QUANTITATIVE FINANCE
RESEARCH CENTRE

UNIVERSITY OF
TECHNOLOGY SYDNEY

QUANTITATIVE FINANCE
RESEARCH CENTRE

UTS

THINK.CHANGE.DO

Quantitative Finance Research Centre

Research Paper 220 **March 2008**

Modelling Adverse Selection on Electronic Order-Driven Markets

Louis R. Mercorelli, David Michayluk and Anthony D. Hall

ISSN 1441-8010

Modeling Adverse Selection on Electronic Order-Driven Markets

Louis R. Mercorelli

David Michayluk*

Anthony D. Hall

March 17, 2008

Key words: bid-ask spread models, adverse selection, anonymity

JEL Codes: G10, G15

We are grateful for helpful comments from seminar participants at the Reserve Bank of Australia and at the Santa Fe Institute. All errors remain our responsibility.

*corresponding author. All authors are in the School of Finance and Economics, University of Technology, Sydney, PO Box 123, Broadway, NSW, Australia.
Fax +61-2-9514-7711.
Michayluk: +61-2-9514-7761, david.michayluk@uts.edu.au,
Mercorelli: +61-2-9514-7732, louis.mercorelli@uts.edu.au,
Hall: +61-2-9514-7777, tony.hall@uts.edu.au

Modeling Adverse Selection on Electronic Order-Driven Markets

The vast majority of models that decompose the bid/ask spread assume the quote-driven, specialist structure of the NYSE. This paper critically evaluates these models to construct a model specific for an electronic order-driven exchange. The model not only captures adverse selection and the impact of order flows on price discovery but it includes the imbalance of supply and demand inherent in the public limit order book. With this new model we investigate the change to anonymity on the Australian Securities Exchange (ASX). Following the change to anonymity, both adverse selection and the demand/supply imbalance have an increased impact on prices while order flow has a decreased influence, suggesting the change to anonymity has improved market efficiency. The model also uncovers a change in traders' behavior once their fear of front-running is reduced. We show that the model is stable and robust across high liquidity stocks as well as stocks with as few as 5 trades per day.

Models that decompose the bid-ask spread into the three theoretical components of adverse selection, inventory holding and order processing costs all have as their foundation the structure of the New York Stock Exchange (NYSE), which is a quote-driven, specialist/market maker exchange. These models have been used to examine electronic order-driven markets, but the theoretical origin is inconsistent with a market where there is no specialist and the observable price dynamics are determined by all market participants instead of a single specialist or a few market makers. Therefore, the development of a model that is applicable in electronic order-driven markets is warranted.

The research in this field, although applicable primarily to the NYSE, is extensive. Comparing the main models indicates that one main finding is consistent throughout the literature: information moves stock prices via trading. We show that each model uses a trade indicator as the primary variable to extract information asymmetry.[1] Generalizing these models provides a framework that can then be used to adapt a model to incorporate the additional information available in electronic order-driven markets. Whereas the observable depth in a quote-driven exchange is under the control of the specialist, in an order-driven market the depth reflects the aggregate information from traders. We use this information in our model via a demand and supply ratio to determine the impact on prices. The order flow is also included since it may be particularly important when traders use strategies that do not include limit orders when they are attempting to hide their trading[2]. We then test this new model around the shift to anonymity on the Australian Securities Exchange (ASX) since one or more of the trading costs are expected to change with this event.

On Monday November 28th 2005, the ASX eliminated broker IDs from brokers' trading screens. This change to anonymity significantly decreased transparency in the market. Before the change, brokers could observe the broker ID for each limit order awaiting execution in the public limit order book. In comparison, institutional traders and retail investors did not have access to this information. Although brokers were prohibited from

[1] Information asymmetry is defined as that portion of a trade that impacts the true value, V_t, of the stock (i.e. excluding noise, inventory management costs, overhead costs, etc.). This term will be used interchangeably with adverse selection and price impact. See Glosten and Harris (1988) for a more detailed discussion.

[2] Traders may use off-market trades where available in some markets or may use market orders to take liquidity from the other side of the limit order book.

passing this information on to their clients, this rule was allegedly often broken. Informed traders were aware of the impact of their IDs on the market and those wishing to purchase (sell) large blocks of shares would not submit limit orders because they did not want to announce their trading intentions via the public limit order book. Instead, they would use market orders or negotiate off-market trades to conceal their intentions and remain anonymous. The ASX and other market participants presumed, after broker IDs were removed, that informed traders would reduce their use of these preventive strategies and increase their use of limit orders since the public limit order book was now anonymous.

In addition to the informational inequity or lack of a level playing field, large traders were worried about front-running. Front-runners are predatory traders who trade with large (typically institutional) traders, driving up stock prices artificially then selling quickly after the share price has risen. They make short-term gains with minimal risk while the large traders end up paying more for their long-term investment than they would have otherwise.[3] With broker IDs visible in the order book, large traders incur higher costs whether they use market orders (i.e. from crossing the bid/ask spread) or if they expose their trades to front-running via the public order book. With the removal of broker IDs, large traders would not only be attracted to limit order strategies but they would also be encouraged to trade more because fewer front-runners results in lower costs. This change to anonymity is an ideal event with which to test our model.

[3] The same profits are possible in a selling scenario only the process is reversed.

The number of articles examining anonymity is sparse. From an empirical standpoint there are only a few examples where a market has provided a natural experiment to test the effects of anonymity. Examples include the Toronto Stock Exchange, Euronext-Paris and the ASX where IDs were removed from limit orders and the Korean Stock Exchange where IDs were introduced on limit orders. From these natural experiments, a handful of papers were published. For example, in analyzing a shift to pre-trade anonymity on Euronext-Paris, Comerton-Forde et al. (2005) identify a statistically significant decline in the bid-ask spread. Foucault et al. (2007) also found that the change to pre-trade anonymity on Euronext-Paris was associated with a decrease in the bid-ask spread, although their results were not statistically significant. In contrast, Waisburd (2003) found that a change in post-trade anonymity on Euronext-Paris caused the spread to increase although his results were also not statistically significant.[4] Comerton-Forde and Tang (2008) examine the shift to anonymity on the ASX and they found that the level of order aggressiveness declined following the removal of broker IDs. Extending the research in the field to measure adverse selection will help clarify how changes in anonymity affect markets.

This paper contributes to the literature by first generalizing several of the quote-driven models to explain trading across any type of market. Then a new model is tailored to explicitly account for the specific characteristics of electronic order-driven markets. The new model is then used to investigate the effect on trading when broker IDs were

[4] Theissen (2003) also prepared a paper concerning anonymity. He examined the Frankfurt Stock Exchange which is a quote-driven exchange with market makers. He found that non-anonymity produces lower adverse selection costs and he argues that anonymity comes with the risk of a higher adverse selection cost.

removed from the ASX. The model captures the increase in information asymmetry and the impact of demand and supply imbalances in the order book following the shift to anonymity. It also identifies the decline in importance of order flow as traders no longer have to use strategies to hide their intentions. The model is shown to work well across all levels of liquidity and robustness checks confirm that the model is stable within the Pre- and Post-anonymity periods.

The remainder of this paper is as follows. The first section reviews the relevant literature highlighting the contributions of this paper. The second section critically reviews earlier spread decomposition models and then develops the general model. The third section develops the specific model for electronic order-driven markets. The fourth section describes the data. The fifth section details the empirical estimates using the electronic order-driven model on the ASX and the sixth section presents robustness tests. The final section concludes.

1. Bid Ask Spread Decomposition Literature

1.1 Quote-driven models

There is a rich literature on decomposing the spread for quote-driven markets that is based on modeling the actions of the specialist on the NYSE. This section introduces the relevant research since we will provide a more detailed description and analysis of quote-driven models in section 2. One of the earliest articles was by Glosten and Harris (1988), who used a trade indicator variable, Q,[5] and a trade volume variable, v, to decompose the spread. The Glosten and Harris (hereafter GH) model decomposed the spread into a transitory component and an adverse selection component on the NYSE. They found that adverse selection was a significant part of the bid-ask spread and the spread increased with increasing trade size. But as they noted, their results were hampered by limitations in the data. Specifically, when large trades are broken up and filled by several opposing limit orders, their independent trade assumption did not hold and therefore biased their results.

