.F. and Lunde, A. (2003). Trades and Quotes: A Bivariate Point Process. *Journal of Financial Econometrics,* 2, 159-188.

Engle, R.F. and Russell, J.R. (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica,* 66, 1127-1162.

Foucault, T. (1999). Order Flow Composition and Trading Cost in a Dynamic Limit Order Market. *Journal of Financial Markets,* 2, 99-134.

Glosten, L. (1994). Is the Electronic Open Limit Order Book Inevitable?. *The Journal of Finance,* 4, 1127-1161.

Glosten, L. and Milgrom, P. (1985). Bid, Ask and Transactions Prices in a Specialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics,* 14, 71-100.

Grammig, J. and Wellner, M. (2002). Modeling the Interdependence of Volatility and Inter-Transaction Duration Processes. *Journal of Econometrics,* 106, 369-400.

Granger, C. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica,* 37, 424-438.

Kalev, P.S., Liu, W.M., Pham P.K. and Jarnecic, E. (2004). Public Information Arrival and Volatility of Intraday Stock Returns. *Journal of Banking & Finance*, 28, 1441-1467.

Kulldorff, G. (1961). *Contributions to the Theory of Estimation from Grouped and Partially Grouped Samples.* PhD thesis. Almqvist & Wicksell, Stockholm.

McKenzie, E. (1985). Some Simple Models for Discrete Variate Time Series. *Water Resources Bulletin, 21*, 645-650.

McKenzie, E. (1986). Autoregressive Moving-Average Processes with Negative Binomial and Geometric Marginal Distributions. *Advances in Applied Probability*, 18, 679-705.

Mitchell, M.L. and Mulherin, J.H. (1994). The Impact of Public Information on the Stock Market. *The Journal of Finance,* 49, 923-950.

Rock, D. (1996). The Specialist's Order Book and Price Anomalies. Working paper Harvard University.

Seppi, D. (1997). Liquidity Provision With Limit Orders and Strategic Specialist. *Review of Financial Studies*, 10, 107-150.

Spierdijk, L., Nijman, T.E. and van Soest, A.H.O. (2002). Modeling Comovements in Trading Intensities to Distinguish Sector and Stock Specific News. Unpublished manuscript, Tilburg University.

Tsay, R.S. (2002). *Analysis of Financial Time Series.* New York: John Wiley & Sons, Inc.

Discretized Time and Conditional Duration Modelling for Stock Transaction Data

Kurt Brännäs and Ola Simonsen
Department of Economics, Umeå University
SE-901 87 Umeå, Sweden

Abstract: The paper considers conditional duration models in which durations are in continuous time but measured in grouped or discretized form. This feature of recorded durations in combination with a frequently traded stock is expected to negatively influence the performance of conventional estimators for intra-day duration models. A few estimators that account for the discreteness are discussed and compared in a Monte Carlo experiment. An EM-algorithm accounting for the discrete data performs better than those which do not. Empirical results are reported for trading durations in Ericsson B at Stockholmsbörsen for a three-week period of July 2002. The incorporation of level variables for past trading is rejected in favour of change variables. This enables an interpretation in terms of news effects. No evidence of asymmetric responses to news about prices and spreads is found.

Key Words: Grouped data, Maximum likelihood, EM-algorithm, Estimation, Finance, News.

JEL Classification: C12, C22, C41, C51, G12, G14.

1. Introduction

This paper considers the discrete nature of empirical duration data obtained from electronic trading systems. The output from such a system typically records transactions to prevailing second and for frequently traded stocks the resulting durations then contain a large fraction of zeros and very short durations. We take a continuous underlying duration density to be discretized and study the consequences of this discretization econometrically. A second objective of the paper is to give empirical evidence on the reaction to news in a specific stock.

The initial argument for studying time in the market microstructure literature is due to Glosten and Milgrom (1985). In their information based model there are informed and uninformed traders. The Easley and O'Hara (1992) extension highlights the importance of time in distinguishing between the two types of traders. No trading, i.e. long durations, means that there is no available information (uninformed trading), while short durations indicate new information (informed trading). In addition, short durations are associated with small price volatility. Recently, Chiang and Wang (2004) confirmed that price volatility is related to durations and to duration related variables. For the modelling of durations, Engle and Russell (1998) suggest conditional duration models for high frequency or intra-day time series data. Among the questions of interest in this field is the reaction to news as reflected, e.g., by indicators of recent transactions.

Engle and Russell (1998) emphasize the appealing properties of the quasi-maximum likelihood (QML) estimator based on the exponential duration model. An increasing amount of empirical research builds on this QML estimator (e.g., Bauwens and Giot, 2001). When the data are available in only a discretized form the QML estimator looses some of its appeal. In essence, the consistency property of this estimator requires a correct conditional mean specification and when data is discretized it is much harder to correctly specify the conditional mean as it then depends on a true but unknown density. Kulldorff (1961, ch. 2) shows in a time invariant case the inconsistency of the maximum likelihood (ML) estimator arising from, e.g., using mid-interval values to represent the interval when data are discretized or grouped. He also demonstrates that discretization and the use of mid-interval values have more serious effects on the performance of ML estimators when the sample records durations to belonging to only a few and wide groups or intervals. For frequently traded stocks the number of groups is small and the width is small. We then expect the inconsistency to remain, but the actual performance of the ML or QML estimators that neglect the discreteness may remain relatively advantageous. The inconsistency remains when the model contains explanatory variables. In addition, if the true duration is viewed as continuous but only discrete time observations are available, any specification containing lagged durations is contaminated by a measurement error in a way to be made clear below.

In this paper we consider estimators that to some extent account for the outlined features of the data. Grouped data ML estimators and EM-algorithm versions are among the studied estimators. We conduct a small Monte Carlo study focusing on the consequences of the various specification choices and the chosen estimators. Empirical results for a three-week period of transaction durations in Ericsson B at the order driven Stockholmsbörsen stock exchange in Stockholm are also reported. Of particular interest, beyond the focus on estimators, is here the reaction to news and whether the responses to positive or negative news are different. We use past price, spread and volume as indicators of new information to the market.

In Section 2 we discuss the model and discuss the ML and EM-algorithm estimators for discretized duration data. Section 3 reports the results from a set of Monte Carlo experiments conducted to study the consequences of the alternative ways of handling the discretized data. Section 4 reports the empirical results and the final section concludes.

2. Model and Estimators

Let the tth continuous duration be denoted by D_t. The duration arises as a difference between two real transaction times, τ, indexed by τ_k and τ_{k-1}, i.e. $D_t = \tau_k - \tau_{k-1}$. When transactions are recorded at a second-level scale, the observed duration measure $d_t = [\tau_k] - [\tau_{k-1}]$, where $[\cdot]$ signifies integer-value, is in seconds and hence integer-valued.[1] For a frequently traded stock the durations are on average short and then d_t will take on one value from a set $\{0,1,2,\dots,M_t\}$, where $\max\{M_t\}_{t=1}^{T}$ is a relatively small number and T is the length of the time series sequence of consecutive durations.

[1] An alternative $d_t = [\tau_k - \tau_{k-1}]$ is of interest if there is direct access to the primary data source. We believe this to be the less likely alternative, and abstain from giving detailed results for this case. The required changes to the given results are rather straightforward.

Figure 1 illustrates how the transaction times, τ_k, the continuous durations, D_t, and the discretized durations, d_t, are related. Note that, except for $d_t = 0$, the d_t-values represent mid-interval values. For $d_t = 0$ the mid-interval value is 0.5.

Given an assumption about the continuous and conditional distribution of D_t and given values in the set $\Delta_{t-1} = \{D_1, D_2, \ldots, D_{t-1}\}$ it is straightforward to obtain the probability for d_t equal to some integer k given Δ_{t-1}, as

$$\begin{aligned}
\Pr(d_t = 0 \mid \Delta_{t-1}) &= \Pr(D_t \in [0,1] \mid \Delta_{t-1}) \\
&= \Pr(D_t \le 1 \mid \Delta_{t-1}) \\
\Pr(d_t = k \mid \Delta_{t-1}) &= \Pr(D_t \in (k-1, k+1] \mid \Delta_{t-1}) \\
&= \Pr(D_t \le k+1 \mid \Delta_{t-1}) - \Pr(D_t \le k-1 \mid \Delta_{t-1}), \quad k \ge 1
\end{aligned} \tag{1}$$

with $\Pr(D_t \le 0 \mid \Delta_{t-1}) = 0$.

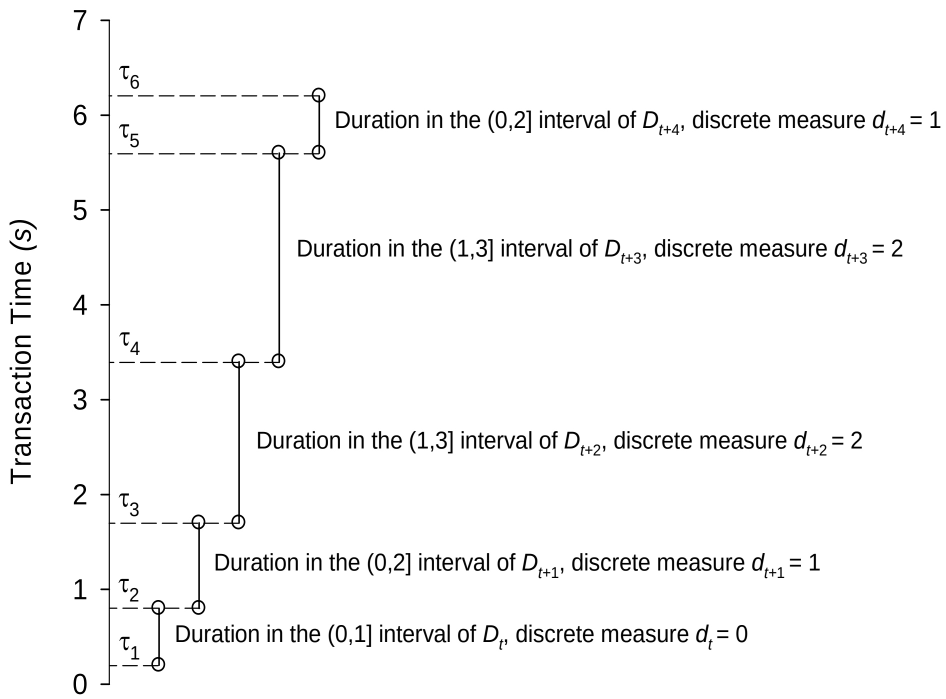

Figure 1: Examples of potential transaction times τ_i and the registered duration variable d_t with the intervals of the underlying continuous duration variable D_t.

In the sequel of this and the next section we only consider the conditional exponential duration model, but any other reasonable duration distribution could have been considered instead. By focusing on the exponential model the technical aspects are kept simpler than for most other models, and the main ideas remain unaltered. For the exponential conditional duration model with conditional mean $E(D_t \mid \Delta_{t-1}) = \theta_t > 0$ and conditional variance $V(D_t \mid \Delta_{t-1}) = \theta_t^2$ we get explicit expressions for the probabilities as

$$\begin{aligned} &\Pr(d_t = 0 \mid \Delta_{t-1}) = 1 - e^{-1/\theta_t} \\ &\Pr(d_t = k \mid \Delta_{t-1}) = e^{-(k-1)/\theta_t} - e^{-(k+1)/\theta_t}, \quad k = 1, 2, \dots, M_t. \end{aligned} \tag{2}$$

Since we can write $D_t = \theta_t \varepsilon_t$, with ε_t exponentially distributed with parameter one, we obtain the conditional expectation of D_t given $d_t = k$ as

$$\begin{aligned} E(D_t \mid d_t = 0, \Delta_{t-1}) &= \theta_t \left[1 - \frac{1}{\theta_t (e^{1/\theta_t} - 1)} \right] \\ E(D_t \mid d_t = k, \Delta_{t-1}) &= \theta_t E \left[\varepsilon_t \mid \varepsilon_t \in \left(\frac{k-1}{\theta_t}, \frac{k+1}{\theta_t} \right) \right] \\ &= \theta_t + \frac{e^{1/\theta_t}(k-1) - e^{-1/\theta_t}(k+1)}{e^{1/\theta_t} - e^{-1/\theta_t}}, \quad \text{for } k \geq 1. \end{aligned} \tag{3}$$

It may be shown that $E(D_t \mid d_t = k, \Delta_{t-1}) \leq k$ for $k \geq 1$ and that $E(D_t \mid d_t = 0, \Delta_{t-1}) \leq \frac{1}{2}$, i.e. under the exponential distribution assumption these conditional expectations are smaller than or equal to their corresponding mid-interval values. Equality arises only when $\theta_t \to \infty$.

In the conventional continuous duration framework advanced by Engle and Russell (1998) the θ_t function is of the type:

$$\begin{aligned} \theta_t &= \alpha_0 + \alpha_1 D_{t-1} + \ldots + \alpha_q D_{t-q} + \beta_1 \theta_{t-1} + \ldots \beta_p \theta_{t-p} + \mathbf{x}_t \boldsymbol{\pi} \\ &= \mathbf{z}_t \boldsymbol{\psi}, \end{aligned} \tag{4}$$

where $\mathbf{x}_t$ is a vector of predetermined variables containing, e.g., past prices. Setting $\xi_t = D_t - \theta_t$ in (4) enables us to rewrite the model on the alternative form

$$\begin{aligned} D_t = {} & \alpha_0 + (\alpha_1 + \beta_1) D_{t-1} + \ldots (\alpha_q + \beta_q) D_{t-q} + \beta_{q+1} D_{t-q-1} + \ldots + \beta_p D_{t-p} \\ & + \xi_t - \beta_1 \xi_{t-1} - \ldots - \beta_p \xi_{t-p} + \mathbf{x}_t \boldsymbol{\pi}, \quad \text{for } p \geq q. \end{aligned} \tag{5}$$

This is an ARMAX model in the continuous exponential duration variable. Obviously, other specifications are also feasible (e.g., Bauwens and Giot, 2001, ch. 3).

2.1 Estimators

We first consider estimation that accounts for the discreteness in the conditional variable d_t

to be explained. Later we extend the estimation setup by also considering the discreteness in the lagged durations that serve as explanatory variables in the θ_t function.

Under an exponential distribution assumption for the conditional D_t variable the log-likelihood function for its discrete (grouped data) corresponding d_t variable takes the form

$$\ln L = \sum_{t=r}^{T} \ln(e^{-\eta_t (d_t - 1)/\theta_t} - e^{-(d_t+1)/\theta_t}), \tag{6}$$

where $r = \max(p,q)+1$ and $\eta_t = 0$, for $d_t = 0$, and $\eta_t = 1$, for $d_t \geq 1$. The associated score vector can be expressed

$$\frac{\partial \ln L}{\partial \psi} = \sum_{t=r}^{T} \frac{\mathbf{z}_t'}{(\mathbf{z}_t\psi)^2} \left[\frac{\eta_t(d_t - 1)e^{-\eta_t(d_t-1)(\mathbf{z}_t\psi)^{-1}} - (d_t+1)e^{-(d_t+1)(\mathbf{z}_t\psi)^{-1}}}{e^{-\eta_t(d_t-1)(\mathbf{z}_t\psi)^{-1}} - e^{-(d_t+1)(\mathbf{z}_t\psi)^{-1}}} \right].$$

For (4)-(5) lagged continuous $D_{t-i}, i = 1,\ldots,q$, variables are assumed observed. Obviously, if durations are measured in discrete form, d_{t-i} rather than D_{t-i} is observed. As a consequence there are measurement errors. The ML estimator based on (6) is consistent and asymptotically normal when $\alpha_1 = \alpha_2 = \ldots = \alpha_q = 0$ (no measurement errors as $d_{t-i}, i = 1,\ldots,q$, are not included in θ_t) and the $\beta_1,\ldots,\beta_p$ parameters are such that the $\{D_t\}$ sequence is stationary. If there is, at least, one $\alpha_i \neq 0$ the ML estimator is inconsistent. An early proof of the asymptotic results for a scalar case is due to Kulldorff (1961), who also studied the loss in efficiency that results from discretizing the time scale. Engle and Russell (1998) consider the case of explanatory variables, continuous durations and the QML estimator. Given a correctly specified stationary model they show the QML estimator to be consistent and asymptotically normal.

Obviously, other duration densities such as Weibull or log-logistic can also be applied. In the absence of strong a priori arguments for a particular model one avenue would be to specify an even wider class of densities such as the generalized gamma. The Appendix gives expressions for the Weibull and Burr models, which also are used in the empirical study below.

Consider as a simple example of the inconsistency an underlying exponentially distributed variable and observations falling into either of only two intervals [0,1] and $(1,\infty)$ with the indicator variable $d_t = 1$ for the latter interval. Let $\theta_t = \alpha D_{t-1}$ in the true case and $\theta_t = \alpha d_{t-1}$

in the assumed case. The score for the assumed model is

$$\frac{\partial \ln L}{\partial \alpha} = \sum_{t=2}^{T} \frac{d_t - e^{-1/(\alpha d_{t-1})}}{\alpha^2 d_{t-1}[1 - e^{-1/(\alpha d_{t-1})}]},$$

with $d_t = 0$ or $d_t = 1$. As $E(d_t \mid \Delta_{t-1}) = \Pr(d_t = 1 \mid \Delta_{t-1}) = \exp[-1/(\alpha D_{t-1})]$, a first order expansion of the score $\partial \ln L / \partial \hat{\alpha}$ around the true parameter value and manipulation shows that the bias depends on $D_{t-1} - d_{t-1}$. As this difference can be expected to be larger than zero, the ML estimator $\hat{\alpha}$ can be expected to be too large.

The EM-algorithm (Dempster, Laird and Rubin, 1987) provides a general framework for dealing with aspects of the limited information in the $\{d_t\}$ sequence that we have in this case. If we take the lagged d_{t-i} variables as is and focus only on the grouped d_t indicator, we can easily extend the constant parameter and grouped exponential model of Little and Rubin (1987, ch. 11). The M-step maximizes the conditional expectation of the log-likelihood function for D_t given d_t with respect to $\tilde{\psi}$, i.e. $Q(\psi,\tilde{\psi}) = E_{\tilde{\psi}}[\ln L(D) \mid d, \Delta] = \sum_{t=r}^{T}\left[-\ln \mathbf{z}_t\psi - E_{\tilde{\psi}}(D_t \mid d_t, \Delta_{t-1})/\mathbf{z}_t\psi\right]$. The required conditional expectation is given in (3) and should be evaluated at $\tilde{\psi}$ (the E-step). The E and M steps are iterated until convergence. Note that the M-step uses the continuous exponential variable log-likelihood function and should therefore be computationally straightforward. The score vector is $\partial Q / \partial \psi = \sum_{t=r}^{T} \mathbf{z}_t'\left[E_{\tilde{\psi}}(D_t \mid d_t, \Delta_{t-1}) - \mathbf{z}_t\psi\right]/(\mathbf{z}_t\psi)^2$ and the Hessian is $\partial^2 Q / \partial\psi\partial\psi' = \sum_{t=r}^{T} \mathbf{z}_t'\mathbf{z}_t/(\mathbf{z}_t\psi)^2$.

Consider next an EM-algorithm that attempts to account also for the presence of d_{t-i} when the true D_{t-i} would have been preferred in θ_t. Little and Rubin (1987, ch. 8) consider a problem of missing observations in a Gaussian AR(1) model, but there appears to be no reported research on the type of problem we have in mind. Recall that the density is a conditional one so that conditioning on past d_{t-i} and not on future d_{t+j} appears reasonable.

Then $Q(\psi,\tilde{\psi}) = \sum_{t=r}^{T}\left[-E_{\tilde{\psi}}(\ln \mathbf{z}_t\psi \mid \bar{d}_t, \Delta_{t-1}) - E_{\tilde{\psi}}(D_t/\mathbf{z}_t\psi \mid \bar{d}_t, \Delta_{t-1})\right]$, where $\bar{d}_t = (d_1,\ldots,d_t)$. This expression is a difficult one to use as it involves taking expectations of past D_{t-i} with respect to more recent d_t, because it involves nonlinearities, and because different time periods are interwoven in the final expectation expression. Therefore, no attempt is made at obtaining an exact EM-algorithm in this paper.

An ad hoc EM-algorithm could use $E_{\tilde{\psi}}(D_t \mid d_t, \Delta_{t-1})$ to replace d_t for all t in an E-step and hence also in the $\mathbf{z}_t$ vector. For the M-step the continuous exponential distribution is employed. The performance of this estimator is studied by Monte Carlo simulation in the next section and is also used empirically.

The simulated maximum likelihood (SML) estimator (Gourieroux and Monfort, 1991) offers an interesting approach to coping with the discreteness of the data. Unfortunately, the results of Cappé et al. (2002) suggest that the SML estimator may be a very time demanding exercise as the number of replications should increase with the number of observations. For high frequency data the sample size is usually large. Note also that the current context differs from the dynamic limited dependent variable model considered by, e.g., Lee (1997).

An obvious way of attempting to avoid the bias arising from measurement errors in lagged variables would be to specify the joint distribution of $\{\Delta_t\}$. It would then be possible to avoid the measurement error problem by accounting for the discretization for all t. Unfortunately, such modelling would also be subject to even larger risks of distributional misspecification as multivariate distributions can come in many alternative shapes. Computationally it is potentially a difficult problem. Closely related to this idea and more directly focusing on the discrete data would be a direct specification of an inhomogeneous Markov chain with transition probabilities depending, e.g., on $\mathbf{x}_t$.

The geometric distribution is the discrete time equivalent of the exponential distribution and with a conditional interpretation the lagged d_{t-i} presents no problem. When data is genuinely continuous this distribution can at most be regarded as an approximation. It can be demonstrated that the ML estimator is biased and inconsistent when data is generated according to a continuous exponential model.

3. Monte Carlo Study

In this section we study the properties of the estimators when data are artificially generated according to conditional exponential and Weibull models.

We specify the conditional mean function that is used in generating the underlying D_t data as

$$\theta_t = \alpha_0 + \alpha_1 D_{t-1} + \beta_1 \theta_{t-1}.$$

Engle and Russell (1998) give the following moment results for the exponential model

$$E(D_t) = \frac{\alpha_0}{1-\alpha_1-\beta_1}$$

$$V(D_t) = \frac{1-\beta_1^2-2\alpha_1\beta_1}{1-\beta_1^2-2\alpha_1\beta_1-2\alpha_1^2} \times E^2(D_t).$$

The parameters should satisfy $2\alpha_1^2+\beta_1^2+2\alpha_1\beta_1<1$ and $\alpha+\beta<1$. For Weibull distributed durations we employ standardization to obtain the mean and variance of the exponential model, cf. the Appendix.

The study uses $\alpha_1 = 0.2,\ 0.3$ and 0.4, $\beta_1 = 0.15,\ 0.2$ and 0.25, and $\alpha_0 = 2.5,\ 5$ and 10, to give mean durations in the range $3.8-28.6$ seconds with variances in the range $4.2-50.5$. For the Weibull model we use $\gamma = 0.8$, which corresponds to negative duration dependence. The time series length is set at $T = 5\,000$ and 50 000. The $T = 5\,000$ case corresponds to a short time series length for frequently traded stocks, and $T = 50\,000$ is used only for the shortest durations $(\alpha_0 = 2.5)$ and exponential data. In each design cell 1000 replications are generated starting from the same initial seed. In generating the series an initial part of 100 observations is dropped. Data are next discretized in accordance with the discussion in Section 2.

