Predicting stock price movements: An ordered probit analysis on the ASX

JOEY WENLING YANG§* and JERRY PARWADA±

§UWA Business School, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia
±Australian School of Business, The University of New South Wales, UNSW Sydney, NSW 2052, Australia

Using stocks from a wide range of industry sectors on the Australian Securities Exchange, this 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 framework. Other elements considered include depth, trade imbalance, and volume. The results show that trade imbalance has a positive effect on the probability of price change. Durations have a negative effect. In-sample and out-of-sample forecasting analyses reveal that in 71% of the cases the system successfully predicts the direction of the subsequent price change.

Keywords: Stock prices, Ordered probit; Autoregressive Conditional Duration; Trade imbalance

[*] Corresponding author. Email: j.parwada@unsw.edu.au

Predicting Stock Price Movements: An Ordered Probit Analysis on the ASX

Using stocks from a wide range of industry sectors on the Australian Securities Exchange, this 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 framework. Other elements considered include depth, trade imbalance, and volume. The results show that trade imbalance has a positive effect on the probability of price change. Durations have a negative effect. In-sample and out-of-sample forecasting analyses reveal that in 71% of the cases the system successfully predicts the direction of the subsequent price change.

Keywords: Stock prices, Ordered probit; Autoregressive Conditional Duration; Trade imbalance

1. Introduction

Security price dynamics are an important part of microstructure research that has become a fundamental framework for a number of financial markets analyses.[1] Earlier theories of information asymmetry propose that private information deduced from trading activities causes market price changes (see, for example, Kyle 1985). Since then investigations into stock price changes have included a variety of market attributes to proxy for information. However, some unique characteristics of stock price series are not accounted for in many prior studies. It is now well recognised that security price series do not follow a continuous path. Ignoring price discreteness potentially causes contamination of results, especially in studies that use intraday prices. In such fine samples discreteness limits the extent of price changes (Hausman *et al.* 1992). This feature, however, is excluded by the commonly used stochastic processes with continuous state spaces. Based on the Australian Securities Exchange (ASX), this paper examines intra-day stock price dynamics by predicting the direction of price movements from market attributes while specifically incorporating characteristics of prices such as discreteness, even spacing of trades, and the conditional mean and variance of price changes. The analysis utilizes an ordered probit model with the power to account for discreteness. We show in our specification that the ordered probit framework can also capture the impact of order flow and market momentum measured by a number of independent variables on price dynamics.

[1] See, for example, Barclay and Litzenberger (1988), Almgren and Chriss (1998) and Bertsimas and Lo (1998).

This study extends the existing literature on market attributes that are related to price changes. Apart from including the variables found significant in other market microstructure studies on transaction data, such as trade size, trade indicator, and bid-ask spread, this paper investigates the explanatory power of depth and trade imbalance, also referred to as order imbalance in studies on quote driven markets.[2]

The use of depth is motivated from the finding of Heflin and Shaw (2005) that depth is a predominant indicator of informed trading when used to scale raw trade size. Thus, the level of information asymmetry should be assessed from trade size relative to the depth measured as the number of shares available for trade at the best bid and ask prices. Inferred from this definition, the intuition is that the consequent price changes should be positively (negatively) correlated with bid (ask) depth.

The inclusion of trade imbalance is derived from the theoretical work of Kyle (1985) that demonstrated the information content of net order flow. Consistent with theoretical model predictions, empirical studies show evidence that order imbalances have a positive impact on returns.[3] We therefore expect to obtain the same relation from our estimations.

We also contribute to the literature on the role of inter-trade arrival time. Past findings suggest that the time between two consecutive trades, referred to as duration here, is not exogenous, but dependent upon other market attributes (Easley and O'Hara (1992)). In this study the exogeneity of durations is tested formally in an Autoregressive

[2] For examples of microstructure studies using transaction data, see Hausman *et al* (1992) and Hasbrouck (1991).

[3] See Huang and Chou (2007) for applications of order imbalances to analyses of order and quote driven markets.

Conditional Duration (ACD) specification. Following the theoretical work of Easley *et al.* (1996), and the empirical finding of Dufour and Engle (2000) that there is more information based trading and hence greater price impact in intensively traded stocks, we conjecture that the duration is negatively related to price changes. Our work contributes to the growing literature applying ACD models to between-trade durations and stock price predictions (see, for example, Hafner 2005 and Fernandes and Grammig 2006). In particular, we extend Hafner (2005) by adding more factors than volume in predicting price changes.

The forecasting results from estimated parameters in the modelling framework provide potentially useful trading guidance to market practitioners as to the direction of stock price movements. The out-of-sample forecasts indicate an average 71% of correctly predicted trades. The distribution of the actual and estimated cases reveals that using the model, even in cases of prediction error, the risk of adverse selection is ruled out.

The remainder of this paper is as follows. The next section provides a review of the literature. Section 3 outlines model specifications. In Section 4 we provide data and variable descriptions. Section V presents our empirical results. Section 6 concludes.

2. Literature Review

The fact that security prices are quoted in increments of one cent or half a cent means that price series are not continuous. Gottlieb and Kalay (1985) were the first to model discrete transaction prices. Variations and generalisations of the model include Ball (1988), Glosten and Harris (1988), Harris (1990), Dravid (1991), and Hasbrouck (1999). In these models discrete transaction prices are allowed to evolve as a rounded random walk. Depending on the market-making cost and rounding methods used, these models generally have incomparable implications for the behaviour of the transaction prices. Ordered probit is another type of model that can capture the discreteness in transaction prices using standard statistical techniques, with relative ease of application. Initially proposed by Aitchinson and Silvery (1957) and developed in Gurland *et al.* (1960), the ordered probit model is simply a generalisation of a linear regression model when the dependent variable is discrete. Hence it can be used to investigate the distribution of transaction prices conditional on the sequence of past price changes and market conditions.

Early theoretical studies on price adjustments to information stressed the importance of the arrival time of trades (Diamond and Verrechia 1987, Admati and Pfleiderer 1988, 1989). On the assumption that the time between transactions, also known as the duration, is exogenous, Hausman *et al.* (1992) and Fletcher (1995) find significant correlation between duration and price changes. However, the exogeneity of time remains a question of interest. 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), where reciprocal interactions of the price, trade and time are discovered on the NYSE. The conditional mean of the duration is modelled using an Autoregressive Conditional Duration (ACD) framework by Engle and Lange (2001) who find the coefficient of durations significant in their model of liquidity.

