The effects of trade size and market depth on immediate price impact in a limit order book market

Manh Cuong Pham[1], Heather Margot Anderson*[2], Huu Nhan Duong[2], and Paul Lajbcygier[2]

[1] *Lancaster University Management School, Bailrigg, Lancaster, LA1 4YX, United Kingdom*

[2] *Monash Business School, Monash University, Wellington Road, Clayton, VIC 3800, Australia*

September 10, 2020

Abstract

We compare trade size to the prevailing market depth at the best level in the limit order book to detect and account for zero impact trades in an immediate price impact model. Our model also incorporates standard trade attributes (trade size, market capitalization and volatility) in a dynamic setting. The incorporation of market depth information reduces the mean absolute/squared forecast error of an immediate price impact prediction by about 60%. After controlling for trade attributes, market depth, price impact dynamics and intra-and inter- day periodicities (in order of relative importance) all improve the prediction of a trade's price impact. We demonstrate the value of our model by showing that splitting a big order into a series of smaller trades results in a reduction of between 60% and 82% of the immediate price impact cost of the big order. We also find that our depth indicator helps with the prediction of order flow and permanent price impact.

Keywords: Immediate Price Impact; Market Depth; Order Flow; Forecasts.

JEL classification: C53, C55, G10, G17.

Declarations of interest: None.

*Corresponding author. Tel: +61 3 9905 8462, Fax: +61 3 9905 5474.

Email addresses: m.c.pham@lancaster.ac.uk (M.C. Pham); heather.anderson@monash.edu (H.M. Anderson); huu.duong@monash.edu (H.N. Duong); paul.lajbcygier@monash.edu (P. Lajbcygier).

Acknowledgements: We are grateful for useful comments and suggestions from Neil Burgess, Pallab Dey (discussant), Joel Hasbrouck, Petko Kalev, Maureen O'Hara, Oliver Linton, Tālis Putniņš, Elvira Sojli, Avanidhar Subrahmanyam, Farshid Vahid, the Editor, an Associate Editor, anonymous referees and conference participants at the 2019 IAAE meetings (Cyprus), the 2019 ESEM meetings (Manchester), the 2019 FMA Asia/Pacific conference (Ton Duc Thang University), the 2018 International Workshop on Time Series and Panel Data Analysis (Melbourne, 2018), the 2018 Monash Financial Markets Workshop, the 10th Annual SoFie Conference (New York University, 2017), the 2016 Finance Research Symposium (La Trobe University), the 4th PhD Workshop in Economics (University of Minho), the 2015 CSIRO-Monash Superannuation Research Conference, the 1st Conference on Recent Developments in Financial Econometrics and Applications (Deakin University), and seminar participants at Cambridge University, Monash University, and RMIT University. We also thank Rohan Fletcher for research assistance. Anderson gratefully acknowledges the hospitality provided by Nuffield College at Oxford and CREI (Centre de Recerca en Economia Internacional) at Universitat Pompeu Fabra, where some of this work was undertaken. This research was funded by the CSIRO-Monash Superannuation Research Cluster, a collaboration across institutions and industry, among stakeholders of the retirement system in the interest of better outcomes for all. We would like to acknowledge the support of the Monash node of the NECTAR research, an initiative of the Australian government's Super Science Scheme and the Education Investment Fund cloud who provided IT infrastructure. We are grateful to the Securities Industry Research Centre of Asia-Pacific (SIRCA) for providing the data. All errors are our own.

1 Introduction

The ability to design optimal trades that minimize trading costs is of great interest to financial market participants. Explicit components of trading costs include bid-ask spreads and commission fees, but an implicit and much larger component of trading cost is market or price impact, which is the change in an asset price that results from a trade and is unknown ex-ante (Keim and Madhavan, 1996, 1998). Recent work on trading costs (e.g. Eisler et al., 2012, Zhou, 2012, Cont et al., 2014, Wilinski et al., 2015, Pham et al., 2017) has focused on the impact caused by a single trade *immediately* after its execution, and it demonstrates that trading volume, market capitalization and volatility are all important determinants of market impact. However, little effort has been made to explicitly incorporate market depth information into empirical models of immediate price impact. We study the empirical relationship between market depth and immediate price impact in this paper.

Our investigation is motivated by theoretical predictions and existing empirical findings on price impacts of trades. Theoretical models such as that by Kyle (1985) suggest that price impacts of orders are increasing in order size, as explored further by Foucault et al. (2013b). Theoretical and empirical work by Kraus and Stoll (1972) and Keim and Madhavan (1996) shows that price impacts are much larger for transactions that exceed available market depth at the best level. Knez and Ready (1996) find that the expected price improvement (measured as the difference between the transaction price and the prevailing bid or ask quote) is mostly dependent on "excess depth" (measured as the difference between quoted depth and order size). Expected price improvement is larger (in magnitude) the more negative "excess depth" becomes, but it essentially drops to zero once "excess depth" turns positive. Knez and Ready (1996) note that most of their observations fall into the latter category. Related later work such as that by Dufour and Engle (2000) has documented a high incidence of zero-impact trades.

The above cited work leads us to use information on quoted depth to identify zero-impact trades. We consider trades conducted in a limit order book market, and note that if a trade is smaller than the prevailing quoted depth at the best level on the opposite side of the order book, then it can be completely absorbed by the market depth at the best bid or ask price, and hence have zero immediate market impact. Trades that are larger than the available quoted depth will move the best bid or ask level after consuming all liquidity at the current depth, and will result in non-zero price impact. Thus, the quoted market depth information can be used in conjunction with trade size to form a zero-impact trade "detector", that equals zero when market depth does not support price impact and equals one otherwise. We relabel this zero-impact trade detector as a market depth indicator, and argue that its inclusion in a price impact model that is applied to tick by tick data is intuitively justified. Further, we show that the use of this "indicator" in an empirical model improves the out-of-sample forecast accuracy of immediate price impact by reducing mean absolute/squared errors by about 60%. Such (statistically significant) reductions translate into a total decrease of approximately $AUD 100 million per annum in the forecast uncertainty of price impact costs, offering clear potential for more accurate projections of the costs and profits of trading strategies.

We then use price impact models that incorporate our market depth indicator to investigate and quantify the effects of order splitting strategies. Kyle (1985), Easley and O'Hara (1987), Barclay and Warner (1993), Chakravarty (2001), Choi et al. (2019) and others have argued that traders often split a large order into several smaller orders to hide their information and avoid adverse price impact against their large orders. Also, as discussed in Hasbrouck and Saar (2013), Hendershott and Riordan (2013) and O'Hara (2015), institutions now use algorithms to implement dynamic trading strategies that split and sequence orders so as to minimize execution costs. Along with this, Chordia et al. (2011) have documented that institutions now resort to splitting orders in response to decreased market depth at the prevailing quotes. In our analysis, we assess the effectiveness of order splitting strategies by comparing the price impact of a series of small trades to the immediate price impact of an artificial trade that aggregates these consecutive small trades.

Our empirical modelling framework is built on a threshold principle, that sets expected immediate price impact equal to zero whenever our market depth indicator is zero. When our indicator is equal to one, our model follows a specification that incorporates standard regressors such as the trade characteristics used by Lillo et al. (2003) and Zhou (2012). We include Corsi (2009) type heterogeneous autoregressive (HAR) variables to account for underlying dynamics that drive orders, trade and price impact, together with day of the week and diurnal variables as well. The "switch" between zero price impact and (potentially) non-zero price impact is accomplished by treating the product of the indicator and a linear specification of the other variables as "the model", as we explain in further detail in Section 2.

We examine an Australian dataset of stocks drawn from the S&P/ASX200 index.[1] We find that a model that incorporates the market depth indicator, other theoretically motivated variables (trade

[1] Australian data has characteristics that make it particularly appropriate for a study of this nature. See subsection 3.1 for more detailed discussion.

size, market capitalization and volatility), price impact dynamics, and time of day/day of the week patterns consistently outperforms all other models that we consider (including a naive model that always predicts zero market impact), based on both out-of-sample mean squared error (MSE) and mean absolute error (MAE). We further show that, in addition to traditional trade attributes, (1) the information from market quoted depth, (2) the dynamics of immediate price impact, combined with (3) intra- and inter-day periodicities (in order of relative importance) contribute to the performance of our favored immediate market impact model.

We assess the effectiveness of order splitting strategies by using the most accurate of our set of price impact models to estimate the immediate price impact of an artificial unobserved large trade that aggregates a series of consecutive same-sign small trades on the same trading day - and find that the observed price impact of these split small trades is much lower than that associated with a single big trade. More specifically, splitting a big order into smaller trades can reduce the immediate price impact of the former by between 60% and 82% on average. This result confirms the effectiveness of the order splitting strategies that may be employed by informed traders in order to hide their information and reduce their price impact costs, as suggested in the literature (e.g. Kyle, 1985, Easley and O'Hara, 1987, Dufour and Engle, 2000), or by institutional investors who use order-splitting strategies to minimize their trading costs while rebalancing their portfolios (e.g. Keim and Madhavan, 1995, Forsyth et al., 2012, O'Hara, 2015, Choi et al., 2019, Korajczyk and Murphy, 2019, van Kervel and Menkveld, 2019).

We also study the relation between market depth information and future order flows, and highlight the use of depth information to predict future order imbalance. This imbalance is an important aspect of the dynamics of incoming orders that provides a link between the immediate price impact studied here and the permanent price impact discussed in Bessembinder and Venkataraman (2010). Our analysis shows that the use of depth information also helps with the prediction of permanent price impact.

We contribute to the literature on immediate price impact (for example Lillo et al., 2003, Zhou, 2012, Wilinski et al., 2015) by demonstrating that the use of our proposed depth indicator in immediate price impact models enhances their forecast accuracy by large and statistically significant margins. We measure some of the economic implications of the use of our indicator in immediate price impact models and find substantial benefits, both in terms of the reduction in the forecast error of price impact models and in terms of potential savings made by splitting large orders. We also find that our depth indicator is informative about future order flows and the gap between immediate price impact and permanent price impact. Given the importance of minimising trading costs, such findings have considerable practical relevance.

The rest of the paper is organized as follows. Section 2 outlines our model of immediate price impact and then discusses our depth indicator and other aspects of model specification that lead to greater precision in immediate price impact estimates. Section 3 reviews our data, relevant institutional detail and our research methodology. Comprehensive out-of-sample comparative results are discussed in Section 4, followed by discussions of the economic implications of a more accurate immediate price impact model in Section 5, and some analysis of how well our depth indicator signals future order flow and permanent price impact in Section 6. We include additional discussion of results relating to a nonparametric extension of our model in an appendix, and Section 7 concludes.

2 Immediate Price Impact Models

This section specifies our model of immediate price impact and discusses how it accounts for depth in the order book as well other characteristics of orders and trades. We use $\Delta p_{i,t}$ to denote the immediate price impact caused by the t-th transaction of stock i, calculated as the logarithmic change in the mid-prices of the bid and ask quotes immediately after (p_{i,t^+}) and before (p_{i,t^-}) that trade, i.e. $\Delta p_{i,t} = \ln p_{i,t^+} - \ln p_{i,t^-}$, and write our model as

$$\Delta p_{i,t} = \left[G(X_{i,t}, T_{i,t}) \times I_{v_{i,t} \geq depth_{i,t}}\right] + \eta_{i,t}, \tag{1}$$

where $X_{i,t}$ consists of "traditional" trade attributes (trading volume, market capitalization and volatility), $T_{i,t}$ consists of variables that capture the dynamics properties of price impact (such as flow on effects from previous transactions and time of day/week effects) and $I_{v_{i,t} \geq depth_{i,t}}$ is our market depth indicator that distinguishes our immediate price impact model from others in the literature. We define and discuss the variable $I_{v_{i,t} \geq depth_{i,t}}$ in subsection 2.1, and provide further details about the sets of variables that we include in $X_{i,t}$ and $T_{i,t}$ in subsections 2.2 and 2.3. We specify $G(.,.)$ as a linear function of $X_{i,t}$ and $T_{i,t}$ and highlight the features of the resulting model in subsection 2.4. The error term $\eta_{i,t}$ captures independent disturbances that have a mean of zero.

2.1 Market depth

We build on the intuition that a trade will not move the best ask (bid) price when the share volume of an aggressive buy (sell) is less than the quoted depth at the best ask (bid), and hence such a trade will have zero price impact. On the other hand, if a trade has volume greater than or equal to the available (other side) market depth, it will erase the currently quoted depth at the best opposite-side level, widen the bid-ask spread and generate a non-zero price impact. Thus, information about the quoted market depth in the opposite order book right before a trade is useful for predicting the price impact caused by that trade, and it can be used to detect zero and non-zero impact trades. We incorporate this idea into our price impact model by using a threshold-type dummy variable denoted by $I_{v_{i,t} \geq depth_{i,t}}$, which equals 1 if the volume of the t-th trade in stock i is greater than or equal to the quoted depth at the best level of the opposite side of the limit order book immediately before this trade, and 0 otherwise.

Our depth indicator enters Equation (1) multiplicatively, so that if $I_{v_{i,t} \geq depth_{i,t}} = 0$ then $G(X_{i,t}, T_{i,t})$ is not activated and we predict no immediate price impact. On the other hand, if $I_{v_{i,t} \geq depth_{i,t}} = 1$, then the trade consumes all depth at the best level on the opposite side of the order book, widens the bid-ask spread and leads to an expected price change of $G(X_{i,t}, T_{i,t})$. As detailed below, all the variables in $X_{i,t}$ and $T_{i,t}$ are measured before the trade takes place, so that when $I_{v_{i,t} \geq depth_{i,t}} = 1$, then $G(X_{i,t}, T_{i,t})$ is activated and delivers a forecast of the immediate price impact of the trade.

The inclusion of the depth dummy variable in a price impact model implicitly implies a negative dependence of immediate price impact on the prevailing market depth, as documented in Chan (2000), Engle and Lange (2001), Engle and Patton (2004), Hautsch and Huang (2012) and Brogaard et al. (2015). Further, our model specification in Equation (1) is able to deliver predictions that are consistent with previous analyses that have found that observed immediate price impact is often zero. Dufour and Engle (2000) report large proportions of zero impacts (from 72 to 92%) for their US stock sample dating from November 1, 1990 to January 31, 1991, and Zhou (2012) observes a 91% (89%) of filled buys (sells) with no price impact in his Chinese stock sample in 2003. Summary statistics of the variables in our study are presented in Table 1, and they suggest that about 60 to 80% of immediate price impact costs in our data are zeros. Given this, we expect that a naive model that always predicts zero price impact may have greater practical utility than classical immediate market impact models such as those due to Lillo et al. (2003), Almgren et al. (2005) and Zhou (2012), and indeed we demonstrate this in subsection 4.2.

2.2 Trade attributes

Standard predictors in price impact models include volume, market capitalization and volatility, and we use the collective term $X_{i,t}$ when referring to all three of these variables. We measure volume using $v_{i,t}/\overline{v}_{i,t}$, where $v_{i,t}$ is the share volume of the t-th trade, and $\overline{v}_{i,t}$ is the average volume of all trades in the trading day that are prior to (and *including*) the t-th trade (and have the same direction as that trade). Our market capitalization variable $M_{i,t}$ is calculated as the product of the mid-quote price and the number of shares outstanding for stock i just prior to the t-th trade, and our volatility measure $\sigma_{i,t}$ is calculated as the standard deviation of the mid-quote returns from the first trade of the day until just before the t-th trade. An important feature of these variables is that they are measurable ex-ante with respect to the t-th trade, and hence they are useful for forecasting immediate price impact.[2]

2.3 Time series characteristics

Pham et al. (2017) find that a model developed by Zhou (2012) forecasts price impact better than models that simply include trade attributes, and they attribute this forecasting superiority to the inclusion of an average price impact variable ($\overline{\Delta p}_{i,t}$) in Zhou (2012) that incorporates lags of $\Delta p_{i,t}$ and thereby captures some of the dynamic properties of price impact. Price impact dynamics can provide insights into market resiliency or how the limit order book replenishes after trades. For example, an observation of consecutive trades with high price impact might indicate that the liquidity supply in the order book is not replenishing sufficiently quickly, so that future trades might also be expected to incur high price impact.

Given these considerations, it is of interest to build some time series characteristics into our models of immediate price impact. ARMA models are standard in the literature, and these include heterogeneous autoregressive (HAR) structures developed by Corsi (2009) for modeling daily volatility. Corsi (2009) allows for heterogeneity in investors' trading horizons by incorporating moving averages of daily volatility over the past week and the past month into his volatility specifications. Here, we adapt this structure to our transaction-data context and include moving averages of past price impacts in our variable set $T_{i,t}$. Specifically, we define $\overline{\Delta p}_{i,t,n} \equiv \frac{1}{n}\sum_{j=1}^{n} \Delta p_{i,t-j}$ for $n \in \{1, 5, 20, 50\}$, where each trade

[2] In contrast, similar variables $\overline{v}_{i,t}$ and $M_{i,t}$ used in Lillo et al. (2003) relate to the entire year in which the t-th trade occurs, and such variables cannot be measured at the time of the trade since information about subsequent transactions in the remainder of the year is not yet known. Similarly, the volatility measure $\sigma_{i,t}$ used in Torre (1997) and Almgren et al. (2005) relates to the entire year of the t-th trade, and is therefore less useful for forecasting.

in the moving average occurs before the t-th transaction and has the same sign as the t-th transaction. The use of these variables is consistent with an assumption that there are various groups of traders who might close or reverse their trading position after 1, 5, 20 and 50 transactions. These transaction times correspond to about 0.4 (2), 2 (10), 8 (40), 20 (100) minutes for the biggest (smallest) stock group listed on the S&P/ASX200 index in 2007 as shown in the descriptive statistics for trade durations reported in Table 1.

Further, many researchers have shown that the frequency and aggressiveness of trades differ throughout the day (Admati and Pfleiderer, 1988, Bloomfield et al., 2005) and that they also vary over the week (French, 1980, Foster and Viswanathan, 1990). These intra- and inter-day periodicities reflect more distant price impact dynamics than the $\overline{\Delta p}_{i,t,n}$ variables discussed above. We capture their effects by adding a day of the week categorical variable ($day_{i,t}$) and a time of day categorical variable ($block_{i,t}$)[3] into our variable set $T_{i,t}$.

2.4 Our immediate price impact model

Our version of Equation (1) combines the details discussed in earlier subsections and is given by

$$\begin{aligned}\text{HARX}_{depth}: \qquad \Delta p_{i,t} = \Big[& a + \alpha\left(\frac{v_{i,t}}{\overline{v}_{i,t}}\right) + \beta M_{i,t} + \gamma \sigma_{i,t} + \delta_1 day_{i,t} + \delta_2 block_{i,t} + \phi_1(\Delta p_{i,t-1}) \\ & + \phi_5(\overline{\Delta p}_{i,t,5}) + \phi_{20}(\overline{\Delta p}_{i,t,20}) + \phi_{50}(\overline{\Delta p}_{i,t,50})\Big] \times I_{v_{i,t} \geq depth_{i,t}} + \eta_{i,t}, \qquad (2)\end{aligned}$$

where the "HAR" portion indicates that the model includes all time series variables in $T_{i,t}$, the "X" indicates that the model includes all trade attributes in $X_{i,t}$ and the *depth* subscript indicates that the model incorporates the market depth indicator and hence allows for non-zero price impact when the volume traded is larger than the quoted depth at the best level on the opposite side of the order book.