Huang and Stoll (1997) (hereafter HS) resolved the issue with independent trades by explicitly building the serial correlation of trades or order flow into their model. In addition, their work was significant because they showed how several earlier models including Roll (1984), Choi et al. (1988), Stoll (1989), George et al. (1991)[6], Glosten and Harris (1988) and Madhavan et al. (1997) could be reconciled by their own model. The HS model was also the first to fully decompose the spread on the NYSE into order processing, inventory holding and adverse selection. One big drawback of the HS model is the large percentage of theoretically implausible estimates.[7] Henker and Wang (2006) presented several timing corrections for the HS model (hereafter HS2). Their corrected HS2 model produced 16 plausible estimates out of a sample of 18, which is a significant

[5] The trade indicator, Q, can take +1 for a buy and -1 for sell and it is a variable commonly used in this research.
[6] George, Kaul and Nimalendran (1991) developed a model using the covariance of past returns with quotes. This model uses daily return and quote data. It is not useful with intraday quote or trade data.
[7] Van Ness et al (2001) found over 50% outside the theoretical range, while Clarke and Shastri (2000) found approximately 60% outside the theoretical range.

improvement from the original HS model. The work by Henker and Wang highlights how a small misspecification can have a large impact on results.

Around the same time, but independently, Madhavan et al. (1997) developed a similar model to HS, the MRR model, and used it to examine intra-day price formation. Their model enabled them to decompose the U-shaped pattern of the bid-ask spread across the trading day. One of their more interesting findings is that the information asymmetry component of the spread decreases over the day on the NYSE. Finally, Hasbrouck (1991) used a vector auto regression model (HVAR) to examine the information content in a stock trade that helped prove that trades have a permanent but delayed impact on price on the NYSE.

Two things stand out about the previous literature on spread decomposition models. First, the variety of situations where the models are usefully applied is large. Second, the NYSE and its specific market mechanics serve as the foundation for all the models.

1.2 Order-driven (electronic exchange) empirical results

The first empirical research decomposing the spread on an order-driven market was done by de Jong et al. (1996) (hereafter DJG model). They disentangled the spread on the Paris Bourse and reported that price impact[8] increases with trade size, similar to the work done on the NYSE. They used two models to verify earlier NYSE findings, one model they

[8] The terms ‘price impact’, ‘information asymmetry’ and ‘adverse selection’ represent the same effect and are used interchangeably throughout the article. They represent the amount of new, unknown information entering the market via an informed trader.

developed from Glosten (1994) and the other was the HVAR model from Hasbrouck (1991). The two models yielded the same qualitative results but differed significantly from a quantitative perspective. For instance, under the DJG model they found that adverse selection, or price impact, was 60% of the total spread, while under the HVAR model, adverse selection was 115%. They suggested that their DJG model might provide incorrect estimates because the one period model neglects lagged effects which result in underestimates of information asymmetry. However, the results from other researchers are similar in size with the results of the Glosten model and not the HVAR model.

On the Tokyo Stock Exchange (TSE), Ahn et al. (2005) employ four existing models to decompose the spread. They used the same two models that were used by de Jong et al. (1996) as well as the MRR and GH models. Their results indicated that the bid-ask spread has a general U-shape on the TSE and the adverse selection costs increase with trade size similar to the NYSE and the Paris Bourse results. They made no attempt to justify their use of the models on the TSE although they did recognize that they were originally designed for quote-driven markets.

The only other article that decomposes the spread for a quote-driven market is by Chan (2000). He reported results for the Hong Kong Stock Exchange (HKSE), and his findings were dissimilar to the NYSE. The asymmetric information was found to have a U-shape pattern instead of decreasing across the trading day. To reach these conclusions, Chan developed a variant of the GH model that takes into account the features of the HKSE, a pure limit order market. One potential problem with his modifications is that they were not made on a theoretical basis but from an empirical data standpoint. In addition, contrary to other research on limit order markets, he found that the inventory holding costs were significant and only slightly smaller than information asymmetry costs. Considering there is no market maker on the HKSE, the results indicating a significant cost due to the non-existent market maker's inventory appear suspect.

In all these articles the models were derived from research based on the NYSE. Some authors made modifications, recognizing that the models were based on market maker exchanges (quote-driven markets) and would need adjustments before they could be used on an electronic, public limit order book exchange (order-driven market). A bigger concern arises when authors used the models without any modification or justification. The fact that results are dramatically different depending on which model is used, seems to confirm this suspicion.

1.3 Do the models really work?

The first article to question the decomposition models appeared shortly after the models themselves. Neal and Wheatley (1998) compared the GH and GKN models.[9] They analyzed a group of continuously traded closed end funds and a control group of stocks on the NYSE. The closed-end funds are interesting because they do not allow new investors, they carry a set of diversified holdings and they report mark-to-market results weekly. All these attributes point to a small amount of adverse selection. Although the results for the closed-end funds do have smaller asymmetric information (19% versus

[9] The GKN model needed modifications so that it could handle intraday data.

34% for the GH model and 52% versus 65% for GKN model) the authors argue that the difference is less than expected and significantly different between the two models. Based on these results they claim the models are most probably mis-specified (an alternative is that the macro level proxies were poor substitutes).

Clarke and Shastri (2000) compare several decomposition models against proxy variables for asymmetric information. They investigate the MRR, HS, GKN and LSB models. Their proxies are of three types: corporate financial figures, deviations from analysts' forecasts and the probability of an informed trade, PIN[10]. The models were highly correlated with each other, negatively correlated with PIN, and somewhat correlated with analysts' forecasts. They also had some ability to detect information asymmetry measures on an annual basis. The authors conclude that since the models use high frequency data around an event and they have some correlation with other information asymmetry measures, the models are useful but they do have significant inconsistencies.[11]

The previous articles that compared models were all tested on NYSE data. Majois and De Winne (2003) compared decomposition models against 19 stocks on Euronext Brussels, which is an order-driven market. The authors concede that the models were originally developed for quote-driven markets but they proceed to test them against the order-driven

[10] An estimate developed by Easley et al (1996).

[11] Another article that compares the decomposition models against NYSE stocks is by Van Ness, Van Ness and Warr (2001). Similar to Clarke and Shastri (2000), they compare GH, GKNmod, HS, LSB and MRR models against a variety of corporate finance variables including analyst forecast errors. Only three of the models have the expected relation to the informed trader proxies suggesting the models are weak measures of information asymmetry at best.

Brussels market, relying on the consistency of results to determine whether the models work. They found the same inconsistencies across models that were found in the quote-driven markets but also noticed that the model that separated the inventory holding component out, the HS model, and the model that used covariance of returns, the GKN model, produced the least reasonable results. It is worth mentioning that they did not use the corrections for the HS model developed by Henker and Wang (2006) and they did not make the modifications that Neal and Wheatley (1998) made to the GKN model for high frequency data.

It is debatable whether the spread decomposition models are applicable under different market dynamics such as automated order-driven markets like the ASX. For example, the HS model was recently applied to the TSE by Ahn et al. (2002 and 2005). One concern is that the original model made several assumptions about the underling market microstructure that may not hold for the TSE. For example, the HS model assumes that a single specialist is quoting the bid/ask spread which represents his expectation of future price movements, while the spread in an automated exchange is more accurately characterized as the aggregate of traders' supply and demand for a stock. In addition, the HS model also assumes negative serial correlation in trades. This assumption is needed to decompose the inventory holding costs which may not be appropriate on an automated exchange. In fact, Farmer et al (2004) showed that there was positive serial correlation in trades on the Euronext, which is an automated exchange. From a data perspective, trades may execute between the bid/ask spread on a specialist exchange while they will always execute at the best bid/ask on an electronic exchange. This structural feature of the exchange will carry through to the data and will affect the dependent variable (i.e. change in price) in many of the models which, in turn, determines the estimate for information asymmetry. Another example of data differences between the two markets is when small limit orders are submitted between the best bid and ask price. On an electronic exchange, these limit orders will immediately change the spread but there may be no effect on a market maker exchange since the specialist sets these prices based on overall demand and may not react to small changes over a short time span. These concerns highlight that a model's suitability should be confirmed before applying it to a different market structure.