The existing market microstructure literature has developed theory for the key role of information asymmetry in the formation of security prices. Theoretical models suggest that market makers always lose in trading with informed traders who possess private information. They therefore adjust bid and ask quotes in trading activities, through which public and private information is impounded in the price and other market attributes. This proposition is verified by evidence of a close relation between stock returns and variables such as spread, trade size and trade indicator in ample empirical research.[4] For example, Barclay and Warner (1993) argue that price is driven mainly by the mid-sized trades on NYSE. Hasbrouck (1991) demonstrates that the trade indicator is an appropriate proxy for classifying the source of price changes into one due to a permanent informational effect and the other resulting from a transient market-related effect. The positive relation between trade size and returns in the theoretical framework of Easley and O'Hara (1987) is backed up by the empirically findings of Hasbrouck (1991), among others.

Order flow is another element that receives continuing interest in academic research. Theoretical accounts of order flow trace back to Kyle (1985). In Kyle's dynamic model, a single informed trader possessing private information acts strategically in a market with dealers and uninformed noise traders to move the market price. The model shows that price change is linear in the order flow '$x + \mu$', where x is the trade quantity of the informed trader and μ is the trade quantity of noise traders. The price will show a sustained change if the order flow is concentrated on one side of the market.

[4] For evidence on the information content of these market variables in theoretical and empirical studies, see for example, Bagehot (1971) Easley and O'Hara (1987), Hasbrouck (1991), and Barclay and Warner (1993).

There is empirical evidence of net order flow being an important measure of trading activity. Chordia and Subrahmanyam (2004) argue that the net order flow measured by order imbalance is more valuable in explaining the direction of future price moves than volume. This is an outcome of market makers' continuous readjustment of their inventories at each trade. Using neural networks, Plerou *et al.* (2002) find that the expected price change is a concave function of trade imbalance. In a fully automated order driven trading system, Huang and Chou (2007) also demonstrate the significant information content of trade imbalance. The usefulness of the net order flow in the absence of market makers is manifest in the evidence that trade imbalance is predictive of future price direction (Chordia and Subrahmanyam 2004).

Many prior studies are dedicated to unveiling the relation between stock price changes and volume. Although there is evidence that trade size, measured as number of transactions or aggregated trading volume, explains future stock price changes (Hafner, 2005), it is only in recent studies that the depth is found also instrumental to returns (Heflin and Shaw 2005). Generally, depth denotes the volume of orders available to be traded at a particular price level, for example at the best bid and ask prices. The depth of the market and its dynamics are investigated by Muranaga and Shimizu (1999) in simulated markets. Heflin and Shaw (2005) argue that since the market maker's price-quantity schedules evolve over time, which reflects in the shift in quoted depth, the level of information asymmetry embedded in trade size should be assessed relative to the quoted depth on the market at the time of the transaction.

In an order driven trading system such as the ASX, where the matching of orders is fully automated and the price-quantity schedules are available to all market participants, the usefulness of depth is more intuitive. For instance, the consequent price movement of a given trade, initiated, for instance, by a market order to buy, will depend on the size of the trade relative to the depth at the best ask price at the time of the trade. A medium trade in its raw size might cause a large price change in a thin market depth, and a large trade in its raw size might cause a small price change in a thick market depth.

3. Model Specification

3.1. The ordered probit model

The ordered probit framework employed in this paper is a variation of Hausman *et al.* (1992). The dependent variable of an ordered probit model is a latent (unobservable) continuous variable, denoted dp^* for instance, 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 $P_0, P_1, \ldots, P_k$, 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$)). In the 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 $q\times 1$ vector of predetermined explanatory variables that determines the conditional mean of $dp_k{*}$. The observed price changes dp_k relate to the continuous variable $dp_k{*}$ using the 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 defined as:

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

According to Hausman et al. (1992), the dependence structure of the observed process Z_k is induced by that of $Z_k{*}$ and definitions of the A_k's as follows:

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

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 for Gaussian ε_k's is

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

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

where $\Phi(.)$ is the standard normal cumulative distribution function. Equations (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 in (7) 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.

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

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. Hausman *et al.* (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.

3.2. 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.

Second, 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 use 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 are utilised 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 excluded.

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 the normal distribution. Our estimating results are therefore likely immune to the underlying distribution functions of the price series.

3.3. The ACD model for the time between trades

In the literature time between two consecutive trades, or 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-effects and to be conditionally ordered when the current arrival rate is dependent upon the times of 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 $\{\tau_1, \tau_2, \ldots\}$ with $\tau_1 < \tau_2 < \ldots < \tau_k < \ldots$, we can express the duration between two consecutive transactions that occur 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, $E(\tilde{x}_k \mid \tilde{x}_{k-1},...,\tilde{x}_1) = \psi_k$, 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

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

Relying on proof provided in Engle and Russell (1998), we show in 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,

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

are independently and identically distributed (i.i.d.) for all *k*'s. This assumption implies that all temporal dependence in the duration series is captured by the defined mean function. As far as the distribution function of durations is concerned, we consider the widely utilized Weibull distribution. The Weibull distribution is generally preferred to the exponential distribution for duration data which tend to show over-dispersion with extreme values (as in very shot and long durations). The conditional density function of adjusted durations, $g(\tilde{x}_k)$, based on the Weibull distribution is given by

$$g(\tilde{x}_k) = \frac{\lambda}{\psi_k^{\lambda}} \tilde{x}_k^{\lambda-1} \, exp\left[-\left(\frac{\tilde{x}_k}{\psi_k} \right)^{\lambda} \right] \text{ for } \lambda,\ \psi_k > 0, \tag{11}$$

where λ is the scale parameter. The above equation also gives the density function for exponential distribution. The density function is obtained as a special case when $\lambda = 1$. In our estimated results clearly λ is unity for none of the stocks.

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

$$L(\eta) = -\sum_{k=1}^{N} ln\left(\frac{\lambda}{\tilde{x}_k} \right) + \lambda ln\left(\frac{\tilde{x}_k}{\psi_k} \right) - \left(\frac{\tilde{x}_k}{\psi_k} \right)^{\lambda} . \tag{12}$$

4. Data and statistics

4.1. Data and Variables

We obtained data for this study from the Securities Industry Research Centre of Asia-Pacific (SIRCA) in Australia. The database contains detailed information relative to orders and trades for all stocks listed on the ASX. Our sample encompasses nine stocks from each of the nine major industry sectors in Australia (Consumer Discretionary, Consumer Staple, Energy, Financials, Health Care, Industrials, Information Technology, Materials, Telecommunication Services, and Utilities). We choose a sample period of 1 April 2002 to 31 July 2002 when there are no significant structural changes in these firms. For each transaction, our sample contains the following information: the date, time, size, price, spread, trade indicator, depth, and the sector index corresponding with each stock.

We note that there may be 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 (Engle and Russell (1998)). 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.[5] 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)

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

variable are also set to zero as it is unrealistic to calculate today's trade imbalance from trades that occurred yesterday.