Relative to other immediate price impact models that typically just include trade attributes as explanators, our HARX_{depth} model embodies two novel features. The first of these is our depth indicator, that allows for zero price impact when market depth is sufficiently high. Although depth has previously been used as a standard explanatory variable, it has not been used as an indicator, nor has it been used in multiplicative form to ensure that no price impact is predicted when observed market conditions

[3]Each trading day is partitioned into six blocks (10:10-11:00, 11:00-12:00, 12:00-13:00, 13:00-14:00, 14:00-15:00 and 15:00-16:00), and the variable $block_{i,t}$ is the time block during the day in which the t-th trade in stock i occurs. All trades in the first 10 minutes of each trading day are excluded from the analysis to avoid the effects of the ASX opening procedure.

do not support a change in price. A second novel feature is our context specific adaptation of HAR lag structures to ensure that our modeling of price changes takes account of underlying dynamics in prices that reflect order flows, and how these flows change over time. HAR variables are common in the financial volatility literature, but to our knowledge they have not yet been used in the price impact literature.

3 Data and Research Methodology

3.1 Data

Most previous empirical work on price impact uses US data. In contrast, we work with Australian data provided by the Securities Industry Research Centre of Asia-Pacific (SIRCA). We choose the Australian market for several reasons. First, unlike the US stock market which has a high degree of market fragmentation with eleven equity exchanges and many alternative trading systems (O'Hara, 2015), the Australian stock market was unfragmented until October 2011 and is still relatively unfragmented.[4] The use of Australian data mitigates issues associated with aggregating order flows across different trading venues, and enables a (near-) complete analysis of the trades of interest. The low level of market fragmentation in Australia also alleviates the potential issue of vulnerability of liquidity supply in a severely fragmented market during periods with large liquidity demand, as highlighted in Menkveld and Yueshen (2019).

Second, the Australian stock market is an electronic limit order book market, rather than a quote driven market. Trades in the Australian stock market cannot be executed within the bid-ask spread unless they are executed against hidden orders queued inside the bid-ask spreads,[5] and the proportion of within bid-ask spread trades over our period of analysis is very small (less than 1%). This level of transparency provides an ideal setting for using quoted market depth to detect zero-impact trades, and since most major financial markets around the globe are also limit order book markets (Goettler

[4]All Australian listed stocks were traded on one exchange (the Australian Securities Exchange (ASX)) until 31 October 2011, when a second trading platform called Chi-X Australia was launched. The on-exchange trade share of Chi-X remained below 2.0% for the first six months, but it had reached 5% by late 2012 and 10% by late 2013. This change in market structure was not accompanied by a "no trade-through" policy such as that in force in the USA, and traders have had the freedom to place orders on either (or both) trading platform(s), (see Aitken et al. (2017)).

[5]We only examine trades on the lit Australian Securities Exchange and do not consider either block trades manually negotiated in an upstairs market or dark trades executed in Australian dark pools. For investigations of block and dark trading, which are beyond the scope of our study, see Bessembinder and Venkataraman (2004), Zhu (2014), Boulatov and George (2013), Comerton-Forde and Putniņš (2015), Kwan et al. (2015), amongst others.

et al., 2009, Malinova and Park, 2013), our findings are likely to have widespread applicability.

Third, the Australian data contains explicit information on whether each trade is buyer or seller initiated, which allows precise determination of the sign of each price impact and avoids the need to use an indirect method such as the widely used Lee and Ready (1991) algorithm which has an accuracy rate of only about eighty-five percent (Odders-White, 2000). It follows that findings based on the Australian data are likely to be more reliable.

Finally, some equity exchanges operate a make/take fee structure that, on top of charging the usual bid-ask spreads, charges liquidity takers additional fees while giving rebates to liquidity makers (Foucault et al., 2013a, Malinova and Park, 2015). This allows the passive party of a trade (who submitted a limit order and provided liquidity) to enjoy a very small or even negative price impact if associated rebates are sufficient to offset the adverse price change. In such markets, the observed ex-post price impact might not fully represent the cost an investor has to pay for immediate execution (net of other explicit trading cost components such as bid-ask spreads or commission fees), and one has to adjust for the make/take fees to determine net price impact. The Australian stock market does not feature make/take fees, which simplifies our analysis by making such adjustment unnecessary.[6]

Similar to most other electronic limit order book markets, orders submitted to the ASX follow a price-time priority. In particular, limit orders are queued and ranked in the limit order book first by price priority and then in the time sequence that they arrive at the market. Meanwhile, market orders, which are orders with the highest price priority, are executed at the best available prices immediately upon their submission. The limit order book is updated instantaneously whenever an order submission, revision, cancellation or execution occurs, and the submitted price of an order must be in multiples of the minimum tick size, which is pre-specified by the exchange and is dependent on the price level of the security.[7]

A typical trading day consists of two sessions: a pre-market session from 7:00am to 10:00am Australian Eastern Standard Time (AEST), and a normal trading session from 10:00am to 4:00pm AEST. The first 10 minutes of the normal trading session are opening auctions. There is also a closing single price auction between 4:10pm and 4:12pm during which the daily closing price for each stock is

[6]We note that make/take fees are typically fixed and known to traders, so that adjustment for these fees is quite straightforward. Thus, our work could be adapted to study a limit order book that imposes make/take fees.

[7]The tick size is $AUD 0.001, 0.005, and 0.01 for stock prices that are below $AUD 0.1, from $AUD 0.1 but below $AUD 2, and from $AUD 2, respectively. The relevant tick size for most trades in our sample (93%) was 0.01, with the relevant tick size for the remainder being 0.005.

determined (see https://www.asx.com.au/about/trading-hours.htm).

We study stocks in the S&P/ASX200 index, which provides a broad representation of the Australian equity market and covers more than 80% of Australian equity market capitalization. The set of stocks in this index changes when the index is rebalanced, and we focus on those stocks that remain in the index (with an unchanged ticker) over the entire time-span of our data (2007 - 2013).[8] Our resulting sample consists of 92 stocks, and we divide them into five groups according to their average market capitalization over the sample period. Group 1 consists of the 12 stocks with the largest market capitalization in our sample, and Groups 2 - 5 contain 20 stocks each and consist of successively lower capitalized stocks.

We collect two datasets from the SIRCA database. The first dataset records details on every order submitted to the central limit order book, including stock code, order type (submission, revision, cancellation and execution), date and time, order price, order volume (number of shares), order value (dollar value), and order direction (buy or sell order). Each new order is assigned a unique identification number (ID) so that the order can be tracked from its initial submission through to any revision, cancellation or execution. Since one large buy (sell) order can be matched against several orders on the sell (buy) side and therefore result in multiple simultaneous transactions, trades executed at the same time and initiated by the same order are aggregated into one trade, as is standard in the literature (see, e.g. Jondeau et al., 2015). We classify trades into buyer-initiated and seller-initiated transactions based on the directions of the (marketable) orders that initiate each trade.

Our second dataset contains detailed intra-day information on stock code, date, time, and the best bid and ask quotes in the limit order book. We remove all observations with a negative bid quote or ask quote, any observations with zero volume and any observations with a higher bid quote than ask quote. We merge the transaction dataset with the bid and ask quotes data to determine the bid-ask midpoint before and after each transaction. Finally, we collect daily data from the DatAnalysis Premium database (https://datanalysis.morningstar.com.au/) on the numbers of shares outstanding for each stock.

[8]We do this because we require long and complete time series of data for each stock to undertake appropriate forecast analysis. Our sampling mechanism raises the possibility of "survival bias" in our analysis, but our results can be validly viewed as being conditional on stock survival. We discuss this further in Section 4.2.5.

3.2 Research methodology

We assess the value of our HARX$_{depth}$ model in Equation (2) by comparing its ability to forecast immediate price impact with a set of other models that are nested within the HARX$_{depth}$ specification. Our comparator models differ from each other in that some incorporate the depth indicator $I_{v_{i,t} \geq depth_{i,t}}$ while others do not and some include the HAR variables while others do not. This allows us to assess the separate contributions of the depth indicator and price impact dynamics towards forecasts of immediate price impact. The full specifications of all models that we use for this purpose (and the acronyms that we use for them) are given in Appendix A. We include the models by Zhou (2012) and Lillo et al. (2003) in our comparison as well, since these traditional models specify a nonlinear power-law dependence of immediate price impact on trading volume, and hence they provide additional well known nonlinear comparators for our linear models. In addition, we include a "naive" model that always predicts zero immediate price impact for each trade. This last model acts as a benchmark for other models that include predictors.

Hendershott et al. (2011) note significant reductions in trading frictions, higher trading frequencies and improvements in stock liquidity over recent years. As a result, we expect that price impact models calibrated using data for one period may not produce accurate predictions for price impact of trades for another, especially if both periods are long and there have been significant changes in market microstructure within each period. Therefore, we use rolling window estimation and forecasts to address this issue. Specifically, we use the first 9 months (January-September 2007) to estimate different price impact models and then use the estimated models to make out-of-sample predictions for all transactions in October 2007. We then roll our data sample one month forward, using 9 months of data starting from February 2007 to re-estimate the models and produce predictions of price impacts in November 2007. The one-month-ahead rolling window procedure is repeated 75 times until we have forecasted price impacts for December 2013. Note that all continuous variables in the study are winsorized at the 1$^{\text{st}}$ and 99$^{\text{th}}$ percentiles to avoid the effects of outliers on in-sample fitting and out-of-sample prediction. The winsorization is done on a stock-by-stock basis, separately for buys and sells, and separately for each rolling window.[9]

[9] We obtain qualitatively similar results when the winsorization is applied to the larger panel dataset for each stock group in each rolling window.

(a) In-sample fitting

We estimate separate price impact models for buys and sells (for each stock group in each rolling window) to allow for potential asymmetries between buys and sells. In particular, all transactions for all stocks in each group over each nine month period are combined into a pooled dataset that is used to estimate the parametric models via least squares.

(b) Out-of-sample comparisons

Each of the estimated models based on each nine-month rolling window is used to make out-of-sample predictions of price impact ($\widehat{\Delta p_t}$) for the following month, with separate models being used for buys and sells of each stock group. We assess the out-of-sample predictive accuracy of one model relative to another using their forecast errors - i.e. the difference between their predicted out-of-sample and realized price impacts. We employ the approach developed by Giacomini and White (2006) (GW) to test the relative *conditional* predictive accuracy between any two models. More specifically, we examine whether the relative performance of these two models changes with some conditioning variable that is set to be the lagged relative performance as in Giacomini and White (2006). The GW conditional test is implemented in a pairwise fashion for all price impact models and for each rolling window. In addition, pairwise tests are carried out for the whole out-of-sample period which combines all out-of-sample one-month periods together.[10,11]

It is well known that statistical inference based on a comparison of p-values obtained from multiple pairwise tests may not be valid, and that the overall size of combining multiple pairwise tests may not coincide with the size of each individual test. We address this issue by using "Model Confidence Set" (MCS) tests due to Hansen et al. (2011) that allow a given set of competing models to be compared simultaneously against one another in a way that circumvents size distortion. The test identifies a non-null subset of models (known as the MCS) that contains the unknown best model with a pre-specified level of confidence by iteratively filtering out weaker competitors, so that the implicit null hypothesis is that the models in the MCS outperform the other competitors. The number of models included in an MCS depends on the informativeness of the data. An MCS might contain just a single dominating model if the data is informative, but it might also contain several or even all models. We

[10] It is the whole out-of-sample GW test that motivates us to set the out-of-sample length to be one month. If a two (or more) month period was chosen for the out-of-sample period instead, we would have to roll the window ahead by the same number of months which would reduce the number of rolling windows significantly.

[11] We also analyze, but do not report, the relative *unconditional* predictive accuracy between any two models using the unconditional version of the Giacomini and White (2006) test and the Diebold and Mariano (1995) test. Results for these tests are qualitatively similar to the conditional tests and are available upon request.

implement the MCS test to assess the predictive accuracy of the set of all price impact models in each out-of-sample month.[12]

4 Results and Discussion

4.1 Descriptive statistics

Table 1 provides summary statistics for individual trades for the 92 stocks that we study. In total, we examine 232,419,333 individual trades, consisting of 119,749,509 buyer-initiated and 112,669,824 seller-initiated transactions. There are more than 79 million trades in the group with the largest market capitalization (Group 1) and almost 21 million trades in the smallest market capitalization group (Group 5). Although Group 1 has only 12 stocks while other groups have 20 stocks each, the former still has many more observations than do any of the latter. Price impact is largest for the last (and least frequently traded) of these groups, and smaller for more highly capitalized stocks, which is consistent with the findings documented in Lillo et al. (2003), Lim and Coggins (2005) and Bouchaud et al. (2009). For all five stock groups, price impact is highest during 2008-2009 as a consequence of the Global Financial Crisis. Furthermore, the price impact of a sale (tabulated in Panel B) is on average larger in absolute value than for a purchase (tabulated in Panel A), as found by Karpoff (1987), who suggested that this might be due to short sale constraints. We also note that for both buys and sells, the absolute value of price impact is smaller when the average (opposite side) prevailing quoted depth at the best level is larger.

<<INSERT TABLE 1 ABOUT HERE>>

Trading volume per trade, either in dollar value or in the number of shares, generally decreases over time, and a similar pattern is observed for trade durations. These observations are consistent with the increasing presence of algorithmic and high frequency traders who tend to trade more frequently with a smaller trading volume per trade (e.g. ASX, 2010, Chordia et al., 2011, Brogaard et al., 2014). Moreover, the average scaled volume is less than one, showing that trading volumes are positively

[12] The huge number of observations over 75 out-of-sample months makes the calculation of the Hansen et al. (2011) MCS test challenging, because this test relies on simulation. Thus, we do not run the MCS tests for the whole out-of-sample period. Instead, we conduct the MCS test for each out-of-sample month by employing the Politis and Romano (1994) stationary bootstrap method with 1500 bootstrap replications, and an average block-size of 10 to allow for serial correlation in the loss series produced by each model. We set the significance level at 5%, and hence work with a confidence level of 95%.

skewed, as noted elsewhere (e.g. Andersen, 1996, Manganelli, 2005, Menkhoff et al., 2010). Similar to price impact, volatility is smaller for the more highly capitalized stock groups. Volatility increased substantially during 2008-2009, but has become relatively stable since 2010. Prevailing quoted depth at the best price tended to increase over 2007 - 2010, but has reduced substantially since 2010.

The majority of trades (between 55% and 85%) do not incur any immediate price impact for buys or for sells, with the proportions of zero-impact trades being larger for the less capitalized stocks. Meanwhile the market depth indicator shows that a larger proportion of trades (between 59% and 85%) are smaller than the available quoted depth at the best level in the opposite book and should, in theory, incur zero market impact. Across all stock groups, the difference between the percentage of trades for which the depth dummy equals zero and the percentage of trades for which there is no price impact is negligible between 2007 and 2011, but then it increases sharply to about 3% in 2012 and 7.3% in 2013 on average.

We call transactions that move the midquote and have a non-zero immediate price impact despite having less volume than the prevailing opposite-side depth "anomalous trades", and we have examined them in detail. We find that about 80% of such trades in 2012 and 2013 are a consequence of multiple (groups of) trades that occur at the same millisecond time instance in our sample but are not initiated by the same order.[13] These trades, which are not aggregated into a big single trade in our algorithm, have lower individual volumes than the prevailing opposite-side depth, and thus their depth indicators are all zero. However, the total volume of these trades is larger than or equal to the prevailing depth, suggesting that there is a new best price after the execution of these trades. Even though only the last (few) trade(s) move(s) the price, all of these trades are recorded as non-zero impact trades, since the midquote right after their execution millisecond is different from the previous midquote.

Simultaneous entries at the same millisecond time stamp also lead to another 15% of our "anomalous trades" in 2012 and 2013. This is because there are other delete (enter) orders occurring at the same millisecond which remove (add) liquidity from (to) the market and hence introduce noise when we identify the mid-points before and after the trades. Hidden orders (such as iceberg orders) that are queued *inside* the bid-ask spread when the bid-ask spread is larger than one tick can also lead to

[13]It is likely that these trades were executed at different time instances that were finer than a millisecond, but due to the millisecond resolution of our data, they were recorded at the same millisecond time stamp. The observation that there are more such "simultaneous" trades in 2012 and 2013 is consistent with the rise in algorithmic and high-frequency traders in recent years, with trading latency that is much smaller than a millisecond (O'Hara, 2015). We thank a referee for pointing out that Menkveld (2018) observed groups of "simultaneous" trades (on the NASDAQ), and suggested that they stem from high frequency races on common signals.

anomalous trades. However, such instances occur much less frequently and account for less than 1.5% of the anomalous trades in 2012 and 2013.

Meanwhile, a comparison of the last two columns of Table 1 shows that trade sizes are bigger than the available depth with no price impact for less than 1% of all observations. This is due to trades arising from hidden orders queued *at* the best bid and ask prices, with such orders being omitted from the calculation of the depth indicator. Although these trades have a larger volume than the visible prevailing depth, they are smaller than the total visible and hidden depth at the best price. Consequently, they do not move the best price and have a zero immediate price impact.

4.2 Empirical results

Comparisons of different market impact models tend to focus on measures of in-sample goodness of fit. However, such comparisons can be misleading because a good in-sample fit can sometimes result from over-fitting, and inference based on overfitted models need not necessarily be valid (Hurvich and Tsai, 1990). Therefore, we compare the out-of-sample performance of our various price impact specifications with natural benchmarks (such as the naive model) to provide reassurance that they are not simply artifacts of our estimation sample. See Appendix A for full specifications of all models that we study, and the acronyms that we use for them. We provide a thorough out-of-sample forecast evaluation, complete with statistical tests, of our immediate price impact models over the whole out-of-sample period (subsection 4.2.1), as well as across different out-of-sample months (subsections 4.2.2 and 4.2.3). Subsection 4.2.4 then provides a detailed analysis that disentangles the contribution of the various features of the HARX_{depth} model that lead to its superior forecast performance,[14] and subsection 4.2.5 discusses the outcomes of our robustness checks.

4.2.1 Summary out-of-sample predictive accuracy comparisons of all models.

Table 2 reports the out-of-sample MSEs and MAEs of various immediate price impact models for all five S&P/ASX200 stock groups over the full out-of-sample period, which spans October 2007 - December 2013. Since Group 1 is the most highly capitalized and heavily traded stock group while Group 5 is the least capitalized and traded group, price impact and its prediction uncertainty (quantified by MSEs and MAEs) are typically smallest for trades in Group 1 and largest for trades in Group 5.

[14] Measures of in-sample fit are qualitatively similar to the out-of-sample outcomes. They are omitted for brevity but are available upon request.