Which of the existing decomposition models is the best is still uncertain and an even more important question is whether they are applicable at all for order-driven markets. The previous research has not adequately addressed these issues. Some researchers have attacked the issue by comparing estimates from the models against macro level proxies. This method suffers because the value that they are estimating, adverse selection, is unobservable leaving them with the problem that they do not know whether the models or the macro-level proxies are incorrect. In the next section we will address these issues by taking a different approach and developing a general model that can be used in any market setting.

2. Qualitative Analysis of Models

Although each spread decomposition model may start from a different initial point, they do have some similarities. Most models use the change in mid-point price as a proxy for the underlying true value, V_t, of the stock. Another similarity is that the models use trade and/or quote data for their estimates. Even though the data format is slightly different across exchanges and data providers, it has enough similarities that each model can be applied using regression analysis without much difficulty. We believe that "ease of use" has encouraged early researchers to overlook the need to first prove that the models are a good structural fit before empirically using the models. In this section, instead of comparing empirical results with proxies, we compare spread models by examining each model at a basic structural level. We do this by looking at a common equation from each model; the change in mid-point price, ΔM_t. This examination will help illuminate whether the models are sensible for an order-driven market like the ASX.

2.1 Huang and Stoll (HS) model

Equations 1 and 2 are the full HS model (1997)[12]. Equation 2 contains ΔM_t which is the change in mid-point price.[13] Other variables include S which is the bid/ask spread (i.e. best ask – best bid) and Q which is the trade indicator.

$$(1) \qquad E(Q_t \mid Q_{t-1}) = (1-2\pi)Q_{t-1}$$

$$(2) \qquad \Delta M_t = (\alpha+\beta)\frac{S_t}{2}Q_t - \alpha(1-2\pi)\frac{S_{t-1}}{2}Q_{t-1} + e_t$$

Equation 1 is used to estimate the probability of a trade reversal, π, which also represents the serial correlation in trade flows. Huang and Stoll's model is based on the

[12] See Huang and Stoll (1997) equations 26 and 21. We have modified the timing convention to match the timing used throughout this article.

[13] The mid-point price, M_t, is defined as the (best bid + best ask)/ 2.

assumption that the serial correlation in trade flows will be negative. In other words, buys are followed by sells and vice versa. For Equation 2, the first term on the right hand side of the equation is the impact from a trade at time t which includes the adverse selection, α, and inventory holding, β. Both are unobservable and impact the spread as well as the change in price. The second term represents the amount of known information (serial correlation) in the trade at time t-1 and is therefore subtracted from the first term. Huang and Stoll argue that trade reversals happen because of inventory rebalancing by the market maker and since the information is known by traders, it should be subtracted from the first term. Notice that if π is equal to 0.5, then equation 1 will be zero and therefore the second term in equation 2 will also equal zero. Or taken another way, if the probability of a trade reversal is 50% then there is no information known by the traders to subtract. Also notice, a π less than 0.5 (trade continuation), is confusing because the second term reverts sign and adds information asymmetry instead of subtracting. This negative information is not intuitively pleasing since any prior known information, either positive or negative, should be removed from the incoming trade. In addition, Goodhart and O'Hara (1997) identify that trade continuations, or stocks with π less than 0.5, are often found empirically and even Huang and Stoll report many stocks with positive serial correlation in their results. This is the main criticism of the HS model.[14] It is important to note that serial correlation in trade flow is used in most decomposition models and some models use positive correlation (trade continuation) instead of negative correlation (trade reversal) to decompose the costs.

[14] Henker and Wang (2006) identified corrections for the timing of the HS equation and found that the number of stocks with π less than 0.5 dropped from 50% to below 10%.

Huang and Stoll (1997) assumed that the market maker changed the spread because of inventory adjustments over the day. They used equation 2 to estimate the inventory component, β. In a quote-driven market, the market maker is the trader of last resort. She must buy or sell at the quotes she provides and it is reasonable to assume that as her inventory increases (decreases) she will adjust the quotes to induce a buy (sell) to help balance her inventory. But, in an order-driven market, this inventory argument is difficult to make since the inventories of all market participants, large and small, contribute to the spread and no one is required to make the market (i.e. there is no trader of last resort).

For these reasons, we do not believe that the Huang and Stoll (1997) model is consistent with the mechanics of an electronic order-driven market like the ASX.

2.2 Hasbrouck model

The Hasbrouck (1991) model (equations 3 and 4, using our naming and timing conventions) has several unique characteristics. First, there are potentially an infinite number of lags for the two equations. As a practical matter, Hasbrouck cuts off the model at the fifth lag but it is worth mentioning that the coefficients beyond the first lag are not statistically significant for equation 4 (change in mid-point price equation) and are not significant beyond the third lag for equation 3 (trade indicator equation). By including the lags for the change in midpoint price and trade indicator, Hasbrouck was able to show, among other things, that the full price impact of a trade arrives with a protracted lag and information asymmetries are more significant for smaller firms. The second finding is similar to other models but the first finding identifies a potential problem for models that do not include lags beyond the first.

$$\text{(3)} \qquad Q_t \quad = c_1 \Delta M_{t-1} + c_2 \Delta M_{t-2} + \ldots + d_1 Q_{t-1} + d_2 Q_{t-2} + \ldots + e_{2,t}$$
$$\text{(4)} \qquad \Delta M_t = a_1 \Delta M_{t-1} + a_2 \Delta M_{t-2} + \ldots + \alpha Q_t + b_1 Q_{t-1} + \ldots + e_{1,t}$$

At the time of the Hasbrouck article, a quote-driven market like the NYSE reported trade and quote data to market participants when the specialist made the information available. In contrast, on most order-driven markets, traders can view the order book and can see limit orders enter the book almost immediately, providing significantly more information than the quote-driven markets. On the ASX, any trader can observe the public limit order book and the incoming quotes. E*Trade Australia, for example, supplies quotes up to 10 price levels deep for the ASX. Since this additional information is observable to traders, it seems obvious that they will use this information and incorporate it into their estimate of the true value of the stock and therefore the change in mid-point, ΔM_t. Since Hasbrouck's model only contains prior trade indicators and prior mid-point price changes it would seem it is missing a valuable piece of information, the public limit order book information.

Hasbrouck's model assumes that the quote revision and trades are not determined simultaneously (i.e. the quote revision follows the trade) and his Granger-Sims causality analysis provides a strong case for causality running from trades to quote revisions. On the automated ASX this causality between trade and quote revisions is stronger. With the order book visible to even the smallest trader, the impact of his or her trade on the quotes is immediately visible and therefore little doubt would exist on causality.

Although this model is missing some order book information, it appears to be the closest fit to an electronic order-driven market from a theoretical viewpoint.

2.3 Glosten and Harris model

The Glosten and Harris (1988) model (equation 5, using our naming and timing conventions) has a few different characteristics compared to the previous models. First, in addition to the asymmetric information coefficient, α, captured by the trade indicator variable, Q_t, the volume of the trade, v_t, is also included. Volume was omitted from the previous models intentionally because the authors found that it added noise without adding additional information[15]. Also, the dependent variable is the change in traded price, ΔP_t, instead of the change in mid-quote, ΔM_t. This inclusion was a result of the data available to Glosten and Harris, since they did not have access to bid and ask quotes but only traded prices.

$$(5) \qquad \Delta P_t = c_0 (Q_t - Q_{t-1}) + \alpha Q_t v_t + e_t$$

In previous models the change in mid-point price was used because it was a better proxy for the true value of the stock, V_t. In a market maker exchange, Huang and Stoll (1997) indicate that the true value is estimated by the market maker and reveals itself through the

[15] Hasbrouck (1991) and Huang and Stoll (1997).

mid-point of the bid/ask spread. In an electronic exchange the true value would also be measured using the midpoint of the bid/ask spread, but for a different reason. The midpoint in an order-driven market represents the aggregate price estimated by all market participants. Replacing the change in traded price with the change in mid-quote price is simple, but may be an inappropriate proxy for V_t.