To examine 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. The market attribute dp_k is the price change at trade k from trade k-1. Since for all stocks the minimum price change allowed is 1¢, multiplying price difference by 100 gives price change in ticks (cents) as an integer,

$$dpk = 100 \times (P_k - P_{k-1}). \tag{13}$$

$Sprd_{k-1}$ is the bid/ask spread immediately before trade k occurs. It is calculated in units of cents. $LBBV_{k-1}$ is the natural logarithm of the number of shares at the best bid price immediately before trade k occurs. One share is added to each $LBBV_{k-1}$ so the logged value of zero volume at the best bid also returns a zero,

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

ε_{k-1} is the standardized transaction duration estimated in an ACD(2,2) model from diurnally adjusted conditional and unconditional durations in equation (10). $LVol_{k-1}$ is the natural logarithm of the size of $(k\text{-}1)th$ trade. TI_{k-1} is a trade indicator of the $(k\text{-}1)th$ trade, where 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} 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 on the same day,

$$TIB_{k-1} = \frac{\sum_{j=1}^{30}(TI_{(k-1)-j} \times Vol_{(k-1)-j})}{\sum_{j=1}^{30} Vol_{(k-1)-j}}. \tag{15}$$

Finally, ΔIDX_{k-1} is the return on index for the k-1th trade calculated from

$$\Delta IDX_{k-1} = Ln(IDX_{k-1} - IDX_{k-2}). \tag{16}$$

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

$$\begin{aligned} dp_k^* &= c_1 dp_{k-1} + c_2 dp_{k-2} + c_3 dp_{k-3} + c_4 TI_{k-1} + c_5 Sprd_{k-1} + c_6 LBAV_{k-1} + c_7 LBBV_{k-1} \\ &+ c_8 TIB_{k-1} + c_9 TI_{k-1} * \varepsilon_{k-1} + c_{10} TI_{k-1} * \psi_{k-1} + c_{11} LVOL_{k-1} + c_{12} \Delta IDX_{k-1} \end{aligned}, \tag{17}$$

where the conditional duration, ψ_{k-1}, and the standardized transaction duration, ε_{k-1}, are estimated from an ACD(1,1) specification

$$\begin{aligned} \psi_k &= \omega + \alpha_1 \psi_{k-1} + \beta_1 \tilde{x}_{k-1} \\ \varepsilon_k &= \tilde{x}_k / \psi_k \end{aligned}. \tag{18}$$

To accommodate the heteroscedasticity in the conditional variance of residuals as in Hausman *et al.* (1992), we consider a GARCH(2,2) specification

$$h_k^2 = \upsilon + \theta_1 h_{k-1}^2 + \theta_2 h_{k-2}^2 + \delta_1 \varepsilon_{k-1}^2 + \delta_2 \varepsilon_{k-2}^2. \tag{19}$$

All the unknown parameters in the system are estimated using maximum likelihood method.

4.2. Sample statistics

Table 1 provides the summary statistics of the variables to be used in the ordered probit-GARCH system. The price levels of the stocks range from A$0.78 for KAZ to A$44.10 for CSL, showing a relatively wide dispersion on the ASX. The degree of volatility is measured as the percentage change in the daily high and low prices. With a high of 262% for a small stock KAZ, most stocks have price volatility of less than 24%. The proportion of buyer-initiated and seller-initiated trades is very similar for all stocks.[6] This property of the data, combined with the observation of zero values in the mean returns on both stocks and indices to, suggests that throughout our sample period there were no major news arrival events which could cause trades and abnormal returns concentrated on one side of the market.

The average time between trades, a measure of trading frequency, varies substantially amongst stocks. Some liquid stocks such as TLS, BHP and NCP are traded every 20 seconds or less, while for others the trading frequency is approximately one minute. KAZ is traded only every two hours. Depth is measured as the number of shares at the best bid and ask price immediately before a transaction. This variable is closely associated with the price impact cost. Comparing it across the sample shows that TLS provides the deepest market, hence for trades in this stock the potential price impact cost is minimized. For this reason, we see that the average trade size in TLS is more than ten times larger than that of other stocks.

[6]An order initially entered by a trader to buy shares and then executed is classified as a buyer-initiated trade; an order initially entered by a trader to sell shares and then executed is classified as a seller-initiated trade. A third type of trades is ‘undefined’, including all trades that are neither buyer initiated nor seller initiated. Examples of this type of trades include crossings, or off-market trades.

The definition of trade imbalance given above in equation (15) implies that its value has a range of –1, when all past 30 trades are seller-initiated trades, and 1, when all past 30 trades are buyer-initiated trades. Having zero in between denotes cases where there are more seller-initiated trades than buyer-initiated trades ($TIB_k < 0$) from those where there are more buyer-initiated trades than seller-initiated trades ($TIB_k > 0$). In Table 1, five stocks have a small negative trade imbalance, indicating that there are slightly more selling transactions in those stocks in the past 30 trades, whereas the other four stocks have slightly more buying transactions in the past 30 trades.

5. Model Estimation Results

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% in an Augmented Dickey-Fuller test of unit roots. For sake of brevity, the results of unit root tests are not presented herein. The maximum likelihood estimates of the ACD(1,1) model on the diurnal adjusted durations, the ordered probit model on price changes, and the GARCH(p, q) on the residual series are computed using the BHHH algorithm proposed by Berndt *et al.* (1974). The coefficients are estimated using the first 16 weeks in the sample for all stocks. The last week starting on 25th July until 31st July 2002 is left for an out-of-sample forecast.

5.1. Estimation results from duration models

Table 2 contains estimating results from an ACD(1,1) model in equation (18). The model is estimated assuming a Weibull distribution for the diurnally adjusted durations of the stocks. The time-of-the-day effect is eliminated by regressing durations on hour-of-the-day dummy variables.

The large values of *t*-statistics of the estimated coefficients in Table 2 provide strong support of diurnally adjusted duration following a Markov process. The past behaviour of conditional and unconditional durations, implies they have explanatory power on current durations. The standardized durations, $\varepsilon_k \sim i.i.d$, are then obtained from the ACD estimation as in Equation (10). Dufour and Engle (2000) explored the degree of exogeneity in the standardized durations. Using NYSE data, they find that returns, trades and volume all have feedback effects on standardized durations. In the case of Australian stock data, we performed a regression of standardized durations on those variables and the results show that most of the parameters are not statistically significant. Therefore, standardized durations are deemed weakly exogenous. That all values of the scale parameter are statistically significant and different from unity indicates the superiority of using the Weibull distribution over the exponential distribution.

5.2. Estimation results from price changes models

The estimates from the probit-GARCH system that also incorporates the ACD model estimates 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 presented below each coefficient is used to measure the level of significance. It is calculated as the estimated mean of a coefficient divided by its asymptotic standard error.[7] Overall, the *z*-

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

statistics indicate that the estimated coefficients are significantly different from zero for most of the explanatory variables.