Table 2 shows that the HARX_{depth} model that incorporates market depth and other theoretical and time series information, produces more accurate predictions for immediate price impact than any other model, regardless of stock grouping and direction of trade. LX_{depth}, the baseline (static) analogue of HARX_{depth}, is the second most accurate price impact model. It is noteworthy that each of these depth augmented models outperform their respective HARX and LX counterparts by very large margins. In addition, a naive model that always predicts zero price impact throughout the entire out-of-sample period significantly outperforms all other models that do not use the market depth information with respect to MAE.[15] We conclude that our market depth indicator provides a very effective modelling tool. It reflects economic intuition that underlies the incidence of observed zero price impact trades in an appropriate way, and it not only improves the forecasting performance of immediate price impact models by statistically significant margins (shown later, in Table 3), but it also leads to an economically significant reduction of about 60% in a model's MSE and MAE (see Section 5.1).

<<INSERT TABLE 2 ABOUT HERE>>

Table 2 shows that after the HARX_{depth} and LX_{depth} and (for MAE) naive models, those models that incorporate time series aspects of immediate price impact into their specifications such as HARX or ZHOU, have considerably higher predictive accuracy than the remaining models. Further, consideration of the three linear models without the depth indicator (i.e. HARX, LXb and LX) shows that HARX has the highest predictive accuracy for immediate price impact while LX is the least accurate model. These results, together with the observation that HARX_{depth} outperforms LX_{depth}, suggest that one or both types of time-series variables (i.e. HAR type dynamics and/or intra- and inter-day periodicities) are important predictors for price impact modeling. This agrees with previous studies such as Admati and Pfleiderer (1988), Foster and Viswanathan (1990), Bloomfield et al. (2005) and Pham et al. (2017).

Comparison between the LX and LFM2 models shows that the latter predicts out-of-sample immediate price impact considerably more accurately than does the former for all (most) stock groups according to MAE (MSE). Given that the LFM2 model allows for nonlinear relationships between price impact and the traditional trade attributes while the LX model does not, we speculate that further

[15] The observation that the naive model has higher MSE relative to models that do not incorporate the depth indicator is attributable to the 30% of trades that have non-zero impact (see Table 1). The naive model often makes substantially large forecast errors in cases that have non-zero impact, whereas the errors produced by other models are usually less extreme. The observation that MSE ranks the naive model quite differently from MAE is then easily explained by the fact that MSE penalizes large forecast errors very heavily whereas MAE does not.

allowance for nonlinearities in the data might also lead to improved forecasts of immediate price impact. We develop and estimate some nonparametric extensions of our models to study this question in Appendix B, and find that these models lead to improved forecasts of immediate price impact (relative to parametric models).

We also note that the out-of-sample MAEs and MSEs are markedly lower for buys than for sells for all models and stock groups, and t-tests indicate that these differences are statistically significant. This result is quite interesting, but it may reflect the observation that the immediate price impact of a purchase is on average of a smaller magnitude than that of a sale, as shown earlier in Table 1.

Of the models that we study here, we find that the HARX_{depth} model provides the most accurate out-of-sample predictions of immediate price impact. After controlling for trade attributes that are typical in the extant literature, we demonstrate that the new features that we include in our HARX_{depth} specification, namely market depth, price impact dynamics and intra-and interday periodicities, all improve the prediction of a trade's price impact. We establish the statistical significance of these improvements in the following subsections.

4.2.2 Out-of-sample MSE/MAE over time

We have shown that the HARX_{depth} model has the highest predictive accuracy among candidate models over the whole out-of-sample period, and it is now of interest to determine whether this strong performance is observed over shorter sub-samples of time. To answer this question, we plot the MSE and MAE series of different price impact models for Groups 1, 3 and 5 in Figures 1 and 2, respectively (and obtain qualitatively similar figures for Groups 2 and 4).

<<INSERT FIGURES 1 AND 2 ABOUT HERE>>

The left and right panels of Figure 1 (Figure 2) show the MSE (MAE) performances for buys and sells respectively. We see that the MSEs and MAEs produced by alternative price impact models are greater than those generated by the HARX_{depth} model (solid line) for most of the 75 forecast windows, showing that the latter model is nearly always a superior forecasting model. The series for LIN_{depth} (dashed line) is closest to that for HARX_{depth}, showing that LIN_{depth} is the second best model[16] and that the results discussed in the previous subsection hold over different periods of time.

[16] Although it is hard to visually differentiate the series for HARX_{depth} and LX_{depth} in the graphs, (unreported) tests indicate that HARX_{depth} produces significantly smaller forecast MSEs and MAEs than does LX_{depth} in more than 90% of the out-of-sample months.

While the naive model (dash-dotted line with circle markers) produces the largest MSEs (compared to all other models) throughout the entire out-of-sample period, it is the third best model according to MAE, outperforming all models that do not incorporate the market depth indicator. Meanwhile, the performance of the three models, namely HARX, LX and ZHOU track each another quite closely. It is hard to determine which of these three models is best according to MSE, although it is clear that LX ranks behind these other two models according to MAE.

The MSEs and MAEs of all price impact models were largest during the Global Financial Crisis (GFC) but they have mostly decreased since then, except briefly in August 2011, when many stock markets (including Australia's) reacted to heightened fear that the European sovereign debt crisis might spread.[17] Interestingly, the predictive accuracy of all models improved slightly during the first few months after the second equities trading platform in Australia (Chi-X) was launched on 31 October 2011 (which is indicated by a vertical dashed blue line in each graph).[18] However, the forecast performance of all price impact models tended to decline after April 2012, and although this performance has fluctuated somewhat since then, the decline might be attributable to increased market fragmentation as Chi-X gained market share in Australia.[19] Another possible contributor to forecast deterioration of depth-augmented price impact models towards the end of our sample might have been increased trading frequency, which we associated with "anomalous trades" and less precision in our depth indicator in Section 4.1.

Overall, it is quite clear that the incorporation of the market depth information into different price impact models such as HARX and LX leads to a strong improvement in out-of-sample predictive accuracy, both in terms of MSE and MAE. Further, models that employ the market depth indicator outperform the naive model in terms of MSE and MAE. Section 4.2.3 shows that the improvements in both MAE and MSE contributed by the market depth information are statistically significant across all 75 out-of-sample months, despite the slight deterioration in the performance of our depth indicator towards the end of our sample.

[17]See: https://www.theguardian.com/business/2011/aug/04/stock-markets-exchange-plunge-business.

[18]Improved predictability in market depth due to increases in algorithmic trading post 2011 (see Jovanovic and Menkveld (2019)) might be a contributing factor here.

[19]The inserts in Figures 1 and 2 focus on the performance of the HARX and HARX_{depth} model over the four year window that is centered around the Chi-X launch. Tests based on the MSE (or MAE) of forecasts from HARX and HARX_{depth} models using windows of $\pm$ 1 month, $\pm$ 3 months and $\pm$ 6 months around the launch date for each of our five stock groups, all found statistically significant improvements in forecasting accuracy after the launch, whereas the same tests based on windows of $\pm$ 1 year and $\pm$ 2 years found statistically significant deteriorations.

4.2.3 Percentage of outperformance across all rolling windows

We check the statistical robustness of the superior performance of the model utilizing market depth information and price impact dynamics (HARX_{depth}) across 75 rolling out-of-sample periods of one-month, by performing Giacomini and White (2006) conditional predictive ability tests and Hansen et al. (2011) MCS tests at a 5% significance level. The GW conditional tests compare the relative accuracy of price impact predictions produced by different models in a pairwise fashion. Thus, we use $n(n-1)/2$ pairwise GW tests to compare n models. In each out-of-sample month, we record the model that strictly outperforms all remaining models with respect to MSE or MAE, and by construction, there is at most one winning model in any out-of-sample month. We summarize the performance of all models over 75 out-of-sample months by tabulating the percentages of each price impact model that strictly outperforms all other models. As noted earlier, the multiple use of GW tests is subject to size distortion, so we also use the MCS test developed by Hansen et al. (2011) to simultaneously compare all n out-of-sample price impact series predicted by n potential models, and present our results in a different way. Only one (iterative) test is needed in each out-of-sample month and this returns a set of the best model(s) with a 95% confidence level and its complement of statistically under-performing models. Since there is at least one model contained in an MCS for each month, we aggregate the results over 75 out-of-sample months by reporting the percentages of months for which each model belongs in the set of the best models.

The results of this analysis are reported in Table 3, which provides further forecast analysis relating to the models studied in Table 2. The analysis relates to all n=*9* models specified in Appendix A, but we only report the results for two models because the results for the remaining models were all zero. Panel A reports the results based on GW conditional tests, while MCS results are shown in Panel B. Since in each out-of-sample month there is at most (at least) one best model according to the GW (MCS) testing procedure, the percentage of outperformance for each model based on the GW tests is no greater than that implied by the MCS tests. Moreover, for each category (e.g. buys of Group 1 or sells of Group 3) the sum of the outperformance percentages for all competing models under consideration (i.e. the row sum of the numbers in each Panel) according to the GW tests is no greater than 100%, while that based on the MCS tests is at least 100%.

<<INSERT TABLE 3 ABOUT HERE>>

Recalling that Table 3 only reports forecast results for those models that provide evidence of outperformance, we see from Panel A that the model incorporating depth information and price impact dynamics ($HARX_{depth}$) generates the most accurate price impact forecasts in more than 93% (86%) of the 75 out-of-sample windows when assessing forecasts using MSE (MAE). LX_{depth} is the second best model, although the numbers of times it produces the lowest MAEs or MSEs are far less than those for $HARX_{depth}$. Panel B shows that the $HARX_{depth}$ model belongs in the MCS for more than 97% (90%) of the forecast windows when MSE (MAE) is used, and although the LX_{depth} model is sometimes present in these MCS's as well, the proportion of times that it does so is much lower. These statistical results underlie the patterns that we noted when discussing Figures 1 and 2, and they show strong support for the model with depth and time series information ($HARX_{depth}$) relative to all other immediate price impact models that we consider.

4.2.4 Contributors of the superiority of the $HARX_{depth}$ model

There are several reasons why the $HARX_{depth}$ model might outperform standard models that only use trade attributes as predictors of price impact, since in addition to these predictors, the $HARX_{depth}$ model also incorporates market depth information, price impact dynamics and intra-/inter-day seasonal effects. We consider which of these factors makes the greatest contribution to $HARX_{depth}$'s overall strong performance below.

(a) Importance of the market depth indicator

We compare the performance of those models that incorporate the market depth indicator ($HARX_{depth}$ and LX_{depth}) with that of their non-depth analogues (HARX and LX) over the 75 out-of-sample months to see how the market depth indicator contributes to the superiority of $HARX_{depth}$, and more generally, to immediate price impact modeling. Our results are summarized in Panel A of Table 4.[20]

<<INSERT TABLE 4 ABOUT HERE>>

We see that the addition of the market depth indicator to the HARX and LX models leads to a statistically significantly improvement in out-of-sample MSEs and MAEs in all out-of-sample months considered, reducing out-of-sample MSEs and MAEs by about 60% on average.[21] Thus, the use of

[20]Table 4 reports the results of GW tests only, since each comparison relates to just two price impact models at a time.

[21]We obtain qualitatively similar results when we add the market depth indicator into our other models that do not account for depth.

market depth data to detect zero-impact trades is of high relevance for modeling and forecasting the immediate price impact of individual trades. This, together with results from Table 3 and Figures 1 and 2, shows that it is the use of market depth information that makes the largest contribution to the overall superiority of the HARX_{depth} model.

(b) Importance of price impact dynamics

A comparison of the predictive accuracy of HARX and LXb based on results in Table 2 shows that the modeling of price impact dynamics makes a strong contribution to the prediction of immediate price impact costs over the entire out-of-sample period, while Panel B of Table 4 shows that this contribution is statistically significant in all of the 75 rolling window periods, for both buys and sells. This holds for each stock group. Overall, the addition of price impact lags into LXb to obtain HARX leads to reductions of about 5-6% in MSEs and MAEs. While these reductions are smaller than those brought about by including the market depth information, they are statistically and economically significant. Therefore, as in Pham et al. (2017), we conclude that the autocorrelation in immediate price impact is an important factor when modeling price impact.

(c) Importance of intra- and inter-day periodicities

Previous studies have shown that the time of trades possesses informational content that contributes to the evolution of prices (Diamond and Verrecchia, 1987, Easley and O'Hara, 1992, Dufour and Engle, 2000). In particular, there are diurnal and inter-day patterns in stock returns (Admati and Pfleiderer, 1988, Foster and Viswanathan, 1990). Panel C of Table 4 lends support to these theories and demonstrates that taking the intra- and inter-day patterns into account enhances the performance of an immediate price impact model. In particular, the inclusion of the time of the day and day of the week effects statistically improves the predictive accuracy of LX in the vast majority (more than 94% (89%)) of the 75 out-of-sample months according to MSE (MAE), even though the reductions in these measures are on average relatively small.

Overall, although the incorporation of the market depth information brings the biggest contribution to the forecast accuracy of an immediate price impact model, the allowance for time series indicators (i.e. price impact dynamics and intra- and inter-day seasonalities) provides non-trivial enhancements to the forecast performance of an immediate price impact model.

4.2.5 Robustness

In this section, we briefly comment on the robustness of our study. Details of our sensitivity analyses are not reported here for brevity, but they are available upon request.

(a) Different in-sample rolling window length

The above forecast analysis is based on nine-month in-sample rolling windows, but we have also used different in-sample rolling window lengths (of 6 months and 12 months) to produce series of one month ahead out-of-sample forecasts, and have found that our results are qualitatively the same, with the model with the depth indicator (HARX_{depth}) dominating the other models that we consider. We also find that in addition to economic variables (i.e. trading volume, market capitalization and stock volatility), the time series variables improve the prediction of price impact.

(b) Augmentation of depth information to other non-depth models

The above analysis examines only two depth-augmented models, namely HARX_{depth} and LX_{depth}. Nevertheless, we have extended our analysis to include four additional depth augmented models, based on specifications in Appendix A. All our key empirical findings in Section 4.2.1 remain qualitatively unchanged. In particular, the HARX_{depth} is the most accurate immediate price impact model, followed by the LXb_{depth} and LX_{depth}.[22] All depth-augmented models significantly outperform their non-depth analogues and the naive model. Further, the LX_{depth} outperforms all remaining depth-augmented models and it also dominates the HARX model. This latter observation suggests that the depth indicator provides a proxy for the effects of past orders and trades on current immediate price impact.

(c) Analysis of big trades

We have also re-conducted our comparative analysis on a subsample of "big trades" - that we define to be trades that are larger than or equal to the prevailing quoted depth, and have found qualitatively similar results. Most noticeably, HARX, the most comprehensive model which is identical to HARX_{depth} when we focus on big trades whose depth indicator is equal to 1, produces more accurate predictions of the immediate price impact of big trades than any of the other models. Also, the forecast accuracy of an immediate price impact model is improved once one brings the time series variables into its specification.

[22] We also compare our nonparametric versions of the HARX_{depth} and HARX models in Appendix B and find that the former are more accurate.

(d) Exclusion of trades executed against hidden orders

We have analyzed immediate price impact for trades that were not executed against hidden orders, to assess the extent to which such orders might have affected our conclusions relating to all trades. A priori, we expect that hidden orders at the best level or within the bid-ask spread will lower the accuracy of our depth indicator and weaken the forecasting ability of those models that employ this indicator. This is what we find. In particular, models with market depth information (i.e. HARX_{depth} and LX_{depth} models) produce smaller forecast MSEs and MAEs (on average) when trades transacted against hidden orders are excluded. However, trades executed against hidden orders account for less than 1% of our total sample, and the hidden orders reduce the predictive MSE (MAE) measures for the depth indicator models in the total sample only slightly. Overall, we find qualitatively similar results, in terms of relative model rankings, to those based on all trades.

(e) Sample composition

Our sample of stocks consists of those that remained in the S&P/ASX200 index from 2007 to 2013, and although this selection removed several higher-cap stocks because of ticker changes, it still favoured the retention of relatively higher-cap stocks.[23] Therefore our numerical results are subject to this caveat. However, our presentation of (conditionally valid) results for each of five market-cap classifications, allows the reader to see how immediate price impact and the characteristics of its predictions change with different levels of market capitalization, and have confidence that the depth indicator and the HARX_{depth} model will be useful for a wide range of market capitalizations.

5 Economic implications of more accurate price impact models

Prior literature highlights that although it is unobserved ex-ante, price impact makes up the biggest component of total trading costs, especially for large trades (e.g. Keim and Madhavan, 1996, 1998, Almgren et al., 2005). Therefore, a more accurate price impact model will bring great benefits to market traders and fund managers, since it provides more accurate prediction of price impact costs of trades, which in turn will facilitate the design, implementation and evaluation of optimal investment

[23] We note that the average (median) level of market capitalization for excluded stocks was $AUD1,850m ($AUD868m), while the average market capitalization for our Group 5 stocks ranged from $AUD542m to $AUD1,441m, with a median of $AUD1,015m.

strategies in order to minimize trading costs.

In this section, we study in deeper detail two economic implications brought about by the incorporation of the market depth information into an immediate price impact model. These are (i) by how much can one reduce the forecast error of price impact costs after accounting for market depth, which is discussed in subsection 5.1; and (ii) how much price impact one can save by splitting a large order into smaller trades, which is discussed in subsection 5.2.

5.1 Reduction in the forecast error of price impact costs

After the market depth dummy is incorporated into the specifications of the HARX and LX models, their forecast MSEs and MAEs fall by about 60% (see Panel A of Table 4). There were more than 232 million trades over the seven years in our sample, and if each trade had an average size of $AUD 13,320 (which approximates the observed mean dollar value), the total turnover for the five stock groups from 2007 to 2013 was roughly $AUD 3,090 billion. If a model with market depth information such as HARX_{depth} or LX_{depth} had been employed, the average reduction in the MAE of predicting immediate price impact would have been about 2.2 (1.3) basis points compared to the non-depth analogue (naive model) (see Panel B of Table 2). This translates into a total $AUD 679.8 (401.7) million decrease in the forecast error or forecast uncertainty of price impact costs over seven years, or a reduction of about $AUD 97.1 (57.4) million per annum. This reduction can be reflected in more accurate projections of the costs and profits of investment strategies. Thus, the cumulative effect of using market depth information in a price impact specification could save investors millions of dollars per annum - showing that this information is potentially very useful for measuring and predicting immediate price impact.

5.2 Price impact savings from splitting large orders

It is a stylized fact that larger trades have a higher impact on prices than smaller trades (e.g. Hasbrouck, 1991, Lillo et al., 2003, Gabaix et al., 2006, Hautsch and Huang, 2012), and thus traders, especially informed traders, may have incentives to split their large orders into smaller trades in order to hide their information from market makers and reduce their price impact (e.g. Kyle, 1985, Easley and O'Hara, 1987, Dufour and Engle, 2000, Forsyth et al., 2012, O'Hara, 2015, Toth et al., 2015, Choi et al., 2019, Korajczyk and Murphy, 2019, van Kervel and Menkveld, 2019). Clearly, the impact of a series of trades on prices will not be the same as, and is often smaller than, the immediate price impact of a large trade of the same aggregated size, since the quoted depth may be replenished during the course of the smaller trades. It is interesting to quantify how much price impact order splitting strategies might save an investor.