There are additional issues with the GS model. First, as Hasbrouck (1997) and Huang and Stoll (1997) both point out, lagged information will be known by the market and therefore will influence the quoted spread. Also, since the GS model does not include public limit order book information, there is the same issue that exists with models previously discussed.

2.4 George, Kaul and Nimalendran model

The George, Kaul and Nimalendran (1991) model (hereafter GKN and shown in equations 6 and 7 using our naming and timing conventions) uses the difference between the trade price and quote price, RD_t, to estimate the spread. GKN use this approach to provide unbiased and efficient estimators of the spread components. Their main contribution is to prove that using traded price, as GH did, induces bias to the estimates and thus support using the change in mid-point price, ΔM_t, instead.

$$(6) \qquad s_q = {}^{2}\!/_{c}\sqrt{-Cov(RD_t, RD_{t-1})}$$

$$(7) \qquad RD_t = \pi\left({}^{s_q}\!/_{2}\right)[Q_t - Q_{t-1}]$$

One issue with the GKN model is that they assume that trade price is a function of order processing costs and can be captured by the serial correlation in trade flows. This differs from previous models where serial correlation in trade flows is a proxy for inventory management by the specialist. GKN argue that price smoothing by market makers contributes to serial correlation and therefore proxies as order processing costs for the market maker. In an electronic market, market participants would have very different order processing costs and it seems unlikely that these costs would result in serial correlation in trade flows. In addition to this issue, GKN do not include lagged information nor do they include depth information from the public limit order book. These issues lead us to conclude that the GKN model is not a good fit for an electronic order-driven market.

From the analyses several points can be derived.

1. The trade indicator, Q_t, is commonly used as an exogenous variable to estimate adverse selection.
2. The change in mid-point price, ΔM_t, is commonly used and is the most appropriate dependent variable to represent the change in true value of the underlying stock.
3. Lagged variables are commonly included to ensure that serial correlation is accounted for as well as any information previously known by market participants.

Using these basic points the following general model is constructed:

$$\Delta M_t = aQ_t + \sum_{j=1}^{N} \sum_{i=1}^{M} b_{i,j} A_{i,t-j} + k_t \qquad (8)$$

In this model the change in mid-point price is determined by new information entering the market via the trade. Information asymmetry is captured by the coefficient, a. A series of prior variables contain the known information, $A_{i,t-j}$, which contributes to the change in price. The disturbance term, k_t, captures public information and errors in price discreteness[16]. The double summation can include one or more lagged trade indicators, Q_{t-1}, or a lagged change in mid-point price, ΔM_{t-1}, or new variables like the public limit order book information or, in fact, any variable which is observable and is known in advance of the trade. The inner summation identifies different variables at a particular trade time t-j, while the outer summation identifies different trade times. This set of variables should be thought of as the predictable part of the change in price.

3. The electronic order-driven model

Equation 9 represents the electronic order-driven model and will be used for all the OLS regression results presented throughout the article.[17] It was developed specifically for the ASX although it should migrate easily to any electronic order-driven market.[18]

---

[16] Equation 8 is similar theoretically to the set-up by Hasbrouck's (1991).

[17] Additional lags were tested in an earlier model but were found to be statistically insignificant and did not improve estimation R-squares.

(9) $$\Delta M_t = aQ_t + b_1(DS^l_{t-1} - DS^f_{t-1}) + \mathrm{b}_2 DS^f_{t-1} + b_3\Delta M_{t-1} + b_4 Q_{t-1} + k_t$$

In this equation, information asymmetry is captured by ‘a’. The new variable in this model, DS, represents the information in the public limit order book and is referred to as the “demand and supply imbalance” at a market microstructure level (hereafter DS imbalance). It is calculated as the log ratio of the demand volume (i.e. total volume at the best buy price) to the supply volume (i.e. total volume at the best ask price). A positive DS imbalance would indicate an excess demand of shares and a negative DS imbalance would indicate an excess supply of shares.

Serial correlation in transaction flows, ΔM_{t-1} or Q_{t-1}, is captured by b_3 and b_4. The latter, b_4, represents serial correlation in trade flow. As mentioned earlier, in market maker exchanges this coefficient is assumed to be negative since market makers rebalance their inventory. But in an automated exchange like the ASX this coefficient is expected to be positive as large orders are split into several small orders and create positive serial trade correlation. The coefficient b_3, measures what we term the “value bounce” and measures order flow. This measure is similar to the bid/ask bounce in a quote-driven market and we would expect, once serial trade correlation is taken into account, that the value of a stock would bounce around the true value as market participants’ supply and demand requirements push back and forth within the limit order book.

[18] From a theoretical standpoint, equation 9 is similar to the Hasbrouck model in equation 4 and should be more accurately considered a modification of the Hasbrouck model than a new theoretical model.

Figure 1 helps to explain the sequence and timing of the variables represented in equation 9. First, we assume that everything starts at trade time, t, represented by Q_t in Figure 1. Immediately after the trade the market participants view the new order book and observe the mid-point price, $M_{t,}$ and the first DS imbalance represented by DS^f_t. Using this information and their own experience, market participants then determine the true value of the stock, V_t (not shown nor observable) and they submit limit order changes which adjust the public limit order book and therefore the DS imbalance until a new trade hits the market at time t+1 and the whole process starts again. Note, immediately before the new trade hits the order book the last DS imbalance is observed, DS^l_t.

Insert Figure 1 about here

There are two new variables in equation 9 that require some explanation. They are $(DS^l_{t-1} - DS^f_{t-1})$ and DS^f_{t-1}. The latter variable represents the imbalance in supply and demand within the order book immediately after the trade at time t. A positive ratio implies excess demand in the order book resulting in upward pressure on price and a negative ratio means an excess supply resulting in downward price pressure. The other variable, $(DS^l_{t-1} - DS^f_{t-1})$, represents the change in the supply/demand ratio from the last trade to immediately before the trade at time t. It is similar to the earlier variable since a positive difference would indicate an immediate increase in demand prior to the trade while a negative difference implies an immediate increase in supply prior to the trade. The difference in the change variable is that it captures more recent order book information. Since broker IDs are eliminated from Pre- to Post-anonymity the change in DS imbalance should contain less information in the Post-anonymity period and therefore it is expected that the coefficient should decrease with increasing anonymity.

4. The ASX, Data and Descriptive Statistics

The ASX is a fully automated order-driven market and the automating trading is handled by the Stock Exchange Automated Trading System (SEATS). It operates from 10am to 4pm, Monday to Friday. It has an opening call auction that initiates trading and typically lasts for 10 minutes in the morning resulting in regular trading starting at 10:10[19].

The data used for our analysis was sourced from Reuters International and was provided by the Securities Industry Research Centre of Asia-Pacific (SIRCA). For the main analysis all trades and quotes for the top 300 stocks was used. The orders (i.e. trade and quote information) were marked to the nearest millisecond and, similar to earlier research, trades at the same time and same price were merged into one trade. The sample runs from November 6th 2005 to December 16th 2005, giving a total of 30 trading days. The data is further filtered by requiring each stock to have at least 50 trades over the sample period which eliminated 40 stocks bringing the total number of stocks to 260.[20] The data was then split into a Pre-anonymity period (15 trading days ending on November 25th) and a Post-anonymity period (15 trading days beginning on November 28th). The Post-anonymity dataset is chosen to end before the Christmas holiday period

[19] See www.asx.org.au for more detail.
[20] There is no minimum requirement on the number of quotes.

and neither period contains any holidays in Australia. This restriction was chosen because liquidity and daily volume is typically lower during the holiday period. There is a similar rationale for starting the period on November 6 since there is a company reporting season in Australia that finishes at the end of October. Since the earnings reporting season is usually associated with information entering the market, the period chosen for the analysis minimizes this confounding activity.