We first examine the coefficients of the three lags of the dependent variable. These coefficients are significant with a negative sign across all stocks, indicating a consequent price reversal from past price changes. For example, holding 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 1.36+0.83+0.42 = 2.61 ticks, which means dp_k= 1 as 2.61 is greater than the upper boundary. This negative relation is consistent with findings in the literature. Second, the coefficients of those conventional variables, the spread, trade size, and the index returns, are statistically significant for approximately half of the stocks. The trade indicator is significant for eight stocks

Third, consistent with our initial conjecture, the substantially significant *z*-statistics of the depth variables, the number of shares at the best bid price (LBBV) and the number of shares at the best ask prices (LBAV), show a close relation to price changes. Moreover, the signs of these variables provide interesting insights into the relation. The positive sign observed for all coefficients of LBBV imply a positive effect of the volume at the best bid price on the conditional mean of dp_k*, indicating that LBBV increases the probability of a price rise. On the other hand, the negative sign observed for all coefficients of LBAV indicates a negative effect of volume at the best ask price on the conditional mean of dp_k*, indicating that LBAV increases the probability of a price fall.

Fourth, the trade imbalance (TIB), used to measure the degree of buy-sell imbalance, is also positively related to price changes with statistical significance for all stocks. This evidence implies that if there have been more buyer-initiated trades in the past 30 trades, resulting in a positive value of trade imbalance, then there is a greater probability of a subsequent price rise. This finding is intuitive because a greater number of buyer-initiated trades put pressure on the buy side that may eventually push the price up. The opposite is true with negative trade imbalance when there are more seller-initiated trades and the pressure is placed on the sell side of the stock. This finding confirms our initial inference as well as the theory.

Fifth, the coefficients on the signed conditional durations (also called expected durations), $TI_{k-1} * \psi_{k-1}$, and the signed standardized durations (also called unexpected durations), $TI_{k-1} * \varepsilon_{k-1}$, are examined. For half of the stocks either coefficient is significant. For instance, we see that the conditional duration drives the subsequent price changes in BHP, but it is the standardized duration that determines the future price change in the case of WPL. Both durations are significant for TLS, CBA and CSL. Recall that these two durations are two components of the diurnally adjusted duration in the equation: $\tilde{x}_k = \psi_{k-1} * \varepsilon_{k-1}$. Hence if one of the variables is positive while the other is negative, the product will be negative.[8] For all stocks we find that the aggregated or diurnally adjusted duration has a negative sign, indicating its negative correlation to price changes. This finding is in line with the theoretical model prediction of Easley *et al.* (1996). It is also similar to the finding of Dufour and Engle (2000) that a buying transaction arriving after a long period of time causes less price changes than a buying transaction arriving right after a previous trade. Absent market makers on an order driven market such as the ASX, the reason for this phenomenon could be that traders perceive a higher likelihood of

[8] Note that the trade indicator, TI_k, won't affect the sign of the product in the equation, as $TI_k * TI_k$ will always be positive.

informed trading when the trading is intense. The presence of heavy trading may also deter the uninformed from transacting, further increasing the proportion of informed trading, and therefore increasing the probability of price changes.

To further assess the significance of durations on price changes, we perform Wald tests of the null hypotheses that the coefficient of the conditional duration is zero, the coefficient of the standardized duration is zero, and both coefficients are jointly zero. The results are presented in Table 4. For five out of nine stocks, the *f*-statistics suggest that we reject the null hypotheses.

Presented in the second panel of Table 3 are the partition boundaries computed in the estimation to partition the differing directions of price changes. Given the three possible directions of price changes, $dp_k < 0$, $dp_k = 0$ and $dp_k > 0$, these boundaries are 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. Pseudo-R^2 values are presented at the bottom of the first panel.

Finally, the conditional variance is parameterized. We in this case presume that the conditional variance follows an ARMA process and thereby apply a GARCH(p, q) to the residual series of the probit model. The coefficients are included in the bottom panel of Table 3. The order of p and q for each individual stock is determined by Akaike information criteria. The estimation results of GARCH models show that all coefficients are significantly different from zero in the first or second lag of both or either of the conditional variance and the squared residuals.

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, serial correlation is detected for the generalized residuals along the lines of Hausman *et al.* (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, the theoretical value of the cross correlation should be zero, or very close to zero. Using 20 lags, the cross correlation coefficients of the generalized residuals and the lagged generalized fitted values $d\hat{p}_k^*$ are very small and close to zero. To save on space, the results are not presented here.

5.3. Forecasting Analysis

Turning to the forecasting power of the modelling framework, we undertake in-sample and out-of-sample forecasting tests. Data from the last week (week 16) of the estimation period from 18 July to 24 July are used for the in-sample forecasts. For out-of-sample forecasts, the next seven days' data from 25 July to 31 July that are not included in the original estimation are used. For each trade, the fitted values of $d\hat{p}_k$ are computed from $d\hat{p}_k^*$ using estimated parameters and partition boundaries. Then with the estimated and actual price movements, we evaluate the forecasting effectiveness from the forecast error statistics computed and reported in Table 5 below. We first present the widely used forecast error statistic known as the root mean square error (RMSE), computed as follows:

$$RMSE = \sqrt{\sum_{1}^{n} (\hat{y} - y)^2 / n}, \tag{21}$$

where n represents the number of observations in the forecasting sample. Using this statistic we find that NCP yields the best forecasting estimates in both in-sample and out-of-sample tests, as RMSE has the smallest value in this stock. With an RMSE of 0.58, stock AGL performs the least effective in both forecasts.

A second measure of forecasting error is the mean absolute error (MAE), computed as follows:

$$MAE = \sum_{1}^{n} | \hat{y} - y | / n, \tag{22}$$

where n is defined as above. From the closeness of the two measures, MAE's statistics presented in Table 5 are consistent with those from RMSE, pointing at NCP as the best performer, and AGL the worst.

It is however noted that in this study the dependent variable is discrete and only takes three numbers of -1, 0, and 1, the usual forecast error statistics above may not give a thorough test for the forecasting ability of the model. We therefore consider a second method. We pair and compare the fitted $d\hat{p}_k$ (forecasted from estimated coefficients) with the actual dp_k (in the original sample data) and find the degree of difference in the forecasting samples. The percentage of correctness is calculated from the number of correct forecasts on total forecasts. The in-sample forecasting results are illustrated in Panel 1 of Table 6, and the out-of-sample forecasting results are provided in Panel 2 of

Table 6. The top half of each panel gives information on the size of the forecasting sample and the proportion of correct forecasts for each stock.

Contrary to general finding that the *ex ante* forecast is worse than the *ex post* forecast, for four out of our nine sample stocks, the out-of-sample forecast is better than the in-sample forecast. The average proportion of correct forecast is approximately 72%, implying strong forecasting power in the ordered probit system. As expected, the ranking of the stocks' performance differs from that based on the earlier forecast error tests. In this case we find that KAZ is the best performer followed by BHP.

