Predicting Stock Price Movements: An Ordered Probit Analysis on the Australian Stock Market

Joey Wenling Yang+

Financial Studies Discipline
UWA Business School
The University of Western Australia
35 Stirling Hwy Crawley WA 6009 Australia
Ph: +61 8 6488 2916
Fax: +61 8 6488 1047
E-mail: jyang@biz.uwa.edu.au

Abstract

The present paper examines the conditional distribution of intra-day stock prices and predicts the direction of the next price change in an Ordered-Probit-GARCH framework that accounts for the discreteness of prices. The analysis also incorporates the endogeneity of the time between trades in an ACD model. Other elements considered include depth, spread, trade imbalance, etc. The results show that all variables are significant with trade imbalance and standardized durations having positive effect on the probability of price changes. The in-sample and out-of-sample forecasting analyses reveal that in 80% cases the system successfully predicts the direction of the consequent price change.

Key words: Ordered Probit; Autoregressive Conditional Duration; Trade Imbalance;

Jel classification: C52; G1

+ The author is greatly indebted to Security Industry Research Centre of Asia-Pacific and Capital Markets CRC for providing data and ABN AMRO Sydney office for financial support of this study. The editor and anonymous referee get special thanks.

Predicting Stock Price Movements: An Ordered Probit Analysis on the Australian Stock Market

I Introduction

The price dynamics is long recognized as an important part in microstructure research. It has not only received substantial interest in its investigation, but also become an imperative framework for other types financial markets analyses to be based on (Barclay and Litzenberger (1988), Almgren and Chriss (1998) and Bertsimas and Lo (1998)). In an empirical analysis, this paper is concerned with examining intra-day stock prices and predicting the direction of price movements in an ordered probit-GARCH system of conditional mean and variance of quoted price changes and attributes associated with trades such as market liquidity measured by spread and depth, trade indicator, volume, trade imbalance, market condition, and standardized transaction durations derived from an autoregressive conditional duration (ACD) model. The estimates are then generalised in both in-sample and out-of-sample forecasts.

The present study accounts for the discreetness of intra-day price series. Previously stock prices were modelled by many competing specifications, from the simple random walk or Brownian motion (Cho and Frees (1988), to the Vector Autoregressive Models (Hasbrouck (1991)), and other more complicated non-linear specifications. One distinct feature of price series is that it is not continuous. In other words, it is quoted in increments of eighth of a dollar on the NYSE, and in one cent or half of a cent on the

Australian Stock Exchange (ASX). This feature of price series is however not captured by the commonly used stochastic processes with continuous state spaces. The outcome of this mis-specification is even worsened for high-frequency intra-day transaction data[i]. To rectify this problem, in this paper we undertake an ordered probit analysis proposed by Aitchinson and Silvery (1957) and developed by Gurland, Lee and Dahm (1960) to investigate the intra-day stock price movements for individual stocks listed on the ASX. By its specification, the ordered probit model has the power to account for the discreteness in dependent variables.

The present paper extends the published literature on the market attributes that affect price changes. In most microstructure studies, the commonly used market elements associated with trades to detect consequent price changes include volume, trade indicator, spread, and index return (e.g., Hausman *et al* (1992); Hasbrouck (1987, 1991)). However, the results of the present study suggest that the depth that measures the width of market liquidity and the trade imbalance are also substantially explanatory to price movements. The use of depth is motivated from the close relation of liquidity and price process. The inclusion of trade imbalance is originated from the argument of (Fetcher (1995), Fishman and Longstaff (1992), etc. that price changes are affected by the information about the sequence of trades, part of which can be manifested in the imbalance of past trades.

This paper also contributes to the literature on the role of inter-trade arrival time. Recent findings suggest that the time between trades is not exogenous, but dependent upon other trade variables (Diamond and Verrecchia (1987); Easley and O'Hara (1992) or its past process (Dufour and Engle (2000), Engle and Lange(2001)). If the time and price are determined simultaneously, the parameter estimates are generally inconsistent. In view of this, we consider using standardized durations that are derived from the decomposition of the time between two consecutive trades in an Autoregressive Conditional Duration (ACD) specification. The standardized duration is thereby deemed exogenous to price changes.

The results from the forecasts of the present paper provide informative trading guidance to stockbrokers as to the direction of stock price movements. The out-of-sample forecasts indicate an average 80% of correctness, and this percentage is even higher for frequent traded stocks. The distribution of the actual and estimated cases reveals that even in the case of mis-prediction, the risk of adverse selection is ruled out.

The remainder of this paper is as follows: section II provides a review of the literature, section III is concerned with model specification; section III provides data description and estimation results and the paper concludes in section IV.

II Literature Review

The ordered probit modelling is widely applied in social economics where the dependent variable is naturally ordered with finite number of values. In an ordered probit framework

O'Donnell and Connor (1996) show that variations in the attributes of road users can lead to variations in the probabilities of sustaining different levels of injury in motor vehicle accidents. This technique is also gaining mounting popularity in finance. Bolleslev and Melvin (1994) find that the size of the bid-ask spread is positively related to the underlying exchange rate uncertainty in an ordered probit analysis that captures the discreteness in the spread distribution. Haan and Hinloopen (2003) use the ordered probit analysis to estimate a corporation's preference in choosing incremental financing means among internal finance, bank loans, bond issues and share issues conditional on a number of firm specific explanatory variables. The discreteness of transaction prices in financial markets is recognized and accounted for by Hausman, Lo and MacKinley (1992) and Flecture (1995). Hausman, Lo and MacKinley (1992) use an ordered probit model to estimate the conditional distribution of trade-to-trade price changes of NYSE stocks on a number of market attributes.

As an important determinant in the price formation and market behaviour, liquidity has received extensive research in the social science literature. The study of liquidity goes back to that of Keynes (1930) and Hicks (1962) who chose 'market liquidity' as a subject in economics and relate it to the 'future volatility of market prices' or the 'possibility of immediate execution of a transaction'. In the market microstructure theory, liquidity is considered with factors such as the existence of adverse selection effects due to information asymmetry and the price impact of trades (see Admati and Pfleiderer (1988), Chowdhry and Nanda (1991) and Dufour and Engle (1999)). Over the years the complexity of liquidity is recognized by a number of authors including Kyle (1985), Harris (1990), O'Hara (1995), Seppi (1997) and Muranaga and Shimizu (1999), etc. and is identified from at least two main dimensions: the width (tightness), which measures the extent of the transaction prices' divergence from the mid-market price irrespective of the level of market prices; and the depth, which denotes the amount of orders on either side of the order book that can be traded without price moving away from the current level.

While some studies use the bid-ask spread as a proxy of liquidity, (Bagehot (1971), Amihud, Mendelson and Wood (1990) and Lee, Mucklow and Ready (1993)), Aitken, Berkman and Mak (2001) adopted the price difference of daily high and low to measure market liquidity. Nevertheless, price changes that show the difference between trade prices and quoted prices or daily highs and lows only captures the width (tightness) dimension of liquidity. In this paper, we attempt to include the effect of liquidity on the consequent direction of price changes with liquidity being measured from both tightness and depth. Generally, tightness that measures how far transaction prices (i.e. bid or ask prices) diverge from the mid-market prices is captured by the bid-ask spread. On the other hand, the dynamics of market depth is more difficult to examine. Depth denotes the volume of trades able to be traded at a particular price level, which can be measured by the amount of orders in the order book. Muranaga and Shimizu (1999) investigated the dynamic of market depth by constructing simulated markets. To give a proxy for the market depth, we follow the definition of it to use the number of shares sitting at the best bid and best ask price respectively before a trade occurs. This measure captures the number of shares in the order book that the market can absorb before moving the price away from the current level of best bid/ask price. It does not account for the latent liquidity which represents the market orders entered at any instant of time, but it provides a dynamic picture of how the volume in order book evolves through trades.