The following model variants and estimation algorithms are used: (i) The continuous time exponential model (based on $f(D) = \theta^{-1}\exp(-D\theta^{-1})$ and indicated by C) serves as a base case and is estimated by ML and a scoring algorithm. All other data sets are based on discrete duration $\{d_t\}$ sequences. (ii) The same ML algorithm as in (i) is used with discrete data (indicated by D). Note that for $d_t = 0$ we use $d_t = 0.5$ so that all d_t can be interpreted as mid-interval values. (iii) The grouped data ML estimator (based on (6)) with d_{t-1} in the θ_t function is estimated by a BHHH algorithm (indicated by G); (iv) The EM-algorithm with $\hat{D}_t$ replacing D_t for all t is estimated by alternating between a ML and an E-step (based on $f(D)$ and expectations in (3), and indicated by EM). Hence, even if data is generated as Weibull distributed the employed density underlying the estimators is throughout the exponential. Note that all estimators are conditional ones, as estimation is throughout conditional on initial observations. The number of iterations is limited to 100 and true parameter values are used to initialize iterations. All computations are performed using Fortran code on a 1.9 GHz Laptop.

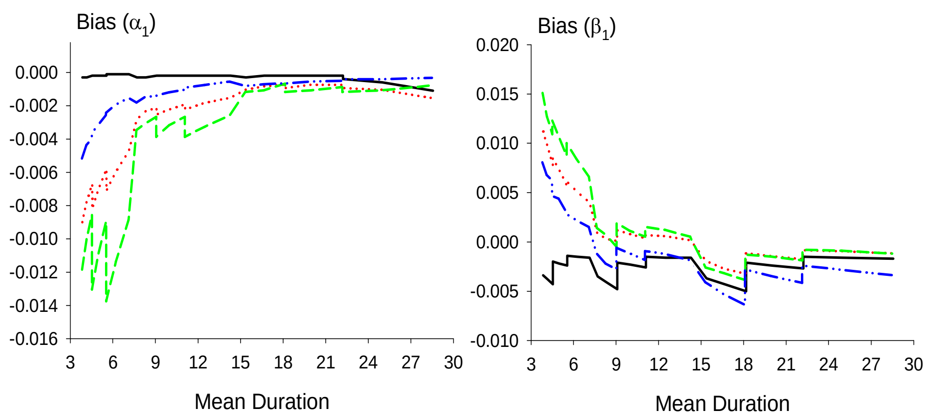

Figure 2: Biases for the estimators of α_1 and β_1 (T = 5 000 and 1 000 replications; Solid line indicates continuous duration ML estimator (C), dotted line discrete data with the continuous duration ML estimator (D), dashed line indicates the grouped data ML estimator (G) and the dot-dashed line the EM-algorithm (EM)).

The biases of the estimators of α_1 and β_1 are displayed in Figure 2 for $T = 5\,000$ and exponential data. It is quite obvious from the patterns for both parameters that the ML estimator based on continuous data has small bias. All other estimators are based on discretized data and manifest some bias for short durations, while bias is much smaller for longer durations. The largest bias for $\alpha_1 = 0.2$ and the shortest duration of 3.8 seconds is noted for the grouped data ML estimator and amounts to 6 percent. For the EM-algorithm the corresponding bias is less than 3 percent. For both parameters there is a clear-cut ranking of the estimators, in particular for the short mean durations. The biases of the EM-algorithm are smaller than the biases of the discretized data ML and grouped data ML estimators. It appears that the grouped data ML estimator has the weakest performance. As the EM-algorithm in this particular case also is rather fast to calculate it is our tentative choice of a best estimator. The results for $T = 50\,000$ reiterate the main conclusions derived from Figure 2. It is quite apparent that all estimators but the continuous data ML estimator (C) have a bias and that the EM-algorithm comes out as the least biased estimator for discretized data. For Weibull data the internal ranking between estimators remain relatively unaltered.

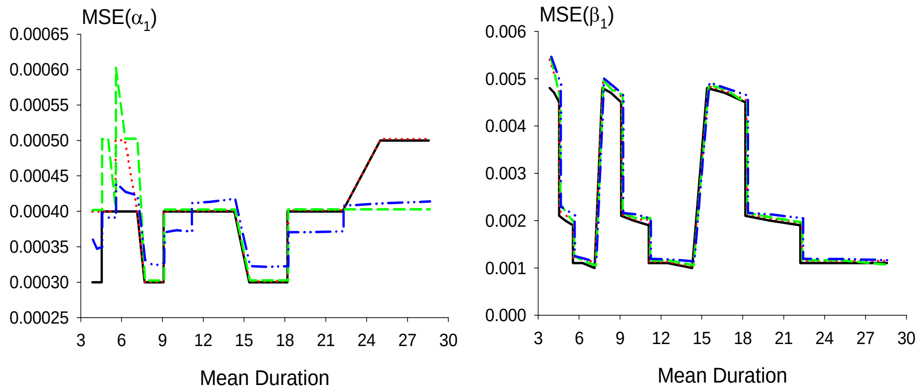

Figure 3: MSEs for the estimators of α_1 and β_1 (T = 5 000 and 1 000 replications; Solid line indicates continuous duration ML estimator (C), dotted line discrete data with the continuous duration ML estimator (D), dashed line indicates the grouped data ML estimator (G) and the dot-dashed line the EM-algorithm (EM)).

The MSE results of the α_1 and β_1 estimators are exhibited in Figure 3 for $T = 5\,000$ and the exponential data. When it comes to the MSEs for β_1 the most apparent feature is their striking similarity across mean durations. For this parameter the MSE is then completely dominated by the variance component. For α_1 there is some variation for short durations and for the long ones. The MSE of the EM-algorithm is not much different from those of the continuous duration ML estimator based on discretized data and the grouped data ML estimator. Among the latter two, the grouped data ML estimator has the weaker performance for short mean durations. For $T = 50\,000$ there is an expected drop in MSEs due to sample size, but the ranking between estimators remains unaltered. As expected the MSEs of the base case ML estimator for the exponential model (C) are the smallest in most cases, and also for the Weibull generated data.

In summary, among the estimators accounting for discretized data model the EM-algorithm is the preferred estimator in terms of bias. With respect to MSE it is not worse than the two competitors, though differences are quite small. No estimator manages to completely avoid bias for short mean durations.

4. Empirical Results

4.1 Data and Descriptives

Empirical results are reported for 15 days of trading (July 2, 2002 - July 22, 2002) in Ericsson B at the order driven Stockholmsbörsen stock exchange in Stockholm. The data were downloaded from the Ecovision system and processed further by the authors. The number of observed durations or the time series length is, after some reduction due to day changes, 57735. On average there are 3 849 durations per day. Figure 4 gives a histogram of the durations. The estimated average of the integer-valued duration is 7.4 seconds with a standard deviation of 11.2 seconds. The average varies between 2.9 to 13.5 seconds over the 15 days. About 79 percent of the durations are 10 seconds or shorter and the longest duration is 403 seconds. There appears to be a weak increase in average duration over the day so that trading is slightly less frequent towards the end of the trading day. However, there appears to be no strong reasons for deseasonalizing the series as done in some previous studies.

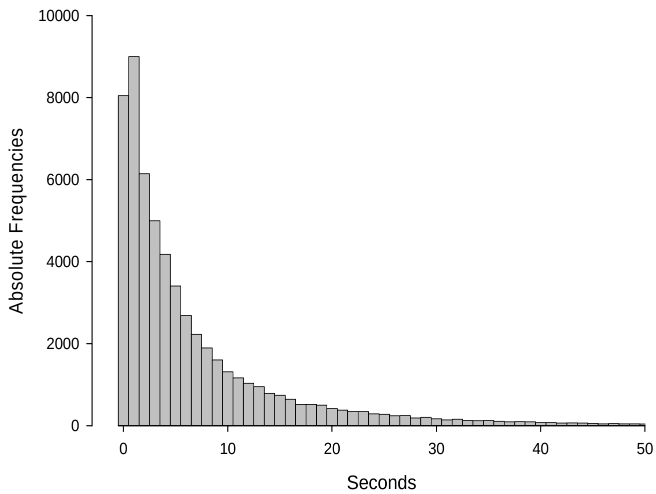

Figure 4: Histogram of discrete durations d_t (T = 57 735, the one percent of durations exceeding 50 seconds is excluded).

To give an indication of the trading volume, the number of traded stocks during the first day of the sample period is 12 596 496 with a closing price of 14.90 SEK. The trading volume in the major summer vacation month of July is usually smaller than during other months. Figure 5 gives the autocorrelation function for the time series of successive durations. The autocorrelations are quite small but the function decreases only slowly. Note that all autocorrelations are positive. The partial autocorrelations decrease rather quickly and are approximately zero after 5-6 lags. The patterns of Figure 5 and the partial autocorrelation function indicate that the model should be able to capture low order both autoregressive and moving average effects.

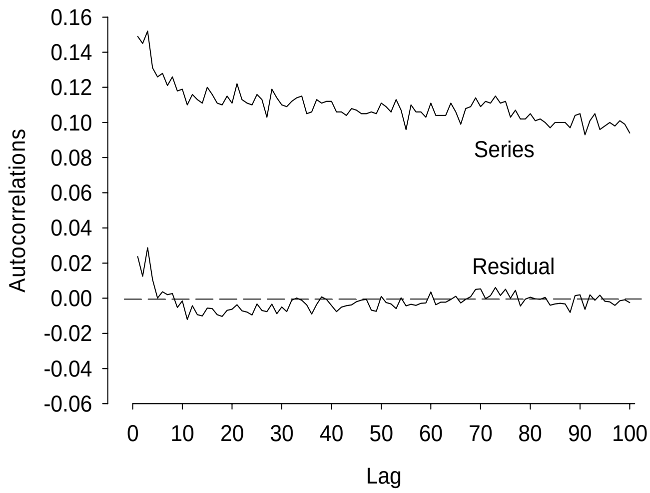

Figure 5: Autocorrelation function for durations d_t and residuals (calculated as $(d_t - \hat{\theta}_t)/\hat{\theta}_t$) from the exponential model.

For a pure time series analytical approach (i.e. $\pi = 0$ in (4)) a reasonable starting point is to search for a model with $p \leq 3$ and $q \leq 3$. In addition, in the final models we include as explanatory variables the price (mean 14.61, standard deviation 1.90), the spread (0.10, 0.02) and the number of traded stocks (2134.5, $1.48 \cdot 10^6$) ending the previous duration. Including the first two variables as changes instead of as levels was rendered empirical support, see below. For example, the price part of the model, $\pi_1 p_{t-1} + \pi_2 p_{t-2}$, was used and empirically we found $\hat{\pi}_1 \approx -\hat{\pi}_2$. This suggests the use of a restricted $\pi(p_{t-1} - p_{t-2}) = \pi \nabla p_{t-1}$ specification, i.e. in terms of a change.

4.2 Estimation Results

To estimate the parameters we assume three parametric density specifications that have been used previously - the exponential, the Weibull and the Burr (see the Appendix for a brief account of the latter two distributions). The Weibull contains the exponential model as a special case. These two densities were also included in the Monte Carlo study of the previous section. The Burr model is more flexible than Weibull in that it has more parameters and then a more flexible hazard function. The Burr model does not nest neither the exponential nor the Weibull models, so that straightforward use of, e.g., likelihood ratio tests for model selection is ruled out. We employ two versions of the EM-algorithm for the exponential model.

The continuous exponential model served as a tool for determining the model specification.[2] The best model has $R^2 = 0.1$. There is some remaining serial correlation in all models to be reported and this could not be eliminated, cf. Figure 5 for the autocorrelation function corresponding to column one of Table 1.[3] Note that no serial correlations are determined for the discretized models. No serial correlation remains in the squared residuals, except for in the Burr model and for the model of the final column of Table 1. Individual correlations are, however, quite small and the Ljung-Box statistic is obviously influenced by the large sample size.

The estimation results are presented in Tables 1-3. The parameter estimates are throughout almost exclusively of the same sign, roughly of similar sizes and when significant this happens across models and estimators. Note that there are more lags in these models than in most previous models.

Table 1 reports results based on an assumed, continuous variable exponential model. It is found that the estimated models of this table (and other tables) satisfy the stationarity condition on the α and β parameters, albeit with a rather narrow margin. Initially, alternative lag structures (different p and q values) were tried. Table 1 also reports on how explanatory variables should be included. There is strong support throughout for utilizing

---

[2] In this case and whenever continuous variable methods are employed 0 is replaced by 0.5 seconds to reflect mid-interval-value in the same way as for longer durations.

[3] The residual is defined as $r_t = (d_t - E(D_t \mid \Delta_{t-1})/V^{1/2}(D_t \mid \Delta_{t-1})$, where $E(.)$ and $V(.)$ are different for the different distributions. The squared residual is r_t^2.

change variables for the price and spread. If, e.g., the price follows a random walk the change corresponds to the innovation or the unpredicted new information over the previous duration. A positive price change leads to a longer duration. The effect of the spread change is negative but not significant. A higher trading volume prolongs the next duration but not significantly so. The final column suggests that separate inclusion of v_{t-1} and v_{t-2} is preferable judging by the log-likelihood values. However, the serial correlation properties speak against this specification. The change variables will be retained in all further model estimations.

Table 1: Maximum likelihood estimates for alternative specifications of the continuous exponential model.

Variable	Coeff	s.e.	Coeff	s.e.	Coeff	s.e.	Coeff	s.e.
$d_{t\text{-}1}$	0.0378	0.0029	0.0381	0.0029	0.0408	0.0030	0.0356	0.0030
θ_{t-1}	0.8962	0.0603	0.9162	0.0756	0.8712	0.0524	0.9905	0.0149
θ_{t-2}	-0.2634	0.0803	-0.3072	0.0966	-0.3350	0.0697	-0.3120	0.0185
θ_{t-3}	0.3282	0.0564	0.3506	0.0645	0.4818	0.0041	0.2834	0.0275
Price change	2.8274	0.3564	--		3.1143	0.3419	1.6284	0.2899
p_{t-1}	--		0.4985	0.0090	--		--	
p_{t-2}	--		-0.4954	1.3161	--		--	
Spread change	-1.7938	1.1494	-0.8952	0.8908	--		1.0475	1.0689
s_{t-1}	--		--		-1.0680	0.8481	--	
s_{t-2}	--		--		1.2123	1.3161	--	
Volume	0.6266	2.6438	-0.7224	1.1746	0.8041	1.1797	--	
v_{t-1}	--		--		--		0.6973	0.0048
v_{t-2}	--		--		--		-0.6945	1.3161
Constant	0.0103	0.0354	-0.0250	0.1116	-0.0038	0.0935	0.0066	0.0531
LB_{100}	235.4		233.1		234.9		382.5	
LB^2_{100}	6.7		6.2		6.7		828.6	
ln$L$	-165860		-165860		-165859		-165743	

Notes: Volume pertains to the previous transaction, while v_t is the accumulated (within the day) trading volume. Both are throughout devided by 10 000 000. LB_{100} is the Ljung-Box statistic of the standardized residual over 100 lags. LB^2_{100} is the same statistic for squared residuals.

In Table 2 a comparison within the exponential model of using continuous or grouped data is reported. There are no substantial differences between the ML estimators based on the two data types. The two versions of the EM-algorithm are quite similar, too. Given this result, arguments supportive of the conventional QML estimator are strengthened even if data are discretized.

Table 2: Maximum likelihood estimates for alternative specifications of the continuous exponential model.

	ML-Continuous		ML-Grouped		EM-Grouped		EM-full	
Variable	Coeff	s.e.	Coeff	s.e.	Coeff	s.e.	Coeff	s.e.
d_{t-1}	0.0378	0.0029	0.0380	0.0029	0.0379	0.0029	0.0403	0.0029
θ_{t-1}	0.8962	0.0603	0.9172	0.0548	0.8800	0.0532	0.8841	0.0530
θ_{t-2}	-0.2634	0.0803	-0.3036	0.0742	-0.2681	0.0712	-0.2693	0.0704
θ_{t-3}	0.3282	0.0564	0.3473	0.0522	0.3492	0.0507	0.3337	0.0496
Price change	2.8274	0.3564	2.9936	0.3350	3.1907	0.3330	3.1266	0.3303
Spread change	-1.7938	1.1494	-3.0104	1.0554	-2.9918	1.0574	-1.4293	1.0879
Volume	0.6266	2.6438	0.8230	2.5724	0.9936	2.5879	0.7678	2.5562
Constant	0.0103	0.0354	0.0092	0.0317	0.0078	0.0311	0.0096	0.0314
LB	235.4		--		--		--	
LB^2_{100}	6.7		--		--		--	
ln$L$	-165860		-131050		-165096		-165105	

Notes: Volume pertains to the previous transaction, while v_t is the accumulated (within the day) trading volume. Both are throughout devided by 10 000 000. LB_{100} is the Ljung-Box statistic of the standardized residual over 100 lags. LB^2_{100} is the same statistic for squared residuals.

Table 3 studies this issue further; if the QML is to be useful we would expect no large changes in parameter estimates even if the exponential model is not the 'true' one. The qualitative conclusions correspond to those of the exponential model, though sizes of estimates are slightly different. The exponential model is nested within the Weibull model and the exponential can be rejected against the Weibull model ($\hat{\gamma}$ is significantly smaller than one). The Weibull model is not nested within the more general Burr model, though the γ and λ estimates of the latter model may indicate that the shape of the Weibull hazard is not supported by data. The Weibull duration dependence parameter γ is significantly smaller than one, which implies a decreasing hazard function and that the exponential model can be rejected. In a similar way the form of the Burr hazard function is an indication against the exponential model.

Table 3: Maximum likelihood estimates for alternative model specifications.

	Weibull				Burr			
	Continuous		Grouped		Continuous		Grouped	
Variable	Coeff	s.e.	Coeff	s.e.	Coeff	s.e.	Coeff	s.e.
d_{t-1}	0.0393	0.0425	0.0437	0.0042	0.0386	0.0033	0.0418	0.0042
θ_{t-1}	0.8875	0.0678	0.7792	0.0698	0.9994	0.0629	0.9990	0.0771
θ_{t-2}	-0.2967	0.0979	-0.2748	0.0860	-0.5001	0.0848	-0.5159	0.1045
θ_{t-3}	0.3663	0.0629	0.4455	0.0631	0.4434	0.0589	0.4679	0.0709
Price change	3.0200	0.4241	3.1342	0.4640	2.9754	0.4057	2.5722	0.4725
Spread change	-1.2306	1.3314	-1.1535	1.4182	-0.7020	1.0991	-0.5374	1.3905
Volume	0.6712	3.0271	0.2148	3.3013	-0.9633	2.2659	-0.0063	3.1934
Constant	0.0132	0.0425	0.0045	0.0470	0.0286	0.0405	0.0044	0.0426
γ	0.9130	0.0023	0.8054	0.0005	1.1263	0.0019	0.8286	0.0044
λ	--		--		0.3600	0.0083	0.0350	0.0032
LB_{100}	229.7		--		201.3		--	
LB^2_{100}	7.1		--		240.9		--	
ln$L$	-165409		-129178		-164497		-129152	

Note: See Table 3 for explanations.

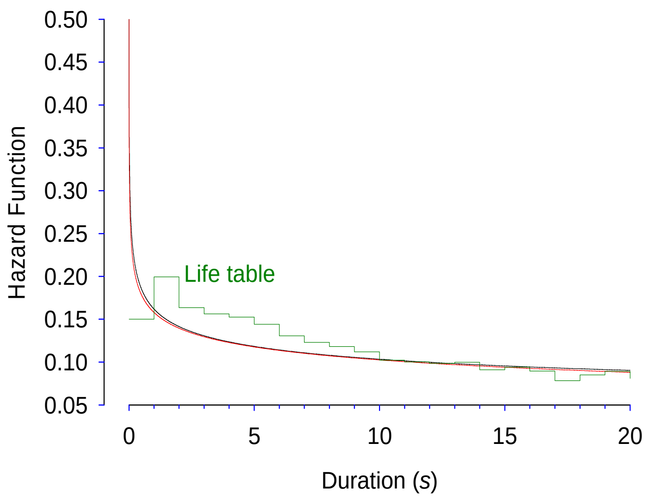

Figure 6: Hazard functions based on grouped ML estimates of Table 4 and evaluated at the sample mean of the duration variable, and a life table estimated hazard function.

Figure 6 shows the Burr and Weibull hazard functions, when θ is replaced by the sample mean and estimates from Table 3 are used for γ and λ. The hazard functions are hardly distinguishable and decrease rapidly within the first second, but are roughly constant thereafter. Hence, these hazards differ the most from the life table estimate in the (0,1) interval and discrimination between the two parametric models would obviously be much strengthened if short and continuous duration data in the (0,1) interval were available.

We also studied whether the response to news is symmetric in the sense that positive and negative news affect subsequent durations in the same way. The potentially asymmetric response to news (the variables are constructed as $\nabla x_t^+ = \max(0, \nabla x_t)$ and $\nabla x_t^- = \min(0, \nabla x_t)$) is studied in terms of the price and spread changes within the framework of the grouped data Weibull and Burr models. By likelihood ratio tests we find no evidence of asymmetric response to price changes and the two estimates for positive and negative changes are quite similar. There are different responses to spread changes depending on their signs, but not significantly so. Individually neither of the spread change effects appear to have a significant effect.

5. Conclusions

The paper has discussed the discrete nature of duration measures between transactions in stocks and studied the consequences of this discretization of a continuous time scale.

Grouped maximum likelihood and EM-algorithm estimators were discussed. In the small Monte Carlo study the EM-algorithm that accounts for the discrete nature of the data both in the outcome and the lagged explanatory variables comes out as the best estimator of the compared ones. In the empirical study the differences between estimators are generally quite small, and the EM-algorithm and ML estimators based on discrete data are not too different from ML based on grouped data and Weibull and Burr models.

When it comes to the effects of explanatory variables the study provided support for using changes rather than levels to reflect news. There is throughout a significant and positive effect of news about prices and a negative effect of a change in the spread. The spread effect is not significant, however. A higher volume has an insignificant but prolonging effect in most cases. We could not find statistically significant support for an asymmetric response to news about spreads nor about prices. The log-likelihood function value of the Burr is larger than for other models but the models are not nested. In addition, the serial correlation properties of the exponential and Weibull models speak in favor of these two models. A generalized gamma was also employed and provided a better fit to the data than both the exponential and Weibull models. A reason for not reporting generalized gamma results is the numerical problems we faced in obtaining standard errors.

Appendix

Weibull

Using the specification of Bauwens and Giot (2001, pp. 98-99) the Weibull model has hazard function $\lambda(D) = \gamma D^{\gamma-1} / \theta^{\gamma}$, which gives the integrated hazard function $\Lambda(D) = (D/\theta)^{\gamma}$. From this follows the distribution function $F(D) = 1 - \exp(-\Lambda(D))$, expected value $E(D) = \theta\,\Gamma(1+\gamma^{-1})$ and variance $V(D) = \theta^2\left[\Gamma(1+2\gamma^{-1}) - \Gamma(1+\gamma^{-1})\right]$. Random durations can be generated according to $D = \theta\left[-\ln(1-u)\right]^{\gamma^{-1}}$, where u is a uniform [0,1] random deviate. A standardization of D to get the moments of the exponential model is obtained by $D_* = a^{-1/2}(D - \theta b)$, where $a = \left[\Gamma(1+2\gamma^{-1}) - \Gamma(1+\gamma^{-1})\right]$ and $b = \Gamma(1+\gamma^{-1}) - a^{1/2}$.

