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:

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

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

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

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

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

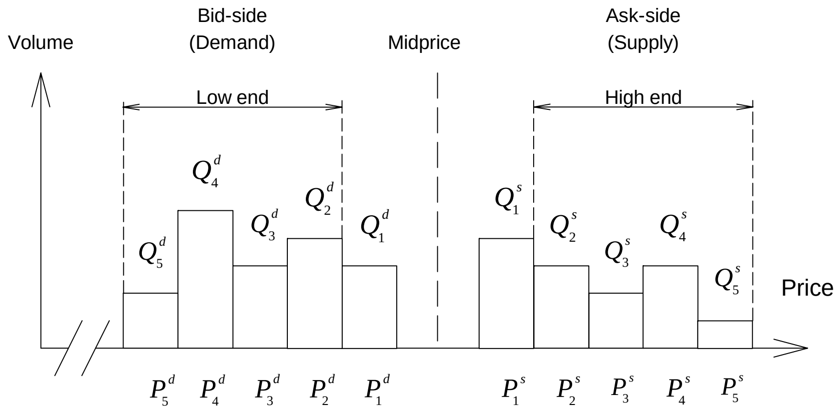

Figure 1: Illustration of the limit order book.

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

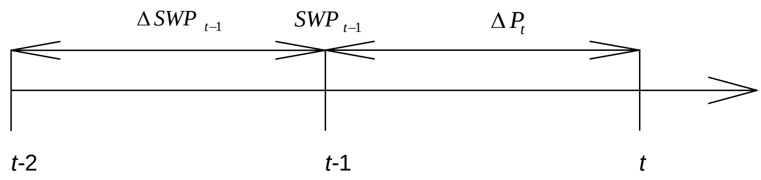

Figure 2: The structure of the order book variables.

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.

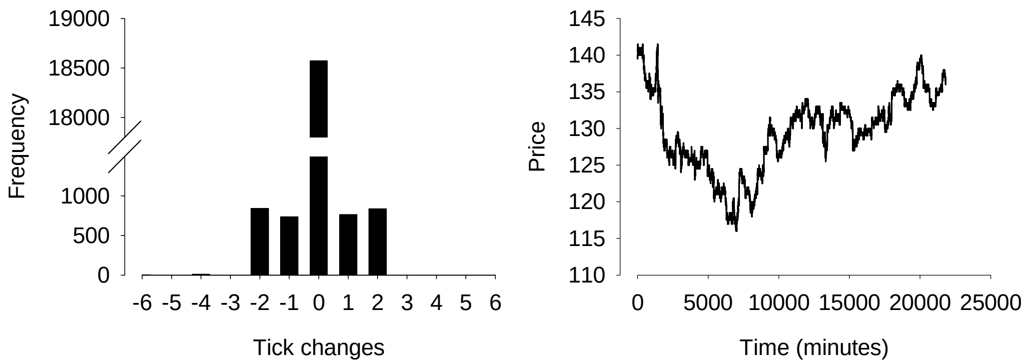

Figure 3: Price change in number of ticks and the price serie for the Nokia share.

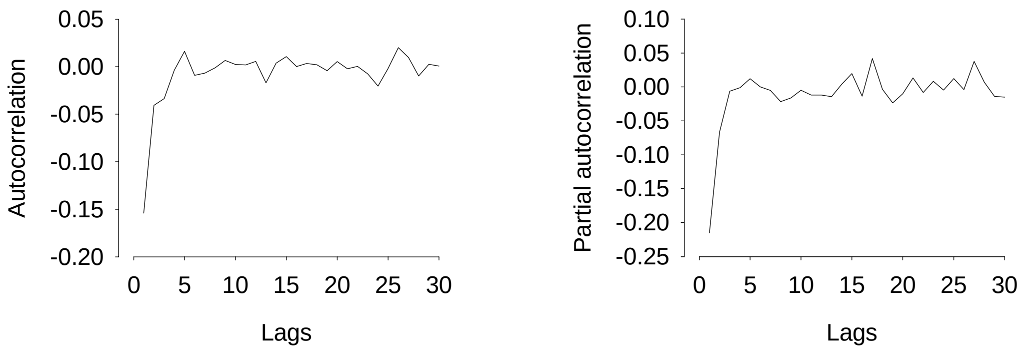

Figure 4: Autocorrelation and partial autocorrelation functions for price tick changes in the Nokia share.

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.