Early theoretical studies on price adjustments to private information develop an informational role for the arrival time between trades (Diamond and Verrechia (1987), Admati and Pfleiderer (1988, 1989), etc.). Based on the assumption that the time between trades is exogenous, Hausman, Lo and MacKinley (1992) and Fletcher (1995) find correlation relationship between the time between trades and price changes. The exogeneity of time however remains a question. Easley and O'Hara (1992) propose that if the time between trades is related to other market covariates, then time is no longer exogenous to the price process. This prediction is confirmed in the empirical study of Dufour and Engle (2000), who find reciprocal interactions of the price, trade and time. In this study, instead of using logged term of transaction durations, we decompose durations to deterministic and stochastic components in the Autoregressive Conditional Duration (ACD) model of Engle and Russell (1998), and derive an exogenous time variable - standardized durations. Engle and Lange (2001) also consider the standardized duration as exogenous and use it as an explanatory variable in their VNET model.

The degree of buy-sell trade imbalance from the sequence of trades is another important element that affects the consequent direction of price changes but is often ignored in the literature. Using neural networks, the recent empirical study of Plerou, Gopikrishnan, Gabaix and Stanley (2002) show that the expected price change is a concave function of trade imbalance, which is defined as the difference in the number buyer and seller initiated trades. This implies that if we have observed a few consecutive buyer-initiated trades that would put pressure on the buy side, then the price is expected to go up as a result, and vice versa for seller initiated trades. In the current paper, we track 30 trades down the past to compute the value of trade imbalance for the current trade.

III Model Specification

1. The ordered probit model

The ordered probit framework employed in this paper is a variation of Hausman, Lo and MacKinley (1992). The dependent variable of an ordered probit model is a latent (unobservable) continuous variable, say dp^*, whose conditional mean is a linear function of a number of explanatory variables. Although dp^* is unobserved, it is related to an observed discrete variable dp, whose value is dependent upon the values that dp^* takes. The ordered probit model requires that the dependent variable should be in the form of integer with natural ordering. In the case of a sequence of quoted prices denoted as P_0, P_1, …, P_k, the price change in dollars from trade $k-1$ to k, i.e. $P_k - P_{k-1}$, is therefore multiplied by 100, that is, $dp_k = 100 \times (P_k - P_{k-1})$ to obtain an integer that denotes price change in ticks. For example, if the price of a given stock rises from A$3.04 to A$3.05, we say it has moved one tick up ($dp_k = 100 \times (3.05 - 3.04) = 1$)). By definition of ordered probit specification, we denote dp_k* as a latent continuous random variable that is determined by a number of explanatory variables such that:

$$dp_k* = X'_k\beta + \varepsilon_k,$$

$$E\,[\varepsilon_k \mid X_k] = 0, \varepsilon_k \text{ i.n.i.d. } N\,(0, \sigma^2_k), \tag{1}$$

Where ε_k's are independently but not identically distributed with a mean of zero and a conditional variance of σ^2, and X_k is a $q\times1$ vector of predetermined explanatory variables that determines the conditional mean of dp_k*. The observed price changes dp_k are relating to the continuous variable dp_k* using following rules:

$$dp_k = \begin{cases} s_1 \text{ if } dp_k^* \in A_1, \\ s_2 \text{ if } dp_k^* \in A_2, \\ \vdots \quad \vdots \\ s_m \text{ if } dp_k^* \in A_m, \end{cases} \tag{2}$$

Where the sets A_k form a partition of the state-space of $Z*_k$. In our current application, s's denote price changes in ticks, *-2*, *-1*, *0*, *1*,… and so on. Thus the state-space partitions A's can be further defied as:

$$\begin{cases} A_1 \equiv (-\infty, \alpha_1] \\ \vdots \\ A_i \equiv (\alpha_{i-1}, \alpha_i] \\ \vdots \\ A_m \equiv (\alpha_m, +\infty) \end{cases} \quad (3).$$

According to Hausman, Lo and MacKinley (1992), the dependence structure of the observed process Z_k is induced by that of Z_k* and the definitions of the A_k's as the following:

$$P(dp_k = s_i \mid dp_{k-1} = s_j) = P(dp_k^* \in A_i \mid dp_{k-1}^* \in A_j)\,, \tag{4}$$

and the conditional distribution of dp_k on X_k is determined by the partition boundaries and the particular distribution of ε_k. The conditional distribution is as following for Gaussian ε_k's,

$$\begin{aligned}
&P(dp_t = s_i \mid X_k) \\
&= P(X_k^{'}\beta + \varepsilon_k \in A_i \mid X_k) \\
&= \begin{cases} P(X_k^{'}\beta + \varepsilon_k \le \alpha_1 \mid X_k) & \textit{if } i = 1, \\ P(\alpha_{i-1} < X_k^{'}\beta + \varepsilon_k \le \alpha_i \mid X_k) & \textit{if } 1 < i < m, \\ P(\alpha_{m-1} < X_k^{'}\beta + \varepsilon_k \mid X_k) & \text{if } i = m, \end{cases}
\end{aligned} \tag{5}$$

$$= \begin{cases} \Phi\left(\dfrac{\alpha_1 - X_k^{'}\beta}{\sigma_k}\right) & \text{if } i = 1, \\ \Phi\left(\dfrac{\alpha_i - X_k^{'}\beta}{\sigma_k}\right) - \Phi\left(\dfrac{\alpha_{i-1} - X_k^{'}\beta}{\sigma_k}\right) & \text{if } 1 < i < m, \\ \Phi\left(\dfrac{\alpha_{m-1} - X_k^{'}\beta}{\sigma_k}\right) & \text{if } i = m, \end{cases} \tag{6}$$

where $\Phi(.)$ is the standard normal cumulative distribution function. Equation 5 and 6 show that the probability of a particular observed price change is determined by the location of the conditional mean X_k', relative to the partition boundaries. For a given conditional mean, a shift in the boundaries will change the probabilities of observing the initial states. On the other hand, given the partition boundaries, a higher conditional mean suggests a higher probability of observing a more extreme positive state. Therefore, by allowing the data to decide the appropriate partition boundaries, i.e. the α's, the β coefficients of the conditional mean, and the conditional variance σ^2_k in a log-likelihood function as shown below, the ordered probit model can capture the relation between the observed discrete price changes dp_k and the unobserved continuous process dp_k* as a function of a number of financial market attributes X_k.

$$L(dp \mid X) = \sum_{k=1}^{n} \left\{ \begin{array}{l} Y_{1k} \cdot \log \Phi\left(\dfrac{\alpha_1 - X_k^{'}\beta}{\sigma_k}\right) + \sum_{i=2}^{m-1} \{ Y_{ik} \cdot \log\left[\Phi\left(\dfrac{\alpha_i - X_k^{'}\beta}{\sigma_k}\right) - \Phi\left(\dfrac{\alpha_{i-1} - X_k^{'}\beta}{\sigma_k}\right)\right]\} \\ + Y_{mk} \cdot \log\left[1 - \Phi\left(\dfrac{\alpha_{m-1} - X_k^{'}\beta}{\sigma_k}\right)\right] \end{array} \right\} \tag{7}$$

Recall that the residual series ε_k from the estimation is not identically distributed with a time-varying conditional variance of σ_k^2. A GARCH (2,2) specification is then applied to accommodate this heteroscedasticity in residuals. MacKinley, Lo and Hausman (1992) find that σ_k^2's depend on the time between trades and the trade indicator. In this context, the dependent variable dp_k* is expected to be fully explained by the explanatory variables, so the residual series is independent of the explanatory variables.