We investigate this question in a novel way by comparing the observed price impact of a series of consecutive trades in the same direction and on the same trading day to the immediate price impact of an *artificial* trade that aggregates these consecutive trades. We focus on consecutive same-sign trades on the same day since (i) smaller transactions split from a large order should follow the same direction as the direction of the parent order; and (ii) the avoidance of cross-day and/or non-consecutive trades originating from an informed big order reduces the computational complexity of constructing our artificial trades. We further restrict our attention to consecutive same-sign same-day trades that (i) are each smaller than their prevailing depths; and (ii) whose total volume (which equals the volume of an artificially aggregated transaction) is bigger than or equal to the prevailing depth of the first trade. This identifies examples of series of trades that could be components of a split order. None of the individual trades that comprise this "split order" should have immediate price impact on its own (because each individual trade has less volume than available depth), whereas the (artificially) aggregated order should.

We compute the observed price impact of a series of consecutive same-sign same-day trades as the logarithmic change in the mid-prices of the best bid and ask quotes immediately after the last trade and right before the first trade of the series.[24] Since the artificial trade that aggregates these consecutive trades is unobserved, we estimate its immediate price impact using the HARX_{depth} model, which is the most accurate price impact model in our study. The HARX_{depth} model is (re-)estimated in each month for buys and sells of each stock group, and it is used to predict the price impact of all artificial trades in that month.[25]

Table 5 summarizes the price impact savings results from order splits. We report four different values for the number of consecutive same-sign trades: $K = 2$, 5, 10, and all, where the latter refers to all observed series of consecutive same-sign trades that satisfy the two aforementioned restrictions (i.e. we have series of consecutive trades of different lengths as observed in the sample). Δp_t^K and Δp_t^a

[24]Note that the overall price impact of a sequence of consecutive same-sign (individually zero-impact) trades can be non-zero in limit order book settings because these trades consume liquidity and reduce quoted depth, and new orders can be placed on the order book while the sequence of trades is taking place, shifting the bid-ask spread and hence the mid-quote price that is observed immediately after the last of these zero-impact trades.

[25]We obtain qualitatively similar results using the HARX_{depth} model to predict the price impact of the artificial trades, but with the model estimated over the nine month rolling windows described in Subsection 3.2.

respectively denote the price impact of K consecutive same-sign same-day trades and the immediate price impact of their artificially aggregated trade predicted by the HARX$_{depth}$ model. $\%\Delta p_t^S = (\Delta p_t^a - \Delta p_t^K)/\Delta p_t^a \times 100\%$ computes the proportion of the immediate price impact that can be saved by splitting an order. Similar to that observed for single trades, the price impact of a series of consecutive trades is inversely related to stock market capitalization and is generally larger in magnitude for sells than for buys. As expected, the number of series of consecutive same-sign trades decreases with the series length. Interestingly, the average observed price impact of a series of consecutive trades (Δp_t^K) does not necessarily increase with the series length, although the average predicted immediate price impact of the associated artificial trade generally does. This is because new orders can either increase or decrease market depth (while a series of small trades is taking place), so that the resulting price impact does not simply depend on the series length, whereas the volume (and hence price impact) of an aggregate trade will generally increase with the number of underlying trades.

<<INSERT TABLE 5 ABOUT HERE>>

Consistent with our expectations, the observed price impact of a series of consecutive same-sign same-day trades is much lower than the predicted immediate price impact of their artificially aggregated trades. On average, splitting a big order that is larger than the prevailing depth into a series of smaller trades reduces the immediate price impact of the former considerably by between 60% and 82%, and an order split into $K = 10$ small trades appears to bring about the largest price impact reductions. Such reductions are attainable because (i) the smaller trades typically have zero or much smaller immediate price impact, and (ii) market depth may be replenished as the smaller trades occur, as a result of order submissions from other traders. This finding supports the effectiveness of order splitting strategies suggested in the literature (see, eg, Kyle, 1985, Easley and O'Hara, 1987, Dufour and Engle, 2000) and provides institutions with a rationale to use algorithms to split and sequence orders to minimize execution costs (Hendershott et al., 2011, Forsyth et al., 2012, Hasbrouck and Saar, 2013, O'Hara, 2015, van Kervel and Menkveld, 2019).[26]

Figure 3 illustrates the implied savings due to order splitting strategies over time, by plotting time series of monthly price impact savings $\%\Delta p_t^S$, for our stock groups 1, 3 and 5, for our four different values of K over 2007-2013.[27] Complementing the results in Table 5, order splitting strategies

[26] Institutions can also split their orders and randomize their order size to avoid "back running" from sophisticated traders (see, for example, van Kervel and Menkveld (2019), Yang and Zhu (2019)).

[27] We obtain qualitatively similar plots for Groups 2 and 4.

consistently reduce the immediate price impact of big orders by large proportions. The proportions of the price impact that can be saved by order splitting are more than 80% (and even close to 100% in the few months before 2011), but they then fall and range between 20% and 75% by the end of 2013. This result is consistent with a strong declining pattern in trading volumes over the same period as shown in Table 1. It suggests that traders have employed order splitting strategies more extensively in recent years (e.g. Chordia et al., 2011, Friederich and Payne, 2014) to get a better execution price for their orders and reduce their price impact costs. Deeper and more liquid markets in recent years have also contributed to the decreasing pattern of price impact savings, since they help reduce the price impact costs of trades.

<<INSERT FIGURE 3 ABOUT HERE>>

6 Market depth, order dynamics and price impact gaps[28]

Prior literature (e.g. Hasbrouck, 1991, Dufour and Engle, 2000, Bessembinder and Venkataraman, 2010, Obizhaeva and Wang, 2013) has used permanent price impact to measure the information content of trades. This measure is defined as the change in the price of an asset over some fixed time interval after a trade. If neither new orders nor updates such as revisions or cancellations enter the order book during a specified time period after a trade, then the immediate price impact of that trade is equal to its permanent price impact over that period. However, if the trade induces subsequent trading activities, then immediate and permanent price impacts may differ, and the gap between these two measures will reflect the dynamics of the information and order flows after the initial trade. Given that market depth information is particularly useful for predicting immediate price impact, as we demonstrate in Section 4, we conjecture that market depth also conveys important information about subsequent incoming order flows and the consequential price impact gap.

This section provides some analysis of how informative our depth indicator $I_{v_{i,t} \geq depth_{i,t}}$ is about future order imbalance (a proxy for future order dynamics) and the difference between permanent and immediate price impact following a trade. We use the data relating to 2007 for three different stock groups (Groups 1, 3 and 5 of the ASX200 index), which allows us to observe cross sectional variation

[28]We thank an anonymous referee for suggesting that we investigate this important issue.

of the information content of the depth indicators for stocks with different market capitilization,[29] and we estimate the following regression:

$$y_{i,t,\delta} = c + \alpha_1 I_{v_{i,t} \geq depth_{i,t}} + \alpha_2 LnVolume_{i,t} + \alpha_3 LnDepth_{i,t} + \alpha_4 day_{i,t} + \alpha_5 block_{i,t} + \eta_{i,t,\delta}, \qquad (3)$$

where $y_{i,t,\delta}$ is either (i) the order imbalance ($OIB_{i,t,\delta}$), defined as the ratio of the total buy minus sell traded volume to the total traded volume during a δ-minute interval after a trade in stock i at time t;[30] or (ii) the price impact gap (PIG) defined over a δ-minute interval, as the signed difference between the permanent price impact ($\Delta p_{i,t,\delta}$ in basis points (bps)) and the immediate price impact ($\Delta p_{i,t}$ in bps) of a trade in stock i at time t, i.e. PIG $= \epsilon_{i,t} \times (\Delta p_{i,t,\delta} - \Delta p_{i,t})$, where $\epsilon_{i,t}$ is the sign (+1 for buy and -1 for sell) of a trade. While the immediate price impact $\Delta p_{i,t}$ is already defined in Section 2, the permanent price impact is similarly defined as $\Delta p_{i,t,\delta} \equiv p_{i,t+\delta,-} - p_{i,t,-}$, where $p_{i,t,-}$ ($p_{i,t+\delta,-}$) is the prevailing log midpoint of stock i right before time t ($t + \delta$).

Our depth indicator $I_{v_{i,t} \geq depth_{i,t}}$ signals a large transaction whose volume is larger than or equal to the prevailing opposite-side depth right before the trade. The $LnVolume_{i,t}$ and $LnDepth_{i,t}$ variables are respectively the natural logarithms of the share volume of the trade and the prevailing opposite-side depth right before trade. The $day_{i,t}$ and $block_{i,t}$ variables capture day of the week and time of day effects, as discussed earlier in subsection 2.3, and $\eta_{i,t,\delta}$ is an independent error term with a zero mean.

Tables 6 reports the estimation results. Our dependent variables are either the order imbalance (OIB) or the price impact gap (PIG) over a δ-minute interval, where δ is either 5 or 30 minutes.

<<INSERT TABLE 6 ABOUT HERE>>

We begin by focusing on the order imbalance, OIB. The sign of the depth indicator, $I_{v_{i,t} \geq depth_{i,t}}$, coefficient (i.e., α_1) alternates between being negative for buy and positive for sell trades (with just a single exception that is not statistically significant). This finding suggests that the order imbalance after a large buy (sell) is on average smaller (larger) than that after a similar small buy (sell). In other words, large trades (with volume that is greater than or equal to the prevailing opposite-side depth) tend to induce more opposite-direction transactions than same-direction trades in the future.

The above result suggests that in the next δ-minute interval after a large initial trade (whether it is a buy or sell), the subsequent aggregate trades are in the opposite direction to the initial trade.

[29] We focus on a sub-sample of our dataset to reduce computation time, but point out that this sample contains more than 10 million observations.

[30] We set $OIB_{i,t,\delta}$ equal to zero if there are no transactions during a δ-minute interval after a trade.

Consequently, the stock price after a δ-minute interval is likely to be lower (higher) than the price immediately after a large buy (sell), implying a negative permanent versus immediate price impact gap after a large trade, as well as a negative relation between the price impact gap and the depth indicator. This is confirmed by the regression results reported for the price impact gap (PIG) in Table 6, in which the depth indicator, $I_{v_{i,t} \geq depth_{i,t}}$, coefficient (i.e., α_1) is statistically significantly negative for all groups, trades and intervals. This is consistent with replenishment of the order book after a large trade.

Table 6 also reveals a noticeable cross-sectional difference among the three stock groups. In particular, a large trade in a smaller-capitalized stock (e.g. in Group 5) leads to markedly larger (average) changes in both the order imbalance and the price impact gap than does a large trade in a bigger-capitalized stock (i.e. in Groups 3 and 1). This is because smaller-capitalized stocks are less frequently monitored and traded than bigger-capitalized stocks, so a large liquidity shock to the smaller-capitalized stocks after a large trade appears to affect future orders and prices more strongly, leading to a bigger price overshoot followed by stronger price reversals than a similar shock to the larger-capitalized stocks.

Overall, the above results highlight the ability of the depth indicator to predict order imbalance as well as the price impact gap. Thus, it is potentially beneficial to incorporate the depth indicator into a model to predict order flow dynamics, price impact gap and permanent price impact after trades. We leave further work on this issue for future research, but point out that our empirical observation that the depth indicator provides a useful link between temporary and permanent price impact makes a novel contribution.

7 Conclusion

Measuring immediate price impact is important for accurately estimating the costs associated with immediate trade as well as measuring impediments to the formation of capital in financial markets (Duffie, 2010). This research extends the prior literature on price impact modeling by highlighting the significance of market depth information and price impact dynamics for estimating and predicting immediate market impact. It also provides a comprehensive analysis of out-of-sample performance for different price impact models.

We find that the inclusion of market depth information as a threshold-type indicator variable to detect zero-impact trades in an immediate price impact model can lead to reductions in the MAE and MSE measures of out-of-sample predictions by about 60% on average. Furthermore, the inclusion of an autoregressive component in an immediate price impact model can lead to reductions in the MAE and MSE of out-of-sample predictions that amount to about 5-6%. Given that immediate price impact is the biggest component of total trading costs, it is important to quantify this impact accurately, and we show that the cumulative effects of such improvements in accuracy could save investors millions of dollars per annum in projecting the costs and profits of investment strategies. We also show that splitting a big order into a series of smaller trades can reduce the immediate price impact cost of the former considerably by between 60% and 82%, and thus order splitting strategies can provide significant economic benefits to traders.

Our work focuses on the immediate component of price impact. Other studies (e.g. Dufour and Engle, 2000, Bouchaud et al., 2009, Obizhaeva and Wang, 2013, Jondeau et al., 2015) have shown that price impact exhibits permanent characteristics. That is, the execution of a trade affects the dynamics of incoming order flow, the prices of subsequent trades, and the permanent price impact (i.e. how prices move to a new equilibrium following a trade). An exploratory analysis in our paper shows that the depth indicator is informative about the order imbalance and the difference between permanent price impact and immediate price impact following a trade. Building a model that incorporates market depth information and price impact dynamics into the forecasting of order flow dynamics and permanent price impact is an important and interesting question for future research.

References

Admati, A. R. and Pfleiderer, P. (1988). A theory of intraday patterns: Volume and price variability. *Review of Financial Studies*, 1(1):3–40.

Aitken, M., Chen, H., and Foley, S. (2017). The impact of fragmentation, exchange fees and liquidity provision on market quality. *Journal of Empirical Finance*, 41:140–160.

Almgren, R., Thum, C., Hauptmann, E., and Li, H. (2005). Equity market impact. *Risk*, 18:57–62.

Anand, A., Chakravarty, S., and Martell, T. (2005). Empirical evidence on the evolution of liquidity: Choice of market versus limit orders by informed and uninformed traders. *Journal of Financial Markets*, 8(3):288–308.

Andersen, T. G. (1996). Return volatility and trading volume : An information flow interpretation of stochastic volatility. *The Journal of Finance*, 51(1):169–204.

ASX (2010). Algorithmic trading and market access arrangements. *ASX Review*, available at https://www.asx.com.au/documents/media/20100211_review_algorithmic_trading_and_market_access.pdf.

Barclay, M. J. and Warner, J. B. (1993). Stealth trading and volatility: Which trades move prices? *Journal of Financial Economics*, 34(3):281–305.

Bessembinder, H. and Venkataraman, K. (2004). Does an electronic stock exchange need an upstairs market? *Journal of Financial Economics*, 73(1):3–36.

Bessembinder, H. and Venkataraman, K. (2010). Bid-ask spreads: Measuring trade execution costs in financial markets. *Encyclopedia of quantitative finance*, pages 184–190.

Bloomfield, R., O'Hara, M., and Saar, G. (2005). The "make or take" decision in an electronic market: Evidence on the evolution of liquidity. *Journal of Financial Economics*, 75(1):165–199.

Bouchaud, J. P., Farmer, J. D., and Lillo, F. (2009). How markets slowly digest changes in supply and demand. In Hens, T. and Schenk-Hoppé, K. R., editors, *Handbook of Financial Markets: Dynamics and Evolution*, chapter 2, pages 57–160. Elsevier Inc., New York, 1st edition.

Boulatov, A. and George, T. J. (2013). Hidden and displayed liquidity in securities markets with informed liquidity providers. *Review of Financial Studies*, 26(8):2096–2137.

Brogaard, J., Hagströmer, B., Norden, L., and Riordan, R. (2015). Trading fast and slow: Colocation and liquidity. *Review of Financial Studies*, 28(12):3407–3443.

Brogaard, J., Hendershott, T., and Riordan, R. (2014). High-frequency trading and price discovery. *Review of Financial Studies*, 27(8):2267–2306.

Chakravarty, S. (2001). Stealth-trading: Which traders' trades move stock prices? *Journal of Financial Economics*, 61(2):289–307.

Chan, Y. C. (2000). The price impact of trading on the stock exchange of Hong Kong. *Journal of Financial Markets*, 3(1):1–16.

Choi, J. H., Larsen, K., and Seppi, D. J. (2019). Information and trading targets in a dynamic market equilibrium. *Journal of Financial Economics*, 132(3):22–49.

Chordia, T., Roll, R., and Subrahmanyam, A. (2011). Recent trends in trading activity and market quality. *Journal of Financial Economics*, 101(2):243–263.

Comerton-Forde, C. and Putniņš, T. J. (2015). Dark trading and price discovery. *Journal of Financial Economics*, 118(1):70–92.

Cont, R., Kukanov, A., and Stoikov, S. (2014). The price impact of order book events. *Journal of Financial Econometrics*, 12(1):47–88.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Diamond, D. W. and Verrecchia, R. E. (1987). Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics*, 18(2):277–311.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–265.

Dominici, F., McDermott, A., Zeger, S. L., and Samet, J. M. (2002). On the use of generalized additive models in time-series studies of air pollution and health. *American Journal of Epidemiology*, 156(3):193–203.

Duffie, D. (2010). Presidential address: Asset price dynamics with slow-moving capital. *The Journal of Finance*, 65(4):1237–1267.

Dufour, A. and Engle, R. F. (2000). Time and the price impact of a trade. *The Journal of Finance*, 55(6):2467–2498.

Duong, H. N., Kalev, P. S., and Krishnamurti, C. (2009). Order aggressiveness of institutional and individual investors. *Pacific-Basin Finance Journal*, 17(5):533–546.

Easley, D. and O'Hara, M. (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics*, 19(1):69–90.

Easley, D. and O'Hara, M. (1992). Time and the process of security price adjustment. *The Journal of Finance*, 47(2):577–605.

Eisler, Z., Bouchaud, J.-P., and Kockelkoren, J. (2012). The price impact of order book events: Market orders, limit orders and cancellations. *Quantitative Finance*, 12(9):1395–1419.

Engle, R. F. and Lange, J. (2001). Predicting VNET: A model of the dynamics of market depth. *Journal of Financial Markets*, 4(2):113–142.

Engle, R. F. and Patton, A. J. (2004). Impacts of trades in an error-correction model of quote prices. *Journal of Financial Markets*, 7(1):1–25.

Foresi, S. and Peracchi, F. (1995). The conditional distribution of excess returns: An empirical analysis. *Journal of the American Statistical Association*, 90(430):451–466.

Forsyth, P. A., Kennedy, J. S., Tse, S. T., and Windcliff, H. (2012). Optimal trade execution: A mean quadratic variation approach. *Journal of Economic Dynamics and Control*, 36(12):1971–1991.

Foster, F. D. and Viswanathan, S. (1990). A theory of the interday variations in volume, variance, and trading costs in securities markets. *Review of Financial Studies*, 3(4):593–624.

Foucault, T., Kadan, O., and Kandel, E. (2013a). Liquidity cycles and make/take fees in electronic markets. *The Journal of Finance*, 68(1):299–341.

Foucault, T., Pagano, M., and Röell, A. (2013b). *Market liquidity: Theory, evidence, and policy.* Oxford University Press.