Corresponding to the log-likelihood function in (6) we have

$$\begin{aligned}
\ln L &= \sum_{t=r}^{T} l_t = \sum_{t=r}^{T} \ln\left[\exp(-\eta_t\left[(d_t - 1)/\theta_t\right]^{\gamma}) - \exp(-\left[(d_t+1)/\theta_t)\right]^{\gamma}\right] \\
&= \sum_{t=r}^{T} \ln\left[g_{1t} - g_{2t}\right],
\end{aligned} \tag{A.1}$$

where the final step is notational. The derivatives for l_t are

$$\begin{aligned}
\frac{\partial l_t}{\partial \gamma} &= \frac{1}{\theta_t^{\gamma}} \frac{\left[-\eta_t(d_t-1)^{\gamma}\ln\left(\frac{d_t-1}{\theta_t}\right)\right]g_{1t} + \left[(d_t+1)^{\gamma}\ln\left(\frac{d_t+1}{\theta_t}\right)\right]g_{2t}}{g_{1t} - g_{2t}} \\
\frac{\partial l_t}{\partial \theta_t} &= \frac{\gamma}{\theta_t} \frac{\eta_t\left(\frac{d_t-1}{\theta_t}\right)^{\gamma} g_{1t} - \left(\frac{d_t+1}{\theta_t}\right)^{\gamma} g_{2t}}{g_{1t} - g_{2t}} \\
\frac{\partial l_t}{\partial \psi} &= \frac{\partial l_t}{\partial \theta_t} \cdot \frac{\partial \theta_t}{\partial \psi} = \mathbf{z}_t' \frac{\partial l_t}{\partial \theta_t}.
\end{aligned} \tag{A.2}$$

The conditional expectations corresponding to those in (3) and required for EM-algorithms are of the form

$$\begin{aligned}
E(D_t \mid d_t = 0, \Delta_{t-1}) &= c\,\frac{P(1+\gamma^{-1}, 1/\theta_t^{\gamma})}{1 - e^{-1/\theta_t}} \\
E(D_t \mid d_t = k, \Delta_{t-1}) &= c\,\frac{P\left[1+\gamma^{-1}, \left(\frac{k+1}{\theta_t}\right)^{\gamma}\right] - P\left[1+\gamma^{-1}, \left(\frac{k-1}{\theta_t}\right)^{\gamma}\right]}{e^{-(k-1)^{\gamma}/\theta_t} - e^{-(k+1)^{\gamma}/\theta_t}},
\end{aligned} \tag{A.3}$$

where $c = \theta_t \Gamma(1+\gamma^{-1})$ and $P(.,.)$ is the incomplete gamma function (e.g., Press et al., 1992, p. 209).

Burr

Bauwens and Giot (2001, pp. 101-104) give the Burr density function:

$$f(D) = \frac{\gamma}{\theta}\left(\frac{D}{\theta}\right)^{\gamma-1}\left[1+\lambda\left(\frac{D}{\theta}\right)^{\gamma}\right]^{-1-1/\lambda} \tag{A.4}$$

with mean and variance

$$\begin{aligned} E(D) &= \theta\,\frac{\Gamma(1+\gamma^{-1})\Gamma(\lambda^{-1})}{\lambda^{1+\gamma^{-1}}\Gamma(1+\lambda^{-1})}, \quad \text{for } \gamma/\lambda > 1 \\ V(D) &= \theta^2\,\frac{1}{\lambda^{1+2\gamma^{-1}}\Gamma(1+\lambda^{-1})} \\ &\quad \times \left[\Gamma(1+2\gamma^{-1})\Gamma(\lambda^{-1}+2\gamma^{-1}) - \frac{\Gamma(1+\gamma^{-1})\Gamma(\lambda^{-1})}{\lambda\Gamma(1+\lambda^{-1})}\right], \quad \text{for } \gamma/\lambda > 2. \end{aligned} \tag{A.5}$$

The survival and hazard functions are

$$\begin{aligned} \overline{F}(D) &= \left[1+\lambda\left(\frac{D}{\theta}\right)^{\gamma}\right]^{-1/\lambda} = c^{-\lambda^{-1}} \\ h(D) &= \frac{\gamma}{\theta}\left(\frac{D}{\theta}\right)^{\gamma-1}\left[1+\lambda\left(\frac{D}{\theta}\right)^{\gamma}\right]^{-1}. \end{aligned} \tag{A.6}$$

Using (A.4)-(A.6) it is then possible to obtain the log-likelihood function corresponding to (6) and then to obtain ML estimates. The derivatives of $\overline{F}(D)$ with respect to γ, θ (and $\boldsymbol{\psi}$) and λ make up the score vector and are given by

$$\begin{aligned} \frac{\partial \overline{F}(D)}{\partial \gamma} &= -\left(\frac{D}{\theta}\right)^{\gamma}\ln\left(\frac{D}{\theta}\right)c^{-\lambda^{-1}-1} \\ \frac{\partial \overline{F}(D)}{\partial \theta_t} &= \frac{\gamma}{\theta}\left(\frac{D}{\theta}\right)^{\gamma} c^{-\lambda^{-1}-1} \\ \frac{\partial \overline{F}(D)}{\partial \boldsymbol{\psi}} &= \frac{\partial \overline{F}(D)}{\partial \theta_t}\cdot\frac{\partial \theta_t}{\partial \boldsymbol{\psi}} = \mathbf{z}_t'\,\frac{\partial \overline{F}(D)}{\partial \theta_t} \\ \frac{\partial \overline{F}(D)}{\partial \lambda} &= \frac{1}{\lambda}\left[\lambda^{-1}\ln(c) - c^{-1}\left(\frac{D}{\theta}\right)^{\gamma}\right]c^{-\lambda^{-1}}. \end{aligned} \tag{A.7}$$

References

Bauwens, L. and Giot, P. (2001). *Econometric Modelling of Stock Market Intraday Activity*. Kluwer, Dordrecht.

Cameron, A.C. and Trivedi, P.K. (1998). *Regression Analysis of Count Data*. Cambridge University Press, Cambridge.

Cappé, O., Douc, R., Moulines, E. and Robert, C. (2002). On the Convergence of the Monte Carlo Maximum Likelihood Method for Latent Variable Models. *Scandinavian Journal of Statistics* **29**, 615-635.

Chiang, M-H. and Wang, C-H. (2004). Intradaily Relationship between Information Revelation and Trading Duration under Market Trends: The Evidence of MSCI Taiwan Stock Index Futures. *Applied Economics Letters* **11**, 495-501.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion). *Journal of the Royal Statistical Society* **B39**, 1-38.

Easley, D. and O'Hara, M. (1992). Time and the Process of Security Price Adjustment. *Journal of Finance* **47**, 576-605.

Engle, R.F. and Russell, J.R. (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica* **66**, 1127-1162.

Glosten, L. and Milgrom, P. (1985). Bid, Ask, and the Transaction Price in a Specialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics* **13**, 71-100.

Gourieroux, C. and Monfort, A. (1991). Simulation Based Econometrics in Models with Heterogeneity. *Annales d'Economie et de Statistique* **20/1**, 69-107.

Kulldorff, G. (1961). *Contributions to the Theory of Estimation from Grouped and Partially Grouped Samples*. PhD thesis. Almqvist & Wicksell, Stockholm.

Lee, L-F. (1997). Simulated Maximum Likelihood Estimation of Dynamic Discrete Choice Statistical Models: Some Monte Carlo Results. *Journal of Econometrics* **82**, 1-35.

Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. Wiley, New York.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992). *Numerical Recipes*. Cambridge University Press, Cambridge.

An Empirical Model for Durations in Stocks

Ola Simonsen*
Department of Economics, Umeå University
SE-901 87 Umeå, Sweden
email: ola.simonsen@econ.umu.se

Abstract

This paper considers an extension of the univariate autoregressive conditional duration model to which durations from a second stock are added. The model is empirically used to study duration dependence in four traded stocks, Nordea, Föreningssparbanken, Handelsbanken and SEB A on the Stockholm Stock Exchange. The stocks are all active in the banking sector. It is found that including durations from a second stock may add explanatory power to the univariate model. We also find that spread changes have significant effect for all series.

Key Words: Finance, multivariate, transaction data, market microstructure, Granger causality.

JEL Classification: C12, C32, C41, G14.

*The author would like to thank Kurt Brännäs and Jörgen Hellström for valuable comments and suggestions. Comments were also received at the Economic Society of Australia 2005. Financial support from the Wallander-Hedelius Foundation is gratefully acknowledged.

1 Introduction

This paper empirically examines the dependence between durations in stocks from the banking sector traded on the Stockholm Stock Exchange. In this context durations are the times between two consecutive transactions in a stock. To empirically capture dependence between durations in stocks the Autoregressive Conditional Duration model (ACD) of Engle and Russell (1998) is extended by adding lagged durations from a second stock to the conditional mean function.

The economic motivation to studying trade intensity or, the time between transactions originates from the information based model of Glosten and Milgrom (1985). In their model they argue that the information set diffuses among informed and uninformed traders. The informed traders have superior information, public or private, regarding the true value of the asset and make use of their information advantage as long as it is profitable. Uninformed traders, however, do not have private information concerning the value of the asset and trade mainly for liquidity reasons. To distinguish between informed and uninformed traders Easley and O'Hara (1992) highlights the importance of the trade intensity or, the time between transactions. In their model non-trading, i.e. long durations between transaction, is an indicator of no new information, neither good nor bad. Trading is on the other hand an indicator of new information. Thus, the probability is high that transactions involve informed traders when the observed durations are short.

New available information to the stockmarket may, for example, concern a specific stock, an industry sector or even the whole stockmarket. The stock specific news may, however, not only be relevant for the value of the specific stock. Values of related stocks, e.g., in the same sector, may also be affected. The possibility that the new information discovered by the increased trade intensity in a specific stock also is relevant for related stocks may therefore influence trading in, e.g., stocks in the same sector. Thus, new information may lead to dependence between durations in different stocks.

With the availability of ultra high frequency data (i.e., within the day series where every transaction is registered) standard econometric tools may not be appropriate as transactions are irregularly spaced over time. One way of handling the irregularity is to aggregate data into regularly spaced intervals and apply, e.g., count data models. However, aggregation results in information loss which is not always desirable. Albeit the complication of modelling irregularly spaced data one influential approach for handling the data is the ACD model. It models the duration between transactions and conditions the duration on recent durations and other explanatory variables. Within the ACD framework several extensions and applications of the original Engle and Russell (1998) model have been presented (e.g., Bauwens and Giot, 2001).

Models for the dependence between durations in more than one duration series have proven rather complicated, e.g., Engle and Lunde (2003), Russell (1999) and Bauwens and Hautsch (2004). The difficulty is in modelling an expected duration while at the same time accounting for events during the duration. Pairs of durations with the same starting time would be ideal, but are rarely available in transaction data. Both Engle and Lunde (2003) and Russell (1999) advance a model for the dependence between the trade and the quote arrival processes. Bauwens and Hautsch (2004) present a model for the dependence between durations and report results for German stocks. The model of Engle and Lunde (2003) is closely related to the ACD model as it treats one of the duration series as censored and applies the ACD model. In the same line is the model of Mosconi and Olivetti (2005). Russell (1999) and Bauwens and Hautsch (2004) take another approach and instead model the intensities. An application of the model of Russell (1999) for dependences between duration series have been given by Spierdijk et al. (2002). They find positive dependence between stocks active in the same industry sector.

Here, a technically simpler approach to capturing dependence between duration series is suggested. In the ACD model durations from a second stock beyond the focused one are added to the conditional mean function and the result is empirically evaluated. Admittedly, the proposed models are rather ad hoc even though they attempt to capture the main features of the data generating process. However, the empirical evidence of the current type of models may serve a purpose or inspiration for both building a base of empirical evidence and more probabilistic work. Maximum likelihood and conditional least square are suggested as estimators of the unknown parameters. Further, an impulse response function is plotted and Granger causality is studied. Empirical results are presented for 98 trading days in four stocks in the banking sector at the Stockholm Stock Exchange. The stocks are Nordea, Föreningssparbanken, Handelsbanken and SEB.

The outline of the paper is the following. In Section 2 we give a brief account of the ACD model and present the extended version. The section ends with details concerning estimation. Section 3 presents the data, while Section 4 presents the empirical results. The final section concludes.

2 Model and estimation

In this section we first give a brief account of the ACD model and then proceed to discuss the extension which is aimed at accounting for a second transaction series. The estimation of unknown parameters is also considered.

2.1 Autoregressive conditional duration model

In the ACD model of Engle and Russell (1998) the time between consecutive transactions is modelled. The duration, d_i, is the time between two consecutive transactions at t_{i-1} and t_i, i.e. $d_i = t_i - t_{i-1}$. The conditional expectation of a duration d_i is specified as $E(d_i\,|d_{i-1}, ..., d_1; x) = \theta_i$. The d_i is conditioned on past durations and other explanatory variables x, and θ_i is specified in such a way that $\epsilon_i = d_i/\theta_i$ is independent and identically distributed. The θ_i may be parameterized as

$$\theta_i = \omega + \sum_{j=1}^{p} \alpha_j d_{i-j} + \sum_{j=1}^{q} \beta_j \theta_{i-j} + \pi' x_{i-1}. \tag{1}$$

In (1) θ_i is parameterized with p lagged durations and q lags of conditional durations. This is called an ACD(p,q,x) where p and q are the orders of the lags in the mean function and x_i is a vector of explanatory variables such as volume and spread at time i.

The unconditional mean and variance of an ACD(1,1) are easy to obtain. By assuming that ϵ_i follows a standard exponential distribution, $\pi = 0$ in (1), and that d_i is weakly stationary the mean is

$$E(d_i) = \mu = \frac{\omega}{1 - \alpha - \beta}. \tag{2}$$

The time invariant unconditional mean together with the obvious $d_i > 0$ condition imply the restrictions $\omega > 0$ and $\alpha + \beta < 1$. The unconditional variance of the ACD(1,1) is given by

$$Var(d_i) = \sigma^2 = \mu^2 \times \frac{1 - \beta^2 - 2\alpha\beta}{1 - \beta^2 - 2\alpha\beta - 2\alpha^2}. \tag{3}$$

For finite variance $\beta^2 + 2\alpha\beta + 2\alpha^2 < 1$ must hold.

2.2 Extended ACD model

To capture the dependence between durations in stocks we suggest that the ACD model be extended by adding durations from stocks beyond the focused one. The available durations for the focused stock are past finished durations in the same and the other stocks, and the length of the most recent incomplete duration in other stocks. For the presentation of the model we introduce a counting process and then proceed with the suggested formulation of the model.

Let the number of events for the k^{th} duration series by time t be denoted by $N^k(t)$ and $t_0^k < t_1^k < t_2^k < ... < t_i^k$ the transaction times for the k^{th} series. The most recent transaction time is then at $t^k_{N^k(t)}$ and the most recent duration is $d^k_{N^k(t)} = t^k_{N^k(t)} - t^k_{N^k(t)-1}$.

An example with two stocks, 1 and 2, is presented in Figure 1. The expected length of the next duration $d^1_{N^1(t)}$ is conditioned at $t^1_{N^1(t)-1}$ in stock 1. The $d^1_{N^1(t)}$ may be conditioned on its own past durations, e.g., $d^1_{N^1(t)-1}$, $d^1_{N^1(t)-2}$,...,$d^1_{N^1(t)-q}$, and other explanatory variables. If durations from a second stock are added to the conditional mean function, e.g., $d^2_{N^2(t)}$, $d^2_{N^2(t)-1}$ and $d^2_{N^2(t)-2}$, the most recent added duration $d^2_{N^2(t)}$ in stock 2 may not be completed when the conditioning takes place in stock 1. The observed length of the most recent duration $d^2_{N^2(t)}$ in stock 2 is τ_1 where $\tau_1 = t^1_{N^1(t)-1} - t^2_{N^2(t)-1}$ or $\tau_1 = \sum_{k=1}^{N^1(t)} d^1_{k-1} - \sum_{k=1}^{N^2(t)} d^2_{k-1}$, and $\tau_1 \geq 0$.

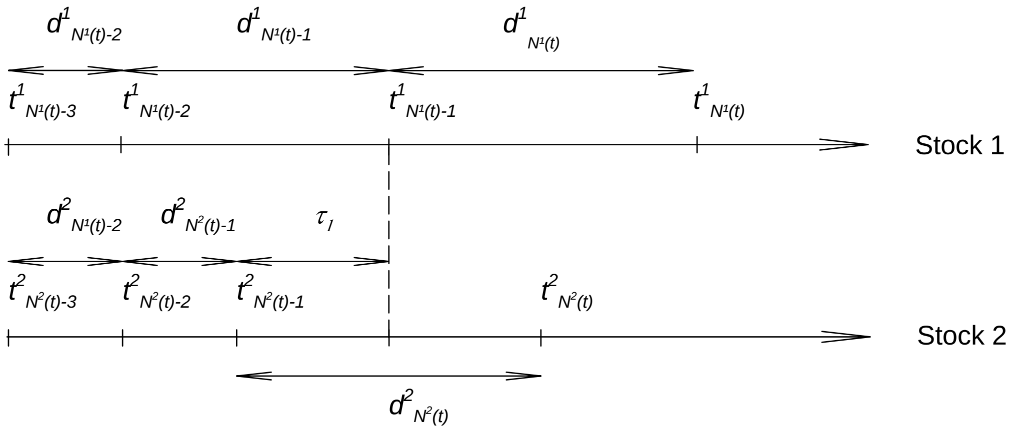

Figure 1: Illustration of two stocks, stock 1 and stock 2, with transaction times t and durations d. τ_1 is the observed length of the most recent duration in stock 2 at time $t^1_{N^1(t)-1}$.

The proposed, extended ACD model is in a bivariate case for stocks 1 and 2:

$$\theta^1_{N^1(t)} = \omega^1 + \sum_{j=1}^{p_1} \alpha^1_j d^1_{N^1(t)-j} + \sum_{j=1}^{q_1} \beta^1_j \theta^1_{N^1(t)-j} + \pi' x^1_{N^1(t)-1} + e^{\delta^1_0 \tau_1}(1 + \sum_{j=1}^{r_1} \delta^1_j d^2_{N^2(t)-j}) \qquad (4)$$

$$\theta^2_{N^2(t)} = \omega^2 + \sum_{j=1}^{p_2} \alpha^2_j d^2_{N^2(t)-j} + \sum_{j=1}^{q_2} \beta^2_j \theta^2_{N^2(t)-j} + \pi' x^2_{N^2(t)-1} + e^{\delta^2_0 \tau_2}(1 + \sum_{j=1}^{r_2} \delta^2_j d^1_{N^1(t)-j}). \qquad (5)$$

In (4) where completed durations from a stock 2 are added with r_1 lags, completed durations are given weights dependent on the size of τ_1, i.e. depending on how far away in time the completed durations are. The intuition behind the parameterization is that a large value of τ_1 in $\exp(\delta^1_0 \tau_1)$ gives low weight to the finished durations from stock 2 and a small value of τ_1 gives a larger weight to the finished durations when $\delta^1_0 < 0$. The conditional mean function (5) for stock 2 with durations from stock 1 added is formulated in a analogous manner.

When durations from a second stock are added to the conditional mean function as suggested the crucial point is the updating of the mean function. For the most recent duration from stock 2 to enter the mean function (4) a transaction in stock 1 must occur. During a long period without a transaction in stock 1 but with transactions in stock 2 information from stock 2 may not be included in the mean function for stock 1.

Obviously, the specifications of the final terms in (4)-(5) could take on other forms. For instance, the final term could be additive with $\theta^2_{N^2(t)-j}$, $j \geqslant 1$, and τ, entering (4) in a linear way and vice versa.

The conditional moment functions in (4)-(5) are easy to interpret with respect to effects of changes in post durations, etc. Obtaining the unconditional mean and variance is, however, quite difficult as it involves taking the expectation of the nonlinear final term in (4)-(5). To illustrate, consider the properties of an ACD(1,1) with one lag of the added duration. The model can be written

$$d^1_{N^1(t)} = \theta^1_{N^1(t)}\epsilon^1_{N^1(t)}, \qquad \theta^1_{N^1(t)} = \omega^1 + \alpha^1 d^1_{N^1(t)-1} + \beta^1\theta^1_{N^1(t)-1} + e^{\delta^1_0\tau_1}(1 + \delta^1_1 d^2_{N^2(t)-1}), \quad (6)$$

where $\epsilon^1_{N^1(t)}$ follows a standard exponential distribution. By taking expectations of both sides of (6), assuming that $d^1_{N^1(t)}$ and $\theta^1_{N^1(t)}$ are weakly stationary the unconditional mean may be expressed as

$$\mu_1 = E(\theta^1_{N^1(t)}) = E(d^1_{N^1(t)}) = \frac{\omega^1 + \varphi^1}{1 - \alpha^1 - \beta^1},$$

where $\varphi^1 = E\left[e^{\delta^1_0\tau_1}(1 + \delta^1_1\theta^2_{N^2(t)})\right]$ can be either positive or negative depending on the sign of δ^1_1. To have positive durations the most natural conditions are $\alpha^1 + \beta^1 < 1$ and $\omega^1 > -\varphi^1$. Obtaining an explicit formulation of φ^1 and the variance appears more difficult than illuminating, and hence detailed expressions are not given.

One useful extension of the original ACD model that ensures positive expected durations is the log-ACD model by Bauwens and Giot (2000). The conditional mean function for a log-ACD(p,q) with explanatory variables is then

$$\theta^1_{N^1(t)} = \exp\left[\omega^1 + \sum_{j=1}^{p_1}\alpha^1_j d^1_{N^1(t)-j} + \sum_{j=1}^{q_1}\beta^1_j \ln\theta^1_{N^1(t)-j} + \pi' x^1 + e^{\delta^1_0\tau_1}(1 + \sum_{j=1}^{r_1}\delta^1_j d^2_{N^2(t)-j})\right]. \quad (7)$$

For the specification log-ACD(1,1) the only parameter restriction is $\beta^1 < 1$ and $\delta^1_0 < 0$.

2.3 Estimation

Engle and Russell (1998) popularized the quasi maximum likelihood (QML) estimator building on the exponential distribution for the estimation of the unknown parameters of the ACD model. By extending the conditioning set to also include observed durations in the other duration sequence, the same QML estimator can be used for our purpose. The estimator maximizes the log-likelihood function,

$$\ln \ell = -\sum_{j=1}^{T} \left[\ln \theta_j^k + \frac{d_j^k}{\theta_j^k} \right]. \tag{8}$$

The QML estimator is consistent when the conditional mean function is correctly specified. A correctly specified distribution may, however, render a more efficient estimator. For the practical estimation of the parameters we employ the log-ACD specification in (7), which we implement in the RATS package using the BFGS algorithm. The standard errors of the parameter estimates are the robust standard errors given by RATS.

Other estimators for the unknown parameters may also be considered, e.g., conditional least squares (CLS) or generalized method of moments (GMM). With these estimators the distribution assumption on the conditional duration is relaxed, though, relaxing the distributional assumption is not a major drawback given that the density is in the exponential family and θ_j^k is correctly specified (Bauwens, Giot, Grammig and Veredas, 2004).

Starting with the CLS estimator the prediction error from (4)-(5) is

$$e_{N^k(t)}^k = d_{N^k(t)}^k - \theta_{N^k(t)}^k.$$

The CLS estimator of the parameters in the vector ψ' minimizes the sum of the squared prediction errors

$$Q(\psi) = e'e$$

where $e' = (e_1', e_2')$. Standard errors of the estimated parameters are obtained from the robust error option in RATS. A major benefit of leaving the QML based on a univariate exponential model is that joint estimation of the two conditional mean functions is feasible. By this it is easier to extend previous conditional mean functions to, e.g., be functions of lagged conditional means from other duration sequences. Naturally, if considering the joint estimation of the conditional mean functions the irregular intervals and updating of the conditional mean functions must be carefully considered.