Trades with special reporting or processing requirements were omitted from the sample. The ASX has many different special trades. Examples of special trades include off-market crossing, on market crossings, and overseas trades. The distinction between trade types is important because different trades are exposed to different procedures, restrictions and levels of disclosure which impacts price discovery, liquidity and anonymity of the market. For example, an off market crossing is prearranged by a broker, must be over $AUD 2 million and must be manually reported by the broker who performed the crossing. These restrictions prohibit participation and make it difficult for the wider trading audience to know when the trade even took place. Another example is the on-market crossing. It is performed automatically by SEATS over a 10 second interval and is initiated by a broker. Although this crossing is reported automatically, the buy and sell sides are prearranged by the broker using a process similar to that on Euronext-Paris, thereby eliminating the opportunity for the wider market to participate in the trade. Most of the special trades are prearranged by a broker and manually reported to the ASX by the trading broker. Since these special trade types have different market procedures and varying levels of anonymity, they were omitted from the sample. In addition, only trades between 10:20am and 3:55pm were included in the sample. Trades outside these times including the open and closing auctions were omitted from the sample because they may include different market procedures that deviate significantly from the original design of the model in equation 9.

Table 1a reports descriptive statistics of the data. Each figure in the middle of the table was calculated by finding the average for each stock and then finding the average across each sample period. The standard deviations are also calculated as an average across the sample period. The t-statistics and p-values reported at the bottom of the table are from a Wilcoxon signed rank test for paired samples.

Insert Tables 1a and 1b about here

Overall the two periods are very similar. The price increased slightly from $AUD 7.015 to $AUD 7.0651. The bid-ask spread slightly increased when measured in dollars (Sprd column) and slightly decreased when measured as a % of the mid-point price (relative spread column). Using the bid-ask spread as a measure of liquidity, these results indicate that liquidity did not significantly change with the removal of broker IDs.

Another way to measure liquidity is by traded volume. The last two columns in the table represent the average number of shares per trade (Trade Vol) and the total number of shares traded per day (Daily Vol). For the Post-anonymity period, the daily volume decreased but the average trade volume increased, but only the average trade volume was statistically different. These statistics imply that liquidity was not significantly different from the Pre- to the Post-anonymity period.

Table 1b further subdivides the sample into six groups ranked by market capitalization. Each group contains 50 stocks with the first group containing the top 50 stocks by market capitalization and the sixth group containing the bottom 50 stocks. Stocks with less than 50 trades across a 15 day sample period were omitted; therefore, the column labeled "Nbr" may be less than 50 especially for the stocks in the smaller market cap groups. Notice that the means do not change across most variables from the Pre- to Post-anonymity period. The two exceptions are the average volume per trade, Trade Vol, which increases from the Pre to the Post-anonymity period and the average daily frequency of trades, Freq which decreases from the Pre- to the Post-anonymity period. Tables 2a, 2b and 2c statistically confirm these results. The fact that neither of these changes has a materially effect on the average daily volume, Daily Vol, may indicate that the traders are trading the same amount from Pre- to the Post-anonymity period but they have somehow change their method of trading.

Insert Tables 2a, 2b and 2c about here

The last column worth noting in table 1c is the minimum trade frequency per group, Min Freq. This column contains the stock with the smallest average trade frequency within each group and helps us understand how well the model works for low liquidity stocks. The smallest in the Pre-anonymity period is 5.6 trades per day and the smallest for the

Post-anonymity period is 4.7 trades per day. There are over 30 stocks with an average trade frequency less than 10 and the R-squared for this low liquidity sample ranges from 0.16 to 0.37 with an average R-squared across the low liquidity sample of 0.25 (low liquidity results provided on request). This provides strong evidence that the model works well for stocks with relatively low liquidity.

5. Regression Results

Table 3 reports the results from estimating the regression model described in equation 9. Coefficients for each stock were estimated for the Pre- and Post-anonymity periods. Means and standard errors were calculated for each coefficient for each stock then averaged across all 260 stocks and reported for each period. The column on the far right represents the probability of equal means between the two paired samples using the Wilcoxon signed rank test. These results indicate that the coefficients changed significantly from the Pre- to the Post-anonymity period.

The first coefficient, a, measuring adverse selection, increased from 6.87 to 7.75 after broker IDs were removed. The column on the far right confirms that the increase was highly statistically significant[21]. This finding is consistent with the results from Foucault et al. (2007) who found that asymmetric information increased when anonymity was adopted.

[21] The test used was a Wilcoxon signed rank to ensure that any deviations from a normal distribution would not bias the results. A paired t-test was also used and the results were qualitatively similar.

Insert Table 3 about here

The coefficient for the short-term change in the order book, b1, increased from 0.0099 to 0.7217 indicating that changes in short-term DS imbalance significantly increased in the Post-anonymity period. The other variable which measures the impact of DS imbalance is b2 and it increased from 2.2587 to 2.5481. The remaining variables b3 and b4 which measure the contribution order flows have on changing prices both decreased in magnitude with b3 dropping from -0.19 to -0.11 and b4 decreasing from 0.6095 to -0.0024. From these results it can be inferred that information asymmetry and the public limit order book both have a stronger impact on prices in the Post-anonymity period while the influence of order flows significantly decreased.

Insert Table 4 about here

In order to determine if the results of the regression are robust across different size companies, stocks are divided into six groups based on market capitalization of the firm, see Table 4. The top row in each group is the average coefficients from the OLS regression model for Pre-anonymity and the bottom row is the average for Post-anonymity coefficients. A Wilcoxon signed rank test was performed to determine if the paired coefficients in each group were significantly different across the two periods. Two stars indicate statistical significance at the 1% level and one star indicates statistical significance at the 5% level. Differences in coefficients with 1% statistical significance are shaded in Table 4 and subsequent tables.

The first column in Table 4 represents the average R-squared across the stocks in the group and the second column lists the total number of stocks in a particular group. Stocks were included only if they had at least 50 observations in both the Pre- and Post-anonymity data sets. Notice that the R-squared is relatively equal across all six groups with a minimum average of 18.4% and a maximum average of 25.6% indicating that the model is relatively stable even for small companies with low liquidity (i.e. group six).

Consistent with Table 3, the coefficients for information asymmetry, DS imbalance and the change in DS imbalance (a, b2 and b1, respectively) increase from Pre- to Post-anonymity. For example, asymmetric information increases from 1.7883 to 2.0068 in the group of largest firms (Group 1) and from 16.0259 to 21.0070 in the group of smallest firms (Group 6). These results support Foucault et al. (2007), who reported that adverse selection increases as anonymity increases. The three coefficients show an increase once the broker IDs are eliminated for all six groups and are statistically significant for all but the group of smallest firms (Group six). The results for the order book variables, b1 and b2, are especially noteworthy since the removal of broker IDs might lead one to conclude that the order book would be less informative regarding future price changes when in fact the order book is more informative in the anonymous environment. One likely cause is that during the Post-anonymity period, informed traders are participating more in the public limit order book via the submission of limit orders. This finding is consistent with the reduction in aggressiveness found by Comerton-Forde and Tang (2008).

Also consistent with Table 3, the coefficients for order flows (b3 and b4) decrease in magnitude with b3 becoming less negative and b4 becoming less positive. The decrease in magnitude is also consistent across all six groups for b3 and b4 with all but one group statistically significant at the 1% level. These results support the earlier conclusion that the first three coefficients have more influence on prices after anonymity while the last two coefficients have less influence on prices after the broker IDs are removed.