Although the results show that even the worst performing stock still achieves over 60% of correct forecasts, it would be useful to further investigate the scenario of prediction error. In practice, a trading strategy can be created from the predicted direction of stock price movements. For instance, if a stock rise is predicted, an investor can take a long position in the stock at the current time and liquidate it later when an actual rise occurs. A short position can be taken if the price is predicted to fall. However, of all the prediction error scenarios the most risky one would be adverse selection, that is, a case when a price fall is predicted but a rise actually occurs, or vice versa. For this reason we examine the percentage of that occurrence in our forecast estimates. As shown in the bottom half of both panels in Table 6, for all stocks, over 99% of forecasts do not fall into predicting opposite price directions.

Prediction errors may come from cases where *no change* in price direction is predicted when there is actually a *change*. To further investigate whether the model has the tendency to predict *no change*, we compute the fitted probability of the estimated $d\hat{p}_k$ falling into each of the categories, -1, 0, and 1 over the forecasting period. We find that, for all stocks, the fitted probability of falling into category 2, or no change, is usually over 50%.[9] The fitted probability of falling in each of the categories sums up to one for each observation, suggesting the model does tend to predict *no change*. For illustration purposes, the fitted probability of falling into all three categories is depicted in Figure 1 for a single stock, KAZ, for 50 observations in the forecast sample.

5.3. Robustness and limitations

We perform robustness checks of the usefulness of an ordered probit model that explicitly accounts for price discreteness in this section. In particular, using the same explanatory variables, we compare the performance of our model with one that is estimated using OLS. Instead of using an integer dependent variable that takes values of -1, 0, and 1, the actual returns are employed as the dependent variable for the regressions. The estimating results show that on average, the OLS estimates generate lower *R*-squares than their equivalent in ordered probit model estimates for all stocks. We are unable to use this model to create a trading strategy and make comparisons as the forecast estimates do not give a broad and general direction of price movements as in the ordered probit model.

To contrast forecasting performance, as an alternative we compared our model with another ordered probit model that does not have any explanatory variable other than the three lags of the dependent variable. Within our expectation, the *R*-square of the ordered probit models decreases by as much as 50% in the sample. For example, the *R*-square of the naïve model for BHP reduced to 9.1% from 17.6% in the original model. In the in-sample and out-of-sample forecasts, we obtain a higher forecast error using both RMSE and MAE tests, and a lower percentage of correctly forecasted trades. For

[9] In the interest of saving space, the results are not presented here, but will be provided upon request.

example, in the case of BHP in-sample forecast, the RMSE is 0.60 with 65% of correct predictions, markedly worse than that from the original model.

As mentioned before, the model has a tendency to predict no change when there is actually a fluctuation in future prices. This means a small number of cases involving movement in prices are not captured by the system. Although this characteristic rules out the risk of adverse selection, an active and growth driven investor may find this model not dynamic and reactive enough to market price changes.

6. Conclusion

In an empirical analysis, this paper is concerned with analysing the intra-day stock prices and predicting the direction of price movements conditional on the past price paths, trade imbalance, depth, durations, trade indicator, volume, spread, and index returns. We use a sample of nine stocks, selected from each of the main industry sectors on the ASX. Unlike many prior studies, the exogeneity of the arrival rate of each trade, referred to as duration, is formally tested in the present study. The conditional mean of diurnally adjusted durations is modelled in an ACD framework of Engle and Russell (1998). The discreteness of price series is also accounted for by an ordered probit-GARCH system.

The results show that all independent variables are statistically significant for most stocks. In particular, the contrary signs of the depth coefficients indicate that volume at the best bid price has a positive effect, and volume at the best ask price a negative effect on the probability of consequent price change. A strong positive relation is found between the trade imbalance and the conditional price changes, suggesting that dominant buying transactions in the past put upward pressure on prices, thereby increasing the probability of a consequent rise in price. Similarly, dominant selling transactions in the past put downward pressure on prices, therefore increasing the probability of a consequent fall in price. This finding is consistent with theory on the information content of net order flow (Kyle (1985)).

We find signed conditional durations and unexpected durations to be significant for five stocks, contributing to the existing informational role of time in the price process. A negative joint sign found in all stocks means that a buying transaction after a long period has less probability of an increase in price than a buying transaction right after the previous transaction. This is consistent with the finding of Dufour and Engle (2000) on NYSE stocks.

The forecasting power of the model is tested for in in-sample and out-of-sample forecasts. In both cases the average percentage of correct forecasts is 72%. The forecasted direction of prices can be used by investors to establish profitable trading strategies. In most prediction error cases, the model tends to forecast no change when there is actually a change in prices. This propensity however is useful to investors for ruling out the risk of adverse selection. In the meantime, active and growth driven investors may find this model of limited use. Increasing the flexibility and adaptability of the model is a potentially fruitful line for future research.

References

Admati, A.R. and Pfleiderer, P., A theory of intraday patterns: volume and price variability, *Review of Financial Studies*, 1988, **1**. 3-40.

Admati, A.R. and Pfleiderer, P., Divide and conquer: a theory of intraday and day-of-the-week mean effects, *Review of Financial Studies*, 1989, **2**. 189-223.

Aitchinson, J. and Silvery, S.D., The generalization of probit analysis to the case of multiple responses, *Biometrika*, 1957, **44**. 131-140.

Almgren, R. and Chriss, N., Optimal liquidation, Working paper, University of Chicago, 1998.

Bagehot, W., The only game in town, *Financial Analysts Journal*, 1971, **27**. 12-14.

Ball, C.A., Estimation bias induced by discrete security prices, *Journal of Finance*, 1988, **43**. 841-865.

Barclay, M.J. and Litzenberger, R.H., Announcement effects of new equity issues and the use of intraday price data, *Journal of Financial Economics*, 1988, **21.** 71-99.

Barclay, M.J. and Warner, J.B., Stealth trading and volatility: which trades move prices? Journal of Financial Economics, 1993, **34**, 281-306.

Bertsimas, D. and Lo, A.W., Optimal control of execution costs, *Journal of Financial Markets*, 1998, **1**. 1-50.

Chordia, T. and Subrahmanyam, A., Order imbalance and individual stock returns: theory and evidence, Journal of Financial Economics, 2004, **72**. 485-518.

Diamond, D.W. and Verrecchia, R.E., Constraints on short-selling and asset price adjustments to private information, *Journal of Financial Economics*, 1987, **18**. 277-311.

Dravid, A.R., Effects of bid-ask spreads and price discreteness on stock returns, Working paper, Wharton School, 1991.

Dufour, A. and Engle, R.F., Time and the price impact of a trade, *Journal of Finance*, 2000, **6**. 2467-2498.

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

Easley, D., and O'Hara, M., Time and the process of security price adjustment, *Journal of Finance*, 1992, **47.** 577-605.