The GMM estimator introduced by Hansen (1982) may also be considered for our purpose by utilizing the orthogonality conditions suggested by Grammig and Wellner (2002) and by

Table 1: Summary statistics for Nordea, SEB, FSB and SHB durations.

	Mean	Std	Nr obs	LB_{100}	Min	Max	Daily turnover million SEK
NDA	35.1	70.9	73723	29614	0	1415	586.9
SEB	45.7	92.9	56590	11763	0	1767	326.4
FSB	58.9	115.6	44083	7106	0	2270	298.9
SHB	50.6	99.5	51094	15454	0	1606	343.4

Note: LB_{100} is the Ljung-Box autocorrelation statistic for 100 lags.

potentially adding a parametrization for the covariation between duration sequences.

3 Data and descriptive statistics

The transaction data were downloaded from Ecovision, a provider of real time financial information from the Stockholm Stock Exchange. Four stocks, Nordea (NDA), SEB, Föreningssparbanken (FSB) and Handelsbanken (SHB), were recorded for 98 trading days (May 3 - September 30, 2005). The companies are active in the same line of business, the banking sector. Every transaction is recorded on a second scale with associated volume, bid, ask, and price variables. Descriptive statistics of durations of the four stocks are presented in Table 1.

The stocks are four of the most traded stocks on the Stockholm Stock Exchange and roughly of the same size. The number of observations indicates that the most frequently traded stock is NDA followed by SEB, SHB and FSB. Also the daily turnover indicates that NDA is the most traded stocks followed by the other three stocks. The tick size, i.e. the minimum amount a price can move, is for FSB, SEB and SHB 0.5 SEK and for NDA 0.25 SEK. The durations show strong serial correlation. The Ljung-Box statistics presented in Table 1 are high for all four stocks. Figure 2 shows the autocorrelation functions for the series. The autocorrelation functions share the same pattern, quite small and the decay is also quite slow.

Transaction data often show strong intraday seasonality patterns (e.g., Bauwens, Galli and Giot, 2002; Engle and Russell, 1998). Figure 2 shows the intraday seasonality pattern for the duration of the current sample. Durations tend to be shorter when the market opens and near the closing time for all stocks. Similar patterns are also present in variables such as volume, spread and trade intensity. To account for the potential seasonality of the data the adjusted duration (Engle and Russell, 1998) is computed as $d_i = D_i/\phi_i$, where D_i is the duration from the dataset and ϕ_i is a cubic spline with nodes on each hour. The adjusted duration d_i is used for estimation. In a similar way the other variables, volume, spread, trade intensity and the added durations are adjusted to account for the diurnal effect.

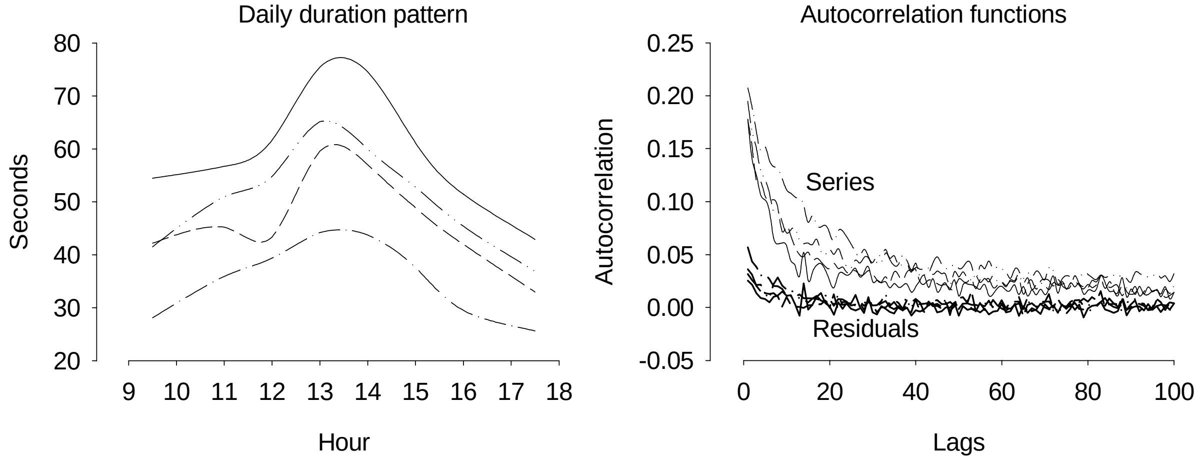

Figure 2: Average duration lengths per hour smoothed with a cubic spline (left) and autocorrelation functions of durations and residuals (right). In the figure are FSB (solid line), SHB (dashed dot dot line), SEB (dashed line) and NDA (dashed dot line).

There are many consecutive zero durations, i.e. multiple transactions within a second, in the data. Suggestions in the literature of the treatment of zero durations are, e.g., aggregate all data within a second (cf. Engle and Russell, 1998) or assume that the distance between transactions within a second are equally spaced (cf. Darolles and Gourièroux, 2000). In this paper the zero durations are replaced with 0.5.

In the empirical part volume and spread changes are used as explanatory variables together with a measure of trade intensity. A variable capturing the intensity of trade (Engle and Russell, 1998) can be constructed by dividing the number of transactions within a price duration with the length of a price duration. Bauwens and Giot (2000) use a similar approach using instead the spread duration. A price duration is calculated as the time it takes for the price to move a predetermined number of ticks. Obviously a new price duration is initiated more often with a low predetermined tick size. The number of ticks is here chosen as 1. Other studies, e.g., Engle and Russell (1998) suggests the number of ticks as 4. The price used for the price duration is defined as the midprice, i.e. the price in the middle of bid and ask at time i, when a transaction occur.

4 Results

The results to be reported are based on the log-ACD model. A main reason for using the log-ACD in favour of the ACD model is that numerical problems were faced when a pure ACD model was implemented. To find a lag structure we employ the AIC criterion. The minimized

AIC gives us the model specification.

Tables 2-5 give the CLS and QML estimation results for the stocks with differences in volume, spread, trade intensity and durations as explanatory variables. The parameter estimates for the explanatory variable volume change are significant for NDA but not for SHB, SEB and FSB. The sign of the volume change variable are throughout positive. For all series the spread change has a significant and negative effect. Trade intensity changes have significant effects for FSB, SEB with added durations from NDA but not for the models of SHB and NDA. The sign is positive except for SHB. The interpretation of a negative parameter sign is that a positive change in the variable shortens the next duration. Brännäs and Simonsen (2006) who also utilize change variables when studying Ericsson B, for a different sample, also find a negative effect of the spread change and a positive one for the volume change.

The estimated parameters α and β in Tables 2-5 are all significant with a few exceptions. Also exclusion of the insignificant parameters were considered. However, such specification is not minimizing the AIC. The parameter β is close to one for all stocks which is also what is found in other studies of transaction data (e.g., Brännäs and Simonsen, 2006). The interpretation of a β parameter estimate close to one is that a short (long) duration tend to be followed by a short (long) duration.

Next, considering the added dependence terms in (4) and (5) we find significant parameter estimates for SEB, SHB and NDA, though not all, while added dependence terms to FSB are all insignificant. More specifically we find significant estimates of the first and third lags of added durations from SEB, FSB and SHB to NDA and also the eighth and ninth lag of durations from SEB (Table 2). In the estimates of added durations to SEB we find significant parameter estimates except for the first and sixth lag of added durations from NDA and the third lag of durations from SHB (Table 4). Added durations to SHB are significant for the first and fifth lags of FSB and fifth lag of NDA, while added durations from SEB are not (Table 5).

Regarding the parameter sign, the expected sign of lagged added dependence terms is positive. Positive signs of the estimates implies that long durations in the added duration series has a positive impact on the conditional mean function i.e., prolongs the next duration. Positive parameter signs is also found for a major part of the estimates. The expected sign of the δ_0 parameter is negative as this implies that the impact of the other stocks is decreasing with the size of τ. However, we find the sign of δ_0 positive except for SEB with durations from NDA, although insignificant (Table 4).

To sum up, we find significant and the expected positive parameter estimates, though not for all lags, of the added durations to NDA and SEB. Also added durations to SHB are significant except durations from SEB. For the added durations to FSB all estimates are insignificant.

Table 2: Estimates of Nordea (NDA) with explanatory variables and durations from SEB A, Föreningssparbanken (FSB) and Handelsbanken (SHB). (Robust t statistics in parenthesis).

Variable		NDA			
	NDA	SEB A	FSB	SHB	CLS
d_{t-1}	0.158	0.157	0.157	0.157	0.041
	(37.67)	(41.66)	(40.86)	(43.62)	(7.61)
d_{t-2}	-0.023	-0.021	-0.021	-0.022	0.0003
	(-4.27)	(-4.31)	(-4.02)	(-4.70)	(0.04)
d_{t-3}	-0.014	-0.012	-0.013	-0.013	-0.004
	(-2.69)	(-2.43)	(-2.48)	(-2.60)	(-0.77)
d_{t-4}	-0.010	-0.009	-0.007	-0.010	-0.001
	(-2.29)	(-2.17)	(-1.77)	(-2.50)	(-0.27)
θ_{t-1}	0.864	0.856	0.856	0.861	0.91
	(130.78)	(121.8)	(131.3)	(142.2)	(96.2)
Volume change	0.097	0.090	0.103	0.097	0.016
	(2.49)	(3.12)	(2.63)	(2.37)	(0.21)
Trade intensity change	0.007	0.007	0.006	0.007	0.003
	(1.36)	(1.61)	(1.88)	(1.51)	(1.44)
Spread change	-0.326	-0.329	-0.316	-0.328	-0.588
	(-5.71)	(-5.83)	(-22.97)	(-5.73)	(-14.0)
δ_0		0.002	0.007	0.003	
		(1.18)	(4.62)	(1.81)	
δ_1		0.002	0.005	0.002	
		(2.02)	(3.88)	(2.08)	
δ_2		0.001	-0.003	0.002	
		(1.32)	(-1.91)	(1.87)	
δ_3		0.003	0.005	0.005	
		(6.07)	(4.29)	(4.22)	
δ_4		0.0001			
		(0.14)			
δ_5		0.002			
		(1.79)			
δ_6		0.0009			
		(1.28)			
δ_7		0.004			
		(7.57)			
δ_8		0.002			
		(2.30)			
δ_9		0.001			
		(1.93)			
Constant	-0.140	-1.163	-1.16	-1.153	-0.039
	(-21.60)	(-148.6)	(-174.0)	(-185.9)	(-11.8)
LB_{100}	919.2	941.7	947.1	924.2	5499.1
LB^2_{100}	114.3	111.2	117.5	110.7	154.9
$\ln l$	-58717.8	-58492.0	-58538.8	-58599.9	

Note: LB_{100} is the Ljung-Box statistic of the residuals over 100 lags. LB^2_{100} is the same statistic for squared residuals. Residuals are calculated as $(d_{N(t)} - \widehat{\theta}_{N(t)})/\widehat{\theta}_{N(t)}$

Table 3: Estimates of Föreningssparbanken (FSB) with explanatory variables and durations from SEB A, Nordea (NDA) and Handelsbanken (SHB). (Robust t statistics in parenthesis).

Variable		FSB			
	FSB	SEB A	NDA	SHB	CLS
d_{t-1}	0.140	0.140	0.140	0.140	0.053
	(25.46)	(30.13)	(26.50)	(31.84)	(16.92)
d_{t-2}	-0.046	-0.032	-0.046	-0.046	-0.015
	(-7.23)	(-2.56)	(-6.77)	(-7.89)	(-2.73)
d_{t-3}	-0.025	-0.015	-0.025	-0.025	0.001
	(-6.28)	(-1.56)	(-3.98)	(-4.09)	(0.10)
d_{t-4}	-0.023	-0.017	-0.023	-0.023	-0.011
	(-3.69)	(-2.03)	(-3.31)	(-3.40)	(-1.87)
d_{t-5}	0.003	0.007	0.003	0.003	0.001
	(0.26)	(0.94)	(0.45)	(0.33)	(0.21)
d_{t-6}	-0.013	-0.011	-0.013	-0.013	-0.006
	(-2.00)	(-1.67)	(-2.11)	(-2.16)	(-1.05)
d_{t-7}	-0.026	-0.008	-0.025	-0.025	-0.019
	(-5.58)	(-0.90)	(-4.78)	(-6.25)	(-4.71)
θ_{t-1}	0.988	0.888	0.985	0.985	0.990
	(297.5)	(11.3)	(138.2)	(151.5)	(363.9)
Volume change	0.001	0.047	0.001	0.003	0.011
	(0.05)	(1.03)	(0.09)	(0.18)	(0.64)
Trade intensity change	0.014	0.047	0.016	0.017	0.013
	(2.72)	(2.69)	(2.02)	(2.47)	(2.86)
Spread change	-0.133	-0.161	-0.134	-0.137	-0.375
	(-2.42)	(-2.79)	(-3.00)	(-2.30)	(-5.89)
δ_0		0.007	0.0009	0.000	
		(1.29)	(1.00)	(0.01)	
δ_1		0.004	-0.00009	0.0006	
		(1.07)	(-0.14)	(1.63)	
δ_2				0.0005	
				(1.18)	
Constant	-0.011	-1.085	-1.014	-1.014	-0.005
	(-4.15)	(-19.74)	(-166.0)	(-186.18)	(-3.67)
LB_{100}	199.0	194.7	194.5	197.3	976.1
LB^2_{100}	122.1	117.8	103.1	106.0	135.5
$\ln l$	-39504.8	-39469.7	-39492.5	-39481.2	

Note: LB_{100} is the Ljung-Box statistic of the residuals over 100 lags. LB^2_{100} is the same statistic for squared residuals. Residuals are calculated as $(d_{N(t)} - \widehat{\theta}_{N(t)})/\widehat{\theta}_{N(t)}$

Table 4: Estimates of SEB A with explanatory variables and durations from Föreningsspar-banken (FSB), Nordea (NDA) and Handelsbanken (SHB). (Robust t statistics in parenthesis).

Variable		SEB			
	SEB	NDA	FSB	SHB	CLS
d_{t-1}	0.140	0.138	0.138	0.139	0.0459
	(33.99)	(36.63)	(38.49)	(37.57)	(11.07)
d_{t-2}	-0.017	-0.015	-0.015	-0.014	0.0011
	(-3.05)	(-2.58)	(-2.96)	(-2.89)	(0.21)
d_{t-3}	-0.023	-0.018	-0.019	-0.020	-0.0150
	(-4.31)	(-3.72)	(-3.75)	(-3.83)	(-3.25)
θ_{t-1}	0.845	0.825	0.832	0.829	0.9160
	(81.84)	(73.97)	(80.81)	(79.66)	(74.79)
Volume change	0.012	0.017	0.007	0.011	-0.092
	(0.20)	(0.31)	(0.13)	(0.20)	(-0.98)
Trade intensity change	0.013	0.014	0.013	0.012	0.003
	(2.02)	(2.24)	(1.60)	(1.56)	(0.47)
Spread change	-0.206	-0.201	-0.208	-0.195	-0.539
	(-3.73)	(-5.76)	(-3.81)	(-3.59)	(-6.31)
δ_0		-0.002	0.013	0.010	
		(-0.90)	(6.71)	(5.31)	
δ_1		0.002	0.003	-0.005	
		(1.36)	(2.28)	(-2.67)	
δ_2		0.007		0.004	
		(4.95)		(2.47)	
δ_3		0.004		0.002	
		(2.07)		(1.56)	
δ_4		0.007		0.004	
		(3.75)		(2.66)	
δ_5		0.003			
		(2.36)			
δ_6		0.003			
		(1.90)			
δ_7		0.005			
		(3.52)			
Constant	-0.120	-1.156	-1.143	-1.143	-0.034
	(-15.57)	(-113.8)	(-139.6)	(-140.7)	(-8.98)
LB_{100}	326.7	323.1	327.7	323.3	1905.4
LB^2_{100}	130.1	137.6	128.9	124.4	189.1
$\ln l$	-49242.9	-48960.0	-49097.7	-49096.4	

Note: LB_{100} is the Ljung-Box statistic of the residuals over 100 lags. LB^2_{100} is the same statistic for squared residuals. Residuals are calculated as $(d_{N(t)} - \widehat{\theta}_{N(t)})/\widehat{\theta}_{N(t)}$.

Table 5: Estimates of Handelsbanken (SHB) with explanatory variables and durations from Föreningssparbanken (FSB), Nordea (NDA) and SEB A. (Robust t statistics in parenthesis).

Variable		SHB			
	SHB	NDA	SEB	FSB	CLS
d_{t-1}	0.143	0.143	0.143	0.1430	0.052
	(37.48)	(29.46)	(35.08)	(31.66)	(20.33)
d_{t-2}	-0.041	-0.041	-0.040	-0.040	-0.013
	(-10.30)	(-5.99)	(-6.50)	(-6.39)	(-3.08)
d_{t-3}	-0.028	-0.028	-0.027	-0.028	0.001
	(-6.08)	(-3.96)	(-8.37)	(-4.30)	(0.26)
d_{t-4}	-0.014	-0.014	-0.013	-0.014	-0.012
	(-2.70)	(-3.13)	(-2.24)	(-2.07)	(-4.12)
d_{t-5}	-0.007	-0.007	-0.008	-0.007	-0.001
	(-1.30)	(-1.32)	(-1.29)	(-1.16)	(-0.22)
d_{t-6}	-0.019	-0.020	-0.019	-0.019	-0.005
	(-3.73)	(-2.74)	(-3.45)	(-2.96)	(-0.93)
d_{t-7}	-0.023	-0.022	-0.020	-0.021	-0.016
	(-5.39)	(-4.44)	(-4.60)	(-4.39)	(-4.04)
θ_{t-1}	0.988	0.985	0.982	0.986	0.990
	(392.0)	(253.7)	(190.7)	(217.1)	(300.0)
Volume change	0.013	0.011	0.011	0.016	-0.016
	(0.68)	(0.54)	(0.48)	(0.78)	(-0.87)
Trade intensity change	-0.004	-0.004	-0.002	-0.004	-0.011
	(-0.89)	(-0.73)	(-0.33)	(-0.69)	(-1.62)
Spread change	-0.356	-0.362	-0.368	-0.365	-0.603
	(-4.62)	(-5.37)	(-5.12)	(-5.11)	(-8.79)
δ_0		0.0003	0.0012	0.0011	
		(0.63)	(1.80)	(1.96)	
δ_1		0.0005	0.0007	0.0001	
		(0.90)	(1.49)	(0.26)	
δ_2		-0.0001	0.0006	0.0008	
		(-0.15)	(1.50)	(2.21)	
δ_3		0.0002	0.0002	0.0003	
		(0.56)	(0.48)	(0.89)	
δ_4		-0.0001	0.0006	-0.0001	
		(-0.28)	(1.46)	(-0.30)	
δ_5		0.0011		-0.0008	
		(2.93)		(-2.63)	
δ_6		0.0006			
		(1.63)			
Constant	-0.0130	-1.0176	-1.0205	-1.0163	-0.006
	(-5.35)	(-248.9)	(-188.0)	(-219.2)	(-3.48)
LB_{100}	388.5	384.3	369.2	389.0	1844.8
LB^2_{100}	186.7	194.5	126.1	231.9	232.7
$\ln l$	-43826.75	-43716.1	-43746.6	-43760.6	

Note: LB_{100} is the Ljung-Box statistic of the residuals over 100 lags. LB^2_{100} is the same statistic for squared residuals. Residuals are calculated as $(d_{N(t)} - \widehat{\theta}_{N(t)})/\widehat{\theta}_{N(t)}$.

When considering the CLS estimator the parameter signs and sizes are roughly the same, though with a few exceptions. Some of the parameter signs change when using the CLS estimator, e.g., the second lag of lagged durations in NDA. The estimated parameters of volume change are negative for SEB and SHB.

One may use a Granger causality test (Granger, 1969) to examine if the inclusion of a duration from other stocks to the conditional mean in (4) and (5) adds explanatory power or not. Considering individual parameter estimates we find significant parameters of adding durations from stocks beyond the focused one to NDA and SEB but not for FSB. Individual parameters of added durations to SHB are significant for NDA and FSB but not for SEB. Consequently it is found that NDA granger causes SEB and SEB granger causes NDA. FSB Granger causes NDA, SEB and SHB, while NDA, SEB SHB is not Granger causing FSB. FSB and NDA Granger causes SHB while SEB is not. Also a likelihood ratio test is applied to test if the individual parameters of the added durations are jointly, significantly different from zero (Tables 2-5). The test statistic has in large sample a χ^2 distribution with the degrees of freedom equal to the number of additional parameters i.e., the number of added lagged durations including δ_0. The 5 percent critical value range from 5.99 with 2 degrees of freedom to 18.31 with 10 degrees of freedom. It is found that there is Granger causality between all the stocks, including FSB and SHB with added durations from SEB.[1] Similar results have also been found for other stocks and markets, e.g., Spierdijk, et al. (2002) and Bauwens and Hautsch (2004).[2]

In the models reported in Tables 2-5 there is serial correlation left that we were unable to reduce further. To test for remaining serial correlation the Ljung-Box statistic is used. The test statistic has asymptotically a χ^2_{100} distribution and a 5 percent critical value of 124.3. For the residuals of NDA, SEB, FSB and SHB the null of no serial correlation is rejected for all models. For the squared residuals the null hypothesis is rejected for all models of SHB and NDA. Note that the reported results are based on deseasonalized data.

The estimated models may be used to evaluate and illustrate the response to a chock in, e.g., the added duration. Figure 3 gives the response in the expected mean duration $\theta_{N(t)}$ to a 50 percent reduction of 10 consecutive durations for SHB and NDA illustrated. The shocks have a small and diminishing effect on $\theta_{N(t)}$. Hence, the estimated models are stationary. The response in percent in NDA of a shortened duration in SHB is roughly twice the size of the response in SHB of a shortened duration is NDA.

[1] The result is also valid for a test statistic with 1 percent critical values.

[2] In an early draft of the paper, from another sample, we found that AstraZeneca is not Granger causing Ericsson B, while Ericsson B Granger causes AstraZeneca

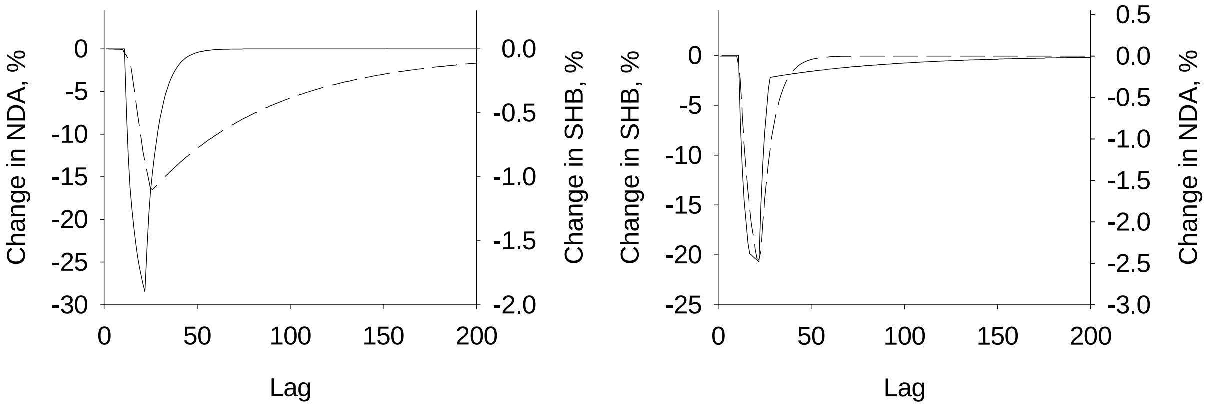

Figure 3: The change in percent of a 50 percent shortened duration in 10 consequtive durations from the mean duration at time $t = 10$ in Nordea (NDA) (left) and Handelsbanken (SHB) (right). The solid line in the left figure shows the response in NDA of a shortened duration in NDA and in the right figure the response in SHB of a shortened duration in SHB. The dashed lines show the responses of a shorthened duration in the other stock.