2. *The empirical specification*

First of all, the number of states, *m*, needs to be chosen for the ordered probit model. As this paper is concerned with predicting the probability of whether the direction of the next price will rise, fall or stay the same, we set m=3 to represent the three states of price changes. In particular, all negative price changes starting from one tick downwards are grouped together into a common event that is denoted by $dp_k = -1$; all price rises starting from one tick upwards are grouped together into a common event that is denoted by $dp_k = +1$; and unchanged prices are denoted as $dp_k = 0$.

Then the dependent variable, the price change, needs to be defined. To eliminate unnecessary autocorrelation and volatility in price series due to price reversals between bid and ask prices, which is usually detected in market microstructure studies, we consider using quote revisions, i.e. changes in quote prices, to measure price movements instead of transaction price changes or mid-point price changes. In particular, the quoted bid revisions is used since a risk-averse investor is more sensitive to adverse changes in the bid price than the ask price. Price changes from one transaction to another that do not result in a change in quoted bid price are not counted in.

It should be noted that the distribution function of the price series is accounted for in the model specification. By shifting the boundaries, the ordered probit can fit other arbitrary multinomial distributions as well as normal distribution. This implies that our estimating results are immune to the underlying distribution functions of the price series.

3. The ACD model for the time between trades

In literature it is commonly found that the time between two consecutive trades, or the transaction duration, is highly irregularly spaced. The Autoregressive Conditional Duration (ACD) approach taken by Engle and Russell (1998) to model this irregularly spaced transaction duration is based on its following a conditional point process. A point process is said to evolve with after-effect and be conditionally orderly when the current arrival rate is dependent upon the times of the prior transactions. The ACD is a type of point process which is suited for modelling characteristics of clustering and over-dispersion in time series. Engle and Russell (1998) suggest a description of such a process in terms of the intensity function conditional on all available past arrival rates. In other words, the conditional intensity function is considered as the conditional probability of the next transaction occurring at τ, being conditioned on the transaction times of previous trades over the interval $[\tau_0, \tau)$.

If the sequence of times of each transaction's occurrence is denoted as $\{\tau_1, \tau_2, \dots\}$ with $\tau_1 < \tau_2 < \dots < \tau_k < \dots$, we can express the duration between two consecutive transactions that occurs at time τ_k and τ_{k-1} as $x_k = \tau_k - \tau_{k-1}$. Following Engle and Russell (1998), we first remove the deterministic diurnal component Φ_{k-1} of arrival times and consider the stochastic component of durations that are diurnally adjusted, $\tilde{x}_k = x_k / \Phi_{k-1}$. Then a linear ACD($p$, q) model parameterises the k^{th} durational conditional mean, $E(\tilde{x}_k \mid \tilde{x}_{k-1}, \dots, \tilde{x}_1) = \psi_k$, in a ARMA-type specification

$$\psi_k = \omega + \sum_{i=1}^{p} \alpha_i \psi_{k-i} + \sum_{j=1}^{q} \beta_j \tilde{x}_{k-j} \tag{8}$$

with ω > 0, > 0, i = 1, 2, …N. If we consider the simplest ACD(1, 1) model with parameters α and β only, the unconditional expectation (μ) and variance (σ^2) of the durations are

$$\begin{aligned} \mu &= E(x_k) = \frac{\omega}{1-\alpha-\beta} \\ \sigma^2 &= \frac{1-2\alpha\beta-\beta^2}{1-(\alpha+\beta)^2-\alpha^2}. \end{aligned} \tag{9}$$

With proof provided in Engle and Russell (1995), it is shown in the equation (9) that σ is greater than μ whenever α > 0, implying that the model can account for over dispersion, which is commonly observed in duration series. It is assumed that the standardized durations computed from conditional and unconditional durations,

$$\varepsilon_k = \tilde{x}_k / \psi_k, \tag{10}$$

is independent and identically distributed (i.d.d.) for all k's. This assumption implies that all temporal dependence in the duration series is captured by the defined mean function. Finally, as far as the distribution function of durations is concerned, we consider the exponential distribution, which is recognized as the reference distribution for duration data as equivalent to the normal distribution for real-value data. So the conditional density function of adjusted durations, $g(\tilde{x}_k)$, on exponential distribution is given by

$$g(\tilde{x}_k) = \frac{1}{\psi_k} exp(-\frac{\tilde{x}_k}{\psi_k}) \tag{11}$$

As the ACD models resemble GARCH models in many properties, a quasi-maximum likelihood approach used for estimating GARCH parameters is also employed to estimate ACD parameters. Given the conditional density function, it is straightforward to derive the log-likelihood function

$$\text{Ln}\, L = -\sum_{k=1}^{N}\left(\log(\frac{1}{\psi_k}) - \frac{x_k}{\psi_k} \right) \qquad (12)$$

and estimate the parameters by maximizing this function.

IV Data and Results

1. Data and Variables

The data sets used in this study are sourced from Security Industry Research Centre of Asia-Pacific (SIRCA), who maintains a database called SMARTS that possesses detailed transaction information at order-level for all stocks listed on the Australian Stock Exchange. To construct the sample, three representative stocks from differing industry sectors are selected such that these three stocks are distinctively different in their level of market liquidity. The sample period is chosen from 1st April 2002 to 31st July 2002 when there are no significant structural changes in these firms. For each transaction, our sample data contain the following information: the date, time, size, prevailing bid and ask price, spread, trade indicator, and the index level and the. In addition, the information of the market depth that shows the number of shares at each level of price immediately before the transaction occurs is extracted from the order book for each stock.

It is noticed in practice that when multiple trades are executed in one lot at the same time, there is market overlapping, which is also called continuous single/double auction period that induces temporary negative bid/ask spread. As soon as the execution of these multiple trades is completed, the market overlapping disappears and the spread gets back to its normal form of displaying the gap between the best bid price at which a buyer is will to offer and the best ask price at which a seller is willing to give up his holdings. To avoid the effect of negative spread that does not have economic meaning, multiple trades that are executed in one lot at the same time are aggregated into one single trade and the spread prevailing immediately before the execution of the first trade in the trade-lot is used. This aggregation involves matching order data on the order book with trade data for every transaction.

Another effect that could affect the estimation results is the abnormal price changes at the opening of the market due to overnight arrival of new information, as well as at the closing of the market when fund managers and stock brokers trade aggressively to achieve the VWAP (volume weighted average price) of the day or to close out their outstanding positions. For this reason, only trading and order information from normal trading hours, i.e. between 10:40am and 15:30pm, are included in the sample[ii]. In addition, price changes are adjusted for date changes. The first observation of the price change at the start of each day is set to zero so that the price change of today does not depend on yesterday's last price. Similarly, the first 30 observations of the trade imbalance (TIB) variable is also set to zero as it is unrealistic to calculate today's trade imbalance from trades that occurred yesterday.