The coefficient representing adverse selection, a, is positive and increases as the company size becomes smaller. This result is consistent with earlier literature on the NYSE. On the ASX, in the Post-anonymity period, information asymmetry increases 2.0068 to 21.0070, an increase of 1050%, from the largest to the smallest companies. The coefficients for b1 and b2 also increase as the company size becomes smaller with b1 increasing from 0.1897 to 5.8943 (or 3107%) while b2 increases from 0.8834 to 6.1616 (or 698%). The trend for b3 and b4 is not as obvious across company size. This implies that past information contained in lagged order flows is less influential on prices. In contrast, information asymmetry, DS imbalance and the change in DS imbalance are more informative about price movements as company size decreases.

The DS imbalance, b2, and the change in DS imbalance, b1, were hypothesized to be positive across all variables because a positive value indicates an excess demand which should forecast a positive price change. However in the Pre-anonymity results, the negative values observed in groups 2, 3, 4 and 5 indicate that excess demand results in a negative price change. Figure 2 illustrates what happened at the market microstructure level. Notice at t2 there is more buying volume (bottom shaded area at t2) than selling volume (top shaded area at t2) which indicates a demand surplus, then at t3 a sell trade enters the market. Informed traders avoiding front-runners can explain these results. For example, if there was an informed trader in the market at t1 who wants to sell and he does not want to reveal his broker ID (i.e. submitting a limit order) then a simple avoidance strategy would be to watch for large buy limit orders and immediately take that liquidity using a market order. For an informed trader this allows him to sell without revealing his identity. Once the broker IDs were removed, there was little incentive to use this strategy and this is one explanation for why the negative coefficients in the Post-anonymity period all but disappeared. This result was investigated further (results provided upon request) by sorting the change in DS imbalance into quintiles and grouping them from negative to positive. After running regressions and comparing average coefficients, the large negative and large positive groups were found to both predict large negative changes in the midpoint price while the middle groups all predicted positive changes. This evidence supports the theory that informed traders were capturing large limit orders that recently entered the market causing the negative coefficients and supports the claim that the fear of front-running was a significant concern for informed traders.

Insert Figure 2 about here

6 Robustness checks

6.1 Testing Pre-anonymity data sets (reporting season)

This section is devoted to testing the robustness of the model described by equation 9. For the first test, two Pre-anonymity periods are examined on the ASX using the electronic order-driven model. A 15 day period in October 2005 (10^{th} to 28^{th}) is compared to the 15 day Pre-anonymity period, November 6^{th} to November 25^{th}. This October period is significant because it is during the annual reporting season for Australian companies and during the fifteen day period approximately one third of all companies reported their annual results. Since this model is testing for information entering the market, the information asymmetry coefficient is expected to be larger in the October period than the November period. Since both periods are before the removal of broker IDs, the other variables in the regression are not expected to change. Table 5 reports results for this robustness check and confirms that information asymmetry declines for four of the six groups and the other variables are not affected.

Insert Table 5 about here

6.2 Testing Post-anonymity data sets (January effect)

For the second robustness test, two Post-anonymity periods are compared. A 15 day period in January 2006 (the 4^{th} to the 24^{th}) is compared to the 15 day Post-anonymity period, November 28^{th} to December 16^{th}. The January period was chosen because it is after the introduction of anonymity and the period contains relatively few companies reporting annual results and no holidays. So we would expect all the coefficients to remain the same across the two Post-anonymity periods.

Table 6 reports that the information asymmetry coefficient dropped significantly in three of the six groups. The cause of this decline is not immediately obvious and two competing explanations warrant further investigation. First, it could indicate that the large number of annual general meetings that occur during the November period convey a significant amount of information to the market. Or, it could indicate that the market is adjusting to the change in anonymity and the January period is a more stable representation of the new market dynamics. Further investigation is required to determine which of these explanations is the most plausible. Regardless, it is encouraging to see that the other variables are relatively stable and information asymmetry for the largest group (i.e. top 50) does not change significantly.

Insert Table 6 about here

7. Concluding Remarks

The development of a model to estimate information asymmetry in an order-driven electronic market is long overdue. Using previous authors' bid-ask spread decomposition models for insights, we compare several models using matching notation which leads to the identification of the trade indicator as the key variable for extracting adverse selection from observable data. A general model is then tailored to an order-driven electronic market structure which has become commonplace in exchanges around the world. The prior information in the public limit order book and in the lagged order flow is recognized as being important in these markets and this information is specifically included in this model.

The new model is used to investigate the effects on intraday price formation during the switch to anonymity on the Australian Stock Exchange in late 2005. An examination of the period immediately before and after the event indicates that the risk of adverse selection increased subsequent to the shift to anonymity. The results also indicate that informed traders shifted their trading strategies once broker IDs were removed from the limit order book. Before anonymity, informed traders used market orders to avoid detection and to mask their trading intentions. While after anonymity, their fear of front-running reduced and they increased their use of the public limit order book via limit orders. This change in behavior not only resulted in information asymmetry increasing but also increased the impact of the order book on prices while the impact from lagged order flow decreased from Pre- to Post-anonymity. In addition, this result supports an increase in market efficiency since prior information from order flow became less important subsequent to the shift to anonymity.

The model is tested for robustness and stability by comparing estimates from the earnings reporting period that occurred immediately before the Pre-anonymity sample period. These results are encouraging for the model's validity since adverse selection was higher during the earnings reporting season, while there was no significant change in the other variables in the model. The January period after the Post-anonymity period is also used to test robustness. Although there was an unexpected decline in adverse selection, the other variables experienced only minor changes confirming the model's robustness and stability. In addition, the reported R-squared results vary around 20% independent of company size, suggesting that the model is well specified for large companies as well as small companies with less liquidity.

This new model is important because it is consistent with the underlying market structure in an electronic order-driven market. The model measures the risk of adverse selection and the factors that affect intraday price formation. The model lends itself to a number of further studies for the ASX and other electronic exchanges including intra-day and weekly seasonality, special trade types, very low liquidity stocks and other market structure changes.

References

Ahn, H.J. Ahn, J. Cai, Y. Hamao, and R.Y.K. Ho., 2002, "The Components of the Bid-Ask Spread in a Limit-Order Market: Evidence from the Tokyo Stock Exchange," *Journal of Empirical Finance*, 9, 399-430.

Ahn, H.J. Ahn, J. Cai, Y. Hamao, and R.Y.K. Ho., 2005, "Adverse Selection, Brokerage Coverage, and Trading Activity on the Tokyo Stock Exchange," *Journal of Banking and Finance*, 29, 1483-1508.

Chan, Y.C., 2000, "The Price Impact of Trading on the Stock Exchange of Hong Kong", *Journal of Financial Markets*, 3, 1-16.

Choi, J.Y., D. Salandro, and K. Shastri, 1988, "On the Estimation of Bid-Ask Spreads: Theory and Evidence," *Journal of Financial and Quantitative Analysis*, 23, 219-230.

Clarke, J. and K. Shastri, 2000, "On Information Asymmetry Metrics," working paper, Georgia Institute of Technology.

Comerton-Forde, C., A. Frino, and V. Mollica, 2005, "The Impact of Limit Order Anonymity on Liquidity: Evidence from Paris, Tokyo and Korea," *Journal of Economics and Business*, 57, 528-540.

Comerton-Forde, C. and K.M. Tang, 2008, "Anonymity, Liquidity and Fragmentation," *Journal of Financial Markets, forthcoming.*

de Jong, F., T. Nijman, and A. Roell, 1996, "Price Effects of Trading and Components of the Bid-Ask Spread on the Paris Bourse," *Quantitative Finance*, 4, 383-397.

Easley, D., N. Kiefer, M. O'Hara, and J. Paperman, 1996, "Liquidity, Information, and Infrequently Traded Stocks," *Journal of Finance*, 51, 1405-1436.

Farmer, J.D., L. Gillemot, F. Lillo, S. Mike, and A. Sen, 2004, "What Really Causes Large Price Changes," *Quantitative Finance*, 4, 383-397.