5 Conclusions

The paper proposed adding durations from a second stock to the conditional mean function in the ACD model of Engle and Russell (1998). By including durations from a second stock dependence between duration series is captured in the model. The model is applied to four stocks, Nordea, Föreningssparbanken, SEB and Handelsbanken, all from the banking sector, traded on the Stockholm Stock Exchange.

In the empirical part, we find Granger causality between all the stocks, Nordea, Föreningssparbanken, SEB and Handelsbanken, when the added durations from stocks beyond the focused one is considered jointly. For the individual parameter estimates we find Granger causality between Nordea and SEB. Föreningssparbanken Granger causes Nordea, SEB and Handelsbanken, while Nordea, SEB and Handelsbanken is not Granger causing Föreningssparbanken. Handelsbanken Granger causes Nordea and SEB. Nordea and Föreningssparbanken are Granger causing Handelsbanken while SEB is not. This result indicates that there may be duration dependence between stocks active in the banking sector on the Stockholm Stock Exchange. The dependence may be caused by new information revealed by the trade intensity. In view of this empirical result the suggested model extension is able of capturing dependence between duration series and of providing an improvement in the econometric specification of the model. The attraction of the extension of the model is the simplicity by which information from other duration series can be added to the ACD model.

The parameter estimates of spread change is significant for all stocks and volume change is significant for Nordea. The expected negative parameter estimates are found for spread change and trade intensity change for Handelsbanken. The descriptive statistics of duration and other variables show presence of time of day seasonality in the stocks from the Stockholm Stock Exchange, cf. the results for other stock markets.

References

Bauwens, L., F. Galli, and Giot, P. (2002). The moments of log-ACD models. unpublished manuscript, CORE, Louvain-La-Neuve.

Bauwens, L., Giot, P. (2000). The logarithmic ACD Model: An application to the Bid/Ask quote process of two NYSE stock. *Annales d'Economie et de Statistique,* 60, 117-149.

Bauwens, L., Giot, P. (2001). *Econometric Modelling of Stock Market Intraday Activity.* Kluwer, Dordrecht.

Bauwens, L., Giot, P., Grammig, J., Veredas, D., (2004). A comparison of financial duration models via density forecast. *International Journal of Forecasting*, 20, 589-609.

Bauwens, L., Hautsch, N. (2004) Dynamic Latent Factor Models for Intensity Processes. Core Discussion Paper 2003/103.

Brännäs, K., Simonsen, O. (2006). Discretized time and conditional duration modelling for stock transaction data. *To appear in Applied Financial Economics.*

Darolles, S., Gouriéroux, C. (2000). Intraday Transaction Price Dynamics. *Annales d'Economie et de Statistique,* 60, 207-237.

Easley, D., O'Hara, M. (1992), Time and the process of security price adjustment. *Journal of Finance,* 47, 577-606.

Engle, R.F., Lunde. (2003). Trades and Quotes: A Bivariate Point Process. *Journal of Financial Econometrics,* 2, 159-188.

Engle, R.F., Russell, J.R. (1998). Autoregressive Conditional duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica,* 66, 1127-1162.

Grammig, J., Wellner, M. (2002). Modeling the interdependence of volatility and inter-transaction duration processes. *Journal of Econometrics,* 106, 369-400.

Glosten, L., Milgrom, P. (1985). Bid, Ask and Transactions Prices in a Specialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics,* 14, 71-100.

Granger, C. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica,* 37, 424-438.

Hansen, L.P., (1982). Large sample properties of generalized method of moments estimators. *Econometrica,* 50, 1029-1054.

Mosconi, R., Olivetti, F. (2005). Bivariate Generalizations of the ACD Models. Unpublished manuscript, Politecnico di Milano.

O'Hara, M. (1994). *Market Microstructure Theory.* Blackwell Publishers Ltd.

Spierdijk, L., Nijman, T.E. and van Soest, A.H.O. (2002). Modeling Comovements in Trading Intensities to Distinguish Sector and Stock Specific News. Unpublished manuscript, Tilburg University.

Russell, J.R. (1999). Econometric Modelling of Multivariate Irregulary-Spaced High-Frequency Data. Unpublished manuscript, University of Chicago.

Does the Open Limit Order Book Reveal Information About Short-run Stock Price Movements?*

Jörgen Hellström and Ola Simonsen

Department of Economics, Umeå University, SE-90187 Umeå

jorgen.hellstrom@econ.umu.se, tel: +46-90-7866143, fax: +46-90-772302

ola.simonsen@econ.umu.se, tel: +46-90-7866085, fax: +46-90-772302

Abstract

This paper empirically tests whether an open limit order book contains information about future short-run stock price movements. To account for the discrete nature of price changes, the integer-valued autoregressive model of order one is utilized. A model transformation has an advantage over conventional count data approaches since it handles negative integer-valued price changes. The empirical results reveal that measures capturing offered quantities of a share at the best bid- and ask-price reveal more information about future short-run price movements than measures capturing the quantities offered at prices below and above. Imbalance and changes in offered quantities at prices below and above the best bid- and ask-price do, however, have a small and significant effect on future price changes. The results also indicate that the value of order book information is short-term.

Keywords: Negative integer-valued data, time series, INAR, finance, stock price, open limit order book.

JEL: C25, G12, G14.

*The authors would like to thank Kurt Brännäs, Carl Lönnbark, Tomas Sjögren and participants in a seminar at Umeå University for valuable comments and suggestions. Financial support from the Wallander Foundation is gratefully acknowledged.

1 Introduction

The purpose of this paper is to empirically study the information contained in the open limit order book about future short-run stock price movements. Specifically, attention is paid to whether changes or asymmetries in the order book concerning offered quantities of a share at prices below the best bid price (low end of the order book, see Figure 1 below) and above the best ask price (high end of the order book) are informative. To assess the information contained in the order book the paper presents new measures as well as extensions to existing measures summarizing order book movements. The integer-valued autoregressive model (e.g., McKenzie, 1985, 1986, Al-Osh and Alzaid, 1987) is utilized to adhere to the discrete nature of high frequency stock price data. Although the model should be seen as an approximation to the underlying price process it offers interpretability concerning parameter estimates in contrast to conventional time series models.

In the literature models of limit order books (e.g., Glosten, 1994, Rock, 1996, Seppi, 1997) build on the assumption that informed traders always use market orders (immediate execution at best bid- or ask-price in the order book) instead of using limit orders. Accordingly there should be limited information of observing offered quantities of a share at other prices than the best bid- or ask-price. At the same time there has been a significant growth of electronic limit order book trading systems, offering additional transparency compared to dealers' markets, around the world. For example, on January 24, 2002, the New York Stock Exchange (NYSE) began to publish aggregated depths (quantities offered) at all price levels on both the bid- and the ask-side of the order book under what is known as the NYSE Open Book program. One may speculate whether this additional information, regarding the quantities offered at prices in the low, respectively, the high end of the order book, contain information concerning short-run price movements. For example, if traders use the order book as a proxy for market demand (bid side) and supply (ask side) and base their short-run trading decisions upon this information an imbalance in offered quantities between the bid- and the ask-side of the order book may contain information concerning future short-run price movements, even if traders are uninformed. Also, uninformed traders (e.g., index fund managers) may view the order book to determine price impact costs (e.g., Keim and Madhavan, 1998). Thus, changes or asymmetries in the quantities offered in the low and in the high end of the order book may reveal information concerning future short-run price movements.

In spite of the success of open limit order book trading systems around the world little research has been done to assess the impact of the information contained in the open limit order book (Jain, 2002). Cao et al. (2004) introduce summary measures of the open limit order book and study their impact on short-run returns on the Australian stock market. Among other things, they find that the high and low end of the order book contain information about future price movements. Recent experimental results of Bloomfield et al. (2005) also suggest that informed traders may use limit orders for trading. Harris and Panchapagesan (2005) provide empirical evidence that the limit order book on the NYSE dealers' market is informative about future price movements.

A prominent feature of transaction stock price data is the discreteness of the prices process. A majority of stock exchanges allow prices to only be multiples of a smallest divisor, called a "tick". The basic idea of fixing a minimum price change is to obtain a reasonable trade-off between the provision of an efficient grid for price formation and the possibility to realize price levels that are close to the traders' valuation. To handle this feature a number of approaches to modeling the discrete price movements have been suggested. Hausman et al. (1992) propose the ordered probit model with conditional heteroskedasticity. The inclusion of conditioning information in the ordered response model is straightforward and is a substantial advantage compared to the rounding and barrier models suggested by Ball (1988), Cho and Frees (1988), and Harris (1998). Two shortcomings of the ordered response model are that parameters are only identifiable up to a factor of proportionality without further restrictions and that price resolution is lost, i.e. price changes larger than the largest specified discrete value are grouped together.

Contrary to the ordered response model no threshold parameters have to be estimated when adopting a count data approach. Hautsch and Pohlmeier (2002) utilize

Poisson and zero-inflated Poisson models to analyze absolute price changes. A shortcoming with a count data approach is that conventional count data models are not able to explain negative discrete price changes. An exception is the recent dynamic integer valued count data modelling approach presented by Liesenfeld et al. (2006).

To account for the discrete nature of stock transaction price data, this paper utilizes a count data time series approach. The approach builds on the integer-valued autoregressive (INAR) class of models (e.g., McKenzie, 1985, 1986, Al-Osh and Alzaid, 1987). Assuming that the stock price (measured in number of ticks) is described by an INAR(1) process and expressing the price change in a differenced form, in such a way that negative price changes are removed from the right hand side of the model, a model allowing for negative discrete price changes, i.e. a negative conditional mean, is obtained.[1] This differenced INAR(1) model is consistent with the underlying assumptions of the conventional INAR(1) model. The differenced INAR(1) model explains a price change with two parts, the mean upward movement in price and the mean downward movement in price. These mean movements may be separately parameterized (Brännäs, 1995) to allow for conditioning information, e.g., a summary measure of the information contained in the open limit order book and lagged price changes. The main advantage of this approach to accounting for discrete price movements, compared with, e.g., the ordered probit model, is that identification of parameters of interest and extensions to multivariate settings are simplified. The model also allows for asymmetric effects which are common in financial time series.

Section 2 presents the econometric model as well as the estimation framework. Section 3 contains a discussion of how to summarize the information displayed in the open limit order book. Section 4 describes the data. Section 5 contains the empirical results, while some final remarks are left for the concluding section.

[1] The current approach is obviously also suitable when interest lies in analysis of net-changes in a count data variable, i.e., count data with negative observations.

2 Econometric model

The objective is to model discrete stock price movements, i.e. changes measured by the number of ticks. Contrary to conventional count data observations these stock price changes produce negative integer-valued count observations. To avoid having to restrict the analysis to absolute price changes (Hautsch and Pohlmeier, 2002) the analysis is built on a differencing of the INAR(1) model. The differenced model allows for a negative conditional mean without violating any of the basic assumptions of the INAR(1) model.

2.1 The differenced INAR(1) model

Denote the stock price at t with $p_t \geq 0$ and the tick size with s. The integer-valued stock price at time t is given by $P_t = p_t/s$ (measured in number of ticks). Since a price change in number of ticks may be a negative integer conventional count data models are not possible to use. To adhere to the discrete nature of the data and facilitate the use of a count data time series modelling framework, consider a slight rearrangement of the INAR(1) model. Consider approximate the price process with an INAR(1) model[2]

$$P_t = \alpha \circ P_{t-1} + \varepsilon_t$$

where $\{\varepsilon_t\}$ is a sequence of integer-valued random variables, and ε_t is independent of P_{t-1}, with $E(\varepsilon_t) = \lambda$, $V(\varepsilon_t) = \sigma^2$ and $Cov(\varepsilon_t, \varepsilon_s) = 0$, for all $t \neq s$. The $\{\varepsilon_t\}$ sequence may be seen as increases in price, in terms of number of ticks, to the series, with λ as the mean price increase. The binomial thinning operator, defined as $\alpha \circ P_{t-1} = \sum_{i=1}^{P_{t-1}} u_i$, where u_i is an independent binary variable with survival probability $\Pr(u_i = 1) = 1 - \Pr(u_i = 0) = \alpha$, $\alpha \in [0, 1]$, represents the price, in number of ticks, at the end of the interval $t-1$ to t. Among the properties of the basic INAR(1) model (e.g., Brännäs and Hellström, 2001) the first and second order conditional and unconditional

[2] Higher order INARMA specifications may also be utilized as starting points. The advantage, however, with an INAR(1) model is that it renders a parsimonious interpretation of the price process, i.e. one parameter describing the upward movement and one parameter describing the downward movement of the price change.

moments are given by $E(P_t) = \lambda/(1-\alpha)$, $V(P_t) = \left[\alpha(1-\alpha)E(P_{t-1}) + \sigma^2\right]/(1-\alpha^2)$, $E(P_t|P_{t-1}) = \alpha P_{t-1} + \lambda$ and $V(P_t|P_{t-1}) = \alpha(1-\alpha)P_{t-1} + \sigma^2$.

A differenced form is obtained by subtracting P_{t-1} from both sides:

$$\begin{aligned} P_t - P_{t-1} &= \varepsilon_t - (P_{t-1} - \alpha \circ P_{t-1}) \\ \triangle P_t &= \underbrace{\varepsilon_t}_{increase} - \underbrace{(P_{t-1} - \alpha \circ P_{t-1})}_{decrease}. \end{aligned} \tag{1}$$

The first part of the model represents an increase in the price, measured in number of ticks, while the second part represents a decrease in the price, measured in number of ticks. Note that the conventional rule for multiplication, i.e. $1 \circ P_{t-1} - \alpha \circ P_{t-1} \neq (1-\alpha) \circ P_{t-1}$ do not hold for the binomial thinning operator. The first part of the parenthesis in (1) is the stock price measured as the number of ticks at $t-1$. The second part in the parenthesis represents the number of ticks remaining at the end of the period $(t-1, t)$. Thus, the difference between the two parts in the parenthesis represents the reduction in the number of ticks. The advantage of stating the price difference in this form is that the thinning operator does not contain the possibly negative $\triangle P_t$.[3] The first and second conditional moments are given by

$$E(\triangle P_t|P_{t-1}) = \lambda - (1-\alpha)P_{t-1} = \lambda - \theta_d P_{t-1}$$

$$V(\triangle P_t|P_{t-1}) = \sigma^2 + \theta_d(1-\theta_d)P_{t-1} \tag{2}$$

Note that the conditional mean is allowed to be negative. As long as the restrictions upon parameters are satisfied the count data features of the basic INAR(1) model are satisfied.

The model specification may be conditioned on explanatory variables, following Brännäs (1995). The parametrization of the mean increase in the number of ticks, λ, and the probability of a decrease in ticks, θ_d, may be accomplished by use of an

[3] The binomial thinning operator is only defined for positive values of a variable, i.e. $\alpha \circ P_t$ is only valid for $P_t \geq 0$.

exponential functional form, $\lambda_t = \exp(\mathbf{x}_t\boldsymbol{\beta}_1)$, and a logistic functional form, i.e. $\theta_{dt} = 1/[1+\exp(\mathbf{x}_t\boldsymbol{\beta}_2)]$. An extension in order to get a more flexible conditional variance is to let the variance σ^2 become time dependent. The variance σ_t^2 may, e.g., be dependent on past values of σ_t^2 and other explanatory variables parameterized the following way (cf. Nelson, 1991)

$$\sigma_t^2 = \exp\left[\phi_0 + \phi_1 \ln \sigma_t^2 + ... + \phi_P \ln \sigma_{t-P}^2 + \mathbf{x}_t'\boldsymbol{\gamma}\right]. \tag{3}$$

When analyzing the effect of order book measures upon future price movements the total (net) effect on the expected price change and conditional variance are of interest. For any explanatory variable the average net effect over all observations on the expected price change by a marginal change in the explanatory variable is given by

$$\begin{aligned} m_{i,t}^E &= T^{-1}\sum_{t=1}^{T}\frac{\partial E(\triangle P_t|P_{t-1})}{\partial x_{it}} = T^{-1}\sum_{t=1}^{T}\left(\frac{\partial\lambda_t}{\partial x_{it}} - \frac{\partial\theta_{td}}{\partial x_{it}}P_{t-1}\right) \\ &= T^{-1}\sum_{t=1}^{T}\left(\beta_i\exp(x_t\beta) + \frac{\beta_i\exp(x_t\beta)}{[1+\exp(x_t\beta)]^2}P_{t-1}\right), \end{aligned} \tag{4}$$

while the average net effect on the conditional variance (via θ_{dt} and σ_t^2) is given by

$$\begin{aligned} m_{i,t}^V &= T^{-1}\sum_{t=1}^{T}\frac{\partial V(\triangle P_t|P_{t-1})}{\partial x_{it}} = T^{-1}\sum_{t=1}^{T}\left[\left(\frac{\partial\theta_{dt}}{\partial x_{it}} - \frac{\partial\theta_{dt}^2}{\partial x_{it}}\right)P_{t-1} + \frac{\partial\sigma_t^2}{\partial x_{it}}\right] \\ &= T^{-1}\sum_{t=1}^{T}\left[\left(\frac{\beta_i\exp(x_t\beta) - \beta_i\exp(x_t\beta)^2}{[1+\exp(x_t\beta)]^3}\right)P_{t-1} + \beta_i\exp(x_t\beta)\right]. \end{aligned} \tag{5}$$

The variance of the marginal effects may be determined by the delta method, i.e., the variance is approximated with $V(m_{i,t}) \approx \mathbf{g}'V(\psi)\mathbf{g}$ where $\psi' = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ and the covariance matrix $V(\psi)$ and $g = \partial m_{i,t}/\partial\psi$ are evaluated at the estimates.

2.2 Estimation

Estimation of the basic INAR(1) model has been studied by, e.g., Al-Osh and Alzaid (1987), Brännäs (1995) and Brännäs and Hellström (2001). Since the conditional first and second order moments are similar for the differenced INAR(1) model estimation may be based on previous results for the INAR(1) model. In the present paper conditional least squares (CLS) and weighted conditional least squares (WCLS) are used to estimate parameters of interest. Weighted or unweighted conditional least squares estimators are simple to use and have been found to perform well for univariate models and short time series (Brännäs, 1995). The conditional mean or the one-step-ahead prediction error can be used to form the estimator. The CLS estimator of θ_{dt} and λ_t minimizes the criterion function

$$Q = \sum_{t=2}^{T} \left[\triangle P_t - \lambda_t + \theta_{dt} P_{t-1}\right]^2 .$$

The σ^2 term is estimated by OLS from the empirical conditional variance expression

$$\hat{\varepsilon}_t^2 = \hat{\theta}_{dt}(1 - \hat{\theta}_{dt}) P_{t-1} + \sigma_t^2 + \eta_t,$$

where $\hat{\varepsilon}_t$ is the residual from the CLS estimation phase and η_t is a disturbance term. The WCLS estimator of the unknown parameters λ_t and θ_{dt} minimize the criterion function

$$Q^W = \sum_{t=2}^{T} \frac{\left[\triangle P_t - \lambda_t + \theta_{dt} P_{t-1}\right]^2 .}{\hat{\theta}_{dt}(1 - \hat{\theta}_{dt}) P_{t-1} + \hat{\sigma}^2},$$

where the conditional variance in the denominator is taken as given.

3 Summarizing the open limit order book

In this section summary measures of the limit order book are presented. The measures summarizing the limit order book capture both the shape (balance/imbalance in offered quantities of a share between the bid and ask side) and activity (changes in offered quantities of a share over time) of the limit order book.

In an order driven market, i.e. with no market makers involved, traders submit their buy and sell orders to a computerized system. A trader may submit two types

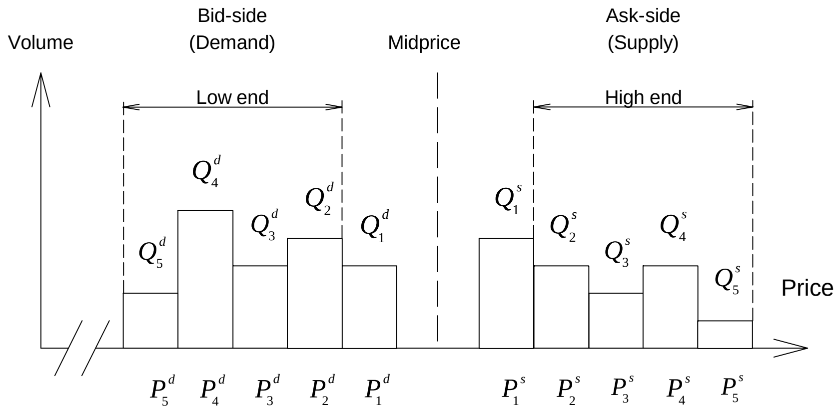

Figure 1: Illustration of the limit order book.

of orders, a market or a limit order. Limit orders are placed in a queue in the order book, i.e. they are not immediately traded, where the price and the time of the order determines the priority of execution. Market orders are executed immediately to the best bid or ask price. A limit order book displays the quantities of a stock that buyers and sellers are offering at different prices. For example, the publicly visible limit order book for a stock listed on the Stockholm stock exchange shows the first five levels on the bid and ask side, respectively. This is illustrated in Figure 1 where P_i^d and P_i^s are the prices on the bid- (demand) and ask- (supply) side of an arbitrary order book for the levels $i = 1, 2...5$. The bid- and ask-volumes contained at level i are denoted Q_i^d and Q_i^s.

The order book can be summarized in different ways, e.g., by capturing the shape or the activity over time of an order book, discriminating activity and shape between different levels and so on. Measures capturing the shape may focus on asymmetry of the order book, i.e. if there are more value on the bid- (ask-) side relative to the ask- (bid-) side or on market depth, i.e. the spread of buy and sell orders. The activity in the order book may be measured with, e.g., the turnover during a predetermined interval of time.

In Cao et al. (2004) the shape of the open limit order book is summarized by the following weighted price measure

$$WP^1 = \frac{Q_1^d P_1^d + Q_1^s P_1^s}{Q_1^d + Q_1^s}$$

for the first level in the order book. The rest of the open limit order book is summarized by

$$WP^{n_1-n_2} = \frac{\sum_{i=n_1}^{n_2}(Q_i^d P_i^d + Q_i^s P_i^s)}{\sum_{i=n_1}^{n_2}(Q_i^d + Q_i^s)},\ n_1 < n_2.$$

To compare different stocks a slightly modified standardized measure can be constructed as

$$\begin{aligned} SWP^1 &= \frac{Q_1^d P_1^d + Q_1^s P_1^s}{Q_1^d + Q_1^s} - \frac{P_1^d + P_1^s}{2} \\ &= -\frac{1}{2}\frac{P_1^d Q_1^s + P_1^s Q_1^d - Q_1^d P_1^d - Q_1^s P_1^s}{Q_1^d + Q_1^s} \end{aligned}$$

and

$$SWP^{n_1-n_2} = \frac{\sum_{i=n_1}^{n_2}(Q_i^d P_i^d + Q_i^s P_i^s)}{\sum_{i=n_1}^{n_2}(Q_i^d + Q_i^s)} - \frac{P_1^d + P_1^s}{2},$$

that are centered (for a symmetric order book) around zero for all stocks. On comparison WP is centered around the bid-ask midpoint. Both WP and SWP are unbalanced towards the buy (ask) side if the values are lower (higher) than the bid-ask midpoint, respectively negative (positive).