French, K. R. (1980). Stock returns and the weekend effect. *Journal of Financial Economics*, 8(1):55–69.

Friederich, S. and Payne, R. (2014). Trading anonymity and order anticipation. *Journal of Financial Markets*, 21:1 – 24.

Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H. E. (2003). A theory of power-law distributions in financial market fluctuations. *Nature*, 423:267–270.

Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H. E. (2006). Institutional investors and stock market volatility. *Quarterly Journal of Economics*, 121:461–504.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578.

Goettler, R. L., Parlour, C. A., and Rajan, U. (2009). Informed traders and limit order markets. *Journal of Financial Economics*, 93(1):67–87.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2):453–497.

Hasbrouck, J. (1991). Measuring the information content of stock trades. *The Journal of Finance*, 46(1):179–207.

Hasbrouck, J. and Saar, G. (2013). Low-latency trading. *Journal of Financial Markets*, 16(4):646 – 679.

Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1(3):297–310.

Hastie, T. and Tibshirani, R. (1990). *Generalized additive models*. Chapman & Hall.

Hastie, T. and Tibshirani, R. (1995). Generalized additive models for medical research. *Statistical*

Methods in Medical Research, 4(3):187–196.
Hautsch, N. and Huang, R. (2012). The market impact of a limit order. *Journal of Economic Dynamics and Control*, 36(4):501–522.
Hendershott, T., Jones, C. M., and Menkveld, A. J. (2011). Does algorithmic trading improve liquidity? *The Journal of Finance*, 66(1):1–33.
Hendershott, T. and Riordan, R. (2013). Algorithmic trading and the market for liquidity. *Journal of Financial and Quantitative Analysis*, 48(4):1001–1024.
Hou, A. J. (2013). Asymmetry effects of shocks in Chinese stock markets volatility: A generalized additive nonparametric approach. *Journal of International Financial Markets, Institutions and Money*, 23:12–32.
Hurvich, C. M. and Tsai, C.-L. (1990). The impact of model selection on inference in linear regression. *The American Statistician*, 44(3):214–217.
Jondeau, E., Lahaye, J., and Rockinger, M. (2015). Estimating the price impact of trades in a high-frequency microstructure model with jumps. *Journal of Banking & Finance*, 61:S205–S224.
Jovanovic, B. and Menkveld, A. J. (2019). Equilibrium bid-price dispersion. Available at https://ssrn.com/abstract=2463066.
Karpoff, J. M. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis*, 22(1):109–126.
Keim, D. and Madhavan, A. (1995). Anatomy of the trading process: empirical evidence on the behavior of institutional traders. *Journal of Financial Economics*, 37:371 – 398.
Keim, D. B. and Madhavan, A. (1996). The upstairs market for large-block transactions: Analysis and measurement of price effects. *Review of Financial Studies*, 9(1):1–36.
Keim, D. B. and Madhavan, A. (1998). The cost of institutional equity trades. *Financial Analysts Journal*, 54(4):50–69.
Knez, P. J. and Ready, M. J. (1996). Estimating the profits from trading strategies. *Review of Financial Studies*, 9(4):1121–1163.
Korajczyk, R. A. and Murphy, D. (2019). High-frequency market making to large institutional trades. *Review of Financial Studies*, 32(3):1034–1067.
Kraus, A. and Stoll, H. R. (1972). Price impacts of block trading on the New York Stock Exchange. *The Journal of Finance*, 27(3):569–588.
Kwan, A., Masulis, R., and McInish, T. H. (2015). Trading rules, competition for order flow and market fragmentation. *Journal of Financial Economics*, 115(2):330–348.
Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335.
Lee, C. M. C. and Ready, M. J. (1991). Inferring trade direction from intraday data. *The Journal of Finance*, 46(2):733–746.
Lillo, F., Farmer, J. D., and Mantegna, R. N. (2003). Master curve for price-impact function. *Nature*, 421(6919):129–130.
Lim, M. and Coggins, R. (2005). The immediate price impact of trades on the Australian Stock Exchange. *Quantitative Finance*, 5(4):365–377.
Malinova, K. and Park, A. (2013). Liquidity, volume and price efficiency: The impact of order vs. quote driven trading. *Journal of Financial Markets*, 16(1):104–126.
Malinova, K. and Park, A. (2015). Subsidizing liquidity: The impact of make/take fees on market quality. *The Journal of Finance*, 70(2):509–536.
Manganelli, S. (2005). Duration, volume and volatility impact of trades. *Journal of Financial Markets*, 8(4):377–399.
Menkhoff, L., Osler, C. L., and Schmeling, M. (2010). Limit-order submission strategies under asymmetric information. *Journal of Banking & Finance*, 34(11):2665–2677.
Menkveld, A. J. (2018). High frequency trading as viewed through an electron microscope. *Financial Analysts Journal*, 74(2):24 – 31.
Menkveld, A. J. and Yueshen, B. Z. (2019). The flash crash: A cautionary tale about highly fragmented

markets. *Management Science*, 65(10):4470–4488.

Obizhaeva, A. A. and Wang, J. (2013). Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1–32.

Odders-White, E. R. (2000). On the occurrence and consequences of inaccurate trade classification. *Journal of Financial Markets*, 3(3):259–286.

O'Hara, M. (2015). High frequency market microstructure. *Journal of Financial Economics*, 116(2):257 – 270.

Pham, M. C., Duong, H. N., and Lajbcygier, P. (2017). A comparison of the forecasting ability of immediate price impact models. *Journal of Forecasting*, 36(8):898–918.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.

Schwartz, J. (1999). Air pollution and hospital admissions for heart disease in eight US counties. *Epidemiology*, 10(1):17–22.

Torre, N. (1997). *BARRA market impact model handbook*. BARRA Inc., Berkeley.

Toth, B., Palit, I., Lillo, F., and Farmer, J. D. (2015). Why is equity order flow so persistent? *Journal of Economic Dynamics and Control*, 51:218–239.

van Kervel, V. and Menkveld, A. J. (2019). High-frequency trading around large institutional orders. *The Journal of Finance*, 74(3):1091–1137.

Wilinski, M., Cui, W., Brabazon, A., and Hamill, P. (2015). An analysis of price impact functions of individual trades on the London stock exchange. *Quantitative Finance*, 15(10):1727–1735.

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Chapman & Hall.

Wood, S. N. (2017). mgcv: Mixed GAM computation vehicle with automatic smoothness estimation. R package version 1.8-22. URL https://CRAN.R-project.org/package=mgcv.

Yang, L. and Zhu, H. (2019). Back-running: Seeking and hiding fundamental information in order flows. *Review of Financial Studies*, 33:1484–1533.

Zhou, W. X. (2012). Universal price impact functions of individual trades in an order-driven market. *Quantitative Finance*, 12(8):1253–1263.

Zhu, H. (2014). Do dark pools harm price discovery? *Review of Financial Studies*, 27(3):747–789.

Table 1: **Descriptive statistics**

This table presents summary statistics of trades for five stock groups on the S&P/ASX 200 index from 2007 to 2013. "Obs." measures the number of transactions. "Price impact" is the change in the log of mid-quote right after and right before a single trade. "Volume" is the number of shares executed in each trade. "Scaled volume" is the ratio of the trading volume of a trade to the average volume of all trades with the same trade direction as the trade from the beginning of the trading day to that trade. "Dollar volume" equals trading volume multiplied by the mid-quote right before the trade. "Market Cap." is defined as the product of the mid-quote price and the number of shares outstanding right before the trade. "Volatility" of a trade is calculated as the standard deviation of the mid-quote returns from the first trade of the day until that trade. "Duration" is the time interval (in seconds) between two consecutive trades with the same direction. "Depth" is the prevailing quoted depth at the best (Level 1) price on the opposite side of the limit order book right before a trade. The last three columns respectively report the percentage of transactions for which (i) trading volume is less than the prevailing market depth; (ii) price impact is zero; and (iii) the depth indicator correctly predicts no price impact. All figures (except observations, and the last three columns) are the average values.

Stock Group/ Year	Obs.	Price Impact (bps)	Volume	Scaled volume	Market Cap. ($m)	Dollar volume ($)	Vola-tility (bps)	Dur-ation (secs)	Depth	% trades for which $I_{v_{i,t} \geq depth_{i,t}} = 0$	% trades for which $\Delta p_{i,t} = 0$	% trades for which $I_{v_{i,t} \geq depth_{i,t}} = \Delta p_{i,t} = 0$
Panel A: Buys												
A.1: Group 1 (12 stocks with the largest market capitalization)												
2007	2,720,943	1.19	2,288.7	0.961	60,211	62,918	3.02	23.11	26,525	65.71	65.69	65.64
2008	5,572,390	1.66	1,574.7	0.950	50,368	33,011	4.48	11.31	21,991	63.42	63.35	63.32
2009	6,064,702	1.21	1,411.7	0.915	50,143	25,471	3.31	10.47	62,093	66.96	66.78	66.74
2010	6,097,604	0.71	1,242.5	0.920	57,145	27,114	2.24	10.41	123,046	72.32	72.18	72.14
2011	6,862,734	0.85	1,074.8	0.949	54,894	23,894	2.26	9.18	131,995	66.19	65.73	65.50
2012	6,286,642	0.85	955.1	0.969	51,746	20,252	2.15	10.04	76,287	66.90	63.09	62.81
2013	7,331,076	0.84	616.6	0.944	63,192	15,305	2.06	8.06	22,722	64.81	58.16	57.42
A.2: Group 2 (20 stocks)												
2007	1,997,220	2.58	2,838.6	0.975	10,947	29,505	6.68	52.22	33,012	68.99	68.99	68.94
2008	3,934,151	3.74	2,001.6	0.980	9,524	16,372	10.21	26.68	11,217	66.01	65.95	65.92
2009	4,391,502	2.47	2,099.9	0.920	8,714	11,243	7.50	24.03	60,356	71.30	71.18	71.16
2010	4,696,783	1.18	1,498.6	0.916	10,485	11,672	4.41	22.39	67,398	78.30	78.21	78.17
2011	4,899,627	1.65	1,453.3	0.956	11,805	11,752	5.00	21.31	29,530	71.23	70.93	70.74
2012	4,353,656	1.75	1,390.7	0.982	10,531	10,383	5.17	24.07	22,649	72.40	69.15	68.90
2013	5,572,328	1.99	1,179.4	0.953	10,275	9,004	5.11	18.86	16,112	73.15	66.04	65.64
A.3: Group 3 (20 stocks)												
2007	1,588,052	3.04	2,219.7	0.988	5,825	20,509	7.84	66.01	11,782	68.40	68.39	68.36
2008	2,872,544	4.29	1,577.5	0.985	4,409	11,284	12.50	36.40	8,899	68.29	68.25	68.23
2009	3,227,901	3.01	2,293.6	0.919	3,622	8,084	9.51	32.63	36,210	76.33	76.25	76.22
2010	3,499,470	1.44	2,113.4	0.897	4,095	7,875	6.01	30.10	68,435	83.04	82.97	82.95
2011	3,584,799	2.10	2,048.5	0.924	4,121	7,563	7.02	29.15	65,794	74.51	74.28	74.10
2012	3,003,560	2.26	2,044.3	0.951	3,878	6,688	7.36	34.75	61,830	75.21	72.20	71.98
2013	3,773,959	2.55	1,306.3	0.942	4,327	5,901	6.96	27.62	29,146	73.36	66.26	65.83
A.4: Group 4 (20 stocks)												
2007	1,180,120	3.55	2,425.3	0.975	3,186	14,219	9.77	88.13	22,194	72.99	72.99	72.95
2008	2,152,585	5.29	1,784.4	0.977	2,161	7,114	16.20	48.75	12,816	72.19	72.16	72.14
2009	2,210,923	4.05	1,864.4	0.931	1,970	5,382	12.41	47.57	35,765	75.61	75.54	75.51
2010	2,692,888	1.65	1,398.2	0.916	2,252	5,020	6.70	39.09	66,467	81.78	81.70	81.67
2011	2,702,456	2.39	1,617.5	0.936	2,051	4,716	8.30	38.50	85,706	76.64	76.44	76.31
2012	2,269,530	2.56	1,547.0	0.948	2,130	4,472	8.49	45.61	42,637	77.44	74.85	74.66
2013	3,177,937	2.87	1,168.6	0.934	3,142	4,341	8.28	32.62	33,223	72.68	64.82	64.41
A.5: Group 5 (20 stocks with the smallest market capitalization)												
2007	875,051	4.57	2,747.2	0.951	1,574	9,585	13.30	118.24	25,990	75.35	75.32	75.28
2008	1,585,290	5.94	1,509.2	0.968	1,227	4,672	20.11	65.61	12,670	75.20	75.14	75.12
2009	1,631,544	4.70	1,951.9	0.916	1,179	4,283	15.27	62.63	34,562	76.97	76.86	76.84
2010	1,816,662	2.25	1,720.3	0.891	1,277	4,002	9.23	57.05	46,284	84.49	84.41	84.39
2011	1,788,842	3.03	1,524.4	0.921	1,169	3,242	10.81	58.07	52,985	80.16	79.97	79.83
2012	1,493,005	3.31	1,679.0	0.929	1,100	3,163	11.28	69.05	30,122	83.18	80.87	80.74
2013	1,841,033	4.18	1,502.3	0.894	1,323	3,039	12.09	56.22	33,698	79.39	73.10	72.80

Continued on next page

Table 1 – *continued from previous page*

Stock Group/ Year	Obs.	Price Impact (bps)	Volume	Scaled volume	Market Cap. ($m)	Dollar volume ($)	Vola- tility (bps)	Dur- ation (secs)	Depth	% trades for which $I_{v_{i,t} \geq depth_{i,t}} = 0$	% trades for which $\Delta p_{i,t} = 0$	% trades for which $I_{v_{i,t} \geq depth_{i,t}} = \Delta p_{i,t} = 0$
Panel B: Sells												
B.1: Group 1 (12 stocks with the largest market capitalization)												
2007	2,460,728	-1.34	2,470.6	0.966	58,825	65,885	3.10	25.56	44,760	61.08	61.09	61.01
2008	4,829,706	-1.89	1,720.5	0.957	50,686	36,953	4.48	13.05	22,174	59.05	58.95	58.92
2009	5,607,341	-1.32	1,481.4	0.923	50,456	26,994	3.30	11.32	51,042	63.95	63.76	63.72
2010	5,569,425	-0.79	1,375.1	0.935	57,238	29,062	2.23	11.40	103,173	69.74	69.59	69.54
2011	6,492,187	-0.92	1,112.9	0.961	54,572	25,142	2.25	9.70	95,914	63.30	62.80	62.57
2012	6,191,443	-0.86	905.3	0.978	51,932	20,222	2.13	10.20	54,152	66.47	62.58	62.25
2013	7,024,460	-0.88	636.9	0.955	63,387	15,876	2.05	8.42	19,365	62.60	55.76	54.79
B.2: Group 2 (20 stocks)												
2007	1,939,120	-2.69	2,981.0	0.980	10,812	30,550	6.74	53.79	39,697	67.39	67.39	67.36
2008	3,456,762	-4.29	2,184.7	0.989	9,586	17,919	10.16	30.36	10,523	61.24	61.17	61.15
2009	4,096,092	-2.69	2,196.7	0.934	8,719	11,811	7.52	25.76	45,448	68.69	68.57	68.54
2010	4,442,312	-1.26	1,617.3	0.938	10,507	12,413	4.45	23.68	62,110	76.94	76.84	76.81
2011	4,929,286	-1.67	1,474.0	0.951	11,823	11,846	5.02	21.18	30,489	70.65	70.37	70.16
2012	4,332,647	-1.77	1,393.6	0.974	10,583	10,431	5.13	24.18	21,428	72.02	68.83	68.54
2013	5,328,297	-2.11	1,236.2	0.953	10,288	9,339	5.12	19.73	16,310	71.36	63.84	63.28
B.3: Group 3 (20 stocks)												
2007	1,553,436	-3.16	2,286.3	0.989	5,775	21,096	7.89	67.48	11,171	66.87	66.87	66.84
2008	2,485,921	-5.05	1,776.7	0.993	4,516	12,729	12.33	42.05	7,953	63.04	62.99	62.97
2009	3,030,564	-3.24	2,390.3	0.930	3,592	8,393	9.49	34.74	33,092	74.46	74.37	74.34
2010	3,255,386	-1.57	2,314.7	0.911	4,087	8,539	6.10	32.35	68,475	81.47	81.40	81.36
2011	3,592,252	-2.12	2,114.8	0.917	4,096	7,722	7.09	29.09	67,963	74.08	73.86	73.67
2012	2,997,577	-2.29	2,034.0	0.938	3,843	6,685	7.39	34.84	78,612	74.70	71.75	71.49
2013	3,649,856	-2.67	1,365.7	0.942	4,320	6,061	6.93	28.56	28,849	72.15	64.60	64.06
B.4: Group 4 (20 stocks)												
2007	1,153,566	-3.71	2,561.4	0.971	3,147	14,849	9.77	90.19	21,754	71.28	71.18	71.14
2008	1,805,427	-6.45	2,067.2	0.983	2,177	8,265	16.12	58.10	11,558	66.24	66.21	66.18
2009	2,118,532	-4.22	1,920.6	0.940	1,949	5,545	12.46	49.62	26,979	74.55	74.48	74.45
2010	2,538,132	-1.79	1,541.9	0.920	2,256	5,425	6.77	41.44	57,319	80.27	80.19	80.15
2011	2,648,773	-2.48	1,688.3	0.931	2,025	4,901	8.38	39.27	78,615	75.53	75.34	75.19
2012	2,205,166	-2.66	1,603.3	0.931	2,091	4,603	8.67	46.97	45,269	76.43	73.76	73.55
2013	2,979,703	-3.19	1,268.2	0.946	3,174	4,693	8.28	34.79	35,032	69.06	60.48	59.90
B.5: Group 5 (20 stocks with the smallest market capitalization)												
2007	826,846	-4.96	2,926.2	0.950	1,564	10,238	13.47	125.10	24,103	72.78	72.70	72.67
2008	1,242,546	-7.90	1,933.2	0.970	1,266	5,810	19.85	83.62	12,480	67.64	67.58	67.55
2009	1,511,958	-5.11	2,085.7	0.929	1,188	4,590	15.18	67.57	34,342	74.85	74.78	74.74
2010	1,668,492	-2.48	1,959.3	0.892	1,274	4,444	9.41	62.09	48,808	82.86	82.78	82.75
2011	1,696,728	-3.23	1,659.4	0.911	1,175	3,439	10.77	61.17	43,826	78.86	78.65	78.53
2012	1,340,304	-3.75	1,893.4	0.901	1,087	3,538	11.55	76.80	33,616	80.85	78.40	78.21
2013	1,668,853	-4.81	1,706.8	0.892	1,313	3,411	12.22	61.95	41,083	76.05	68.92	68.49

Table 2: **Comparison of models with respect to out-of-sample predictive accuracy**