Easley, D., Kiefer, N.M., O'Hara, M. and Paperman, J.B., Liquidity, information, and infrequently traded stocks, *Journal of Finance*, 1996, **51.** 1405-1436.

Engle, R.F. and Russell, J.R., Autoregressive conditional duration: a new model for irregularly spaced transaction data, *Econometrica*, 1998, **66.** 1127-1162.

Engle, R.F. and Lange, J., Measuring, forecasting and explaining time varying liquidity in the stock market, *Journal of Financial Markets*, 2001, **4**. 113-142.

Fernandes, M. and Grammig, J., A family of autoregressive conditional duration models, *Journal of Econometrics*, 2006, **130**, 1-23.

Fletcher, R.A., The role of information and the time between trades: an empirical investigation, *Journal of Financial Research*, 1995, **18**. 239-260.

Glosten, L.R. and Harris, L.E., Estimating the components of the bid/ask spread, *Journal of Financial Economics*, 1988, **21**. 123-142.

Gottlieb, G. and Kalay, A., Implications of the discreteness of observed stock prices, *Journal of Finance*, 1985, **40**. 135-153.

Gurland, J., Lee, I. and Dahm, P.A., Polychotomous quantal response in biological assay, *Biometrics*, 1960, **16**. 382-398.

Hafner, C.M, Durations, volume and the prediction of financial returns in transaction time, *Quantitative Finance*, 2005, **5**, 145-152.

Harris, L., Estimation of stock price variances and serial covariances from discrete observations, *Journal of Financial and Quantitative Analysis*, 1990, **25**. 291-306.

Hasbrouck, J., Measuring the information content of stock prices, *Journal of Finance*, 1991, **46**. 179-207.

Hasbrouck, J., The dynamics of discrete bid and ask quotes, *Journal of Finance*, 1999, **54.** 2109–2142.

Hausman, J., Lo, A.W. and MacKinley, A.C., An ordered probit analysis of transaction stock prices, *Journal of Financial Economics*, 1992, **31.** 319-379.

Heflin, F. and Shaw, K., Trade size and the adverse selection component of the spread: which trades are “big”? Journal of Financial Research, 2005, **28**. 133–163.

Huang, Y.C. and Chou, J-H., Order imbalance and its impact on market performance: order-driven vs. quote-driven markets, Journal of Business Finance & Accounting, 2007, **34**. 1596-1614.

Kyle, A.S., Continuous auctions and insider trading, *Econometrica*, 1985, **53**. 1315-1335.

Muranaga, J. and Shimizu, T., Market microstructure and market liquidity, Working paper, Bank of Japan, 1999.

Plerou, V., Gopikrishnan, P., Gabaix, X. and Stanley, H.E., Quantifying stock price response to demand fluctuations, *Physics Review E*, 2002, **66**, 027104, 1-4.

Fitted probability for 50 observations for stock KAZ

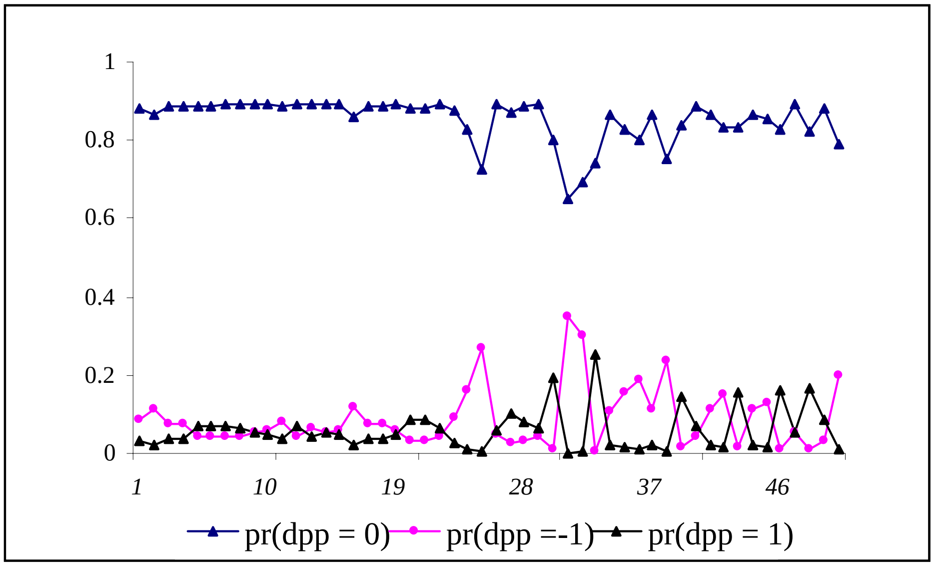

Figure 1. Fitted probability of the estimated $d\hat{p}_k$ falling into each of the categories, -1, 0, and 1 for a single stock, KAZ, as a representative of the other stocks, for 50 observations in the forecast sample. For each observation, the fitted probability of falling in each of the categories sums up to one.

Table 1. Descriptive statistics for all variables. High/low price is the highest or lowest price quoted over the whole sample period. % price change calculates the percentage price change between its highest and lowest level, (high price – low price)/low price. Spread is the bid-ask spread immediately before a trade occurs. Return on the sector index is return calculated from an index of the sector an individual stock belongs to. Time between trades is the time in seconds between two consecutive trades. Trade imbalance is calculated as the number of buyer initiated trades as a percentage of total volume in the past 30 trades.