In Cao et al. (2004) the change in the order book measure $\triangle SWP$ showed little variation at low aggregation levels and the study was instead performed at the 5 minutes aggregation level. A shortcoming with the above measures, as the aggregation level grows, is that information is lost since the measure does not discriminate between different patterns of change during the aggregated period. The value of the order book change measure may be equal for two different periods even if there is large changes at the beginning of the aggregation period in one case or at the end of the period for the second case. To discriminate between these cases this paper proposes a time adjusted

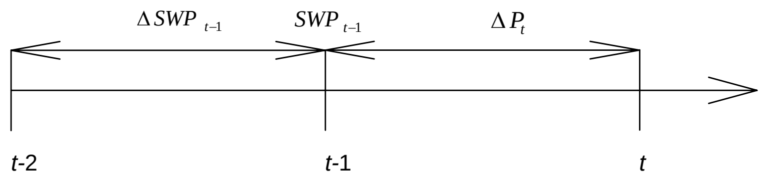

Figure 2: The structure of the order book variables.

standardized weighted price measure

$$\triangle TSWP_t^1 = \sum_{j=2}^{m} g(j) \left(SWP_j^1 - SWP_{j-1}^1 \right), \tag{6}$$

and

$$\triangle TSWP^{n_1 - n_2} = \sum_{j=2}^{m} g(j) \left(SWP_j^{n_1 - n_2} - SWP_{j-1}^{n_1 - n_2} \right), \tag{7}$$

for the rest of the order book. The length of the interval $(t, t-1)$ is divided into m sub-intervals where j is the time of the SWP observation in the j:th sub-interval. To discriminate between observations dependent on where in the interval they are observed the observations are weighted with the function $g(j)$. In the final estimation the weight function $1/\sqrt{j}$ was used since it provided the best fit to the data. Note that the weight function gives more weight to recent changes in the order book in the aggregated interval.

In the empirical part of the paper the above measures are utilized to study whether order book information explains future short-run price movements. To measure order book imbalance (concerning offered quantities), SWP_{t-1} is used. This variable is negative (positive) for skewness towards the buy (ask) side of the order book and is used in order to test whether the shape of the order book influences future short-run price changes. To test whether recent activity in the order book influence future short-run price movements the recent changes in the order book during the previous period is captured by $\triangle SWP_{t-1}$. The alternative weighted measure, $\triangle TSWP_{t-1}$, will also be used for this purpose. The structure of the variables are given in Figure 2.

Two alternative measures to capture order book activity are utilized in the paper. Foucault (1999) argues that an increase in asset volatility increases the proportion of limit order traders and the limit order trader have to post higher ask prices and lower bid prices, i.e. market depth increase. To assess this activity in the order book we suggest the weighted standardized spread measure

$$WSS_t^{n_1-n_2} = \left(\frac{\sum_{i=n_1}^{n_2} Q_i^s P_i^s}{\sum_{i=n_1}^{n_2} Q_i^s} - \frac{\sum_{i=n_1}^{n_2} Q_i^d P_i^d}{\sum_{i=n_1}^{n_2} Q_i^d} \right) / \frac{P_1^d + P_1^s}{2}. \tag{8}$$

To measure the reallocation in the order book during an aggregated interval a total turnover measure is utilized. The idea is that new information may lead to a reallocation in the order book. A high total turnover may then be an indication of new information affecting the price. The measure is calculated as

$$TT_t^1 = \sum_{j=2}^{m} \left| (Q_{j,t}^{d,1} P_{j,t}^{d,1} + Q_{j,t}^{s,1} P_{j,t}^{s,1}) - (Q_{j-1,t}^{d,1} P_{j-1,t}^{d,1} + Q_{j-1,t}^{s,1} P_{j-1,t}^{s,1}) \right| \tag{9}$$

for the first levels of the order book and as

$$TT_t^{n_1-n_2} = \sum_{j=2}^{m} \left| \sum_{i=n_1}^{n_2} (Q_{j,t}^{d,i} P_{j,t}^{d,i} + Q_{j,t}^{s,i} P_{j,t}^{s,i}) - (Q_{j-1,t}^{d,i} P_{j-1,t}^{d,i} + Q_{j-1,t}^{s,i} P_{j-1,t}^{s,i}) \right| \tag{10}$$

for the rest of the order book.

4 Sample data

The data has been downloaded from the Ecovision real time system for financial information and further filtered by the authors. Trading in Nokia[4] from the Stockholm stock exchange was recorded for 50 trading days, September 4-November 13, 2003. The Stockholm Stock Exchange opens at 9.30 am and closes at 5.30 pm. The first and final 15 minutes of the trading day, are deleted from the data. The reason for this is that

[4]Nokia is active in the business of information technology and is one of the most traded stock on the Stockholm stock exchange.

Table 1: Summary statistics of price tick changes and explanatory variables for the Nokia share.

	Mean	Std	Min	Max
Price tick	-0.0006	0.699	-12	12
SWP^1	0.002	0.134	-0.453	0.486
SWP^{2-5}	0.061	0.409	-1.543	1.498
$\triangle SWP^1$	0.000	0.121	-0.611	0.714
$\triangle SWP^{2-5}$	0.000	0.185	-1.499	1.256
$\triangle TSWP^1$	0.0003	0.040	-0.482	0.496
$\triangle TSWP^{2-5}$	0.0001	0.062	-0.685	0.750
WSS^{1-5}	0.014	0.002	0.007	0.025
TT^{2-5}	5595545	17009648	0	284483590

we only focus on studying the price formation during ordinary trading.

The recorded data are the visible part for traders in real time, i.e. hidden orders are not considered. The data consists of every transaction, including volume, price, changes in the limit order book and a measure of time on a second scale. Descriptive statistics of the downloaded variables and the measures presented above are presented in Table 1.

The data is aggregated into fixed intervals during the trading day. A low level of aggregation gives less variation in the order book measures while higher levels of aggregation gives a larger variation. The drawback of a higher aggregation level is that information are lost, i.e. fewer observations. The results in this paper are based on using one minute intervals, but other aggregation levels have also been considered.[5]

The tick size, i.e. the minimum amount a price can move, is for Nokia 0.5 SEK. The price evolution is therefore characterized by discrete jumps. In order to avoid capturing bid-ask bounce effects the bid-ask midpoint is used as the price variable. This renders a discrete price variable with a tick size of 0.25 SEK, i.e. the number of tick doubles compared with using the original price variable. In Figure 3 (left) the price discreteness is illustrated in terms of number of tick changes for Nokia, while the price evolution is shown to the right.

[5] Estimation have also been made on the 2, 5 and 10 minute aggregation levels.

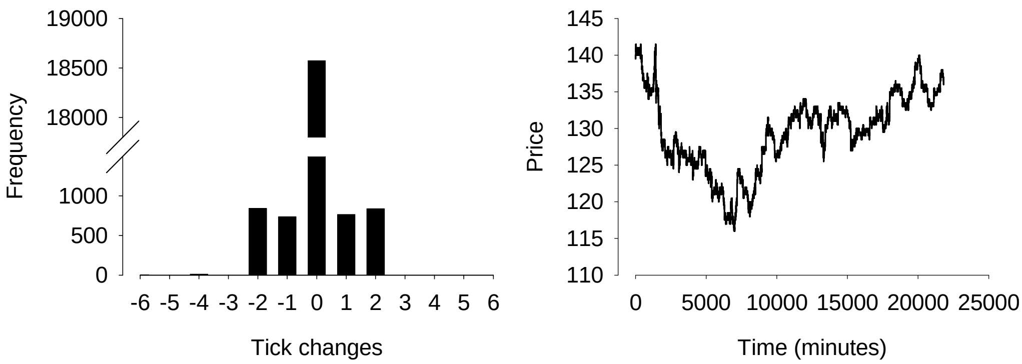

Figure 3: Price change in number of ticks and the price serie for the Nokia share.

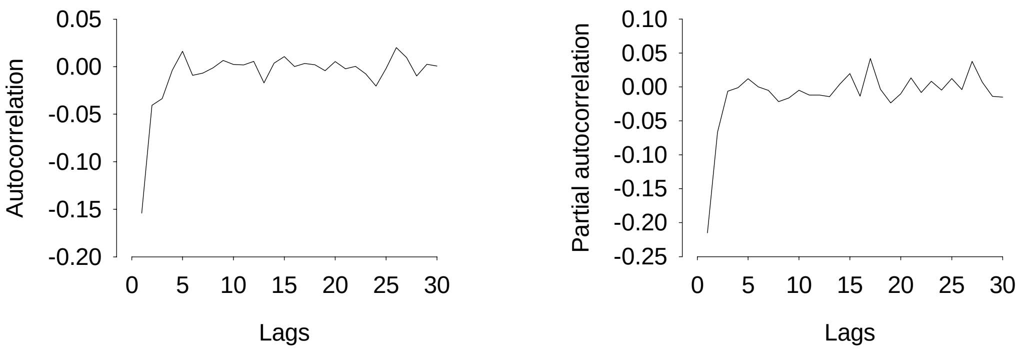

Figure 4: Autocorrelation and partial autocorrelation functions for price tick changes in the Nokia share.

Table 2: Correlation matrix between the five levels of the weighted price measure, SWP, for the Nokia share.

	1	2	3	4	5
1	1.0000				
2	-0.6263	1.0000			
3	-0.2637	0.3256	1.0000		
4	-0.2291	0.2547	0.3775	1.0000	
5	-0.1041	0.0547	0.4736	0.3086	1.0000

The histogram for price tick changes reveals that most changes are of low order and that there is a large proportion of zeros (zero changes) in the sample. Autocorrelation and partial autocorrelation functions for the price tick changes are given in Figure 4. Table 2 show the correlations between the five levels of the weighted price measure. The first levels of the order book (the highest bid price and lowest ask price, see figure 1) is negatively correlated with the rest of the order book (higher and lower end of the order book) while the levels in the higher and lower end of the order book is positively correlated. Since the behavior of the order book seems to be rather homogeneous concerning levels 2-5, these will be treated as aggregates in the empirical application, i.e. order book variables will be constructed for the first level and one measure aggregating the other levels (2-5).

5 Empirical results

The empirical results are presented for the differenced INAR(1) model with explanatory variables at the one minute aggregation level. Estimation results[6] at the 2, 5 and 10 minute aggregation level revealed that the order book measures do not contribute significantly (with minor unsystematic exceptions) in explaining price changes at the higher aggregation levels. The estimations are carried out with CLS and WCLS. For the WCLS, the conditional variance estimated with explanatory variables is employed.

[6] Estimations are throughout performed with the RATS 6.0 econometric software.

Model evaluation against serial correlation in the standardized residuals[7] and squared standardized residuals is performed with the Ljung-Box test statistics.[8] To determine the optimal lag structure the Akaike information criterion has been used.

In Table 3 and 4 estimation results for the different parametrizations concerning the order book measures are reported. The first model show that imbalance in the first levels (between quantities at best bid- and ask-price) of the order book (SWP^1_{t-1}) has a significant (at the 5 percent significance level throughout the paper unless otherwise noted) effect on future price changes. The parameters for the imbalance measure SWP^1_{t-1} has a significant negative impact on the probability for an price increase and a positive impact on the probability for a price decrease (remember that $\lambda_t = \exp(\mathbf{x}_t\boldsymbol{\beta}_1)$ and $\theta_{dt} = 1/[1+\exp(\mathbf{x}_t\boldsymbol{\beta}_2)]$). This means that when there is a positive skewness in the order book concerning the first levels (more on the ask-side of the order book) in the beginning of a period the probability for a decreasing price during the next period increases. The second model tests whether information concerning imbalance in the other levels of the order book (SWP^{2-5}_{t-1}) adds any explanatory power in explaining future price changes. The parameter estimates indicate that an increase in the measure (increasing the ask-side in the higher levels of the order book) has a significant negative impact on the probability of a price increase and increases significantly the probability of a price decrease. The size of the parameters for SWP^1_{t-1} and SWP^{2-5}_{t-1} indicate that imbalance in the first levels of the order book have a larger impact on future price changes than imbalance in the higher levels. The AIC improves as higher levels of the order book is included in the model.

Model 3 and 4 in Table 3 report results for similar models but with the change measures of the order book. The results indicate that changes in the order book measure ($\triangle SWP^1_{t-1}, \triangle SWP^{2-5}_{t-1}$) significantly (at the 10 percent significance level for $\triangle SWP^{2-5}_{t-1}$ in the λ specification) affects future price changes, particularly concerning changes in the first levels. In Table 4 estimation results concerning the weighted change in order book measures $\triangle TSWP^1_{t-1}$ and $\triangle TSWP^{2-5}_{t-1}$ are presented (model 5

[7] $\hat{\varepsilon}_t/V^{1/2}(\triangle P_t|P_{t-1})$.

[8] The critical value at the 5 percent significance level from the test statistic's asymptotic distribution, χ^2_{30}, is 43.7.

Table 3: Results for the Nokia share with explanatory variables estimated by CLS.

	Nokia							
	Model 1		Model 2		Model 3		Model 4	
	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
λ								
ΔP^u_{t-1}	-0.1532	-4.99	-0.1564	-6.18	-0.0093	-0.27	-0.0515	-2.24
ΔP^u_{t-2}	-0.0973	-1.56	-0.1052	-2.22	-0.1028	-7.54	-0.0832	-6.01
ΔP^u_{t-3}	-0.1068	-1.13	-0.1100	-1.37	-0.0950	-4.04	-0.0881	-3.02
ΔP^u_{t-4}	-0.1084	-1.06	-0.1232	-1.40	-0.2200	-4.27	-0.2138	-2.90
ΔP^u_{t-5}	0.0577	0.97	0.0504	0.90	0.0088	0.78	0.0108	1.21
SWP^1_{t-1}	-6.9985	-11.2	-6.4504	-10.6				
SWP^{2-5}_{t-1}			-0.4268	-6.76				
ΔSWP^1_{t-1}					-1.4752	-7.76	-1.4396	-7.78
ΔSWP^{2-5}_{t-1}							-0.1978	-1.73
Constant	-2.5752	-17.89	-2.3184	-16.3	-0.7093	-7.80	-0.6150	-7.18
θ_d								
ΔP^d_{t-1}	-0.1630	-6.65	-0.1608	-7.43	-0.0771	-3.73	-0.0580	-2.46
ΔP^d_{t-2}	-0.0167	-0.44	-0.0218	-0.72	-0.0638	-4.48	-0.0673	-4.90
ΔP^d_{t-3}	-0.0283	-0.63	-0.0278	-0.69	-0.0259	-1.29	-0.0190	-0.88
ΔP^d_{t-4}	0.1447	3.29	0.1445	3.48	0.2216	3.91	0.2155	2.72
SWP^1_{t-1}	-6.5479	-12.2	-5.8709	-11.1				
SWP^{2-5}_{t-1}			-0.1552	-2.12				
ΔSWP^1_{t-1}					-0.4277	-2.22	-0.8634	-3.84
ΔSWP^{2-5}_{t-1}							-0.4821	-3.35
Constant	8.7727	69.9	8.5737	63.6	6.9526	77.0	6.8584	80.2
LB_{30}	83.7		78.1		68.7		68.0	
LB^2_{30}	50.4		50.0		60.2		63.7	
AIC	-0.8197		-0.8222		-0.7620		-0.7681	

Note: LB_{30} and LB^2_{30} are the Ljung-Box statistics of the residuals and squared residuals over 30 lags. Residuals are calculated as $\hat{\varepsilon}_t / V^{1/2}(\triangle P_t | P_{t-1})$.

Table 4: Results for the Nokia share with explanatory variables estimated by CLS.

	Nokia			
	Model 5		Model 6	
	Coeff	t-value	Coeff	t-value
λ				
ΔP^u_{t-1}	-0.0876	-7.20	-0.0903	-6.34
ΔP^u_{t-2}	-0.0763	-4.63	-0.0731	-4.05
ΔP^u_{t-3}	-0.0640	-2.27	-0.0840	-2.22
ΔP^u_{t-4}	-0.1800	-2.96	-0.2187	-2.63
ΔP^u_{t-5}	0.0134	1.38	0.0122	1.13
$\Delta TSWP^1_{t-1}$	-1.2469	-4.02	-2.2476	-6.28
$\Delta TSWP^{2-5}_{t-1}$			-1.4416	-5.14
Constant	-0.4141	-5.16	-0.3800	-3.64
θ_d				
ΔP^d_{t-1}	-0.0872	-7.11	-0.0817	-6.61
ΔP^d_{t-2}	-0.0157	-0.85	-0.0158	-0.77
ΔP^d_{t-3}	-0.0098	-0.49	0.0173	0.47
ΔP^d_{t-4}	0.1848	2.79	0.2259	2.63
$\Delta TSWP^1_{t-1}$	-0.9807	-3.11	-1.4145	-3.71
$\Delta TSWP^{2-5}_{t-1}$			-0.2035	-0.79
Constant	6.6646	82.1	6.6307	62.7
LB_{30}	53.8		58.4	
LB^2_{30}	90.3		82.0	
AIC	-0.7506		-0.7580	

Note: LB_{30} and LB^2_{30} are the Ljung-Box statistics of the residuals and squared residuals over 30 lags. Residuals are calculated as $\hat{\varepsilon}_t / V^{1/2}(\triangle P_t | P_{t-1})$.

and 6). The use of the weighted measures gives similar estimates as for $\triangle SWP^1_{t-1}$ and $\triangle SWP^{2-5}_{t-1}$. This is to be expected since the measures do not differ that much at the one minute aggregation level. Estimation on higher aggregation levels (5 and 10 minutes) showed larger effects from using $\triangle TSWP$ than $\triangle SWP$. The AIC improves as the levels $(2-5)$ of the order book is accounted for (as opposed to only using the first levels) in all the models.

Table 5 reports estimation results using the weighted standardized spread (WSS) and total turnover (TT) measures given in (8) and (9),(10), respectively. The parameter estimates for WSS_{t-1} (model 7) are significant in both the λ and θ_d specifications. However, the signs contradict each other and indicate that a marginal increase in the WSS_{t-1} measure increase the probability for both a price increase (via λ) and a lowered price (via θ_d). The parameter estimate for the total turnover measure (TT_{t-1}), given in model (10), is insignificant and does not explain future price changes.

Table 6 reports the net average marginal effects, i.e. $\partial E(\triangle P_t|P_{t-1})/\partial x_{it}$, of the order book measures (based on the parameter estimates for models 2, 3, 7 and 8) calculated according to (4). The effects indicate a price decrease of 1.62 ticks for a marginal increase in the SWP^1_{t-1} measure (marginal increase in the first level on the ask side of the order book compared to the first level of the bid side). The effect of a marginal change in the higher levels of the order book, i.e. in SWP^{2-5}_{t-1}, is more modest and amounts to a lowered price by 0.08 ticks. The net average marginal effects for the change in the order book measures $\triangle SWP^1_{t-1}$ and $\triangle SWP^{2-5}_{t-1}$ are similar decreasing the price with 0.96 and 0.38 ticks, respectively. The net average marginal effects concerning the total turnover and weighted spread measures are positive (insignificant) and negative (insignificant), respectively.

Table 7 reports parameter estimates of the time-varying specification of the σ^2_t given in (3). Large imbalances (towards the ask-side) in the first level as well as in higher levels of the order book have a negative significant (insignificant for SWP^{2-5}_{t-1}) impact on σ^2. Positive changes in the previous period both at first and higher levels of the order book, i.e. $\triangle SWP^1_{t-1}$ and $\triangle SWP^{2-5}_{t-1}$, also significantly lowers σ^2. The total turnover measure significantly effects σ^2 unlike the weighted spread measure. Note

Table 5: Results for the Nokia share with explanatory variables estimated by CLS.

	Model 7		Model 8	
	Coeff	t-value	Coeff	t-value
λ				
ΔP^u_{t-1}	-0.0862	-4.38	-0.0898	-4.00
ΔP^u_{t-2}	-0.0574	-3.57	-0.0645	-4.43
ΔP^u_{t-3}	-0.0241	-0.99	-0.0219	-0.76
ΔP^u_{t-4}	-0.0020	-0.05	-0.0089	-0.11
ΔP^u_{t-5}	0.0111	1.36	0.0086	0.80
WSS_{t-1}	54.49	8.61		
TT^{2-5}_{t-1}			-0.0831	-0.03
Constant	-0.9855	-24.7	-0.2786	-3.21
θ_d				
ΔP^d_{t-1}	-0.1101	-8.13	-0.1170	-6.67
ΔP^d_{t-2}	-0.0290	-1.81	-0.0278	-1.71
ΔP^d_{t-3}	-0.0380	-1.82	-0.0429	-1.92
ΔP^d_{t-4}	-0.0204	-0.49	-0.0234	-0.29
WSS_{t-1}	-65.43	-11.5		
TT^{2-5}_{t-1}			0.8385	0.26
Constant	7.3872	256.4	6.5235	75.2
LB_{30}	41.6		43.8	
LB^2_{30}	131.9		139.2	
AIC	-0.3916		-0.3677	

Note: LB_{30} and LB^2_{30} are the Ljung-Box statistics of the residuals and squared residuals over 30 lags. Residuals are calculated as $\hat{\varepsilon}_t/V^{1/2}(\triangle P_t|P_{t-1})$. The variable TT is divided by 1000 000 000.

Table 6: Average net effect of explanatory variables for the conditional mean and variance for the Nokia share.

	Conditional mean		Conditional variance	
	Coeff	t-value*	Coeff	t-value*
SWP^{1}_{t-1}	-1.6201	-13.6	-0.5346	-2.54
SWP^{2-5}_{t-1}	-0.0786	-7.01	-0.0793	-1.46
ΔSWP^{1}_{t-1}	-0.9622	-6.41	-0.1473	-0.58
ΔSWP^{2-5}_{t-1}	-0.3766	-3.63	-0.1009	-0.68
TT^{2-5}_{t-1}	1.6808	0.34	5.7887	3.37
WSS_{t-1}	-8.7207	-0.70	58.486	6.63

*Standard errors are calculated by the delta method.

Table 7: Estimates of the conditional variance for the Nokia share.

	Model 9		Model 10		Model 11		Model 12	
σ^2	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
SWP^{1}_{t-1}	-0.9778	-5.93						
SWP^{2-5}_{t-1}	-0.0737	-1.56						
ΔSWP^{1}_{t-1}			-0.6815	-2.93				
ΔSWP^{2-5}_{t-1}			-0.4009	-3.04				
WSS_{t-1}					8.9326	0.91		
TT^{2-5}_{t-1}							8.5527	4.81
Constant	0.3075	14.0	-0.0902	-4.30	-0.4870	-3.78	-0.3485	-15.9
LB_{30}	41.2		60.1		91.0		73.8	
LB^2_{30}	0.3		0.3		0.5		0.5	

Note: LB_{30} and LB^2_{30} are the Ljung-Box statistics of the residuals and squared residuals over 30 lags. Residuals are calculated as $\hat{\varepsilon}_t/V^{1/2}(\triangle P_t|P_{t-1})$. The variable TT is divided by 1000 000 000.

that σ_t^2 is the second part of the full conditional variance given in (2) and that effects from order book measures also effect through the first moment parameters. The full average net marginal effects of the order book measures on the conditional variance have been calculated according to equation (5) and are given in Table 6. The significant net average marginal effects are given by the SWP_{t-1}^1, TT_{t-1}^{2-5} and WSS_{t-1} measures. This means that larger imbalance in the first level of the order book, high total turnover at higher levels and a larger weighted spread measure effects the conditional variance negative, positive and positive, respectively. Autocorrelation is present for all models and have been hard to wipe out.