In an attempt to investigate how the direction of price movements is affected by the sequence of trades, we consider an ordered probit model that allows us to include explanatory variables associated with trades. These market attributes are defined in the following:

dp_k: the price change at trade k from trade $k\text{-}1$. As for all three stocks the minimum price change allowed is 1¢, multiplying price difference by 100 gives price change in ticks (cents) as an integer,

$$dpk = 100 \times (Pk - Pk\text{-}1); \tag{13}$$

$Sprd_{k-1}$: the bid/ask spread immediately before trade k occurs. It is calculated in units of cents;

$LBBV_{k-1}$: the natural logarithm of the number of shares at the best bid price immediately before trade k occurs. 1 share is added to each $LBBV_{k-1}$ so the logged value of zero volume at the best bid also returns a zero,

$$LBBV_{k-1} = Ln\ (1+ LBBV_{k-1}); \tag{14}$$

$LBAV_{k-1}$: the natural logarithm of the number of shares at the best ask price immediately before trade k occurs. 1 share is added to each $LBAV_{k-1}$ so the logged value of zero volume at the best ask also returns a zero,

$$LBAV_{k-1} = Ln\ (1 + LBAV_{k-1}); \tag{15}$$

ε_{k-1}: standardized transaction duration estimated in an ACD(2,2) model from diurnally adjusted conditional and unconditional durations in equation (10);

$LVol_{k-1}$: the natural logarithm of the size of $(k-1)th$ trade;

TI_{k-1}: trade indicator of $(k-1)th$ trade, TI_{k-1} = 1 if it is a buyer initiated trade, TI_{k-1}=-1 if it is a seller initiated trade and TI_{k-1}=0 if it is other types of trades such as crossings;

TIB_{k-1}: the trade imbalance variable, calculated as the number of buyer initiated trades as a percentage of the total trading volume in the past 30 trades,

$$TIB_{k-1} = \frac{\sum_{j=1}^{30}(TI_{(k-1)-j} \times Vol_{(k-1)-j})}{\sum_{j=1}^{30} Vol_{(k-1)-j}}; \tag{16}$$

ΔIDX_{k-1}: return on index for the k-1th trade calculated from

$$\Delta IDX_{k-1} = 100 \times Ln(\ IDX_{k-1} - IDX_{k-2}\ ). \tag{17}$$

An ordered probit model that includes all above explanatory variables is given by the following expression:

$$\begin{aligned} dp_k^* &= c_1 + c_2 dp_{k-1} + c_3 dp_{k-2} + c_4 dp_{k-3} + c_5 TI_{k-1} + c_6 Sprd_{k-1} \\ &\quad + c_7 LBAV_{k-1} + c_8 LBBV_{k-1} + c_9 TIB_{k-1} + c_{10}\varepsilon_{k-1} \\ &\quad + c_{11} LVOL_{k-1} + c_{12}\Delta IDX_{k-1} \end{aligned}, \tag{18}$$

where the standardized transaction duration, ε_{k-1}, is estimated from an ACD(2,2) specification

$$\begin{aligned} \psi_k &= \omega + \alpha_1 \psi_{k-1} + \alpha_2 \psi_{k-2} + \beta_1 \tilde{x}_{k-1} \beta_2 \tilde{x}_{k-2} \\ \varepsilon_k &= \tilde{x}_k / \psi_k \end{aligned} \tag{19}$$

To accommodate the heteroscedasticity in the conditional variance of residuals as in Hausman, Lo and MacKinley (1992), we consider a GARCH(2,2) specification

$$h_k^2 = \upsilon_1 + \theta_1 h_{k-1}^2 + \theta_2 h_{k-2}^2 + \delta_1 \varepsilon_{k-1}^2 + \delta_2 \varepsilon_{k-2}^2 . \tag{20}$$

There are altogether 22 parameters in the system and they are estimated using maximum likelihood method.

2. Sample statistics

To give a general picture of the data sets we have, Table 1 presents the summary statistics of all variables used in the ordered probit model. The price levels of these three stocks range from A$4.45 for TLS to A$32.45 for CBA, showing a wide dispersion on Australian stock market. The degree of liquidity that is measured by its tightness and depth for these stocks can be compared using statistics in Table 1. It is first noted that the spread that measures the tightness of liquidity has the smallest values of mean, standard deviation and the range between the maximum and minimum for TLS, showing a more liquid market for TLS than BHP and CBA. Consistently, the measure of market depth also show that the mean of the number of shares at the best bid and ask prices for TLS is the greatest of the three, implying a deeper market for TLS than the other two stocks. If liquidity is alternatively compared by the costlessness of executing a certain number of trades, with the largest mean trade size of 89,842 incurring smallest degree of (adverse) price deviation of 32.7%, TLS is the most costless stock of the three, followed by BHP and then CBA.

The buy/sell trade indicator that differentiates buyer-initiated trades from seller-initiated ones is provided in the original transaction database, where an order initially entered by a trader to buy shares and then executed is classified as a buyer-initiated trade, and thus has a value of $TI_k = 1$; an order initially entered by a trader to sell shares and then executed is classified as a seller-initiated trade, and thus has a value of $TI_k = -1$; and the rest types of trades are undefined and thus has a value of $TI_k = 0$.

By definition (Equation (16)), the value of the trade imbalance from the last 30 trades, TIB_k, ranges between –1, when all past 30 trades are seller-initiated trades, and 1, when all past 30 trades are buyer-initiated trades. Zero in between –1 and 1 differentiates cases where there are more seller-initiated trades than buyer-initiated trades ($TIB_k < 0$) and where there are more buyer-initiated trades than seller-initiated trades ($TIB_k > 0$) in the past 30 trades. For example, the mean of TIB_k for TLS and BHP are 0.03 and 0.04, respectively, indicating that there are on average more buyer-initiate traders than seller-initiated trades in these two stocks, which is consistent with the finding that the overall buyer initiated trades accounts for a greater percentage, 47.2% and 57.7%, than the seller initiated trades for these two stocks. This is however not the case for CBA that is shown to have more seller-initiated traders over the sample period.