Panel A: MSE		**HARX$_{depth}$**	**HARX**	**LX$_{depth}$**	**LXb**	**LX**	**ZHOU**	**LFM1**	**LFM2**	**naive**
Eqn #		(A.1)	(A.2)	(A.3)	(A.4)	(A.5)	(A.6)	(A.7)	(A.8)	(A.9)
Group 1	Buys	**1.568**	3.415	1.615	3.584	3.592	3.545	3.686	3.613	4.948
	Sells	**1.744**	3.628	1.794	3.809	3.817	3.772	3.940	3.834	5.401
Group 2	Buys	**8.45**	20.92	8.64	21.92	21.98	21.55	22.95	21.75	28.96
	Sells	**9.08**	21.84	9.30	22.94	23.01	22.58	24.16	22.88	30.85
Group 3	Buys	**14.36**	35.34	14.69	37.25	37.38	36.64	38.73	36.97	48.13
	Sells	**15.46**	37.02	15.87	39.16	39.31	38.58	41.04	39.09	51.46
Group 4	Buys	**23.76**	55.99	24.34	59.16	59.32	58.17	61.73	59.15	75.95
	Sells	**26.12**	59.73	26.81	63.40	63.60	62.42	66.62	64.30	82.59
Group 5	Buys	**43.23**	103.30	43.93	109.57	109.87	107.53	113.98	109.17	136.79
	Sells	**50.12**	115.57	51.05	123.20	123.60	121.25	129.41	124.34	157.47
Panel B: MAE		**HARX$_{depth}$**	**HARX**	**LX$_{depth}$**	**LXb**	**LX**	**ZHOU**	**LFM1**	**LFM2**	**naive**
Group 1	Buys	**0.497**	1.199	0.503	1.264	1.268	1.204	1.255	1.216	1.007
	Sells	**0.538**	1.243	0.545	1.310	1.314	1.254	1.311	1.268	1.084
Group 2	Buys	**1.064**	2.652	1.077	2.817	2.837	2.610	2.764	2.673	2.090
	Sells	**1.134**	2.734	1.149	2.911	2.934	2.713	2.876	2.773	2.214
Group 3	Buys	**1.333**	3.353	1.350	3.573	3.598	3.303	3.532	3.396	2.558
	Sells	**1.426**	3.480	1.448	3.728	3.759	3.471	3.715	3.561	2.721
Group 4	Buys	**1.627**	4.028	1.651	4.302	4.324	3.976	4.262	4.070	3.036
	Sells	**1.769**	4.239	1.801	4.555	4.585	4.246	4.550	4.344	3.288
Group 5	Buys	**1.984**	5.277	2.004	5.646	5.679	5.159	5.568	5.307	3.871
	Sells	**2.267**	5.787	2.298	6.240	6.285	5.769	6.237	5.916	4.411

This table reports Mean Squared Error (MSE) and Mean Absolute Error (MAE) for nine models for five stock groups listed on the S&P/ASX200 index over the whole out-of-sample period from October 2007 to December 2013 (i.e. 75 months). Results for MSE (in bps^2) and MAE (in bps) are reported in Panels A and B, respectively. See Appendix A for full specifications of the models (A.1) to (A.9). Although not reported, differences in MSEs and in MAEs between any two models for either buys or sells of any stock group are statistically significantly different from 0 (with p-values less than 10^{-10}) based on the Giacomini and White (2006) conditional predictive accuracy test. Bold format denotes the smallest MSE/MAE within each group. Results are based on the analysis of nine-month in-sample and one-month out-of-sample windows rolled one month ahead.

Table 3: **Percentages of out-of-sample outperformance for all models across all rolling windows**

		Panel A: Giacomini & White test		Panel B: Model Confidence Set Test	
(a) MSE		**HARX**$_{depth}$	**LX**$_{depth}$	**HARX**$_{depth}$	**LX**$_{depth}$
Group 1	Buys	98.67	1.33	98.67	2.67
	Sells	94.67	2.67	98.67	8.00
Group 2	Buys	97.33	2.67	98.67	2.67
	Sells	96.00	2.67	98.67	6.67
Group 3	Buys	100.00	0.00	100.00	0.00
	Sells	97.33	2.67	98.67	5.33
Group 4	Buys	100.00	0.00	100.00	1.33
	Sells	98.67	1.33	100.00	4.00
Group 5	Buys	98.67	1.33	100.00	14.67
	Sells	93.33	6.67	97.33	20.00
(b) MAE		**HARX**$_{depth}$	**LX**$_{depth}$	**HARX**$_{depth}$	**LX**$_{depth}$
Group 1	Buys	88.00	12.00	93.33	13.33
	Sells	88.00	12.00	90.67	13.33
Group 2	Buys	93.33	6.67	96.00	12.00
	Sells	92.00	8.00	94.67	9.33
Group 3	Buys	92.00	8.00	94.67	9.33
	Sells	93.33	6.67	94.67	14.67
Group 4	Buys	98.67	1.33	98.67	4.00
	Sells	98.67	1.33	100.00	5.33
Group 5	Buys	86.67	13.33	97.33	18.67
	Sells	92.00	8.00	94.67	16.00

This table compares the out-of-sample performance for all nine models studied in Table 2, over all rolling windows (from October 2007 to December 2013), for five stock groups on the S&P/ASX200 index. Panel A reports the percentages of times that each model statistically significantly outperforms all other eight models in each out-of-sample one-month window out of 75 out-of-sample months, based on pair-wise Giacomini and White (2006) conditional predictive accuracy tests. Panel B reports the percentages of times that each model belongs in the set of the best models in each out-of-sample one-month window out of 75 months according to the Hansen et al. (2011) model confidence set test. A 5% significance level is chosen for both tests. Only results for the HARX$_{depth}$ (specified in Equation (A.1)) and LX$_{depth}$ (in (A.3)) models are shown. Results for the remaining models are all zero and are not reported. Results, reported in %, are based on the analysis of nine-month in-sample and one-month out-of-sample windows rolled one month ahead.

Table 4: **Contributors of the superiority of the HARX$_{depth}$ model**

		Panel A		Panel B	Panel C
		Market depth information		Price impact dynamics	Intra- & inter-day periodicities
(a) MSE		HARX$_{depth}$ vs. HARX	LX$_{depth}$ vs. LX	HARX vs. LXb	LXb vs. LX
Group 1	Buys	100.00 [54.07]	100.00 [55.05]	100.00 [4.73]	90.67 [0.22]
	Sells	100.00 [51.92]	100.00 [52.99]	100.00 [4.77]	89.33 [0.21]
Group 2	Buys	100.00 [59.60]	100.00 [60.68]	100.00 [4.56]	96.00 [0.29]
	Sells	100.00 [58.43]	100.00 [59.58]	100.00 [4.81]	97.33 [0.31]
Group 3	Buys	100.00 [59.37]	100.00 [60.71]	100.00 [5.13]	94.67 [0.34]
	Sells	100.00 [58.24]	100.00 [59.63]	100.00 [5.46]	94.67 [0.38]
Group 4	Buys	100.00 [57.55]	100.00 [58.97]	100.00 [5.35]	92.00 [0.28]
	Sells	100.00 [56.26]	100.00 [57.85]	100.00 [5.79]	100.00 [0.32]
Group 5	Buys	100.00 [58.15]	100.00 [60.02]	100.00 [5.72]	97.33 [0.28]
	Sells	100.00 [56.64]	100.00 [58.70]	100.00 [6.19]	94.67 [0.32]
Average		100.00 [57.02]	100.00 [58.42]	100.00 [5.25]	94.67 [0.30]
(b) MAE		HARX$_{depth}$ vs. HARX	LX$_{depth}$ vs. LX	HARX vs. LXb	LXb vs. LX
Group 1	Buys	100.00 [58.56]	100.00 [60.33]	100.00 [5.17]	86.67 [0.34]
	Sells	100.00 [56.72]	100.00 [58.53]	100.00 [5.11]	85.33 [0.34]
Group 2	Buys	100.00 [59.88]	100.00 [62.05]	100.00 [5.84]	93.33 [0.72]
	Sells	100.00 [58.54]	100.00 [60.84]	100.00 [6.08]	96.00 [0.78]
Group 3	Buys	100.00 [60.26]	100.00 [62.47]	100.00 [6.17]	84.00 [0.69]
	Sells	100.00 [59.03]	100.00 [61.46]	100.00 [6.65]	88.00 [0.81]
Group 4	Buys	100.00 [59.61]	100.00 [61.82]	100.00 [6.38]	88.00 [0.52]
	Sells	100.00 [58.25]	100.00 [60.73]	100.00 [6.94]	89.33 [0.66]
Group 5	Buys	100.00 [62.39]	100.00 [64.70]	100.00 [6.53]	86.67 [0.58]
	Sells	100.00 [60.83]	100.00 [63.44]	100.00 [7.26]	94.67 [0.72]
Average		100.00 [59.41]	100.00 [61.64]	100.00 [6.21]	89.20 [0.62]

This table investigates the contributors of the superiority of the HARX$_{depth}$ model over 75 out-of-sample months (from October 2007 to December 2013) for five stock groups on the S&P/ASX200 index.

Panel A shows the contribution of market depth information by reporting (i) the percentages of times that each price impact model that incorporates market depth information statistically significantly outperforms its respective analogue that does not use such information in each out-of-sample one-month window out of 75 months; and (ii) the percentage reduction in the whole out-of-sample MSEs/MAEs by including the market depth information (in brackets);

Panel B shows the contribution of including the dynamics of price impact by reporting (i) the percentages of times that the HARX model statistically outperforms the LXb model; and (ii) the percentage reduction in the whole out-of-sample MSEs/MAEs by including the price impact dynamics (in brackets);

Panel C shows the contribution of the time of day and day of week effects by reporting (i) the percentages of times that LXb statistically outperforms LX; and (ii) the percentage reduction in the whole out-of-sample MSEs/MAEs by incorporating intra- and inter-day periodicities into LX (in brackets);

"**Average**" rows report the simple averages for both buys and sells for all five stock groups. For each one-month out-of-sample window, the statistical outperformance (at a 5% level) of one model over another model is judged by the Giacomini and White (2006) conditional predictive accuracy test using (a) MSE and (b) MAE. Results, reported in %, are based on the analysis of nine-month in-sample and one-month out-of-sample windows rolled one month ahead.

Table 5: **Price impact savings from splitting large orders**

		K = 2				K = 5			
		Obs.	Δp_t^K	Δp_t^a	$\%\Delta p_t^S$	**Obs.**	Δp_t^K	Δp_t^a	$\%\Delta p_t^S$
Group 1	Buys	1,052,895	0.883	2.724	67.59	588,804	0.869	2.952	70.56
	Sells	968,208	-0.937	-2.666	64.87	455,629	-0.931	-2.866	67.53
Group 2	Buys	739,059	2.299	7.483	69.28	505,949	2.015	8.769	77.02
	Sells	701,738	-2.381	-7.438	67.98	427,886	-2.051	-8.798	76.68
Group 3	Buys	511,984	3.057	10.502	70.89	378,428	2.717	12.498	78.26
	Sells	495,329	-3.106	-10.366	70.04	326,843	-2.653	-12.281	78.39
Group 4	Buys	407,431	3.511	12.530	71.98	301,012	3.315	14.910	77.77
	Sells	398,438	-3.537	-12.316	71.28	255,167	-3.217	-14.819	78.29
Group 5	Buys	238,565	5.298	19.401	72.69	205,063	4.873	21.590	77.43
	Sells	231,198	-5.318	-19.242	72.36	164,967	-4.641	-21.292	78.20
All Groups	Buys	2,949,934	2.335	7.969	70.70	1,979,256	2.302	10.014	77.01
	Sells	2,794,911	-2.417	-7.975	69.69	1,630,492	-2.303	-10.045	77.07

		K = 10				K = all			
		Obs.	Δp_t^K	Δp_t^a	$\%\Delta p_t^S$	**Obs.**	Δp_t^K	Δp_t^a	$\%\Delta p_t^S$
Group 1	Buys	159,194	0.862	3.238	73.39	391,253	1.172	2.989	60.78
	Sells	110,254	-0.926	-3.145	70.56	361,437	-1.206	-2.933	58.89
Group 2	Buys	211,286	1.888	10.094	81.30	279,900	3.321	8.687	61.77
	Sells	168,604	-1.905	-10.291	81.49	265,748	-3.336	-8.609	61.25
Group 3	Buys	181,756	2.534	14.442	82.45	191,494	4.466	12.459	64.15
	Sells	146,339	-2.588	-14.476	82.12	183,352	-4.416	-12.196	63.79
Group 4	Buys	150,196	3.218	17.271	81.36	146,712	5.388	15.281	64.74
	Sells	117,997	-3.238	-17.258	81.24	140,315	-5.362	-14.877	63.96
Group 5	Buys	119,486	4.754	24.146	80.31	94,183	7.935	22.454	64.66
	Sells	86,078	-4.654	-23.969	80.58	89,858	-7.712	-21.745	64.53
All Groups	Buys	821,918	2.492	13.082	80.95	1,103,542	3.427	9.373	63.44
	Sells	629,272	-2.518	-13.190	80.91	1,040,710	-3.438	-9.249	62.83

This table investigates how much price impact one can save by splitting a large order into a series of smaller consecutive trades for five stock groups on the S&P/ASX200 index in 2007-2013. Δp_t^K denotes the observed price impact, measured in bps, of a series of K consecutive trades in the same direction and on the same trading day, computed as the logarithmic change in the mid-quotes immediately after and before the K trades. Δp_t^a denotes the immediate price impact, measured in bps, of an *artificial* transaction that aggregates the K consecutive trades. Δp_t^a is fitted by the HARX$_{depth}$ model that is (re-)estimated monthly for buys and sells of each stock group. The K consecutive trades satisfy two conditions: (i) each has a volume smaller than its prevailing market depth, and (ii) their total volume (i.e. the volume of the artificially aggregated trade) is larger than or equal to the prevailing depth of the first trade. "Obs." denotes the number of series of K satisfying consecutive trades. $\%\Delta p_t^S = (\Delta p_t^a - \Delta p_t^K)/\Delta p_t^a \times 100\%$ computes the proportion of the immediate price impact that can be saved by the order split. This table reports four different values for K: $K = 2, 5, 10$, and all, where the latter refers to all observed series of consecutive same-sign same-day trades that meet the above two conditions. Figures for Δp_t^K and Δp_t^a are the average values.

Table 6: **Order imbalance and price impact gap regressions**

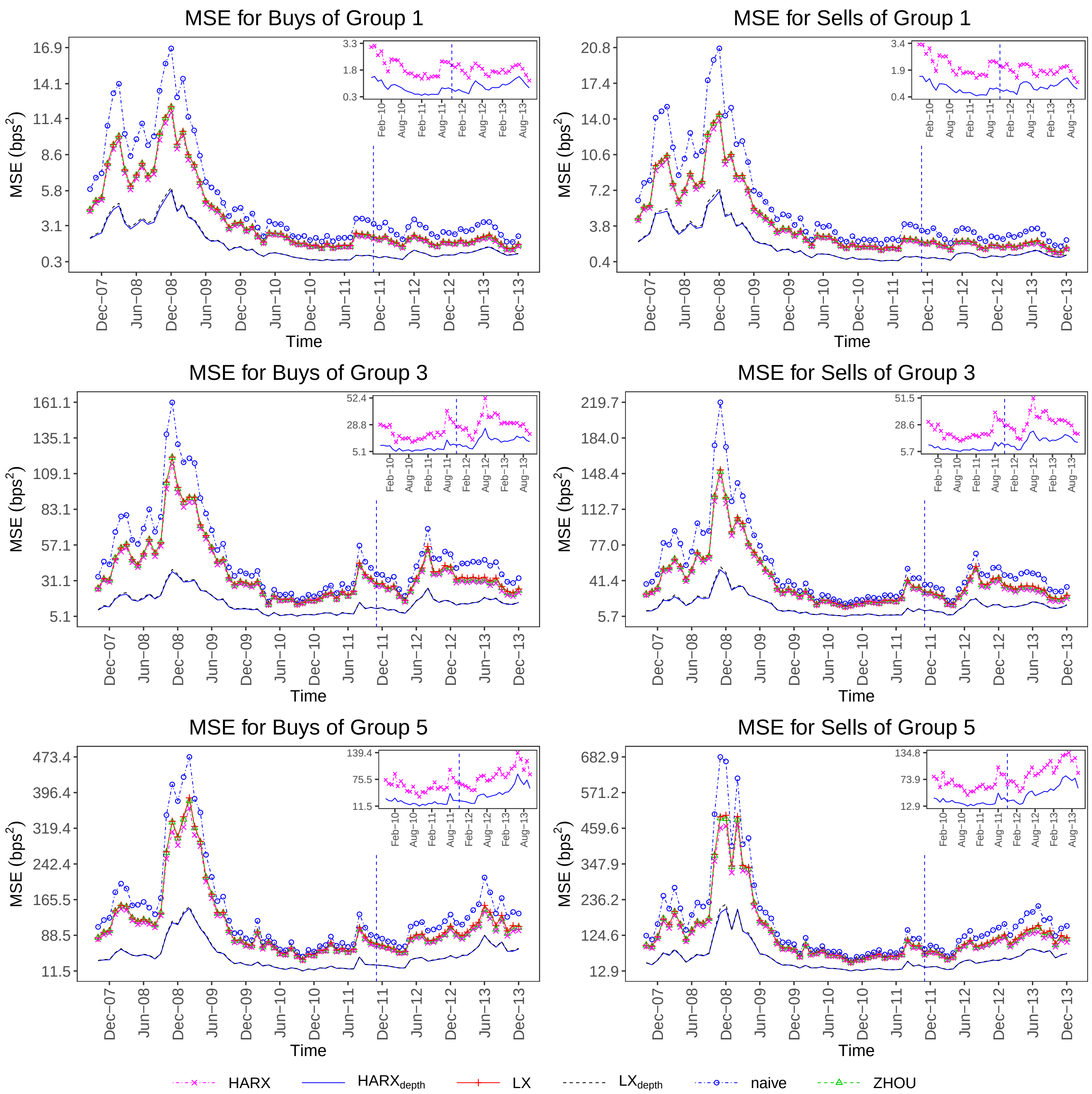
	5-minute window				30-minute window			
(a) Group 1	**Buys**		**Sells**		**Buys**		**Sells**	
	OIB	**PIG**	**OIB**	**PIG**	**OIB**	**PIG**	**OIB**	**PIG**
$I_{v_{i,t} \geq depth_{i,t}}$	-0.034***	-1.633***	0.009***	-1.626***	-0.018***	-1.811***	-0.001	-1.713***
	(-36.91)	(-53.43)	(8.68)	(-52.07)	(-25.89)	(-25.69)	(-0.62)	(-24.73)
LnVolume	0.005***	0.459***	0.002***	0.421***	0.002***	0.486***	0.002***	0.256***
	(16.89)	(54.27)	(6.76)	(46.60)	(10.66)	(24.28)	(7.31)	(11.94)
LnDepth	-0.000	-0.464***	-0.017***	-0.427***	-0.001	-0.452***	-0.011***	-0.301***
	(-1.03)	(-38.30)	(-33.12)	(-35.06)	(-1.35)	(-13.31)	(-20.34)	(-8.72)
Intercept	0.062***	2.517***	0.119***	2.502***	0.058***	3.889***	0.096***	1.365
	(14.34)	(9.99)	(26.13)	(9.41)	(14.41)	(3.45)	(22.48)	(1.21)
Day of week	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time of day	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R2	0.003	0.002	0.010	0.002	0.006	0.003	0.011	0.002
Obs	2,720,943	2,720,943	2,460,728	2,460,728	2,720,943	2,720,943	2,460,728	2,460,728
(b) Group 3	**Buys**		**Sells**		**Buys**		**Sells**	
	OIB	**PIG**	**OIB**	**PIG**	**OIB**	**PIG**	**OIB**	**PIG**
$I_{v_{i,t} \geq depth_{i,t}}$	-0.115***	-5.959***	0.123***	-6.249***	-0.029***	-5.843***	0.032***	-6.522***
	(-74.87)	(-100.44)	(78.90)	(-104.25)	(-30.06)	(-47.41)	(32.97)	(-52.07)
LnVolume	0.005***	1.149***	-0.001*	1.127***	0.000	1.216***	0.003***	1.057***
	(10.96)	(74.14)	(-1.73)	(67.50)	(-0.01)	(35.68)	(10.08)	(29.48)
LnDepth	0.003***	-1.405***	-0.003***	-1.519***	0.006***	-1.363***	-0.008***	-1.588***
	(5.27)	(-65.57)	(-4.33)	(-68.92)	(13.44)	(-26.27)	(-16.59)	(-29.99)
Intercept	0.066***	8.686***	-0.106***	9.979***	-0.001	7.019***	-0.012**	11.175***
	(11.92)	(30.13)	(-18.95)	(33.94)	(-0.16)	(7.06)	(-2.14)	(11.12)
Day of week	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time of day	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R2	0.011	0.011	0.012	0.012	0.006	0.003	0.008	0.004
Obs	1,588,052	1,588,052	1,553,436	1,553,436	1,588,052	1,588,052	1,553,436	1,553,436
(c) Group 5	**Buys**		**Sells**		**Buys**		**Sells**	
	OIB	**PIG**	**OIB**	**PIG**	**OIB**	**PIG**	**OIB**	**PIG**
$I_{v_{i,t} \geq depth_{i,t}}$	-0.193***	-10.736***	0.201***	-11.323***	-0.058***	-11.997***	0.061***	-12.137***
	(-76.87)	(-81.56)	(78.87)	(-87.86)	(-32.15)	(-49.71)	(33.80)	(-50.62)
LnVolume	-0.007***	1.946***	0.015***	2.164***	-0.006***	2.004***	0.015***	2.461***
	(-10.09)	(60.25)	(20.79)	(58.98)	(-11.73)	(30.82)	(25.71)	(34.13)
LnDepth	0.011***	-2.064***	-0.010***	-2.511***	0.012***	-2.015***	-0.015***	-2.486***
	(12.45)	(-56.28)	(-10.82)	(-66.44)	(15.98)	(-24.44)	(-18.57)	(-30.08)
Intercept	0.171***	11.782***	-0.230***	14.282***	0.018**	10.093***	-0.068***	13.004***
	(22.00)	(24.80)	(-28.94)	(29.71)	(2.34)	(7.81)	(-8.81)	(9.93)
Day of week	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time of day	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R2	0.022	0.015	0.025	0.017	0.011	0.006	0.016	0.007
Obs	875,051	875,051	826,846	826,846	875,051	875,051	826,846	826,846