Foucault, T., S. Moinas, and E. Theissen, 2007, "Does Anonymity Matter in Electronic Limit Order Markets," *Review of Financial Studies*, 20, 1707-1747.

George, T.J., G. Kaul, and M. Nimalendran, 1991, "Estimation of the Bid-Ask Spread and its Components: A New Approach," *Review of Financial Studies*, 4, 623-656.

Glosten, L.R., 1994, "Is the Electronic Open Limit Order Book Inevitable?" *Journal of Finance*, 49, 1127-1161.

Glosten, L.R. and L.E. Harris, 1988, "Estimating the Components of the Bid/Ask Spread," *Journal of Financial Economics*, 21, 123-142.

Goodhart, C.A.E. and M. O'Hara, 1997, “High frequency data in Financial Markets: Issues and Applications,” *Journal of Empirical Finance*, 4, 73-114.

Hasbrouck, J., 1991, “Measuring the Information Content of Stock Trades: An Econometric Analysis,” *Journal of Finance*, 46, 179-207.

Henker, T. and J. Wang, 2006, “On the Importance of Timing Specifications in Market Microstructure Research,” *Journal of Financial Markets*, 9, 162-179.

Huang, R.D. and H.R. Stoll, 1997, “The Components of the Bid-Ask Spread: A General Approach,” *Review of Financial Studies*, 10, 995-1034.

Lee, C. and M. Ready, 1991, “Inferring Trade Direction From Intraday Data,” *Journal of Finance*, 46, 733-746.

Madhavan, A., M. Richardson, and M. Roomans, 1997, “Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks,” *Review of Financial Studies*, 10, 1035-1064.

Majois, C. and R. De Winne, 2003, “A Comparison of Alternative Spread Decomposition Models on Euronext Brussels”, *Brussels Economic Review*, 46, 91-135.

Neal, R. and S.M. Wheatley, 1998, “Adverse Selection and the Bid-Ask Spread: Evidence From Closed-End Funds,” *Journal of Financial Markets*, 1, 121-149.

O'Hara, M., 1995, “*Market Microstructure Theory,*” Blackwell Publishing, Malden, MA.

Roll, R., 1984, “A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market,” *Journal of Finance*, 39, 1127-1139.

Stoll, H.R., 1989, “Inferring the Components of the Bid-Ask Spread: Theory and Empirical Tests,” *Journal of Finance*, 44, 115-134.

Theissen, E., 2003, “Trader Anonymity, Price Formation and Liquidity,” *European Finance Review*, 7, 1-26.

Van Ness, B.F., R.A. Van Ness, and R.S. Warr, 2001, “How Well Do Adverse Selection Components Measure Adverse Selection,” *Financial Management*, 30, 77-98.

Waisburd, A., 2003, “Anonymity and Liquidity: Evidence from the Paris Bourse,” working paper, Texas Christian University.

			Price	Sprd	Rel Sprd	Trade Vol	Daily Vol
Pre	Avg	15	7.0150	0.01389	0.00380	5383	219,495,828
	St Dev		9.8827	0.01187	0.00251	5038	43,992,905
Post	Avg	15	7.0651	0.01397	0.00376	6395	218,132,206
	St Dev		9.2214	0.01167	0.00239	6590	26,181,196
	Wilcoxon			2231	404	9715	
	p value			0.07568	0.74866	<0.0001	

Table 1a
Summary Statistics
This table reports daily statistics surrounding the removal of broker IDs on the ASX. Statistics are calculated for each firm on each day and then the mean is reported for all firms within each period. The results from a non-parametric Wilcoxon signed rank test for the Pre- vs Post-anonymity periods are reported in the last two lines of the table.

	Nbr	Price	Spread	Rel Sprd	Trade Vol	Daily Vol	Freq	Min Freq
1	50	15.821	0.01360	0.00147	5,324	2,376,774	450.8	155.2
	50	15.952	0.01402	0.00149	6,108	2,321,776	395.4	134.6
2	50	7.343	0.01420	0.00260	6,345	982,212	181.7	24.1
	50	7.300	0.01419	0.00273	7,106	892,214	154.1	44.2
3	50	6.805	0.01422	0.00336	4,972	536,902	116.6	13.9
	50	6.875	0.01514	0.00334	5,883	583,553	98.5	10.6
4	46	4.076	0.01578	0.00469	3,645	259,267	78.8	8.2
	46	4.160	0.01619	0.00456	4,280	257,506	61.0	7.0
5	40	3.168	0.01454	0.00557	5,575	273,026	51.4	5.6
	40	3.214	0.01602	0.00591	6,350	313,131	43.4	4.7
6	35	2.193	0.01428	0.00753	6,917	155,313	29.6	6.9
	36	2.207	0.01378	0.00716	7,837	183,660	24.3	5.6

Table 1b
Descriptive Statistics
This table reports daily statistics grouped by market capitalization. Stocks with less than 50 observations (i.e. 50 trades) over each 15 day sub-sample were omitted. Statistics are calculated for each firm on each day and then the mean is reported for all firms within each period.

	Price	Spread	Rel Sprd	Trade Vol	Freq	Daily Vol
1	0.32	3.06	0.52	4.86	-7.67	-0.52
2	-0.21	-0.04	3.24	4.24	-11.94	-3.40
3	0.22	2.20	-0.58	7.37	-15.76	2.61
4	0.92	1.01	-2.02	7.47	-15.01	-0.21
5	0.36	2.13	5.28	5.29	-10.56	2.87
6	0.33	-1.73	-3.75	5.23	-9.98	6.29

Table 2a

T-stats for Means with different Variances

The t-stats calculated using the means/standard deviations from the Pre-anonymity period and the Post-anonymity period for each variable in each group. The groups contain the same stocks ranked by market capitalization used in Table 1b.

	Price	Sprd	Rel Sprd	Trd Vol	Freq	Day Vol
1	27.34%	1.29%	27.76%	0.01%	0.00%	79.37%
2	57.75%	79.37%	97.35%	0.02%	0.00%	66.50%
3	0.44%	37.72%	99.62%	0.01%	0.00%	53.26%
4	1.34%	39.10%	30.70%	0.01%	0.00%	65.53%
5	45.06%	27.58%	18.67%	0.21%	0.02%	87.42%
6	25.74%	83.49%	48.92%	7.94%	10.94%	52.02%

Table 2b

Paired T-test

The p-values represented as percentages from the paired t-test across the Pre- vs Post-anonymity groups from Table 1b.

	Price	Sprd	Rel Sprd	Trd Vol	Freq	Day Vol
1	20.91%	1.11%	42.37%	4.64%	0.01%	85.49%
2	58.56%	96.24%	10.63%	1.61%	1.15%	40.48%
3	0.56%	18.72%	74.51%	0.27%	0.03%	55.75%
4	1.86%	32.13%	22.37%	0.02%	0.00%	93.89%
5	47.38%	20.01%	12.11%	1.85%	1.76%	48.17%
6	25.90%	22.93%	10.90%	19.78%	8.17%	20.10%

Table 2c

Wilcoxon Sign-Rank Test

The p-values represented as percentages from the Wilcoxon sign-rank test across the Pre-versus Post-anonymity groups from Table 1b. The Wilcoxon sign-rank test is the nonparametric analog of the paired t-test.

	Pre Anonymity			Post Anonymity		
	Est	Std Err		Est	Std Err	Prob
a	6.8743	0.4215		7.7500	0.4360	<0.001
b1	0.0099	0.0980		0.7217	0.1079	<0.001
b2	2.2587	0.1344		2.5481	0.1146	<0.001
b3	-0.1917	0.0052		-0.1122	0.0046	<0.001
b4	0.6095	0.1391		-0.0024	0.1697	<0.001
k	-0.7030	0.1528		-0.0169	0.1088	<0.001
Total	260			260		

Table 3
Regression model estimates
This table reports regression model estimates for the 15 days before and after anonymity on the ASX. An OLS regression was performed using equation 9 for each stock then the average for each group was reported in the table. A Wilcoxon signed rank test was used to test the mean of the two paired samples.