Stock Code	TLS	AGL	BHP	CBA	CSL	KAZ	NCP	WES	WPL
Industry	Telecom.	Utilities	Resources	Financial	Health Care	Info. Tech.	Media	Industrials	Energy
Panel A: price summary									
High price (A$)	5.42	10.33	11.97	34.94	44.1	0.78	13.89	31.9	14.94
Low price (A$)	4.45	8.98	10.49	29.02	31.31	0.215	8.44	25.75	12.32
% Price change	21.80%	15.03%	14.11%	20.40%	40.85%	262.79%	64.57%	23.88%	21.27%
Panel B: % price change									
Mean	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0001	0.0000	0.0000	0.0000
Median	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Maximum	0.0045	0.0072	0.0036	0.0045	0.01	0.07	0.0061	0.0058	0.0052
Minimum	-0.0044	-0.0072	-0.0054	-0.0042	-0.01	-0.05	-0.0049	-0.0072	-0.0057
Std.Dev.	0.001	0.0007	0.0004	0.0003	0.0008	0.0048	0.0006	0.0006	0.0006
Panel C: return on the sector index									
Mean	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Median	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Maximum	0.41%	0.33%	0.41%	0.11%	0.40%	1.00%	0.29%	0.58%	0.43%
Minimum	-0.18%	-0.39%	-0.18%	-0.16%	-0.34%	1.00%	-0.38%	-0.33%	-0.68%
Std.Dev.	0.02%	0.05%	0.02%	0.01%	0.04%	0.13%	0.02%	0.03%	0.04%
Panel D: time between trades (seconds)									
Mean	12	65	14	20	52	132	16	58	56
Median	3	21	4	7	12	9	5	18	16
Std.Dev.	20	108	27	34	106	357	33	101	101
Panel E: shares at the best bid price									
Mean	227,522	5,905	19,631	8,196	1,489	120,553	12,485	1,833	5,128
Median	84,631	2,833	9,000	2,795	900	39,834	5,000	1,000	3,000
Maximum	18,773,939	753,324	1,100,000	300,000	40,000	819,668	1,000,000	100,153	116,098
Std.Dev.	484,558	14,005	34,836	8,944	1,977	376,799	35,386	3,181	6,552
Panel F: shares at the best ask price									
Mean	254,423	6,054	24,210	5,260	1,589	94,017	11,205	2,454	5,635
Median	84,566	2,790	13,158	3,000	1,000	44,019	6,315	1,441	3,594
Maximum	4,836,780	1,000,000	2,525,000	151,000	52,242	1,000,000	956,501	100,000	165,217
Std.Dev.	385,847	19,717	36,665	7,627	2,097	129,366	18,173	4,198	6,747
Panel G: trade size									
Mean	11,312	1,860	7,494	2,483	4,437	1,666	3,211	788	2,010
Median	2,000	989	2,500	940	186,305	617	1,494	431	1,000
Maximum	1,000,000	746,676	1,200,128	1,500,000	225,000	105,000	200,000	60,000	50,000
Std.Dev.	31,143	6,863	14,780	8,485	6,983	2,694	4,995	1,137	2,699
Panel H: trade imbalance									
Mean	-0.0456	-0.0282	0.0183	0.0011	0.0037	-0.0634	-0.0061	0.0319	-0.045
Median	-0.0578	-0.0189	0.028	0.0029	-0.007	-0.0695	-0.0006	0.0457	-0.0473
Std.Dev.	0.6576	0.4199	0.4553	0.426	0.4508	0.5464	0.4392	0.4149	0.3943
Panel I: trade direction									
Buyer initiated	47.24%	44.93%	47.33%	44.80%	44.54%	40.96%	46.21%	46.93%	44.47%
Seller initiated	44.52%	45.73%	42.23%	47.27%	44.55%	49.16%	45.10%	43.31%	44.95%

Table 2. Descriptive statistics. The table presents ACD(1,1) estimates on a Weibull distribution $\psi_k = \omega + \alpha\psi_{k-1} + \beta\tilde{x}_{k-1}$ for the sample stocks, where $\tilde{x}_k$ represents diurnally adjusted duration for the *k*th trade, and ψ_k represents the conditional mean of duration for the *k*th trade. λ is the scale parameter of the Weibull distribution function. Below the estimated coefficients are *t*-statistics. * significant at 90% level.

	TLS	AGL	BHP	CBA	CSL	KAZ	NCP	WES	WPL
ω	1.49	3.78	1.59	1.83	2.74	5.36	1.57	3.78	3.56
	67.39*	99.70*	78.81*	78.37*	172.43*	114.03*	74.29*	57.24*	54.49*
α_1	-0.16	-0.3	-0.13	-0.15	-0.01	-0.12	-0.13	-0.17	-0.18
	-9.52*	-25.59*	-6.12*	-10.27*	-2.25*	-14.22*	-8.42*	-8.50*	-8.59*
β_1	0.16	0.08	0.15	0.13	0.07	0.05	0.14	0.07	0.08
	65.08*	66.22*	76.05*	73.52*	68.63*	39.86*	71.68*	46.65*	43.96*
λ	5.09	3.61	4.89	4.5	3.74	3.56	4.6	4.11	3.88
	754.15*	357.89*	732.40*	670.88*	385.42*	266.43*	714.06*	400.17*	388.00*

Table 3. Estimation results of price changes in an ordered probit-GARCH system per equations (18) and (20) for sample stocks on the ASX. The dependent variable is dp_k, the price change in ticks; dp_k* is the latent continuous version of dp; dp_{k-i} denotes 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(1,1) 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 an index for the (k-1)th trade multiplied by 100; TI_k is trade indicator. Below the estimated coefficients are t-statistics. * significant at 90% level..

	TLS	AGL	BHP	CBA	CSL	KAZ	NCP	WES	WPL
Panel A: mean equation - ordered probit model									
dp_{k-1}	-1.36	-0.611	-0.851	-0.579	-0.075	-0.735	-0.934	-0.212	-0.492
	-90.13*	-16.50*	-61.47*	-29.25*	-2.38*	-16.12*	-46.37*	-6.96*	-13.75*
dp_{k-2}	-0.826	-0.408	-0.464	-0.217	0.023	-0.407	-0.498	-0.017	-0.17
	-59.25*	-11.38*	-35.10*	-12.21*	0.82	-9.33*	-26.01*	-0.63	-5.02*
dp_{k-3}	-0.422	-0.246	-0.247	-0.084	0.071	-0.275	-0.248	0.052	-0.054
	-36.74*	-7.67*	-20.86*	-4.95*	2.64*	-6.89*	-13.63*	2.03*	-1.77*
TI_{k-1}	-0.065	0.056	0.108	0.115	0.119	0.115	0.153	0.12	0.177
	-2.370*	0.85	3.68*	10.59*	2.04*	1.88*	3.07*	1.93*	2.59*
$Sprd_{k-1}$	-0.504	3.692	-0.379	0.115	-0.296	-8.904	-0.042	-0.226	-0.72
	-2.89*	4.40*	-1.21	2.04*	-1.07	-3.43*	-2.30*	-1.35	-0.5
$LVol_{k-1}$	0.007	0.026	0.02	-0.001	0.008	0.017	0.036	-0.033	0.026
	2.15*	1.58	4.96*	-0.15	0.53	1.25	5.00*	-2.01*	1.67*
$\Delta\Xi\Delta I_{1-\kappa}$	-60.443	71.224	-114.876	235.521	29.673	3.309	43.474	14.005	45.498
	-6.79*	1.81*	-3.65*	2.68*	0.68	0.29	1.05	0.27	1.08
$LBAV_{k-1}$	-0.121	-0.434	-0.211	-0.393	-0.462	-0.198	-0.413	-0.497	-0.477
	-46.07*	-30.54*	-58.02*	-62.62*	-34.39*	-17.58*	-62.99*	-36.24*	-33.94*
$LBBV_{k-1}$	0.116	0.438	0.16	0.409	0.466	0.192	0.356	0.472	0.43
	42.51*	31.57	44.05*	61.98*	34.43*	16.11*	56.34*	33.09*	30.23*
TIB_{k-1}	0.072	0.161	0.06	0.158	0.259	0.079	0.21	0.176	0.156
	9.05*	4.05	5.23*	7.69*	7.02*	2.74*	10.73*	4.77*	3.83*
$\psi_{1-\kappa}$	-0.021	-0.008	-0.007	-0.017	0.014	-0.003	0.001	0.008	0.006
	-4.10*	-1.02	-1.82*	-2.77*	1.83*	-0.4	0.23	1.03	0.74
$\varepsilon_{1-\kappa}$	0.095	0.078	0.018	0.092	-0.117	0.049	-0.002	-0.047	-0.114
	2.58*	1.2	0.6	1.94*	-1.97*	0.91	-0.04	-0.76	-1.79*
no. of obs.	102,273	20,485	90,112	73,309	23,146	10,016	90,058	22,154	22,157
Pseudo-R^2	0.283	0.124	0.176	0.115	0.091	0.153	0. 114	0.098	0.106