This table reports the results for pooled-OLS estimates of

$$y_{i,t,\delta} = c + \alpha_1 I_{v_{i,t} \geq depth_{i,t}} + \alpha_2 LnVolume_{i,t} + \alpha_3 LnDepth_{i,t} + \alpha_4 day_{i,t} + \alpha_5 block_{i,t} + \eta_{i,t,\delta},$$

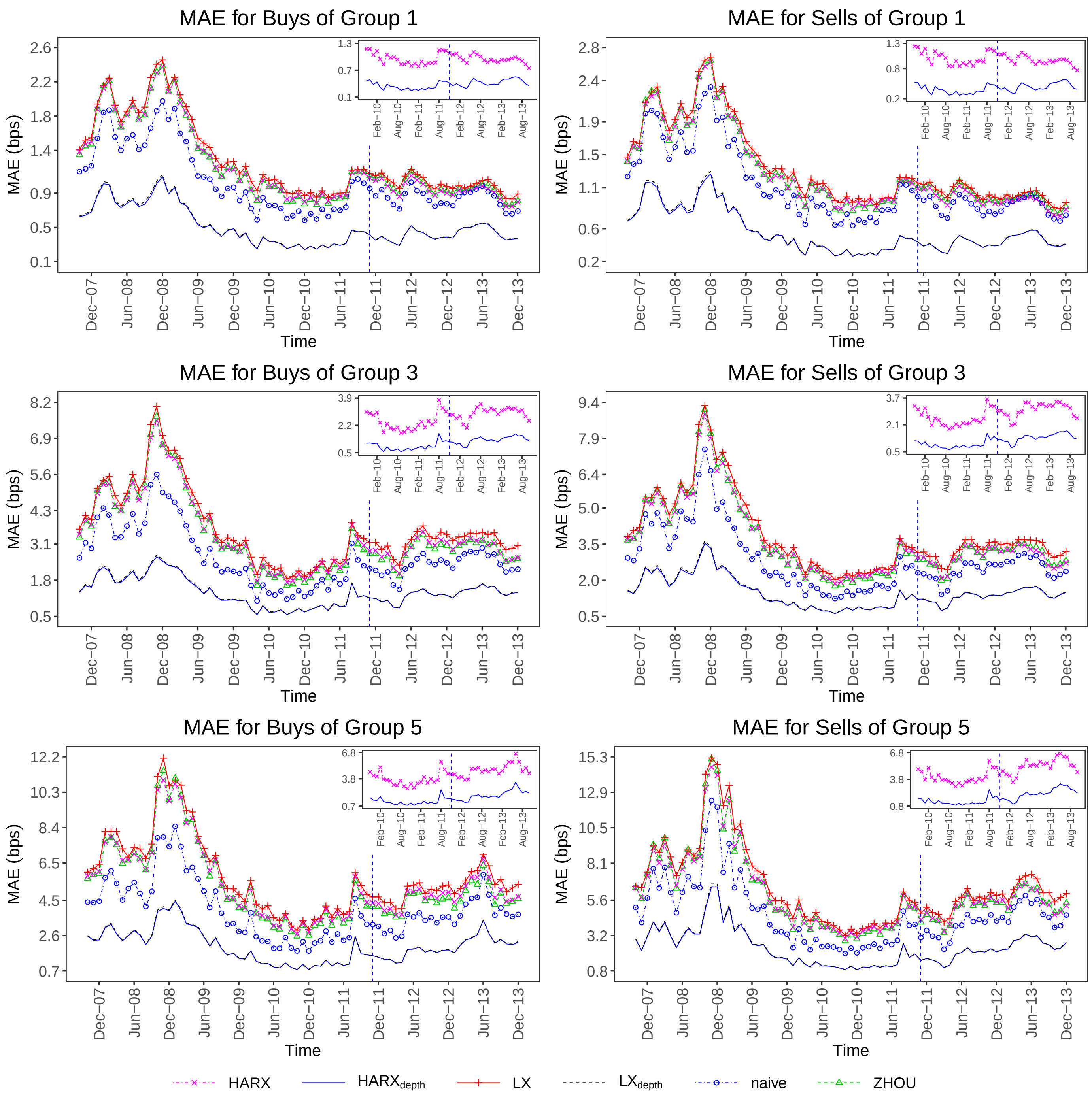
where $y_{i,t,\delta}$ is either (i) the order imbalance ($OIB_{i,t,\delta}$), defined as the ratio of the total buy volume minus sell traded volume to the total traded volume during a δ-minute interval after a trade at time t in stock i, or (ii) the price impact gap (PIG), defined as the signed difference between the permanent price impact ($\Delta p_{i,t,\delta}$, in basis points (bps)) of a trade at time t in stock i over a δ-minute interval and the immediate price impact ($\Delta p_{i,t}$, in bps) of the trade, i.e. PIG $= \epsilon_{i,t} \times (\Delta p_{i,t,\delta} - \Delta p_{i,t})$, where $\epsilon_{i,t}$ is the sign (+1 for buy and -1 for sell) of a trade. $I_{v_{i,t} \geq depth_{i,t}}$ is a dummy variable for a large transaction whose volume is larger than or equal to the prevailing opposite-side depth right before the trade. $LnVolume_{i,t}$ and $LnDepth_{i,t}$ are respectively the natural logarithm of the share volume of a trade and the prevailing opposite-side depth right before the trade. $day_{i,t}$ and $block_{i,t}$ are respectively the set of dummies at time t for day of week and time of day effects. The regressions relate to three stock groups (Groups 1, 3 and 5) listed on the S&P/ASX200 index in 2007. The table reports the coefficient estimates and Newey-West heteroskedasticity and autocorrelation consistent t-statistics (in parentheses) for $\delta = 5$ minutes and 30 minutes. Adj R2 is the adjusted R-squared, Obs is the number of observations and ***, **, and * respectively denote statistical significance at the 1%, 5% and 10% levels.

Figure 1: **Out-of-sample one-month MSE of different price impact models from October 2007 to December 2013 for Groups 1, 3 and 5**



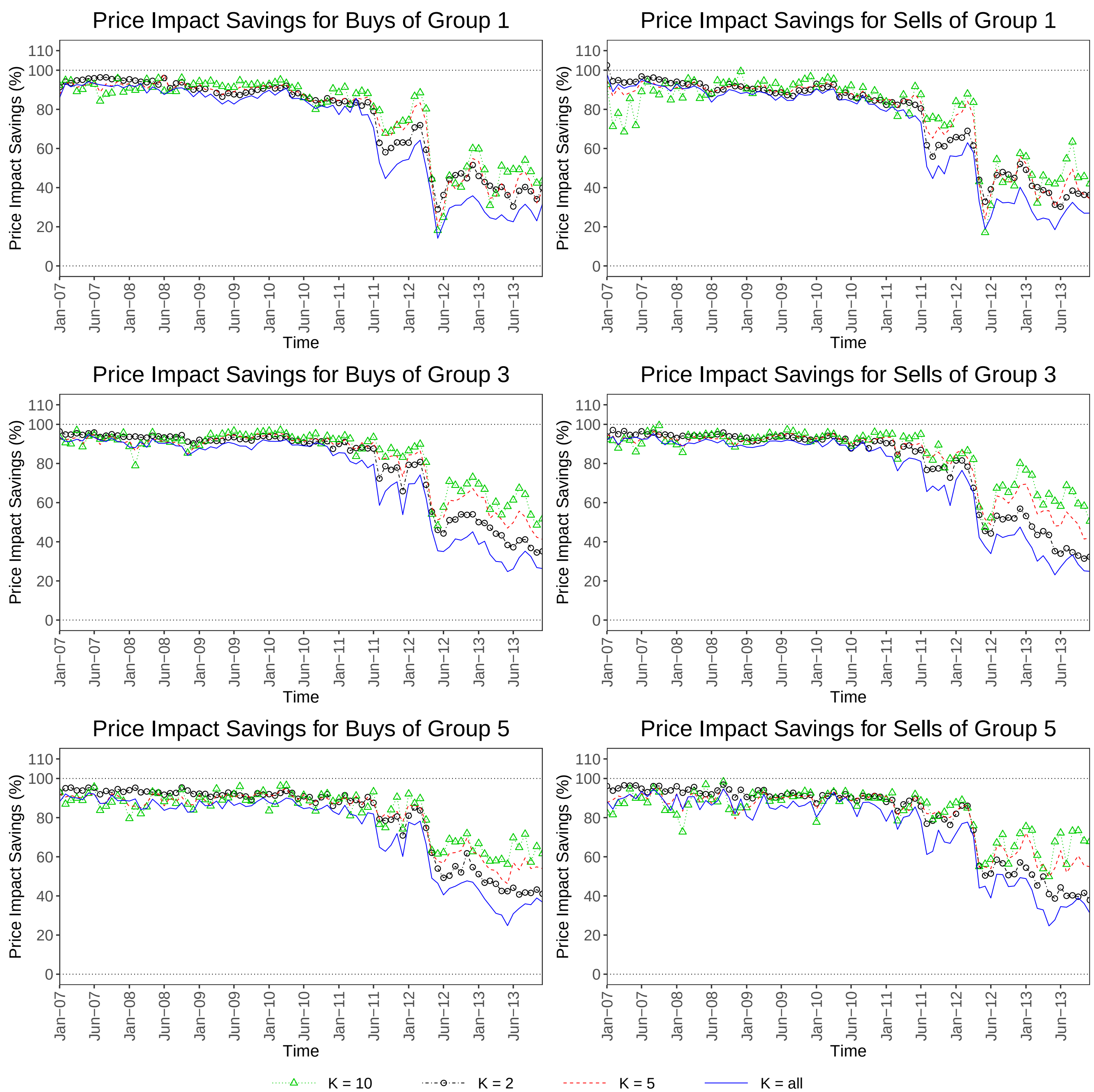
This figure depicts out-of-sample one-month MSEs (measured in bps^2) of different price impact models from October 2007 to December 2013 for Groups 1, 3 and 5. For the ease of graph inspection, only the MSE series for six models including HARX, HARX$_{depth}$, LX, LX$_{depth}$, naive, and ZHOU (the best model amongst LFM1, LFM2, and ZHOU - see Table 2) are plotted. Those for the remaining models which typically have higher MSEs are of less interest. The insert in the top-right of each graph zooms in on the MSE series of the HARX and HARX$_{depth}$ models over a four year period around the introduction of Chi-X on 31 October 2011 (which is marked by a vertical dashed blue line in each graph and its insert).

Figure 2: **Out-of-sample one-month MAE of different price impact models from October 2007 to December 2013 for Groups 1, 3 and 5**



This figure depicts out-of-sample one-month MAEs (measured in bps) of different price impact models from October 2007 to December 2013 for Groups 1, 3 and 5. For the ease of graph inspection, only the MAE series for six models including HARX, HARX_{depth}, LX, LX_{depth}, naive, and ZHOU (the best model amongst LFM1, LFM2, and ZHOU - see Table 2) are plotted. Those for the remaining models which typically have higher MAEs are of less interest. The insert in the top-right of each graph zooms in on the MAE series of the HARX and HARX_{depth} models over a four year period around the introduction of Chi-X on 31 October 2011 (which is marked by a vertical dashed blue line in each graph and its insert).

Figure 3: **Monthly price impact savings in 2007-2013 for Groups 1, 3 and 5**



Notes: Price impact savings are the proportion of the price impact of a big order that can be reduced by splitting it into a series of smaller trades. They are computed as $(\Delta p_t^a - \Delta p_t^K)/\Delta p_t^a \times 100\%$, where Δp_t^K is the observed price impact of a series of K consecutive trades in the same direction and on the same day, and Δp_t^a is the predicted immediate price impact of an *artificial* transaction that aggregates these trades. The K consecutive trades satisfy two conditions: (i) each has a volume smaller than its prevailing market depth, and (ii) their total volume (i.e. the volume of the artificially aggregated trade) is larger than or equal to the prevailing depth of the first trade. The immediate price impact of the artificial trade is estimated by the HARX_{depth} model. Each graph plots the time series of the monthly price impact savings for $K = 2, 5, 10$, and all, where the latter refers to all observed series of consecutive same-sign same-day trades that meet the above two conditions.

Appendix A List of Immediate Price Impact Models

This Appendix provides detailed specifications for each of the models that we consider in this paper. The model names correspond to names that we have used in the text, as do the model numbers.[31,32]

$$\text{HARX}_{depth}: \quad \Delta p_{i,t} = \Big[a + \phi_1 \Delta p_{i,t-1} + \phi_5 \overline{\Delta p}_{i,t,5} + \phi_{20} \overline{\Delta p}_{i,t,20} + \phi_{50} \overline{\Delta p}_{i,t,50} + \delta_1 day_{i,t} + \delta_2 block_{i,t} + \alpha \left(\frac{v_{i,t}}{\overline{v}_{i,t}}\right) + \beta M_{i,t} + \gamma \sigma_{i,t}\Big] \times I_{v_{i,t} \geq depth_{i,t}} + \eta_{i,t}, \tag{A.1}$$

$$\text{HARX}: \quad \Delta p_{i,t} = a + \phi_1 \Delta p_{i,t-1} + \phi_5 \overline{\Delta p}_{i,t,5} + \phi_{20} \overline{\Delta p}_{i,t,20} + \phi_{50} \overline{\Delta p}_{i,t,50} + \delta_1 day_{i,t} + \delta_2 block_{i,t} + \alpha \left(\frac{v_{i,t}}{\overline{v}_{i,t}}\right) + \beta M_{i,t} + \gamma \sigma_{i,t} + \eta_{i,t}, \tag{A.2}$$

$$\text{LX}_{depth}: \quad \Delta p_{i,t} = \Big[a + \alpha \left(\frac{v_{i,t}}{\overline{v}_{i,t}}\right) + \beta M_{i,t} + \gamma \sigma_{i,t}\Big] \times I_{v_{i,t} \geq depth_{i,t}} + \eta_{i,t}, \tag{A.3}$$

$$\text{LXb}: \quad \Delta p_{i,t} = a + \alpha \left(\frac{v_{i,t}}{\overline{v}_{i,t}}\right) + \beta M_{i,t} + \gamma \sigma_{i,t} + \delta_1 day_{i,t} + \delta_2 block_{i,t} + \eta_{i,t}. \tag{A.4}$$

$$\text{LX}: \quad \Delta p_{i,t} = a + \alpha \left(\frac{v_{i,t}}{\overline{v}_{i,t}}\right) + \beta M_{i,t} + \gamma \sigma_{i,t} + \eta_{i,t}. \tag{A.5}$$

$$\text{ZHOU}: \quad \Delta p_{i,t} = \epsilon_{i,t} \times \left(\frac{v_{i,t}}{\overline{v}_{i,t}}\right)^{\alpha} \times \left|\overline{\Delta p}_{i,t}\right| + \eta_{i,t}, \tag{A.6}$$

$$\text{LFM1}: \quad \Delta p_{i,t} = \epsilon_{i,t} \times \left(\frac{v_{i,t}}{\overline{v}_{i,t}}\right)^{\alpha} \times \frac{1}{M_{i,t}^{\beta}} + \eta_{i,t}, \tag{A.7}$$

$$\text{LFM2}: \quad \Delta p_{i,t} = \epsilon_{i,t} \times \left(\frac{v_{i,t}}{\overline{v}_{i,t}}\right)^{\alpha} \times \frac{1}{M_{i,t}^{\beta}} \times \sigma_{i,t} + \eta_{i,t}, \tag{A.8}$$

$$\text{naive}: \quad \Delta p_{i,t} = 0. \tag{A.9}$$

Amongst the above models, HARX is a non-depth analogue of HARX_{depth} that does not discriminate between zero and non-zero price impact trades, LXb is the linear model nested in HARX that excludes the effects of price impact dynamics, and LX further excludes the intra- and inter-day seasonalities from LXb. Meanwhile, the LX_{depth} model is the depth-augmented version of LX and it can also be obtained from HARX_{depth} by excluding time series variables (including price impact dynamics and time of day/week effects) from the latter model.