Table 1. Summary Statistics of the Variables

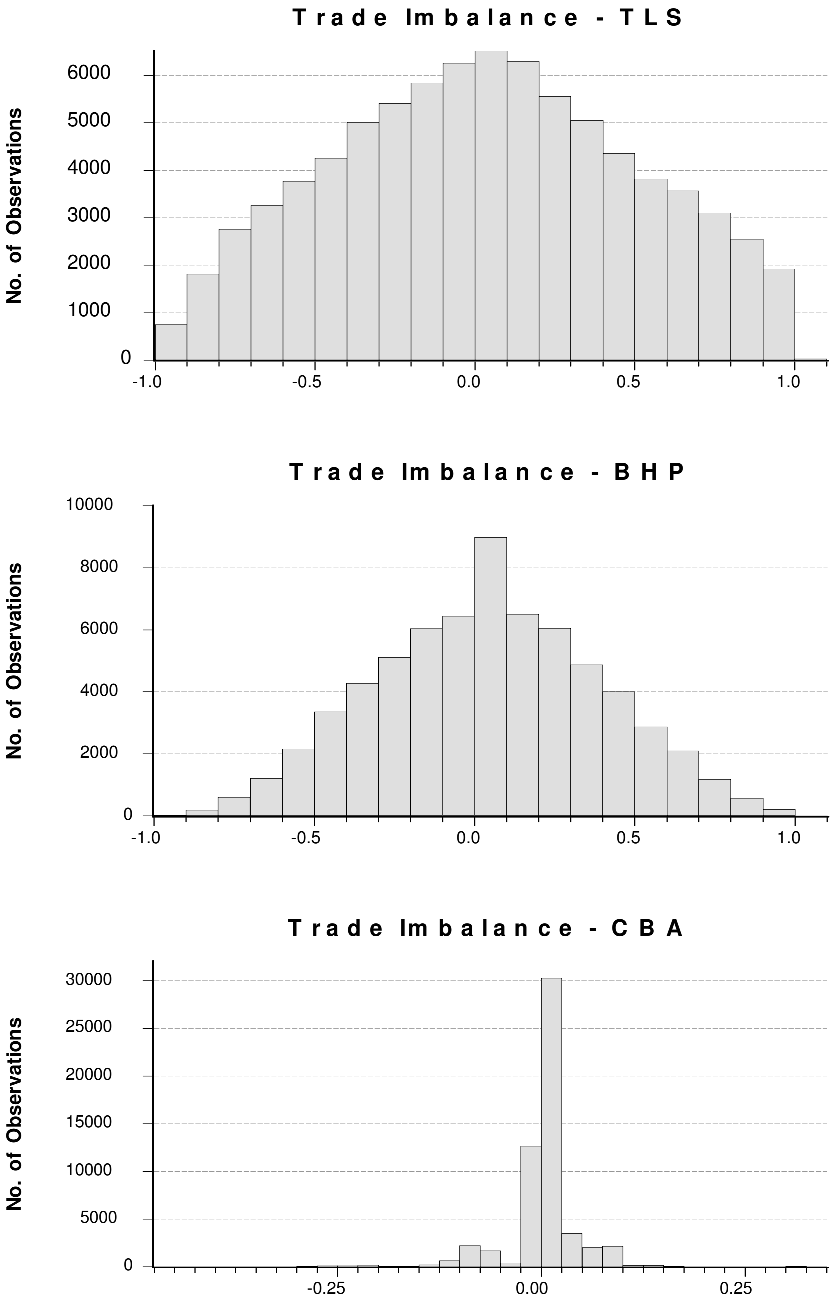
	TLS	BHP	CBA
High price (A$)	5.42	11.97	32.45
Low price (A$)	4.45	10.49	29.01
Net Price change, dp_k (Ticks)			
Mean	0.0036	-0.0017	-0.01975
Maximum	2	6	13
Minimum	-2	-5	-10
Std. Dev.	32.70%	67.91%	161.91%
Bid-Ask Spread, $Spread_{k-1}$ (cent)			
Mean	1.0035	1.0493	1.6628
Maximum	2	6	24
Minimum	1	1	1
Std. Dev.	0.0591	0.2336	1.0694
No. Shares @ Best Bid, BBV_{k-1}			
Mean	716,406.00	44,097.01	6,215.68
Std. Dev.	701,544.00	53,266.08	8,526.28
No. shares @ Best Ask, BAV_{k-1}			
Mean	657,331.60	41,077.65	6,504.01
Std. Dev.	521,708.50	44,857.20	7,333.97
Time between Trades, Δt_{k-1}			
Mean	17.61	21.46	25.74
Std. Dev.	25.52	34.06	39.11
Trade Size, Vol_{k-1}			
Mean	9,842.90	7,493.58	2,483.14
Std. Dev.	32,631.36	14,779.84	8,484.63
Trade Imbalance, TIB_{k-1}			
Mean	0.02976	0.04074	-0.0021
Std. Dev.	0.4676	0.3491	0.0432
Return on Index, $\Delta INDEX_{k-1}$			
Mean	-0.0001%	-0.0002%	-0.0002%
Std. Dev.	0.0207%	0.0227%	0.0247%
% Buyer Initiated Trades	47.24%	57.70%	48.31%
% Seller Initiated Trades	44.52%	37.83%	48.90%

Notes: the above table presents descriptive statistics to all variables to be used in the model for three representative Australian stocks: TLS (Telstra), BHP (BHP Billiton Ltd) and CBA (Commonwealth Bank of Australia). High/low prices are the highest and lowest price quoted in the whole sample period. dp_k is the price change in ticks; $Spread_{k-1}$ is the bid/ask spread immediately before trade k occurs; BBV_{k-1} (BAV_{k-1}) is the number of shares at the best bid (ask) price immediately before trade k occurs; Δt_{k-1} is the time in seconds between trade k and trade k-1; Vol_{k-i} is i lags of the trade size from (k-i)th the trade $i = 1, = 1, 2, \ldots$; TIB_{k-1} is the trade imbalance variable, calculated as the number of buyer initiated trades as a percentage of the total trading volume in the past 30 trades; ΔIDX_{k-1} is Return on index for the k-1th trade; % Buyer/Seller Initiated Trades is calculated using the trade indicator variable, TI_k, is the trade indicator of the kth trade, TI_k = 1 if it is a buyer initiated trade, TI_k=-1 if it is a seller initiated trade and TI_k=0 if it is other types of trade, such as crossings.

To further examine the extent to which the sequence of trades are imbalanced for these stocks, the histograms of TIB_k are presented in figure 1. Generally, a *TIB* value of close to –1 and 1 indicates an extreme imbalance in trades due to consecutive sells or buys of shares in large numbers, and a *TIB* value of close to zero indicates the absence of extreme imbalance in trades as buys are offset by sells along the course of 30 trades. Figure 1 shows that CBA has the fewest trade imbalances over the sample period, as there are no observations falling out of the range of –0.25 and +0.25. On the contrary, extreme trade imbalances occur mostly in TLS, as it is shown to still have approximately 4500 observations of extreme buys at $0.9 < TIB < 1$ and 2600 observations of extreme sells at $-1 < TIB < -0.9$. The values of *TIB* at these ranges for BHP are however close to zero. Comparing this with the range of price changes between its maximum and minimum value for these stocks, the finding coincides with our early conclusion from Table 1: the market for TLS is deeper than that for the other two so that large numbers of trades on one side can be absorbed by the market without inducing significant adverse price changes.

Consistent with findings in previous research that there is less likelihood of non-stationarity in high frequency data points, all variables considered in this context are stationary at 99% even when 50 lags are used in an Augmented Dickey-Fuller test of unit roots. For the purpose of saving space, the results of unit root tests are not presented herein.

Figure 1 Histogram of Trade Imbalance



Notes: the above figure depicts the histogram of Trade Imbalance (*TIB*) variable for three sample stocks: TLS (Telstra), BHP (BHP Billiton Ltd) and CBA (Commonwealth Bank of Australia). TIB_{k-1} is the trade imbalance variable ranging from –1 to 1, and it is calculated as the number of buyer initiated trades as a percentage of the total trading volume in the past 30 trades, see Eq.(13).

3. Model Estimation Results

The maximum likelihood estimates of the ACD(2,2) model, ordered probit model and the GARCH(2,2) are computed using BHHH algorithm proposed by Berndt, Hall, Hall and Hausman (1974). The coefficients estimated over the first 16 weeks of the sample period for all stocks are presented in Table 2. The last seven days 25 July to 31 July is left for an out-of-sample forecast in the next sub-section.

Table 2 ACD Estimation Results

	TLS		BHP		CBA	
Mean Equation	*Coeff.*	*t-statistic*	*Coeff.*	*t-statistic*	*Coeff.*	*z-statistic*
Constant	0.090	3.373*	0.042	1.020	0.058	0.952
α_1	-0.058	-0.237	1.297	3.088*	1.100	1.486
α_2	0.927	4.030*	-0.345	-0.912	-0.168	-0.248
β_1	0.022	3.849*	0.054	3.319*	0.045	2.252*
β_2	0.022	3.349*	-0.041	-2.247*	-0.027	-1.040

Notes: the above table presents ACD(2,2) estimates for three representative Australian stocks: TLS (Telstra), BHP (BHP Billiton Ltd) and CBA (Commonwealth Bank of Australia). * marks significance at 95%.

Table 2 presents estimated parameters from the ACD (2,2) model in equation (19). It shows evidence that the past behaviour of both conditional and unconditional diurnally adjusted durations is explanatory to the expected durations. The standardized durations are then computed from Equation (10) above. The exogeneity of standardized durations are explored by Dufour and Engle (2000). Using data on NYSE, they find that returns, trades and volume all have feedback effects on standardized durations. In the case of Australian data in this context, a regression of standardized durations on those variables shows that most of the parameters are not statistically significant. Therefore, standardized durations are deemed weak exogenous.

The estimates from the Probit-GARCH system are presented in Table 3. The first panel of Table 3 illustrates estimation coefficients from the ordered probit model. For each coefficient, the z-statistic is used to measure the significance of the coefficients. It is calculated as the estimated mean of the coefficient divided by its asymptotic standard error[iii]. Table 3 shows that the z-statistics are statistically significant for all explanatory variables in the model.