6 Conclusion

The results indicates that there is informational value in the order book in particular for the first levels of the bid- and ask-side of the order book. Both the change and the imbalance measures of the order book significantly explains future price changes. The effects are most apparent at a low aggregation level (1 minute) while estimation results for higher aggregation levels (2, 5 and 10 minutes) showed mostly insignificantly results. For the 2 minute aggregation level results were similar but with fewer significant parameters concerning order book measures. At the 5 and 10 minute aggregation level only a few parameters were significant with conflicting signs on the parameters. The results of the paper indicate that the informational content of the order book is very short-term. This can be compared to Cao et al. (2004), who found an informational value of the higher levels of the order book at an aggregation level of 5 and 10 minutes. The use of the SWP and the $\triangle SWP$ measures gave throughout the empirical analysis the most robust results, i.e. gave similar parameter estimates, and seem to capture movements in the order book in a satisfactory way. The use of a discrete time-series integer-valued modelling framework have given reasonable results and was easy to use and implement concerning estimation.

References

Al-Osh, M.A. and Alzaid, A.A. (1987). First-Order Integer Valued Autoregressive Process. *Journal of Time Series Analysis* **8**, 261-275.

Angel, J. (1996). Limit Versus Market Orders. Working paper, Georgetown university.

Ball, C. (1988). Estimation Bias Induced by Discrete Security Prices. *Journal of Finance* **43**, 841-865.

Bloomfield, R., O'Hara, M. and Saar, G. (2005). The "Make or Take" Decision in an Electronic Market: Evidence on the Evolution of Liquidity. *Journal of Financial Economics* **75**, 165-199.

Brännäs, K. (1995). Explanatory Variables in the AR(1) Count Data Model. *Umeå Economic Studies* **381**.

Brännäs, K. and Hellström, J. (2001). Generalized Integer-Valued Autoregression. *Econometric Reviews* **20**, 425-443.

Cao, C., Hansch, O. and Wang, X. (2004). The Informational Content of an Open Limit Order Book. Working Paper. Penn state university.

Cho, D.C. and Frees, E.W. (1988). Estimating the Volatility of Discrete Stock Prices. *Journal of Finance* **43**, 451-466.

Foucault, T. (1999). Order Flow Composition and Trading Cost in a Dynamic Limit Order Market. *Journal of Financial Markets* **2**, 99-134.

Glosten, L. (1994). Is the Electronic Open Limit Order Book Inevitable?. *The Journal of Finance* **4**, 1127-1161.

Gourieroux, C. and Jasiak, J. (2001). *Financial Econometrics: Problems, Models, and Methods.* Princeton Series in Finance, Princeton.

Harris, L. (1998). Optimal Dynamic Orders Submission Strategies in some Stylized Trading Problems . *Financial Markets, Institutions & Instruments* **7**, 1-76.

Harris, L.E. and Panchapagesan, V. (2005). The Information Content of the Limit Order Book: Evidence from NYSE Specialist Trading Decisions. *Journal of Financial Markets* **8**, 25-67.

Hausman, J., Lo, A.W. and MacKinlay, A.C. (1992). An Ordered Probit Analysis of

Transaction Stock Prices. *Journal of Financial Economics* **31**, 319-379.

Hautsch, N. and Pohlmeier, W. (2002). Econometric Analysis of Financial Transaction Data: Pitfalls and Opportunities. *Journal of the German Statistical Society* **1**, 5-30.

Jain, P. (2002). Institutional Design and Liquidity on Stock Exchanges Around the World. Working Paper. Emory university.

Keim, D.B. and Madhavan, A. (1998). The Cost of Institutional Equity Trades. *Financial Analysts Journal* **54**, 50-69.

Liesenfeld, R., Nolte, I. and Pohlmeier, W. (2006). Modelling Financial Transaction Price Movements: A Dynamic Integer Count Data Model. Forthcoming in *Empirical Economics.*

McKenzie, E. (1985). Some Simple Models for Discrete Variate Time Series. *Water Resources Bulletin* **21**, 645-650.

McKenzie, E. (1986). Autoregressive Moving-Average Processes with Negative Binomial and Geometric Marginal Distributions. *Advances in Applied Probability* **18**, 679-705.

Rock, D. (1996). The Specialist's Order Book and Price Anomalies. Working paper Harvard University.

Seppi, D. (1997). Liquidity Provision with Limit Orders and Strategic Specialist. *Review of Financial Studies* **10**, 107-150.

The Impact of News Releases on Trade Durations in Stocks
-Empirical Evidence from Sweden

Ola Simonsen*
Department of Economics, Umeå University
SE-901 87 Umeå, Sweden

Abstract

This paper studies the impact of news announcements on trade durations in stocks on the Stockholm Stock Exchange. The news are categorized into four groups and the impact on the time between transactions is studied. Times before, during and after the news release are considered. Econometrically, the impact is studied within an autoregressive conditional duration model using intradaily data for six stocks. The empirical results reveal that news reduces the duration lengths before, during and after news releases as expected by the theoretical litterature on durations and information flow.

Key Words: Finance, transaction data, intraday, market microstructure, ACD.

JEL Classification: C12, C32, C41, G14.

*The author would like to thank Kurt Brännäs, Jörgen Hellström, Olle Westerlund and Shahid Quoreshi and participants in a seminar at Umeå University for valuable comments and suggestions. Torun Reinhammar from the news agency Direkt is thanked for help with data and describing the work of the news agency. Financial support from the Wallander Foundation is gratefully acknowledged.

1 Introduction

This paper examines empirically the short-run impact of public news announcements on trade durations in stocks traded on the Stockholm Stock Exchange in Sweden. The duration corresponds to the time between two consecutive transactions, and a transaction refers to a trade between a buyer and a seller of a volume of stocks at a given price. The news announcements are the news released from one of the leading news agencies in Sweden. The news announcements are categorized into four categories, Company/Sector news, EU macro, Swedish macro and US macro, and their impacts on durations are studied.

The idea to study the duration between transactions originates in the information based model of Glosten and Milgrom (1985). In this model traders are either informed or uninformed, with information varying with the value of the asset they trade. Uninformed traders mainly trade for liquidity reasons, while informed traders act on their superior information on the value of the asset. Easley and O'Hara (1992) extend this model by highlighting the importance of time to distinguish between informed and uninformed traders. They stress that new information, public or private, to market participants, leads to increased trade intensity, i.e. shorter durations between transactions. This corresponds to an increased number of informed agents trying to exploit their new information. The new public information may, e.g., be news announcements from news agencies or press releases. Indicators thought to carry information of revealed new information are, e.g., larger volumes than expected and higher price volatility. In this paper the public news announcement variables are quantified with data from a news agency that provides economic and financial information in Sweden.

Studies of news arrivals, or from a general point of view the impact of events, may be assigned to the literature of event studies, e.g., Campbell et al. (1997). Previous studies have used various proxies for news announcement. Berry and Howe (1994) use the number of daily newspaper headlines and earning announcements. Ederington and Lee (1993) study macroeconomic news and Mitchell and Mulherin (1994) examine stock market announcements. More recent studies are Bollerslev et al. (2000)

with macroeconomic news, Bauwens et al. (2005) who use headlines released on the Reuters news alert screen and Kalev et al. (2004) who employ firm specific announcements. These previous studies have mainly focused on news announcements impact on returns and volatility in securities, rather than on durations. However, there is a link between trade durations and price formation. For example, Grammig and Wellner (2002), Dufour and Engle (2000) and Engle (2000) study the interdependence between intradaily prices, price volatility and trade durations. Dufour and Engle (2000) find that as the time between transactions become shorter the speed of price adjustment increases. This suggests that an active market with short durations demonstrates presence of informed traders. Consequently, publicly released news announcements that contains price driving information may not only reduce duration lengths but may also affect prices. In this paper we concentrate on trade durations to detect the presence of informed traders after news releases. As far as we are aware this is the first empirical study of the impact of news arrivals on durations.

News announcements are sometimes unanticipated, while in other instances both content and timing are expected. News announcements may therefore have an impact on durations both before and after the actual news are released. For the expected news announcement, e.g., scheduled announcements as annual reports and macro figures, an increased trade activity (shorter durations) before the news may be due to price speculations about the information in the announcement. Trade activity after news announcements may be present both for the expected and unanticipated news announcements. The increased trade activity may be due to, e.g., the incorporation of the new information into the price.

For practitioners it may be of interest to study duration dynamics in relation to unanticipated news releases. According to the definitions of the efficient market hypothesis a market is efficient if current and important information are freely available to all market participants. Hence, the use of nonpublicly released price driving information, or insider trading, is prohibited in most countries, e.g., in Sweden by the Market Abuse Penal Act.[1] A possibility for the authorities to monitor violations of the law may

[1] www.fi.se

be to study the duration dynamics. An increased trade intensity before unscheduled and unanticipated news announcements may be due to leaks of insider information. By analyzing durations before unscheduled and unanticipated news releases it may be possible for the authorities to look for leaks and abuse of insider information.

The study of the impact of news on durations econometrically is based on the autoregressive conditional duration model (ACD) of Engle and Russell (1998). The model is useful for irregularly spaced time series data and may also be used for testing hypotheses about sources of clustering of short durations (e.g., Bauwens and Giot, 2001; Engle and Russell, 1998). The categorized news are incorporated as dummy variables in the mean function of the ACD model. The dummy variables represent intervals of before, during and after the news announcements. Consequently the impact of news releases on durations before, during and after the news announcements can be studied.

The outline of the paper is the following. Section 2 presents the data. Section 3 presents the ACD model with details concerning estimation, followed by Section 4 where the empirical results are presented. The final section discusses and concludes.

2 Data

2.1 Stock transaction data

The transaction data were downloaded from Ecovision (www.ecovision.se), a provider of real time financial information from the Stockholm Stock Exchange, and later filtered by the author. The Stockholm Stock Exchange is an order driven market, i.e. no market maker is involved, traders enter sell or buy orders to the order book, visible to all market participants. The database covers June 7 to August 26, 2005, with opening hours 0930 to 1730. Seven days are missing due to technical problems during the summer holiday, from August 9 to August 18. The database contains every transaction with associated spread, price and traded volume, recorded on a second scale. The opening and closing procedure at the Stockholm Stock Exchange determines the opening and closing price for the day and is not regular trading. As we only consider regular trading the first and last 15 minutes of the trade day is deleted. Six stocks were simultaneously downloaded and recorded. The stocks are; Ericsson, Nordea, SCA, Skanska, TeliaSonera and Volvo. The stocks are active in different lines of business. Ericsson is a provider of telecommunication equipment, Nordea is a banking company, Volvo manufactures among other things trucks and buses, SCA produces paper and hygiene products, Skanska is a construction company, and TeliaSonera is a telecommunication company.

The six selected stocks are presented in Table 1 with associated descriptive statistics. All stocks are among the most traded stocks on the Stockholm Stock Exchange. Table 1 shows that Ericsson is the most traded among the selected stocks, followed by Volvo, TeliaSonera, Nordea, SCA and Skanska in that order.

One feature of transaction data is the diurnal pattern of duration lengths (e.g., Bauwens et al. 2002; Engle and Russell, 1998). Especially after the opening and prior to the closing of the market the trade intensity is high, i.e. shorter durations, and around the lunch-hour the trade intensity is lower, i.e. longer durations (Simonsen, 2005). Several suggestions of how to account for a potential diurnal pattern have been considered. Engle and Russell (1998) suggest the use of time adjusted durations $d_i = D_i/\phi_i$, where D_i is the duration from the data and ϕ_i is a cubic spline with nodes on each hour defined as the expected duration conditioned on the time of day. Obviously, any other flexible function may be used to capture the intraday seasonality in addition to the cubic spline function (e.g., Bauwens et al. 2002). The adjusted duration d_i is used for estimation. By using d_i, autocorrelations are reduced when compared to unadjusted durations, D_i. Still, auotocorrelations are high in the adjusted durations. The deseasonalizing procedure is also used for the explanatory variables used in the estimations.

2.2 News announcement data

The news announcement data consist of news messages released from the news agency Direkt (www.direkt.se). We only consider news announcement data that coincide with the stock transaction data, i.e. from June 7 to August 26, 2005, and opening hours 0945-1715.

Table 1: Summary statistics of the durations. The figures reported are measured in seconds.

	Mean	Std.	Min	Max	Nr obs
Ericsson	8.9	18.7	0	351	184292
Nordea	35.2	74.8	0	1415	40780
SCA	37.3	79.6	0	1491	38487
Skanska	45.0	91.1	0	1489	31941
TeliaSonera	26.0	59.7	0	1048	55209
Volvo	22.7	51.6	0	936	63581

The distributed news are related to Swedish financial markets and include stock specific news, national and international economic macro news, forecasts and expectations in the market. The first release after a news event is a short summary in a so-called flash. It consists of a few sentences and is supposed to be quick and to provide important information to investors. Only information concerning the most important news events is distributed through a flash. A news flash is followed, within a couple of minutes, by additional and more detailed information. The news flash releases are recorded on a minute scale. The news flash releases are rounded to the nearest integer minute, i.e. recorded seconds $s \in [0, 29]$ are assigned to the previous integer minute, and seconds $s \in [30, 59]$ are assigned to the following integer minute. The announcement with additional information is recorded on a second scale.

To capture the effect of different kinds of news, the news data from Direkt are categorized into four groups. Table 2 gives both the frequencies of the news flash and the following additional news releases in the different categories of news. The first group contains both company and sector specific news. The company news contain news directly concerning the company. The sector news are related to companies in the same sector or news related to the sector itself. The three following groups, US, European and Swedish macroeconomic news contain, e.g., employment reports, price indices, GDP reports and other important macro figures from the different regions. Other studies using categorized news announcement data are, e.g., Bauwens et al. (2005) and Bollerslev et al. (2000) with macro news announcement, and Kalev et al.

Table 2: Number of news announcements in the different news categories.

	Company/Sector news					
	Ericsson	Nordea	SCA	Skanska	TeliaSonera	Volvo
Company/sector flash	62	30	67	18	17	26
Company/sector, additional info	32	21	28	16	14	9
	Macro news					
US macro flash	127					
US macro, additional info	79					
Eu macro flash	143					
Eu macro, additional info	103					
Swe macro flash	148					
Swe macro, additional info	69					

(2004) who employ firm specific announcements.

3 Model and estimation

The impact of news announcements is studied within the autoregressive conditional duration model (ACD) of Engle and Russel (1998). For this purpose dummy variables of before, during and after a news announcement are created.

First, we start by presenting the ACD model and a functional form extension of Bauwens and Giot (2001). Let $d_i = t_i - t_{i-1}$ be the duration between transactions at times t_i and t_{i-1} and the conditional expected duration

$$\theta_i = \theta_i(d_{i-1}, ..., d_1; x) = E(d_i \,| d_{i-1}, ..., d_1; x)\,.$$

The conditional expected duration may be parameterized as suggested by Engle and Russell (1998) in a linear form

$$\theta_i = \omega + \sum_{j=1}^{p} \alpha_j d_{i-j} + \sum_{j=1}^{q} \beta_j \theta_{i-j} + \gamma' x_{i-1}$$

where θ_i is specified in such a way that $\epsilon_i = d_i/\theta_i$ is independent and identically distributed. This is called an ACD(p, q) with p duration lags and q lags in the expected duration and x contains explanatory variables such as price and volume. A useful variant of the original ACD model is the log-ACD of Bauwens and Giot (2001). The extension ensures that expected durations remain positive:

$$\theta_i = \exp\left[\omega + \sum_{j=1}^{p} \alpha_j d_{i-j} + \sum_{j=1}^{q} \beta_j \ln \theta_{i-j} + \gamma' x_{i-1}\right].$$

Dummy variables for the news announcements are added as explanatory variables. The dummy variable for a news announcement flash is denoted $N^f_{i,\tau}$, where i indicates the news type and τ is used to indicate the interval before, during or after the news announcement (see Figure 1). The durations considered in the intervals are completed ones, i.e. the terminal point of the durations are in the observation window. $N^f_{i,0}$ is equal to 1 if a news flash is released during the interval and zero otherwise. $N^f_{i,-1}$ and $N^f_{i,1}$ are dummies for the intervals before and after a news flash announcement, respectively. Empirically, the dummies represent observation windows equal to five minutes before the announcement and ten minutes after the news announcement. Other lenghts of the observation windows were also considered, see next section. The observation window during which the news arrives, $N^f_{i,0}$, is equal to 1 minute. The total observation window during and around the news announcement is 16 minutes. To consider the impact of more than 1 news flash announcement of type i during the observation window the dummy $N^f_{i,f>1}$ is equal to 1 if the number of news flash announcements is larger than 1 and zero otherwise.

The impact of the news announcement with additional information, released after the news flash, is considered with a dummy variable (see Figure 1). The dummy $A_{i,1}$ for the additional information is equal to 1 if additional information is released of type i and zero otherwise. The lengths of the observed interval after the additional information, $A_{i,1}$, is equal to 5 minutes.

We may determine the percentage change in the conditional mean duration of a change in the discrete valued dummy variable $N^f_{i,\tau}$. Denote the expected mean duration

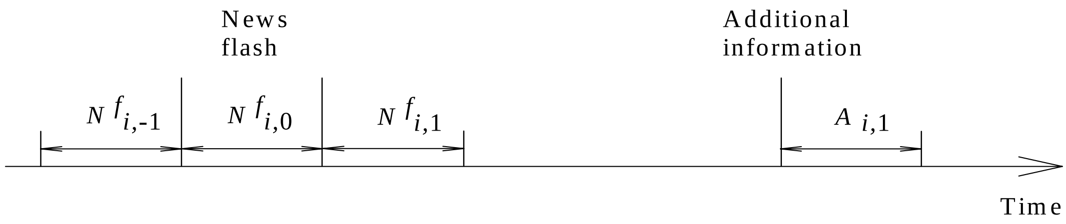

Figure 1: Illustration of the dummy variable structure during and around the news flash and additional news arrivals.

$E(d_i|N^f_{i,\tau}=0)$ as the base level and $E(d_i|N^f_{i,\tau}=1)$ as the mean duration when a news announcement is released during the interval τ. The percentage change, $P^f_{i,\tau}$, of a news flash is then

$$P^f_{i,\tau} = 100\frac{E(d_i|N^f_{i,\tau}=1) - E(d_i|N^f_{i,\tau}=0)}{E(d_i|N^f_{i,\tau}=0)}.$$

The x vector of explanatory variables also contain price, volume and spread.

4 Empirical results

The empirical results are reported for the log-ACD model with explanatory variables. To determine the lag structures we utilize the AIC criterion and for the estimation of parameters we apply the quasi maximum likelihood (QML) estimator based on the exponential distribution (e.g., Engle and Russell, 1998).

The results are presented in Tables 3-5. Table 3 reports the estimates of the dummy variables before, during and after the news flash and the dummy variable of the additional news announcement. Table 4 reports the impact of more than one news flash announcement and the effects of the explanatory variables traded volume, spread and changes in price.

Although not reported in Table 3 but part of the estimates, the parameters α, of lagged durations and β, of lagged expected durations are significant with a few exceptions. Exclusion of the insignificant parameters were also considered, though such a specification is not minimizing the AIC. The β estimates are significant and throughout smaller than one.

Before discussing the individual parameter estimates reported in Table 3 of the news announcement variables of before, during and after the news announcement we may test if the news announcement dummy parameters are jointly significantly different from zero, i.e. if the news variables jointly add explanatory power to the model. A likelihood ratio test is applied for this purpose. It is found that the added news announcement dummies are jointly significantly different from zero for all results. The result holds both when including and excluding the explanatory variables volume, spread and changes in price.

Next, we consider the individual parameter estimates in Table 3. As Easley and O'Hara (1992) claim that new information increases the trade intensity, the expected impact of news before, during and after the release is a reduction of durations, i.e. higher trade intensity. Consequently the expected sign of the parameters of the news announcement dummies is negative. The impact before the news release may be caused by scheduled and anticipated news releases. We find significant results with the expected negative sign for, e.g., US macro news for Ericsson and SCA and Company/Sector news for Ericsson, TeliaSonera and Volvo. The parameter estimates imply that the expected durations lengths are shortened by 5.3 and 2.1 percent for TeliaSonera and Ericsson, respectively, prior to Company/Sector news. The rest of the significant and negative estimates all shorten the durations by less than 5 percent. We find significant and negative parameter estimates also for the 1 minute intervals where the news arrives of, e.g., EU macro news for Ericsson and Nordea and Company/Sector news for SCA, Skanska, Nordea and TeliaSonera. It seems that Company/Sector news reduces duration lenghts, e.g., for Skanska by 19 percent, Nordea by 42 percent and for TeliaSonera by 24 percent.

Table 3: Estimation results of before, during and after news flash and additional news releases for the different stocks (robust t statistics in parentheses).

	US Macro				EU Macro				SWE Macro				Company/Sector				LB_{40}	LB^2_{40}	ln L
Model	$N^f_{i,-1}$	$N^f_{i,0}$	$N^f_{i,1}$	$A_{i,1}$	$N^f_{i,-1}$	$N^f_{i,0}$	$N^f_{i,1}$	$A_{i,1}$	$N^f_{i,-1}$	$N^f_{i,0}$	$N^f_{i,1}$	$A_{i,1}$	$N^f_{i,-1}$	$N^f_{i,0}$	$N^f_{i,1}$	$A_{i,1}$			
Ericsson																			
ACD(8,1)	-0.013	0.010	-0.006	-0.001	-0.003	-0.022	-0.012	0.020	0.002	0.020	-0.003	-0.007	-0.021	0.033	-0.002	-0.008	2631.3	0.6	-133034.4
	(-3.98)	(1.34)	(-2.02)	(-0.35)	(-1.16)	(-3.31)	(-4.98)	(4.85)	(0.28)	(1.10)	(-0.75)	(-1.18)	(-4.88)	(2.29)	(-0.47)	(-1.75)	2495.2	0.6	-133430.4
Nordea																			
ACD(4,1)	-0.025	0.357	0.015	0.008	-0.028	-0.194	0.035	0.055	0.125	0.619	-0.010	-0.071	0.027	-0.544	0.123	-0.150	234.9	128.9	-33286.9
	(-1.52)	(5.41)	(0.72)	(0.29)	(-0.93)	(-3.04)	(2.12)	(2.04)	(3.08)	(4.05)	(-0.26)	(-1.66)	(0.63)	(-9.35)	(3.47)	(-2.64)	237.9	139.6	-33570.4
SCA																			
ACD(7,1)	-0.012	-0.005	0.005	0.009	0.006	-0.009	0.001	0.005	0.0002	-0.024	0.004	0.0004	0.015	-0.045	0.0006	-0.009	269.9	30.0	-31111.9
	(-2.58)	(-0.34)	(1.32)	(1.62)	(1.75)	(-0.62)	(0.64)	(1.79)	(0.04)	(-2.53)	(0.97)	(0.09)	(4.73)	(-11.2)	(0.20)	(-4.06)	255.3	21.9	-31912.0
Skanska																			
ACD(4,1)	0.178	0.014	0.089	-0.144	-0.007	0.050	-0.069	0.152	0.322	0.154	0.149	-0.258	0.028	-0.216	-0.070	0.245	178.9	94.6	-29376.3
	(3.14)	(0.09)	(2.86)	(-3.76)	(-0.18)	(0.39)	(-2.34)	(3.20)	(4.76)	(0.78)	(3.81)	(-5.58)	(0.26)	(-0.69)	(-1.23)	(2.22)	190.8	84.5	-29486.3
TeliaSonera																			
ACD(3,1)	0.016	0.015	-0.042	0.011	-0.021	0.101	-0.062	0.030	0.078	-0.100	0.118	-0.108	-0.054	-0.272	-0.044	-0.090	236.4	284.5	-47742.7
	(1.02)	(0.27)	(-2.00)	(0.36)	(-1.21)	(2.19)	(-4.06)	(1.43)	(2.64)	(-1.59)	(5.53)	(-4.03)	(-3.60)	(-8.83)	(-2.83)	(-2.96)	236.1	254.9	-47922.4
Volvo																			
ACD(8,1)	0.004	0.002	-0.004	0.010	-0.004	0.007	0.0006	-0.002	-0.003	0.001	-0.003	0.006	-0.020	-0.004	0.005	0.004	102.1	11.6	-55843.8
	(1.72)	(0.34)	(-2.96)	(3.51)	(-1.30)	(0.82)	(0.27)	(-0.57)	(-0.82)	(0.09)	(-1.15)	(1.17)	(-2.84)	(-0.50)	(1.43)	(0.58)	263.0	46.4	-56167.6

Notes: Parameter estimates of durations and explanatory variables volume, spread and price changes are not reported in the table, although part of the estimations. Reported LB_{40} and LB^2_{40} are the Ljung box statistic of the residuals and the squared residuals. LB_{40}, LB^2_{40} and loglikelihood are reported both with, uppermost, and without, lower, the news announcement dummies for all stocks.