ZHOU (which is based on Zhou (2012)), LFM1 and LFM2 (which are based on Lillo et al. (2003))

[31] The variable $\epsilon_{i,t}$ in the ZHOU, LFM1 and LFM2 models indicates the direction of the t-th trade in stock i and equals 1 (-1) for buys (sells).

[32] The variable $\overline{\Delta p}_{i,t}$ in the ZHOU model (in Equation (A.6)) is defined as the average price impact of all trades in the trading day that are prior to (but *not* including) the t-th trade and have the same direction. This variable can be measured ex-ante and is in contrast to the ex-post $\overline{\Delta p}_{i,t}$ variable used in Zhou (2012) that measures the average price impact in the entire year in which the t-th trade occurs. Note an important difference in the definitions of ex-ante $\overline{v}_{i,t}$ and $\overline{\Delta p}_{i,t}$ measures. The volume of the t-th trade which is known to the initiator before the execution of the trade is *included* when calculating $\overline{v}_{i,t}$; whereas the price impact of the trade is *excluded* in the calculation of $\overline{\Delta p}_{i,t}$ since it can only be realized ex-post.

are three traditional immediate price impact models that were developed in the literature, and they specify a power-law dependence of immediate price impact on trading volume. Finally, the naive model is the model that always predicts a zero immediate price impact for a trade.

Some additional models (LXb$_{depth}$, ZHOU$_{depth}$, LFM1$_{depth}$ and LFM2$_{depth}$ are briefly considered in subsection 4.2.5. These models are depth augmented versions of specifications (A.4) and (A.6) - (A.8), respectively.

Appendix B Nonparametric Models

We extend our analysis in the main text to study some nonparametric models of immediate price impact that can allow for the power law structures proposed by Lillo et al. (2003) and used by Zhou (2012). Such models can also incorporate nonlinear diurnal and day of the week effects that are often seen in high frequency asset pricing settings. We are interested in whether a more flexible modeling strategy can capture relationships implied by theory, and whether a generalization of our initial model in Equation (1) can lead to forecasting gains.

We work with a class of nonparametric models introduced by Hastie and Tibshirani (1986, 1990) that are collectively called Generalized Additive Models (GAMs). A GAM is a generalized linear model that expresses a dependent variable of interest or some function of the dependent variable as the sum of smooth functions of possible predictors. Specifically, the parameters in linear specifications of the relation between the response variable (y) and one or more explanatory variables ($z_1, z_2, ..., z_n$) are replaced by variable-specific smoothing functions $g_j(z_j)$ that create smooth patterns between y and z_j via "local averaging". Applications of GAMs are widespread in medical and epidemiology literatures (Hastie and Tibshirani, 1995, Schwartz, 1999, Dominici et al., 2002) and they have also been used to study asset pricing (e.g. Foresi and Peracchi, 1995, Hou, 2013). Knez and Ready (1996) used another type of nonparametric model in their study of price improvement. A GAM specification that is useful in our context is given by

$$y_{i,t} = G\left(z_{1,i,t}, z_{2,i,t}, ..., z_{(|X|+|T|),i,t}\right) + \eta_{i,t} = \sum_{j=1}^{|X|} g_j(X_{j,i,t}) + \sum_{k=1}^{|T|} g_k(T_{k,i,t}) + \eta_{i,t}, \tag{B.10}$$

where $y_{i,t} = \Delta p_{i,t}$ is immediate price impact, $X_{i,t}$ and $T_{i,t}$ consist of the same trade attributes and time series variables as before, and $|X|$ and $|T|$ denote the cardinalities of the variable sets $X_{i,t}$ and $T_{i,t}$. We consider generalizations of our HARX$_{depth}$ and HARX models given by the following:

$$\text{GHARX}_{depth}: \qquad \Delta p_{i,t} = \Big[a + g_1\left(\frac{v_{i,t}}{\overline{v}_{i,t}}\right) + g_2(M_{i,t}) + g_3(\sigma_{i,t}) + g_4(day_{i,t}) + g_5(block_{i,t}) + g_6(\Delta p_{i,t-1})$$
$$+ g_7(\overline{\Delta p}_{i,t,5}) + g_8(\overline{\Delta p}_{i,t,20}) + g_9(\overline{\Delta p}_{i,t,50})\Big] \times I_{v_{i,t} \geq depth_{i,t}} + \eta_{i,t}, \tag{B.11}$$
$$\text{GHARX}: \qquad \Delta p_{i,t} = a + g_1\left(\frac{v_{i,t}}{\overline{v}_{i,t}}\right) + g_2(M_{i,t}) + g_3(\sigma_{i,t}) + g_4(day_{i,t}) + g_5(block_{i,t}) + g_6(\Delta p_{i,t-1})$$
$$+ g_7(\overline{\Delta p}_{i,t,5}) + g_8(\overline{\Delta p}_{i,t,20}) + g_9(\overline{\Delta p}_{i,t,50}) + \eta_{i,t}, \tag{B.12}$$

and we fit these nonparametric models via cubic regression spline smoothing.[33]

We demonstrate some of the estimated relationships in Figure B.1, which provides GHARX and GHARX$_{depth}$ estimates of the buy/sell relationships between immediate price impact and volume, volatility, and time of day for Stock Group 5 in 2007.[34] While the GHARX plots are based on all in-sample observations (including the zero price impact trades) and can be directly compared to figures produced by previous studies, the GHARX$_{depth}$ plots are based on just those trades for which $I_{v_{i,t} \geq depth_{i,t}} = 1$.[35]

<<INSERT FIGURE B.1 ABOUT HERE>>

Focusing on the GHARX graphs for buyer-initiated transactions in Panel (a) of Figure B.1, we see a positive concave relation between immediate price impact and trading volume that fits a power-law function well, reaffirming the power-law theory that has been well documented in the literature (e.g. Gabaix et al., 2003, 2006). Similarly, we find strong support for the direct dependence of price impact on volatility as parameterized in prior studies (Torre, 1997, Almgren et al., 2005). The diurnal plot suggests that immediate price impact generally decreases as the trading day progresses, as discussed in Wilinski et al. (2015). Price impact is higher earlier in the day due to institutional investors who are potentially well-informed about news that has accumulated overnight, whereas trading towards the end of the day is lower because it is mainly due to uninformed market participants (Anand et al., 2005, Bloomfield et al., 2005, Duong et al., 2009).

[33]We use the `gam` function in the `mgcv` **R** package (Wood, 2017) to estimate the nonparametric models. We use cubic regression splines because they are simple and lead to "directly interpretable" parameter estimates (Wood, 2006), and use Generalized Cross-Validation (GCV) to select significant variables and estimate each model.

[34]We obtain qualitatively similar plots for other stock groups in other years.

[35]Each plot uses observations that fall within the 5% and 95% quantiles of the relevant predictor. This truncation is used to ensure that plots relate to typical (rather than outlying) values of the predictor, and it is applied to all transactions for the GHARX plots, and only to those trades that have impact (i.e. $I_{v_{i,t} \geq depth_{i,t}} = 1$) for the GHARX$_{depth}$ plots. The truncation is undertaken on a stock-by-stock basis, separately for buys and sells.

The GHARX plots for seller-initiated transactions of Group 5 are depicted in Panel (b) of Figure B.1. As for buys, the size of immediate price impact for sells exhibit a positive and persistent dynamic structure. Since price impact is positive for purchases and negative for sales, the plots in Panel (b) mirror those in Panel (a). Nevertheless, there are subtle asymmetries in price impact between purchases and sales, consistent with work by Jondeau et al. (2015), who find that asymmetries in price impact are related to stocks' liquidity. These asymmetries are very evident from the nonparametric models, but may not be captured easily by tightly specified parametric models.

Comparison of the GHARX and GHARX$_{depth}$ plots (i.e. panels (a) and (b) vs panels (c) and (d) shows how the incorporation of a market depth dummy into the models affects the relation between immediate price impact and its determinants. Not surprisingly, the measured impact of each determinant is now significantly higher. With respect to buyer (seller)-initiated transactions, the overall positive (negative) dependence of market impact on trading volume and volatility is preserved but the concavity between price impact and scaled volume in Panels (a) and (b) become essentially linear in Panels (c) and (d). Meanwhile, the essentially linear relationships between immediate price impact and volatility in panels (a) and (b) are less obvious for data that is far from the mean of the data. Further, the diurnal characteristics implied by the GHARX$_{depth}$ model are more strongly U (inverted U)-shaped in Panels (c) and (d).

The nonparametric specification of some of our models finds strong nonlinear dependencies of immediate price impact on various predictors. Such nonlinearities are generally consistent with those previously documented in the literature when all transactions are examined, but they become quite different when only large trades that are of larger size than the available depth (i.e. those trades that have price impact) are considered.

We compare the forecast performance of the GHARX$_{depth}$ and GHARX models with their parametric HARX$_{depth}$ and HARX analogues in Table B.1. The results show that on average the GHARX$_{depth}$ and GHARX models offer statistically significant improvements over their HARX counterparts in more than 96% (73%) of the forecast windows when assessed using MSE (MAE), and lead to more than an 8% reduction in MSE for the GHARX$_{depth}$ vs HARX$_{depth}$ MSE comparison. The other percentage reductions are smaller in magnitude but are still statistically significant, suggesting that the generalized specifications do not simply overfit the data, but rather capture genuine features of the data that lead to forecasting gains. We note (without reporting details) that the MSE (MAE) measures for the GHARX$_{depth}$ model are all significantly statistically smaller than those for the GHARX model by about 60%, for all stock groups and for both buys and sells, showing yet again the value of the depth indicator for forecasting immediate price impact.

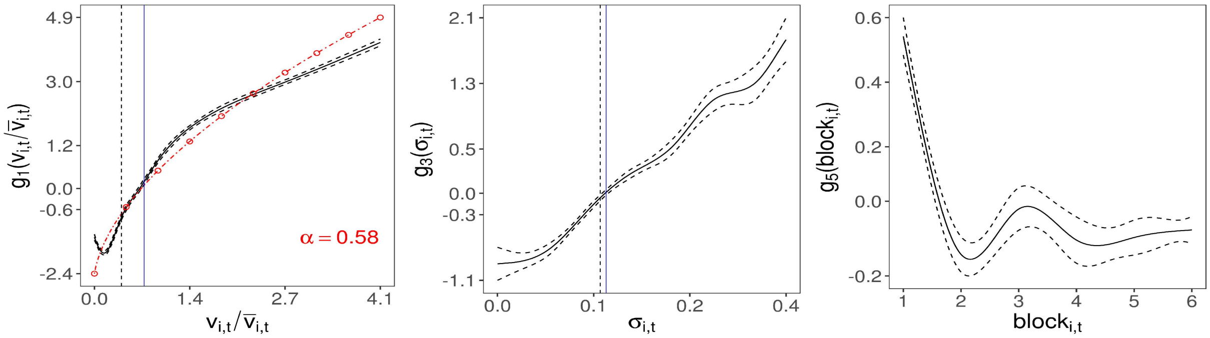
Table B.1: **Performance of nonparametric versus parametric models**

		Panel A: MSE		**Panel B: MAE**	
		GHARX$_{depth}$ vs. HARX$_{depth}$	GHARX vs. HARX	GHARX$_{depth}$ vs. HARX$_{depth}$	GHARX vs. HARX
Group 1	Buys	98.67 [6.80]	100.00 [0.50]	89.33 [2.66]	92.00 [0.92]
	Sells	100.00 [7.11]	98.67 [0.42]	86.67 [2.67]	92.00 [0.79]
Group 2	Buys	97.33 [8.07]	98.67 [0.51]	88.00 [3.45]	97.33 [0.73]
	Sells	98.67 [8.18]	98.67 [0.40]	89.33 [3.45]	96.00 [0.72]
Group 3	Buys	96.00 [10.12]	97.33 [0.51]	85.33 [4.27]	58.67 [0.25]
	Sells	97.33 [10.32]	98.67 [0.41]	90.67 [4.45]	74.67 [0.36]
Group 4	Buys	98.67 [8.98]	93.33 [0.32]	89.33 [4.60]	54.67 [0.02]
	Sells	98.67 [7.90]	92.00 [0.31]	90.67 [4.61]	72.00 [0.24]
Group 5	Buys	96.00 [7.73]	94.67 [0.50]	73.33 [1.97]	44.00 [-0.21]
	Sells	96.00 [7.87]	93.33 [0.33]	65.33 [2.02]	53.33 [0.01]
Average		97.73 [8.31]	96.53 [0.42]	84.80 [3.41]	73.47 [0.38]

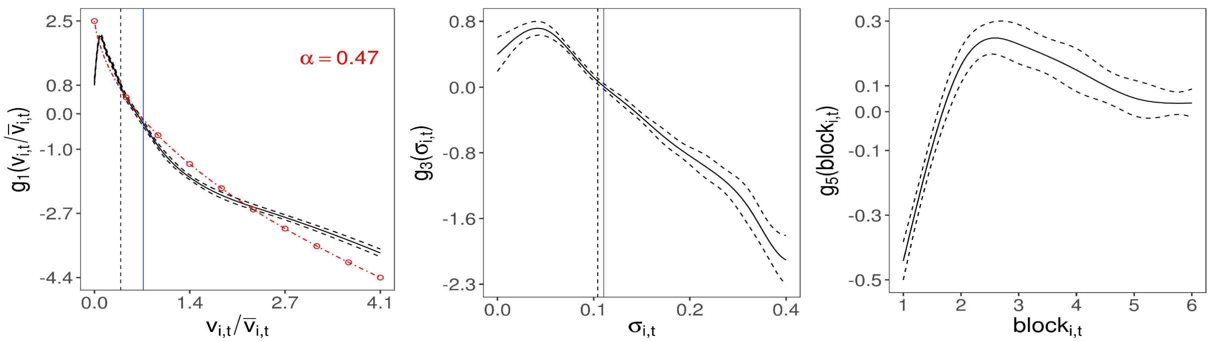
This table compares the performance of nonparametric immediate price impact models with their parametric analogues over 75 out-of-sample months (from October 2007 to December 2013) for five stock groups on the S&P/ASX200 index. It reports (i) the percentages of times that each nonparametric price impact model statistically significantly outperforms its respective parametric analogue; and (ii) the percentage reduction in the whole out-of-sample MSEs/MAEs brought about by nonparametric modeling (in brackets). The "**Average**" row reports the simple averages for both buys and sells for all five stock groups. For each one-month out-of-sample window, the statistical outperformance (at a 5% level) of one model over another model is judged by the Giacomini and White (2006) conditional predictive accuracy test using MSE (in Panel A) and MAE (in Panel B). Results, reported in %, are based on the analysis of nine-month in-sample and one-month out-of-sample windows rolled one month ahead.

Figure B.1: **Price impact implied by predictors in the GHARX and GHARX$_{depth}$ models for Group 5 in Jan-Sep 2007**

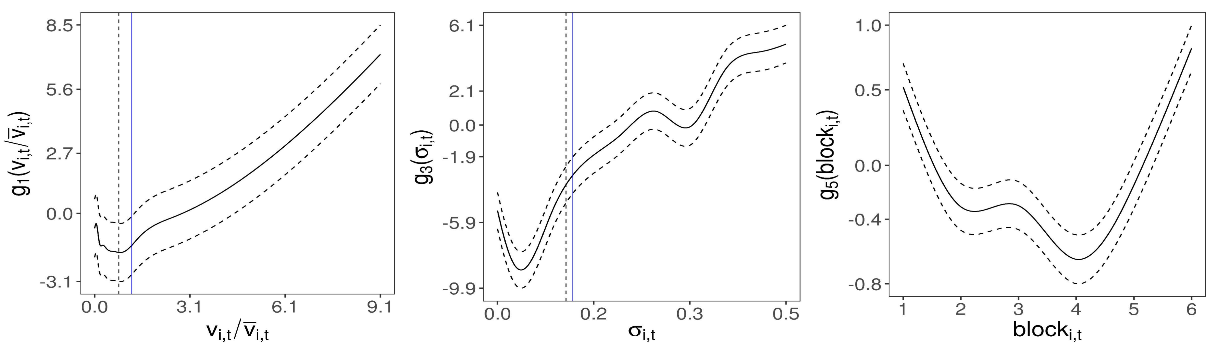
(a) **GHARX model for buys of Group 5 in Jan-Sep 2007**



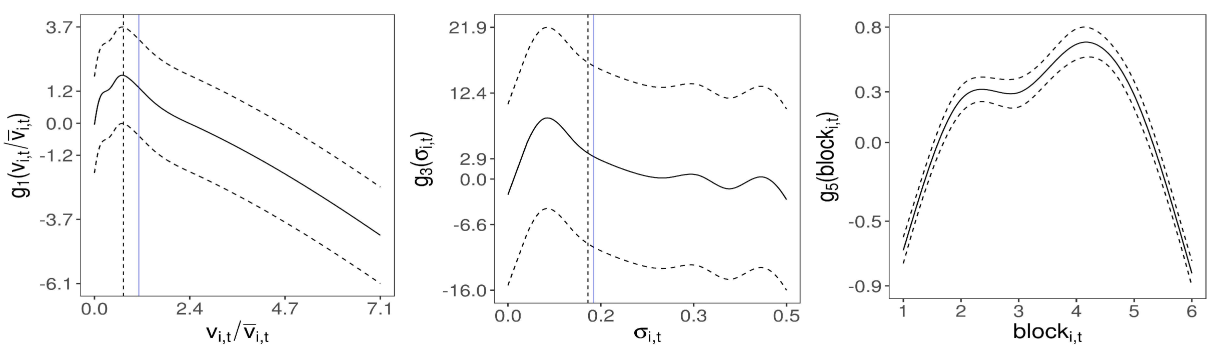
(b) **GHARX model for sells of Group 5 in Jan-Sep 2007**



(c) **GHARX$_{depth}$ model for buys of Group 5 in Jan-Sep 2007**



(d) **GHARX$_{depth}$ model for sells of Group 5 in Jan-Sep 2007**



Notes: In each panel, graphs from left to right in order reveal the dependence of immediate price impact on scaled volume ($v_{i,t}/\overline{v}_{i,t}$), volatility ($\sigma_{i,t}$, in %), and time of day ($block_{i,t}$). Graphs for other predictors are omitted for brevity. Each graph zooms in the central 90% distribution of the relevant predictor, with the solid curve representing the estimated relation and the dashed curves representing the 95% confidence interval. The vertical dashed and solid lines respectively position the median and mean of each predictor, except for time of day. The y-axes are measured in bps. In the scaled volume graph of the GHARX model, the dotted dashed curve with circle markers represents the best fitted power law function whose exponent, α, is reported at the bottom right (for buys) or at the top right (for sells) of the graph.