Firstly, we examine the coefficients of the three lags of the dependent variable. It is noted that these coefficients are significant with a negative sign across all stocks, indicating a consequent price reversal from past price changes. For example, given other variables constant, a one tick downwards in TLS from each of the last three trades will increase the conditional mean of dp_k*, by 1.25 (0.71+0.38+0.16) ticks. This negative relation is consistent with precious findings of Hasbrouck (1991) and MacKinglay, Lo and Hausman (1992).

Secondly, the liquidity measure of the bid-ask spread and the logged volume at the best bid and ask prices are all highly significant, confirming with the effect of liquidity on price movements in the literature. In particular, the opposite signs of LBAV and LBBV imply a positive effect of the volume at the best bid price on the conditional mean of dp_k*, meaning that the more the volume queued on the bid side to buy a stock, the greater the probability of price rises; and a negative effect of the volume at the best ask price on the conditional mean of dp_k*, meaning that the more the volume queued on the ask side to sell as stock, the greater the probability of price falls.

Thirdly, the coefficients of the trade imbalance (*TIB*) that is used to measure the degree of buy-sell imbalance are also significant with a correct sign for all sticks. This positive relation between trade imbalance and the price changes means that if the there has been more buyer-initiated trades in the past 30 trades, resulting in a positive value of *TIB*, then it has a greater probability that the consequent price will rise. This is intuitive because more number of buyer-initiated trades put pressure on the buy side that will eventually push the price up. It is vice versa with negative *TIB* when there are more seller-initiated trades and the pressure is placed on the sell side of the stock.

Fourthly, the use of standardized durations does not reduce the significant relationship between the time and quote changes, which is along the lines of Hausman, Lo and MacKinley (1992) and Flecture (1995). However, the standardized duration, ε_{t-1}, is a weakly exogenous variable with a positive effect on price changes. ε_{t-1} is under control of a trader who trades at a certain speed and wants to know how much the bid price will move in that time to trade a certain volume of stocks. The positive relation between the bid price and waiting time reflects the premium the trader pay for immediate sale of his holdings.

Table 3 Ordered Probit Model Estimation Results

	TLS		BHP		CBA	
Mean Equation	*Coeff.*	*z-statistic*	*Coeff.*	*z-statistic*	*Coeff.*	*z-statistic*
dp_{k-1}	-0.669	-22.262	-0.458	-34.259	-0.204	-42.180
dp_{k-2}	-0.312	-10.214	-0.170	-13.524	-0.074	-15.722
dp_{k-3}	-0.025	-0.794	-0.002	-0.131	-0.010	-2.224
TI_{k-1}	1.249	27.674	1.146	68.667	0.891	116.011
Spr_{k-1}	0.580	7.229	0.092	4.183	0.075	16.044
$LBAV_{k-1}$	-0.510	-64.100	-0.441	-88.355	-0.270	-82.045
$LBBV_{k-1}$	0.466	60.484	0.325	65.888	0.187	56.736
TIB_{k-1}	0.093	4.313	0.164	9.262	0.251	2.193
ε_{k-1}	0.387	6.649	0.107	2.448	0.350	9.991
ΔIDX_{k-1}	1.200	2.823	0.471	1.763	0.455	2.081
$LVol_{k-1}$	0.018	3.571	0.084	22.148	0.026	8.171
Partition Boundaries						
$dp_k <0$	(-∞, -4.506]		(-∞, -2.730]		(-∞, -1.823]	
$dp_k =0$	(-4.506, 2.911]		(-2.730, 2.130]		(-1.823, 1.051]	
$dp_k >0$	(2.911, +∞)		(2.130, +∞)		(1.051, +∞)	
No. of Observations	78,062		61,763		60,899	
Pseudo-R^2	46.22%		40.02%		23.70%	
Variance Equation						
Constant	0.000	166.517	0.026	11.169	0.084	24.266
θ_1	0.709	243.235	0.052	17.948	0.070	33.929
θ_2	-0.689	-243.463	0.052	17.954	0.067	32.124
δ_1	1.192	500.992	-0.082	-16.510	-0.118	-20.866
δ_2	-0.206	-88.665	0.918	182.787	0.865	162.129

Note: the above table presents estimation results of price changes in an Ordered Probit-GARCH system for three stocks listed on the ASX: TLS (Telstra), BHP (BHP Billiton Ltd) and CBA (Commonwealth Bank of Australia). The dependent variable dp_k is the price change in ticks; dp_{k-i} is i lags of the dependent variable; Spr_{k-1} is the bid/ask spread immediately before trade k occurs; BBV_{k-1} (BAV_{k-1}) is the number of shares at the best bid (ask) price immediately before trade k occurs; ε_{k-1} is the standardized transaction duration estimated in an ACD(2,2) model from diurnally adjusted conditional and unconditional durations; $LVol_{k-1}$ is the natural logarithm of the size of (k-1)th trade; TIB_{k-1} is the trade imbalance calculated from past 30 trades; ΔIDX_{k-1} is the return on index for (k-1)th trade multiplied by 100; TI_k is trade indicator.Lastly, as is consistent with other empirical studies, the rest of trade related variables, the size of the trade and the trade indicator, and the measure of the market condition, that is, the return on the index, all have exerted a significantly positive effect on the direction of consequent price changes.

The partition boundaries are also computed in the estimation and illustrated in the second panel of Table 3 to partition the differing directions of the price change. Given the three possible direction of price changes, dp_k<0, dp_k =0 and dp_k>0, these boundaries computed from the ordered probit model using the Maximum Likelihood Estimation is used to determine whether the estimate $d\hat{p}_k$ in ticks has a positive, negative or zero value, depending on the value of the estimated continuous variable $d\hat{p}_k^*$ and which of the three partitioning states it falls in.

The coefficients from a GARCH (2, 2) model for the conditional variance of the residual series are included in the bottom panel of Table 3. Considering the large number of observations, the substantially significant parameters of the lagged conditional variance, σ_{k-i}, and squared error terms, ε_{k-i}, confirm our suspect of heteroscedasticity in the residual series.

In summary, all coefficients utilized in the ordered probit specification are found statistically significant, confirming with our initial hypothesis that they have an impact on the consequent movements of price series. The values of the goodness-of-fit measure, the Pseudo-R^2, indicates that the model performs especially well for liquid stocks, as the most liquid stock in the sample, TLS, has the greatest value of the Pseudo-R^2, followed by the less liquid stock, BHP, and then by the least liquid stock, CBA.

The diagnostic test that examines the properties of the residual series is an important part of statistical estimation because it reveals the validity of the estimates. In this context, the serial correlation is checked for the generalized residuals along the lines of MacKinlay, Lo and Hausman (1992) by computing cross correlation coefficients of the generalized residuals with the lagged generalized fitted values. Under the null hypothesis of no autocorrelation in the residual series, ε_k, the theoretical value of this cross correlation should be zero, or very close to zero. Using 20 lags, the cross correlation coefficients of the generalized residuals and lagged generalized fitted values $d\hat{p}_k^*$ are reported in Table 4 below. Ranging from –0.013 to 0.077, the values of the cross correlation coefficients for all three stocks are quite small considering the large number of observations in the sample.

Table 4 Cross-correlation Tests for Autocorrelation in Residuals

Lag Length	**1**	**2**	**3**	**4**	**5**	**10**	**15**	**20**
TLS	-0.007	0.004	0.013	0.023	0.028	0.040	0.034	0.021
BHP	-0.003	0.001	0.013	0.077	0.039	0.005	0.043	0.056
CBA	-0.013	-0.001	0.003	0.021	0.008	0.012	0.014	0.011

Notes: this table presents the cross-autocorrelation coefficients of the generalized residuals with the lagged generalized fitted price changes obtained from the ordered probit estimation for three stocks listed on the ASX, TLS (Telstra), BHP (BHP Bulliton) and CBA (Commonwealth Bank of Australia). The null hypothesis is that the theoretical value of this cross correlation should be zero, or very close to zero. A lag-length of 20 is used.