For the 10 minute observation window after the news announcement the findings are seven negative and significant estimates. These reflect US macro news for Ericsson, TeliaSonera and Volvo, EU macro news for Ericsson, Skanska and TeliaSonera, and Company/Sector news for TeliaSonera. In particular, the macro related news from US and EU seem to influence durations negatively. The relative reduction in durations is at most 10 percent for the negative and significant estimates of the observation window after the news announcement.

Considering the effect of additional information, i.e. the more detailed news release after the first short and rapid news flash, the plausible impact is a reduction of durations if the additional news release contains important information in additional to the news flash. The sign of the coefficients corresponding to $A_{i,1}$ in Table 3 is therefore expected to be negative. This is also found for, e.g., additional Company/Sector news for SCA, Nordea and TeliaSonera. Also, for SWE macro news negative and significant estimates are found for Skanska and TeliaSonera. Additional US macro news reduces durations for Skanska. These additional news shorten durations by 10 to 25 percent.

Reported in Table 4 is the effect of more than one news flash during the 16 minutes observation window. More than one news flash of the same type during the observation windows may originate from important and complex events. If more than one news flash indicates important news events we expect the impact of more than one news flash to additionally reduce durations. Consequently the expected sign is negative. Negative and significant results are found for, e.g., US macro and SWE Macro news for Skanska and Company/Sector news for and TeliaSonera. A majority of the estimated parameter signs of Company/Sector news are negative but insignificant while for the other news categories a large proportion is positive and significant.

The estimates for traded volume, spread and changes in price are also presented in Table 4 and show negative significant estimates of traded volume while significant positive ones for spread and changes in price. That is durations are reduced after large traded volumes and durations are shorter after negative changes in price. Due to numerical problems during estimation, price change is excluded in all estimates for SCA.

Table 4: The impact of more than one news flash announcement during the total length of the observation windows (5+1+10 minutes) for different news categories (columns 1-4). Estimation results of the explanatory variables (columns 5-7).

	US Macro	EU Macro	SWE Macro	Sector/ Company	Traded vol	Spread	Price change
Ericsson	-0.004	0.007***	0.001	-0.005	-0.001***	0.0003	0.048
Nordea	0.019	0.032	0.057*	-0.031	-0.0009***	0.258***	0.203***
SCA	0.0009	0.006**	-0.001	-0.004	-0.0003***	0.0002	-
Skanska	-0.112***	-0.072*	-0.192***	0.065	0.002	0.156***	0.181***
TeliaSonera	0.063***	0.058***	-0.017	-0.098***	-0.002***	0.163***	1.929***
Volvo	0.002*	0.002	0.004*	0.002	-0.0003**	0.004**	-0.002

Notes: The symbols ***,** and * indicate significance at 1%, 5% and 10% levels, respectively. Standard errors are the robust standard error.

Table 5 summarizes the estimation results for the different news categories, i.e., US macro, EU macro, Swedish macro and Company/sector news. The first and second column reports the impact of the news announcement before and after the announcement, respectively. The third column reports the difference between the parameters of before and after news releases. The total sum of the coefficients of the news windows before, during and after the news releases, totally 16 minutes, is reported in the fourth column.

The expected sign of the parameters reported in the first and second column is, as discussed above, negative. The difference between the parameters reported in column three assesses the relative effect between before and after news releases. A positive sign indicates that durations in the observation window before the news release is shorter relative to the durations in the observation window after the news release. The estimates of the difference between before and after the news are both negative and positive. However, for Company/Sector news a major part of the estimates are positive indicating longer durations after the news releases relative to the length of the durations before the announcement.

The total sum of the coefficients of the news windows before, during and after the news releases is reported in the fourth column. We may test if the parameters in the fourth column is less than zero. Rejection implies that news announcement prolongs duration lengths. The results indicates that news have a negative impact on durations for the total observation lengths. Particularly Company/Sector news influence durations negatively. We find shorter durations of Company/Sector news for SCA, Nordea, TeliaSonera and Volvo. Also EU macro news for Nordea significantly reduces the durations.

Other lenghts of the observation windows than 5 minutes of before, 1 minute during and 10 minutes after news releases were utilized and evaluated. However, both shorter intervals, 30 seconds, 1 minute and 3 minutes and longer intervals, 10 and 20 minutes results in less number of significant parameters, although, both shorter and longer observation windows also results in significant parameter estimates.

Table 5: The impact of news before, after and the total effect of news announcements for the different stocks and news categories.

	$N^f_{i,-1}$	$N^f_{i,1}$	$N^f_{i,1} - N^f_{i,-1}$	$\sum_{\tau=-1}^{1} N^f_{i,\tau}$
		US Macro		
Ericsson	-0.013***	-0.006**	0.007*	-0.009
Nordea	-0.025	0.015	0.040	0.347***
SCA	-0.011***	0.005	0.017**	-0.011
Skanska	0.178***	0.089***	-0.089	0.281*
TeliaSonera	0.016	-0.042**	0.058**	-0.011
Volvo	0.004*	-0.004***	0.009	0.002
		EU Macro		
Ericsson	-0.003	-0.012***	-0.009***	0.037***
Nordea	-0.028	0.035***	0.063**	-0.186***
SCA	0.005*	0.001	0.005	-0.002
Skanska	-0.007	-0.069**	-0.062	-0.026
TeliaSonera	-0.021	-0.062***	0.042*	0.018
Volvo	-0.004	0.0005	0.004	0.004
		SWE Macro		
Ericsson	0.002	-0.003	-0.005	0.019
Nordea	0.125***	-0.010	-0.135***	0.733***
SCA	0.0002	0.004	0.003	-0.020**
Skanska	0.322***	0.149***	-0.173***	0.624***
TeliaSonera	0.078***	0.118***	0.040	0.096
Volvo	-0.003	-0.003	0.000	-0.006
		Company/Sector news		
Ericsson	-0.021***	-0.002	0.019***	0.011
Nordea	0.027	0.123***	0.097**	-0.395***
SCA	0.015***	0.0006	-0.015***	-0.029***
Skanska	0.029	-0.070	-0.099	-0.258
TeliaSonera	-0.054***	-0.043***	0.016	-0.369***
Volvo	-0.020***	0.005	0.025***	-0.020**

The symbols ***,** and * indicate significance at 1%, 5% and 10%, respectively. Standard errors in the first and second column are the robust standard errors. The standard errors reported in the third and fourth column are from the chi-squared distributed Wald statistics.

5 Discussion

This paper examines the impact of news on durations in stocks traded on the Stockholm Stock Exchange in Sweden. The news are categorized into four groups and added as explanatory variables to the autoregressive conditional duration model. The dummy variable structure captures the impact before, during and after the news.

The dummy variable structure of before, during and after the news announcement seem to capture news announcements impact on durations. We find increased trade activity before, during and after news releases. In the current sample the relative reduction in mean duration before the news announcement is less than 5 percent. The impact around the news release of for example Company/Sector related news shortens durations by 20 to 40 percent. After the news release mean durations are at most reduced by 10 percent. The results indicates that particularly during the news release the durations are reduced. After the news release the durations are shortened by at most twice as much as before the news release.

The cause of reduced durations before the actual news releases may originate from anticipated news releases, while the significant result of the impact during and after the news release may be due to both anticipated and unanticipated news events. The support is strong for Company/Sector related news, while weaker for macro related news, although, we find significant results also for macro news. The result supports the predictions of Easley and O'Hara (1992) of shorter durations in connection to news. Recent studies also find significant impacts of news announcements, e.g., Bollerslev et al. (2000) and Bauwens el al. (2005) for macro news, Kalev et al. (2004) for firm specific news. These studies are concentrated with the impact on price volatility.

The result indicates that detailed Company/Sector news and to some extent also macro news followed after a brief news flash reduces the durations. Accordingly, additional Company/Sector news contains important information to market traders. The results show weak support for Company/Sector news releases reducing the durations. Consequently, several news flash releases do not contain additional information reducing the expected durations.

Considering the link between durations and price volatility we may expect the news announcement to also affect volatility. To asses this feature we utilize a AR(1)-GARCH(1,1) model with one minute aggregated price data. The news announcement dummy variables are added as explanatory variables to the model. The findings are significant estimates of the news announcements. In the case of Ericsson, Nordea, SCA, Skanska, TeliaSonera the significant estimates are mainly of the interval after the news announcement. Consequently, it seems that volatility is affected mainly after the news announcement while the durations are affected also before and during the news announcement. This may be the result of uncertainty prior the news announcement that results in trading but not in price movements. After the release of news and possible price driving information durations as well as the market price may be affected, i.e., the volatility is increased. One suggestion for further analysis of the relation between news releases, durations and volatility may be to model the variables simultaneously.

References

Bauwens, L., F. Galli, F. and Giot, P. (2002). The Moments of log-ACD Models. Unpublished manuscript, CORE, Louvain-La-Neuve.

Bauwens, L. and Giot, P. (2001). *Econometric Modelling of Stock Market Intraday Activity.* Kluwer, Dordrecht.

Bauwens, L., Omrane, W.B. and Giot P. (2005). News Announcements, Market Activity and Volatility in the Euro/Dollar Foreign Exchange Market. *Journal of International Money and Finance*, 24, 1108-1125.

Berry, T.D. and Howe, K.M. (1994). Public Information Arrival, *Journal of Finance,* 49, 1331-1346.

Bollerslev, T., Cai, J. and Song F.M. (2000). Intraday Periodicity, Long Memory Volatility, and Macroeconomic Announcement Effects in the US Treasury Bond Market. *Journal of Empirical Finance*, 7, 37-55.

Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. (1997). *The Econometrics of Financial Markets.* Princeton University Press.

Dufour, A. and Engle, R.F. (2000). Time and the Price Impact of Trade. *The Journal of Finance,* 55, 2467-2498.

Easley, D. and O'Hara, M. (1992). Time and the Process of Security Price Adjustment. *Journal of Finance,* 47, 577-606.

Ederington, L.H. and Lee, J.H. (1993). How Markets Process Information: News Releases and Volatility, *Journal of Finance,* 48, 1161-1191.

Engle, R.F. (2000). The Econometrics of Ultra-High-Frequency Data. *Econometrica,* 68, 1-22.

Engle, R.F. and Russell, J.R. (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica,* 66, 1127-1162.

Glosten, L. and Milgrom, P. (1985). Bid, Ask and Transactions Prices in a Specialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics,* 14, 71-100.

Grammig, J. and Wellner, M. (2002). Modeling the Interdependence of Volatility and Inter-Transaction Duration Processes. *Journal of Econometrics,* 106, 369-400.

Kalev, P.S., Liu, W.M., Pham P.K. and Jarnecic, E. (2004). Public Information Arrival and Volatility of Intraday Stock Returns. *Journal of Banking & Finance*, 28, 1441-1467.

Mitchell, M.L. and Mulherin, J.H. (1994). The Impact of Public Information on the Stock Market. *The Journal of Finance,* 49, 923-950.

Simonsen, O. (2005). An Empirical Model for Durations in Stocks, *Umeå Economic Studies,* 657.

Avhandlingar framlagda vid Institutionen för nationalekonomi, Umeå universitet

List of dissertations at the Department of Economics, Umeå University

Holmström, Leif (1972) Teorin för företagens lokaliseringsval. UES 1. PhLic thesis

Löfgren, Karl-Gustaf (1972) Studier i teorin för prisdiskriminering. UES 2. PhLic thesis

Dahlberg, Åke (1972) Arbetsmarknadsutbildning - verkningar för den enskilde och samhället. UES 3. PhD thesis

Stage, Jørn (1973) Verklighetsuppfattning och ekonomisk teori. UES 4. PhLic thesis

Holmlund, Bertil (1976) Arbetslöshet och lönebildning - kvantitativa studier av svensk arbetsmarknad. UES 25. PhD thesis

Löfgren, Karl-Gustaf (1977) En studie i neokeynesiansk arbetslöshets- och inflationsteori. UES 34. PhD thesis

Lundberg, Lars (1976) Handelshinder och handelspolitik - Studier av verkningar på svensk ekonomi. Industriens Utredningsinstitut, Stockholm. PhD thesis

Johansson, Per-Olof (1978) Sysselsättning och samhällsekonomi - En studie av Algots etablering i Västerbotten. UES 53. PhD thesis

Wibe, Sören (1980) Teknik och aggregering i produktionsteorin. Svensk järnhantering 1850-1975; en branschanalys. UES 63. PhD thesis

Ivarson, Lars (1980) Bankers portföljvalsbeteende. En teoretisk studie. UES 64. PhD thesis

Batten, David (1981) Entropy, Information Theory and Spatial Input-output Analysis. UES 92. PhD thesis

Hårsman, Björn (1982) Housing Demand Models and Housing Market Models for Regional and Local Planning. Swedish Council for Building Research, D13:1981. PhD thesis

Holm, Magnus (1983) Regionalekonomiska modeller för planering och samordning i en decentraliserad ekonomi. Byggforskningsrådet, R118:1981 and R5:1983. PhD thesis

Ohlsson, Henry (1986) Cost-Benefit Analysis of Labor Market Programs - Applied to a Temporary Program in Northern Sweden. UES 167. PhLic thesis

Sarafoglou, Nikias (1987) A Contribution to Population Dynamics in Space. UES 179. PhD thesis

Ohlsson, Henry (1988) Cost-Benefit Analysis of Labor Market Programs - Applied to a Temporary Program in Northern Sweden. UES 182. PhD thesis

Anderstig, Christer (1988) Applied Methods for Analysis of Economic Structure and Change. CERUM 1988:2, Umeå University. PhD thesis

Karlsson, Charlie (1988) Innovation Adoption and a Product Life Cycle. UES 185. PhD thesis

Löfström, Åsa (1989) Diskriminering på svensk arbetsmarknad - En analys av löneskillnader mellan kvinnor och män. UES 196. PhD thesis

Axelsson, Roger (1989) Svensk arbetsmarknadsutbildning - En kvantitativ analys av dess effekter. UES 197. PhD thesis

Zhang, Wei-Bin (1989) Theory of Economic Development - Nonlinearity, Instability and Non-equilibrium. UES 198. PhD thesis

Hansson, Pär (1989) Intra-Industry Trade: Measurements, Determinants and Growth - A study of Swedish Foreign Trade. UES 205. PhD thesis

Kriström, Bengt (1990) Valuing Environmental Benefits Using the Contingent Valuation Method: An Econometric Analysis. UES 219. PhD thesis

Aronsson, Thomas (1990) The Short-Run Supply of Roundwood under Nonlinear Income Taxation - Theory, Estimation Methods and Empirical Results Based on Swedish Data. UES 220. PhD thesis

Westin, Lars (1990) Vintage Models of Spatial Structural Change. UES 227. PhD thesis

Wikström, Magnus (1992) Four Papers on Wage Formation in a Unionized Economy. UES 287. PhD thesis

Westerlund, Olle (1993) Internal Migration in Sweden - The Role of Fiscal Variables and Labor Market Conditions. UES 293. PhLic thesis

Bergman, Mats A. (1993) Market Structure and Market Power. The Case of the Swedish Forest Sector. UES 296. PhD thesis

Johansson, Per (1993) Count Data Models - Estimator Performance and Applications. UES 315. PhD thesis

Roson, Roberto (1994) Transport Networks and the Spatial Economy - A General Equilibrium Analysis. UES 340. PhD thesis

Li, Chuan-Zhong (1994) Welfare Evaluations in Contingent Valuation - An Econometric Analysis. UES 341. PhD thesis

Østbye, Stein (1994) Regional Labour and Capital Subsidies - Theory and Evidence of the Impact on Employment under Wage Bargaining. UES 344. PhLic thesis

Westerlund, Olle (1995) Economic Influences on Migration in Sweden. UES 379. PhD thesis

Mortazavi, Reza (1995) Three Papers on the Economics of Recreation, Tourism and Property Rights. UES 396. PhLic thesis

Østbye, Stein (1995) Regional Labour and Capital Subsidies. UES 397. PhD thesis

Hussain-Shahid, Imdad (1996) Benefits of Transport Infrastructure Investments: A Spatial Computable General Equilibrium Approach. UES 409. PhD thesis

Eriksson, Maria (1996) Selektion till arbetsmarknadsutbildning. UES 410. PhLic thesis

Karlsson, Niklas (1996) Testing and Estimation in Labour Supply and Duration Models. UES 413. PhD thesis

Olsson, Christina (1996) Chernobyl Effects and Dental Insurance. UES 428. PhLic thesis

Vredin, Maria (1997) The African Elephant - Existence Value and Determinants of Willingness to Pay. UES 441. PhLic thesis

Eriksson, Maria (1997) To Choose or not to Choose: Choice and Choice Set Models. UES 443. PhD thesis

Widerstedt, Barbro (1997) Employer Change and Migration. Two Papers on Labour Mobility in Sweden. UES 444. PhLic thesis

Lundberg, Sofia (1997) The Economics of Child Auctions in 19th Century Sweden. UES 445. PhLic thesis

Lundberg, Johan (1997) Two Papers on Revenue and Expenditure Decisions in the Swedish Local Public Sector. UES 454. PhLic thesis

Widerstedt, Barbro (1998) Moving or Staying? Job Mobility as a Sorting Process. UES 464. PhD thesis

Bask, Mikael (1998) Essays on Exchange Rates: Deterministic Chaos and Technical Analysis. UES 465. PhD thesis

Löfgren, Curt (1998) Time to Study Students: Two Essays on Student Achievement and Study Effort. UES 466. PhLic thesis

Sjögren, Tomas (1998) Union Wage Setting in a Dynamic Economy. UES 480. PhD thesis

Mortazavi, Reza (1999) Essays on Economic Problems in Recreation, Tourism and Transportation. UES 483. PhD thesis

Rudholm, Niklas (1999) Two Essays on Generic Competition in the Swedish Pharmaceuticals Market. UES 485. PhLic thesis

Olsson, Christina (1999) Essays in the Economics of Dental Insurance and Dental Health. UES 494. PhD thesis

Marklund, Per-Olov (1999) Environmental Regulation and Firm Efficiency. UES 504. PhLic thesis

Berglund, Elisabet (1999) Regional Entry and Exit of Firms. UES 506. PhD thesis

Hellström, Jörgen (1999) Count Data Autoregression Modelling. UES 507. PhLic thesis

Nordström, Jonas (1999) Tourism and Travel: Accounts, Demand and Forecasts. UES 509. PhD thesis

Johansson Vredin, Maria (1999) Economics Without Markets. Four papers on the Contingent Valuation and Stated Preference Methods. UES 517. PhD thesis

Schei, Torbjørn (2000) Natural recreation resources: production and a diversity of interests related to the management of grouse as an outfield resource in Finnmark, Norway, in the Euro-Arctic Barents region. UES 523. PhLic thesis

Backlund, Kenneth (2000) Welfare Measurement, Externalities and Pigouvian Taxation in Dynamic Economies. UES 527. PhD thesis

Andersson, Linda (2000) Job Turnover, Productivity and International Trade. UES 530. PhLic thesis

Ylvinger, Svante (2000) Essays on Production Performance Assessment. UES 531. PhD thesis

Bergkvist, Erik (2001) Freight Transportation. Valuation of Time and Forecasting of Flows. UES 549. PhD thesis

Rudholm, Niklas (2001) The Swedish Pharmaceuticals Market - Essays on Entry, Competition and Antibiotic Resistance. UES 552. PhD thesis

Lundberg, Johan (2001) Local Government Expenditures and Regional Growth in Sweden. UES 554. PhD thesis

Lundberg, Sofia (2001) Going Once, Going Twice, SOLD! The Economics of Past and Present Public Procurement in Sweden. UES 557. PhD thesis

Eliasson, Kent (2001) University Enrollment and Geographical Mobility: The Case of Sweden. UES 558. PhLic thesis

Samakovlis, Eva (2001) Economics of Paper Recycling. Efficiency, policies, and substitution possibilities. UES 563. PhD thesis

Daunfeldt, Sven-Olov (2001) Essays on Intra-Household Allocation and Policy Regime Shifts. UES 570. PhD thesis

Hellström, Jörgen (2002) Count Data Modelling and Tourism Demand. UES 584. PhD thesis

Andersson, Linda (2002) Essays on Job Turnover, Productivity and State-Local Finance. UES 586. PhD thesis

Rashid, Saman (2002) Invandrarinkomster, förvärvsdeltagande och familj. UES 588. PhLic thesis

Hanes, Niklas (2003) Empirical Studies in Local Public Finance: Spillovers, Amalgamations, and Tactical Redistributions. UES 604. PhD thesis

Stenberg, Anders (2003) An Evaluation of the Adult Education Initiative Relative Labor Market Training. UES 609. PhD thesis

Stage, Jesper (2003) Mixing Oil and Water. Studies of the Namibian Economy. UES 611. PhD thesis

Marklund, Per-Olov (2004) Essays on Productive Efficiency, Shadow Prices , and Human Capital. UES 621. PhD thesis

Rashid, Saman (2004) Immigrants' Income and Family Migration. UES 625. PhD thesis

Sandberg, Krister (2004) Hedonic Prices, Economic Growth, and Spatial Dependence. UES 631. PhD thesis

Sjöström, Magnus (2004) Factor Demand and Market Power. UES 633. PhD thesis

Nilsson, William (2005) Equality of Opportunity, Heterogeneity and Poverty. UES 652. PhD thesis

Quoreshi, Shahiduzzaman (2005) Modelling High Frequency Financial Count Data. UES 656. Ph Lic thesis

Ankarhem, Mattias (2005) Bioenergy, Pollution, and Economic Growth. UES 661. PhD thesis

Quoreshi, Shahiduzzaman (2006) Time Series Modelling of High Frequency Stock Transaction Data. UES 675. PhD thesis

Ghalwash, Tarek (2006) Income, Energy Taxation, and the Environment. An Econometric Analysis. UES 678. PhD thesis

Westerberg, Thomas (2006) Two Papers on Fertility – The Case of Sweden. UES683. Ph Lic thesis

Simonsen, Ola (2006) Stock Data, Trade Durations, and Limit Order Book Information. UES 689. PhD thesis