	R^2	Tot	a	b1	b2	b3	b4	k
1st	0.210 0.217	50 50	1.7883 2.0068**	0.0508 0.1897**	0.7746 0.8834**	-0.1914 -0.1083**	0.4830 0.2884**	-0.1700 0.0514**
2nd	0.195 0.205	50 50	3.4967 4.3504**	-0.1080 0.3293**	1.2988 1.5732**	-0.1843 -0.1107**	0.7340 0.2577**	-0.6862 0.0322**
3rd	0.185 0.206	48 48	4.6286 5.5830**	-0.0542 0.4866**	1.8295 2.2910**	-0.1730 -0.0895**	0.6246 0.1579**	-0.8060 -0.1756**
4th	0.205 0.204	44 44	9.1021 10.0126*	-0.2901 0.8212**	2.4608 2.9189**	-0.2281 -0.1247**	0.7391 -0.2669**	-0.7231 0.7534**
5th	0.220 0.230	38 38	9.8480 14.1252**	-0.2821 3.2491**	3.1068 3.9570**	-0.2171 -0.1037**	1.0613 -3.2941**	-0.5367 2.0542*
6th	0.252 0.186	30 30	16.0259 21.0070	1.1969 1.7458*	4.9322 6.1616	-0.2144 -0.0511**	1.6616 1.4429	-1.7689 0.2272

Table 4

Pre- versus Post-Anonymity - Regression results by company size

This table reports regression results after dividing the sample into six groups based on market capitalization. In each group the results represent the average OLS regression results for equation 9. The top line in each group represents the results for the Pre-anonymity period running from November 6th to 25th and the bottom line shows the results for the Post-anonymity period covering November 28th to December 16th. A Wilcoxon signed rank test was used to test for a difference in the paired mean between each group with two stars representing significance at the 1% level and one star representing significance at the 5% level.

	R^2	Tot	a	b1	b2	b3	b4	k
1st	0.210 0.210	50 50	2.2216 1.7883**	0.0099 0.0508	0.8234 0.7746*	-0.1813 -0.1914	0.5440 0.4830*	-0.1034 -0.1700
2nd	0.202 0.195	50 50	4.5700 3.4967**	0.0536 -0.1080	1.4374 1.2988*	-0.1956 -0.1843	0.8069 0.7340*	-0.2431 -0.6862**
3rd	0.189 0.185	48 48	5.3461 4.6286**	-0.1196 -0.0542	1.8906 1.8295	-0.1789 -0.1730	0.8359 0.6246	0.0702 -0.8060**
4th	0.227 0.205	44 44	8.8747 9.1021	-0.2046 -0.2901	2.7569 2.4608*	-0.2288 -0.2281	2.0953 0.7391*	-1.0781 -0.7231**
5th	0.205 0.220	38 38	11.9948 9.8480**	0.2077 -0.2821	3.8002 3.1068**	-0.1891 -0.2171	0.6423 1.0613	0.4578 -0.5367*
6th	0.236 0.252	30 30	17.0285 16.0259	-0.1182 1.1969	4.4252 4.9322	-0.2312 -0.2145	1.1957 1.6616	-0.8124 -1.7689

Table 5
Two Pre-Anonymity Periods – Annual Reporting Season
This table reports regression results across the market capitalization groups in order to test the model across a high adverse selection period when annual reports are released. The annual reporting period covers October 10^{th} to 28^{th} and the Pre-anonymity period covers November 6^{th} to 25^{th}.

	R^2	Tot	a	b1	b2	b3	b4	k
1st	0.217	50	2.0068	0.1897	0.8834	-0.1083	0.2884	0.0514
	0.214	50	1.9226	0.2505*	0.8221*	-0.1073	0.2334**	-0.0384*
2nd	0.205	50	4.3504	0.3293	1.5732	-0.1107	0.2577	0.0322
	0.203	50	3.7304**	0.4466	1.4997	-0.0980	0.1956*	0.1298
3rd	0.206	48	5.5830	0.4866	2.2910	-0.0895	0.1579	-0.1756
	0.198	48	4.7808**	0.4438	2.0473**	-0.0868	-0.0921*	0.1759**
4th	0.204	44	10.0126	0.8212	2.9189	-0.1247	-0.2669	0.7534
	0.191	44	8.8193**	0.5699	2.7473*	-0.0772**	-0.3835	0.3755
5th	0.231	38	14.1252	3.2491	3.9570	-0.1037	-3.2941	2.0542
	0.234	38	11.1244*	1.0171	4.0082	-0.0968	-0.3114	0.0252
6th	0.186	30	21.0070	1.7458	6.1616	-0.0511	1.4429	0.2272
	0.228	30	16.8193	1.5514	5.7254	-0.0992**	0.3270	0.3316

Table 6
Two Post-Anonymity Periods – January Effect.
This table reports regression results across the market capitalization groups immediately after the change to anonymity and in a subsequent period in January. The Post-anonymity period covers November 28th to December 16th and the January period covers January 4th to 25th.

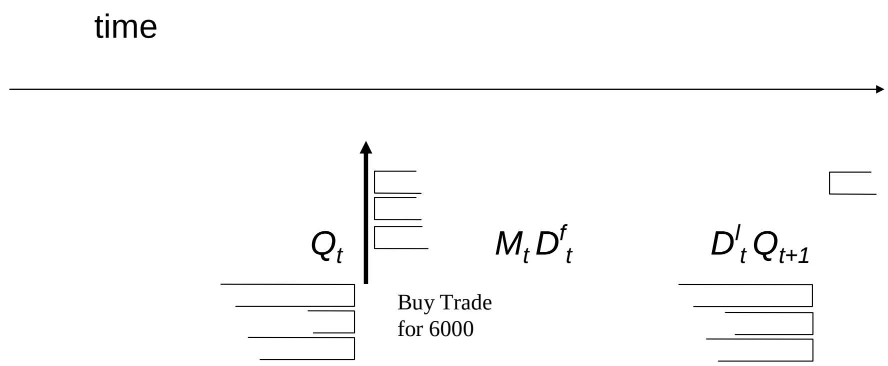
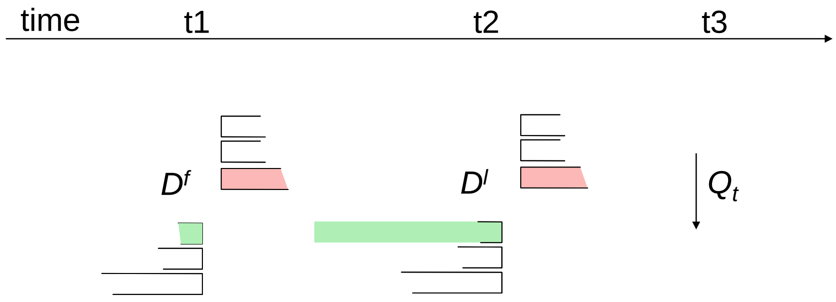


Figure 1
ASX Public Limit Order Book Example
This figure shows the change in the public limit order book following a buy order that takes out two price levels on the ask side of public limit order book highlighting the key observable variables. Q indicates the direction of the trade (+1 for a buy and -1 for a sell), M measures the mid-point price between the best bid and best ask and D is a proxy for the microstructure demand and supply imbalance and is measured by the log(best bid volume / best ask volume).



D^f – Larger Sell pressure therefore negative
D^l – Larger Buy pressure therefore positive

$D^l - D^f$ = large positive

Figure 2
Front Running Avoidance
This figure highlights how an informed trader avoids front-running in the Pre-anonymity period. When a large limit order appears in the order book at time t2 the informed trader who is selling immediately executes a sell at time t3 and avoids revealing his broker id in the public order book. In this example, a large positive DS imbalance or excess demand is followed by a downward movement in price (i.e. a sell) which runs counter to the microstructure theory of supply and demand presented in the article.