4. Forecasting Analysis

To examine the forecasting power of the specified ordered probit model, in this sub-section we undertake forecasting analysis that evaluates the in-sample and out-of-sample predicting ability of the system. The forecasting results are reported in Table 5 below. For in-sample forecasts, the last seven days' data within the estimation period from 18 July to 24 July are used, and for out-of-sample forecasts we use another seven days' data that are originally left out of the estimation ranging from 25 July to 31 July. For each trade, we compute the fitted value of $d\hat{p}_k$ from its continuous counterpart $d\hat{p}_k^*$ using estimated parameters and partition boundaries, and then compare the fitted $d\hat{p}_k$ with the actual dp_k. The percentage of correctness is obtained by comparing the difference between the actual number of observations from dp_k and the estimated number of observations from $d\hat{p}_k$. Table 5 also presents the actual and estimated count of observations under the three cases of price falls, no changes and price rises over the in-sample and out-of-sample periods.

As expected, Table 5 shows that the in-sample forecasts perform better for all stocks than the out-of-sample forecasts. However, even in the *ex ante* forecast we are able to achieve an average percentage of correctly predicted counts of 80%, implying a strong forecasting power of our ordered probit system. This modelling specification can thereby provide immediate trading guidance to stock brokers, who would proceed to get hold of the stock the price is going to rise, or sell out his/her holdings if the consequent price is predicted to fall. Furthermore, the forecasting results indicate that there is a greater chance of correct prediction if the stock is relatively liquid. From previous sections we learn that TLS is a more liquid stock than the other two, so the percentage of correct prediction from TLS is remarkably greater than that from BHP and CBA in both in-sample and out-of-sample forecasts.

Table 5 In-Sample and Out-of-Sample Forecasting of Price Change Directions

	TLS		BHP		CBA	
In Sample Prediction from 18/07 to 24/07						
No. of Observations	3,624		4,422		5,881	
% Correct	93.91%		79.17%		75.24%	
Count of Observations	*Actual*	*Estimated*	*Actual*	*Estimated*	*Actual*	*Estimated*
Fall	127	43	544	262	1021	603
No Change	3336	3,502	3295	3,916	3956	4,685
Rise	131	49	553	214	874	563
Out-of-Sample Prediction from 25/07 to 31/07						
No. of Observations	3,818		4,952		5,749	
% Correct	93.48%		78.48%		73.33%	
Count of Observations	*Actual*	*Estimated*	*Actual*	*Estimated*	*Actual*	*Estimated*
Fall	135	26	681	308	1,018	574
No Change	3523	3,729	3572	4,372	3788	4,582
Rise	130	33	669	242	913	563

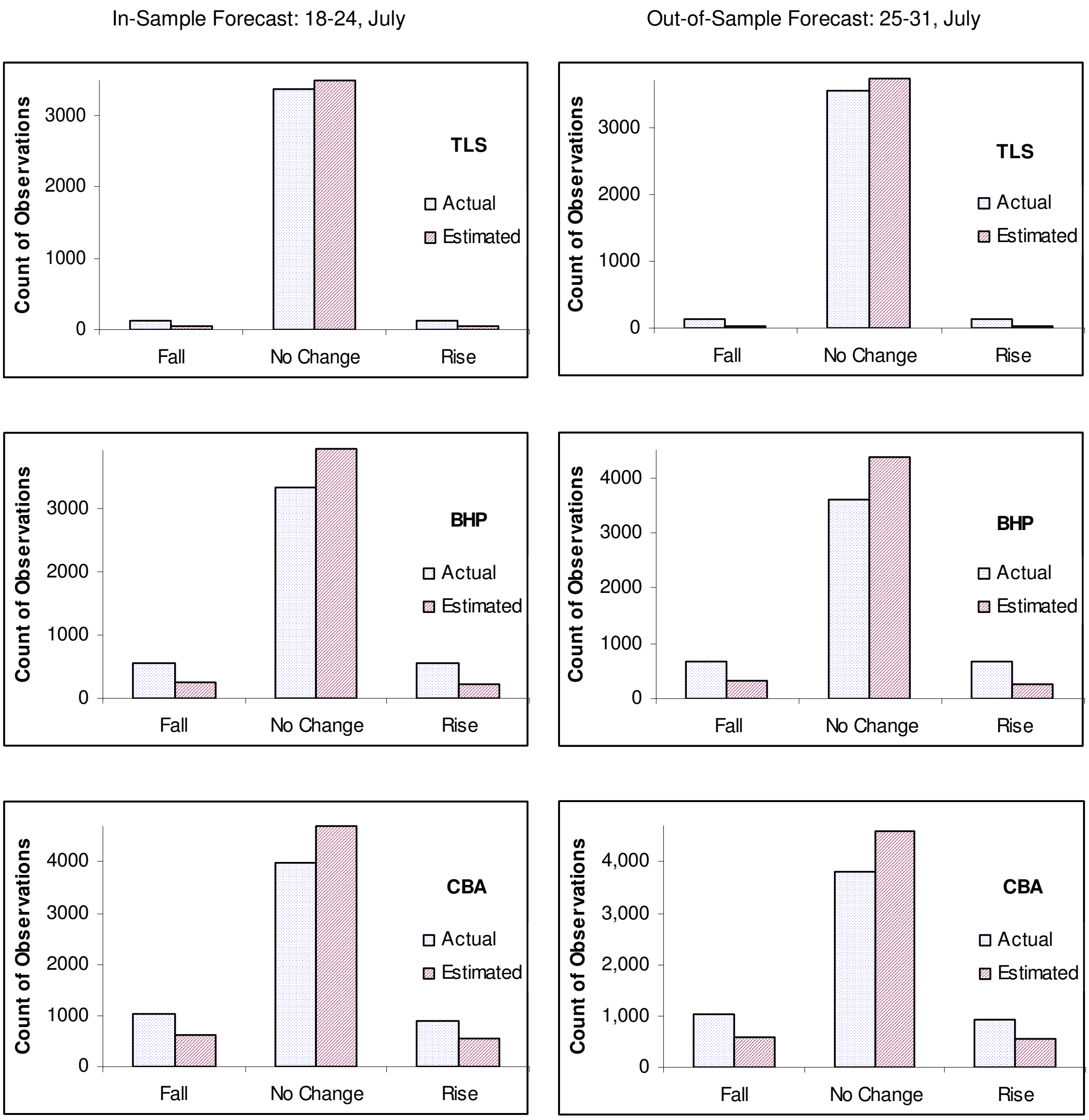
Notes: this table presents in-sample and out-of-sample forecasting count of observations and the percentage of correctness. The actual count of observations are calculated based on dp_k's, and the estimated count of observations are calculated based on the fitted values of $d\hat{p}_k$'s from its continuous counterpart $d\hat{p}_k^*$ using the estimated coefficients of explanatory variables and partition boundaries. % Correct is calculated as the percentage of missed observations on the total number of observations.

It is also noticed that for all stocks, there are approximately 27% chances that the model is unable to predict the direction of price movements correctly. To investigate on the source of this mis- prediction, a comparison of the estimated count of observations with the actual count of observations is plotted by way of column charts in Figure 2 below. The most striking feature Figure 2 reveals is that with both in-sample and out-of-sample forecasts, our model inclines to produce more no-price-change cases than the actual cases where there are virtually price changes. In other words, the model tends to overestimates the *no-change* cases and underestimates the *changed* cases. This means that the mispredicted case occurs when the model predicts *no-changes*, as there may be a change that the model neglects to detect, but when the model predicts *a change*, there is definitely a price change. This rules out the risk of adverse selection that could have been a concern for many investors.