Table 3 (continued)

Panel B: partition boundaries

	TLS	AGL	BHP	CBA	CSL	KAZ	NCP	WES	WPL
$dp_k < 0$	(-∞, -1.68]	(-∞, -1.86]	(-∞, -1.91]	(-∞, -1.83]	(-∞, -1.74]	(-∞, -1.49]	(-∞, -2.54]	(-∞, -2.20]	(-∞, -2.27]
$dp_k = 0$	(-1.68, 1.71]	(-1.86, 2.22)	(-1.91, 1.16)	(-1.83, 2.18)	(-1.74, 1.89)	(-1.49, 1.73)	(-2.54, 1.99)	(-2.20, 1.36)	(-2.27, 1.81)
$dp_k > 0$	(1.71, +∞]	(2.22, +∞]	(1.16, +∞]	(2.18, +∞]	(1.89, +∞]	(1.73, +∞]	(1.99, +∞]	(1.36, +∞]	(1.81, +∞]
Panel C: Variance Equation - GARCH(p, q)									
c	0.017	0.007	0.013	0.013	0.078	0.012	0.007	0.061	0.022
	6.93*	3.91*	4.72*	4.74*	5.92*	2.82*	5.67*	53.83*	3.69*
δ_1	0.12	0.032	0.018	0.018	-0.104	0.017	0.043	0.06	0.045
	33.47*	4.60*	6.57*	6.57*	14.67*	4.56*	12.94*	8.70*	6.59*
δ_2	-0.069	-0.011	0	0	-0.004	-	-0.01	-	-0.015
	-9.80*	-1.5	0.02	0.05	-0.34	-	-1.64	-	-1.85
θ_1	1.59	0.946	0. 321	0.321	0.557	0.958	0.587	0.472	0.865
	13.94*	87.51*	1.5	1.5	8.12*	84.67*	3.63*	3.57*	26.30*
θ_2	-0.144	-	0.636	0.636	-	-	0.34	0.205	-
	-2.19*	-	3.007*	3.07*	-	-	2.24*	1.69*	-

Table 4. Wald tests of coefficient restrictions to the estimated coefficients of two types of durations from the ordered probit model in equation (18) for all stocks in the sample. c(9) is the coefficient of the signed standardized (or unexpected) duration, ε_k, and c(10) is the coefficient of the signed conditional(or expected) duration, ψ_k. * denotes values of *f*-statistics significant at 90%.

	TLS	AGL	BHP	CBA	CSL	KAZ	NCP	WES	WPL
	f-stat	*f-stat*	*f-stat*	*f-stat*	*f-stat*	*f-stat*	*f-stat*	*f-stat*	*f-stat*
H_0: c(9) = 0	16.84*	1.16	3.33*	6.61*	3.32*	0.19	0.07	0.73	0.57
H_0: c(10) = 0	6.66*	1.4	0.36	3.51*	3.85*	0.88	0.01	0.33	3.34*
H_0: c(9) = c(10) = 0	8.43*	0.85	2.72*	3.69*	2.40*	0.45	0.06	0.38	2.68*

Table 5: In-sample and out-of-sample forecasting tests results. The first forecast test calculates root mean square error (RMSE) from the following formula: $\sqrt{\sum_{1}^{n}(\hat{y}-y)^2/n}$, where n is the sample size in the forecasting period. The second forecast test calculates the mean absolute error (MAE) from the following formula: $\sum_{1}^{n}|\hat{y}-y|/n$, where n is as defined above.

Panel 1: in sample prediction from 18/07 to 24/07									
	TLS	AGL	BHP	CBA	CSL	KAZ	NCP	WES	WPL
Total obs. in forecast	4,339	1,162	6,191	8,138	2,122	201	7,810	1,418	1,887
Root Mean Square Error	0.493	0.578	0.464	0.191	0.404	0.457	0.183	0.56	0.418
Mean Absolute Error	0.243	0.332	0.215	0.037	0.163	0.209	0.033	0.31	0.175
Panel 2: out of sample prediction from 25/07 to 31/07									
Total obs. in forecast	4,710	1,433	7,896	8,212	1,344	564	7,713	1,687	1,224
Root Mean Square Error	0.49	0.535	0.455	0.227	0.574	0.266	0.219	0.497	0.584
Mean Absolute Error	0.241	0.359	0.207	0.051	0.328	0.071	0.048	0.291	0.339

Table 6 In-sample and out-of-sample forecasts of price change directions with percentages of correctness. The actual count of observations is based on dp_k's, and the estimated count of observations is 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 the percentage of missed observations on the total number of observations.

Panel 1: In Sample Prediction from 18/07 to 24/07															
		TLS		AGL		BHP		CBA		CSL		KAZ		NCP	
Total obs. in forecast		4,339		1,162		6,191		8,138		2,122		201		7,810	
% correct forecast		75.68%		66.90%		78.46%		69.50%		68.88%		79.21%		77.19%	
Actual	Forecast	*obs.*	*%*	*obs.*	*%*	*obs.*	*%*	*obs.*	*%*	*obs.*	*%*	*obs.*	*%*	*obs.*	*%*
Fall	*Rise*	0	0.00%	1	0.10%	0	0.00%	3	0.00%	1	0.00%	0	0.00%	2	0.00%
Rise	*Fall*	0	0.00%	0	0.00%	0	0.00%	0	0.00%	1	0.00%	0	0.00%	1	0.00%
Panel 2: Out of Sample Prediction from 25/07 to 31/07															
Total obs. in forecast		4,710		1,433		7,896		8,212		1,344		564		7,713	
% correct forecast		74.04%		76.34%		79.33%		68.14%		63.10%		81.71%		74.35%	
Actual	Forecast	*obs.*	*%*	*obs.*	*%*	*obs.*	*%*	*obs.*	*%*	*obs.*	*%*	*obs.*	*%*	*obs.*	*%*
Fall	*Rise*	0	0.00%	0	0.00%	1	0.00%	0	0.00%	1	0.10%	0	0.00%	0	0.00%
Rise	*Fall*	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%