V Conclusion

In an empirical analysis, this paper employs an Ordered Probit-GARCH system to examine the intra-day stock prices and predict the direction of price movements conditional on a variety of market attributes such as past price changes, trade indicator, volume, spread, depth, trade imbalance, index return and standardized durations, for three stocks listed on the ASX. The ordered probit model has the power to account for the discreteness in price series. Apart from the commonly used market attributes in the literature to detect the price process, we incorporate in the depth as an additional measure of liquidity, and the trade imbalance that is a partial representation of the sequence trades. The model also accounts for the endogeneity of the time between trades in an ACD(2,2) specification.

Figure 2 Actual and Estimated Count for In-Sample and Out-of-Sample Forecasts



Notes: this figure depicts the estimated and the actual count of observations from an in-sample (18/7 – 24/7, 2002) forecast and an out-of-sample (25/7 – 31/7, 2002) forecast in column charts for three stocks listed on the ASX: TLS (Telstra), BHP (BHP Billiton Ltd) and CBA (Commonwealth Bank of Australia).

The results show that all independent variables in the ordered probit model are statistically significant for all stocks. In particular, the contrary signs of those depth measures indicate that the volume at the best bid price have a positive effect and the volume at the best ask price have a negative effect on the probability of consequent price change. A positive relation is found between the trade imbalance and the conditional price changes. This suggests a positive value of *TIB* will result in a price rise and a negative TIB will result in a price fall, as more number of buyer-initiated trades in the past put pressure on the buy side of the stock that will eventually push the price up, and on the contrary, more number of seller-initiated trades in the past put pressure on the sell side of the stock that will eventually push the price down. The standardized durations derived from the ACD model is found significant, contributing to the existing informational role of time in the price process. The goodness-of-fit measure of the estimation indicates that the model performs especially well for liquid stocks.

The model is found to have a strong forecasting power in both in-sample and out-of-sample forecasts. Particularly, in an ex ante forecast averagely 80% of the time the model can predict price changes correctly, and the model predicts even better for more liquid stocks. This modelling specification can therefore provide trading guidance to stock brokers, who would proceed to get hold of the stock the price is going to rise, or sell out his/her holdings if the consequent price is predicted to fall. For further research, it would be interesting to extend this methodology to more stocks on other markets.

References

Admati, A. and P. Pfleiderer (1988), ‘A Theory of Intraday Patterns: Volume and Price Variability’, Review of Financial Studies, pp. 30-40.

Admati, A. and P. Pfleiderer (1989), ‘Divide And Conquer: A Theory of Intraday and Day-of-the-week Mean Effects’, Review of Financial Studies, pp. 189-223.

Aitchinson, J. and S. Silvery (1957), ‘The Generalization of Probit Analysis to the Multiple Responses’, Biometrika, pp. 131-140.

Aitken, M., H. Berkman and D. Mak (2001), ‘The Use of Undisclosed Limit Orders on the Australian Stock Exchange’, Journal of Banking and Finance, pp. 1589-1603.

Almgren, R. and N. Chriss (1998), ‘Optimal Liquidation’, Working Paper, University of Chicago, Department of Mathematics.

Amihud, Y., H. Mendelson and R. Wood (1990), ‘Liquidity and the 1987 Stock Market Crash', Journal of Portfolio Management, pp. 65-69.

Bagehot, W. (1971), ‘The Only Game in Town’, Financial Analysts Journal, pp. 12-14.

Barclay, M. and R. Litzenberger (1988), ‘Announcement Effects of New Equity Issues and the Use of Intraday Price Data’, Journal of Financial Economics, pp. 71-99.

Bertsimas, D. and A. Lo (1998), ‘Optimal Control of Execution Costs’, Journal of Financial Markets, pp. 1-50.

Bolleslev, T. and M. Melvin (1994), ‘Bid-Ask Spreads and Volatility in the Foreign Exchange Market: an Empirical Analysis’, Journal of International Economics, pp. 355-372.

Cho, D. and E. Frees (1988), ‘Estimating the Volatility of Stock Prices’, Journal of Finance, pp. 451-466.

Chowdhry, B. and V. Nanda (1991), ‘Multimarket Trading and Market Liquidity’, Review of Financial Studies, pp. 483-511.

Diamond, D. and R. Verrecchia (1987), ‘Constraints on Short-Selling and Asset Price Adjustments to Private Information’, Journal of Financial Economics, pp. 277-311.

Dufour, A. and R. Engle (2000), ‘Time and The Price Impact of a Trade’, Journal of Finance, pp. 2467-2498.

Easley, D., and M. O’Hara (1992), ‘Time and the Process of Security Price Adjustment’, Journal of Finance, pp. 577-605.

Engle, R. and J. Russell (1998), ‘Autoregressive Conditional Duration: a New Model for Irregularly Space Transaction Data’, Econometrica, pp. 1127-1162.

Fishman, M. and F. Longstaff (1992), ‘Dual Trading in Futures Market’, Journal of Finance, pp. 643-671.

Engle, R. and J. Lange (2001), 'Measuring, Forecasting and Explaining Time Varying Liquidity in the Stock Market', Journal of Financial Markets, pp. 113-142.

Fletcher, R. (1995), 'The Role of Information and the Time between Trades: an Empirical Investigation', Journal of Financial Research, pp. 239-260.

Gurland, J., I. Lee and P. Dahm (1960), 'Polychotomous Quantal Response in Biological Assay', Biometrics, pp. 382-398.

Haan, L. and J. Hinloopen (2003), 'Preference Hierarchies for Internal Finance, Bank Loans, Bond, and Share Issues, Evidence for Dutch Firms', Journal of Empirical Finance, pp. 661-681.

Harris, L. (1990), 'Estimation of Stock Price Variances and Serial Covariances from Discrete Observations', Journal of Financial and Quantitative Economics, pp. 291-306.

Hasbrouck, J. (1991), 'Measuring the Information Content of Stock Prices', Journal of Finance, pp. 179-207.

Hausman, J., A. Lo and A. MacKinley (1992), 'An Ordered Probit Analysis of Transaction Stock Prices', Journal of Financial Economics, pp. 319-379.

Kyle, A. (1985), 'Continuous Auctions and Insider Trading', Econometrica, pp. 1315-1335.

Lee, C., B. Mucklow and M. Ready (1993), 'Spreads, Depths, and the Impact of Earnings Information: an Intraday Analysis', Review of Financial Studies, pp. 345-374.

Muranaga, J. and T. Shimizu (1999), 'Market Microstructure and Market Liquidity', mimeo.

O'Donnell, C. and D. Connor (1996), 'Predicting The Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choice', Accid. Anal. and Prev., pp. 739-753.

O'Hara, M. 1995, Market Microstructure Theory, Oxford: Basil Blackwell.

Plerou, V., P. Gopikrishnan, X. Gabaix, and H. Stanley (2002), 'Quantifying Stock Price Response to Demand Fluctuations, Physics Review, E66, 027104-1 - 027104-4, cond-mat 0106657.

Seppi, D. (1997), 'Liquidity Provision with Limit Orders and a Strategic Specialist', Review of Financial Studies, pp. 103-150.

---

[i] See Hausman and Lo and MacKinley (1992).

[ii] Previous papers of Engle and Lange (2001) and Dufour and Engle (2000) also made adjustments to avoid contamination of prices by overnight news arrival.

[iii] *z*-statistic has a null hypothesis of zero estimating coefficient and is asymptotically distributed a normal variate, see MacKinley, Lo and Hausman (1992).