RATIONALITY & COMPETITION
CRC TRR 190

Order Exposure and Liquidity Coordination: Does Hidden Liquidity Harm Price Efficiency?

Gökhan Cebirogly (University of Vienna)
Nikolaus Hautsch (University of Vienna)
Ulrich Horst (Humboldt University Berlin)

Discussion Paper No. 28

April 13, 2017

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

Order Exposure and Liquidity Coordination: Does Hidden Liquidity Harm Price Efficiency?*

Gökhan Cebiroğlu†
University of Vienna

Nikolaus Hautsch‡
University of Vienna
Center for Financial Studies

Ulrich Horst§
Humboldt Universität zu Berlin

April 11, 2017

Abstract

We show that the excessive use of hidden orders causes artificial price pressures and abnormal asset returns. Using a simple game-theoretical setting, we demonstrate that this effect naturally arises from mis-coordination in trading schedules between traders, when suppliers of liquidity do not sufficiently disclose their trade intentions. As a result, hidden liquidity can increase trading costs and induce excess price fluctuations unrelated to information. Using NASDAQ order book data, we find strong empirical support and illustrate that hidden liquidity is higher if bid-ask spreads are smaller and relative tick sizes are higher.

JEL classification: G02, G10, G23
Keywords: hidden liquidity, trade synchronization, trading frictions, counterparty attraction, limit order book

*The paper formerly circulated under the title "Does Hidden Liquidity Harm Price Efficiency?". We thank Torben Andersen, Damir Filipovic, Tony Hall, Joel Hasbrouck, Albert Kyle, Bruce Lehmann, Helmut Lütkepohl, Albert Menkveld and Mathieu Rosenbaum for their helpful comments and suggestions. Hautsch acknowledges financial support from the Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF). Horst acknowledges financial support by the TRR 190 "Rationality and Competition: the Economic Performance of Individuals and Firms".

†Faculty of Business, Economics and Statistics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, *email*: goekhan.cebiroglu@univie.ac.at.

‡Faculty of Business, Economics and Statistics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, *email*: nikolaus.hautsch@univie.ac.at.

§Department of Mathematics, Unter den Linden 6, D-10099 Berlin, *email*: horst@mathematik.hu-berlin.de.

1 Introduction

A growing proportion of traders on financial markets perceive a tangible benefit in concealing their trading intentions from public view. To address the rising demand, exchange operators and markets have introduced a range of order types that allow traders to hide the full extent of their standing limit orders (such as reserve orders, Iceberg orders or hidden orders). As a result, the proliferation of hidden liquidity has grown significantly over the past decade and nowadays accounts for a sizable proportion of overall liquidity supply in electronic equity markets.[1]

Proponents of hidden liquidity argue that the proliferation of hidden volume attracts traders that would otherwise not partake in trading and thus increases market liquidity and lowers transaction costs.[2] Nonetheless over the recent years, the significant growth of hidden liquidity and dark trading practices has amplified the debate about its possible role in generating market frictions for the market as a whole.[3] Critics argue that hiding of trade intentions may have a negative impact in matching trade counterparties.

The general idea is that trade execution requires that counterparties trade synchronously and that, therefore, their trading schedules need to be synchronized. In limit order book markets, this sort of trade synchronization is facilitated through signaling of trade intentions. Typically, liquidity suppliers provide a signal of trade interest by submitting openly displayed orders. Liquidity demanders monitor the market and may initiate a trade as a response. Displayed orders, therefore, act as an instrument to coordinate order flow between liquidity suppliers and demanders.

Hence, if liquidity suppliers hide their presence, some liquidity demanders might miss out on mutually beneficial trades. This scenario is conceivable and most likely prevalent when possible trading counterparties, i.e., liquidity demanders, are strategic and selective in their choice of the timing (and location) of trades. As shown in this paper, in such a case, the decision about order display does not only affect trade execution, but has also wider implications for the price discovery process.

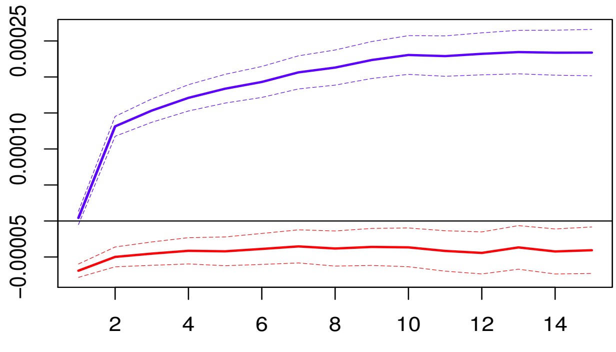
Consistent with this reasoning, our empirical analysis shows that hidden orders have a striking impact on prices. Figure 1 shows the averaged estimated cumulative reaction in one-minute mid-quote returns induced by an increase in hidden (blue line) order submissions. The plot shows that hidden order submissions move prices significantly. In contrast, *displayed* orders (red line) have a negligible effect. In line with this finding, we also show in this paper that – ceteris paribus – markets exhibiting a larger proportion of hidden liquidity tend to be more volatile.

[1] E.g., Bessembinder et al. (2009) show that 44% of volume in Euronext-Paris is hidden, while Frey and Sandas (2009) find a proportion of 16% of hidden liquidity in the German Xetra market.

[2] See Aitken et al. (2001).

[3] A statement by the European Commission is echoing the growing concern about dark trading practices in the context of the new MIFID II framework: "... Strict transparency rules will ensure that [hidden] trading of shares and other equity instruments which undermine efficient and fair price formation will no longer be allowed."

Figure 1: Cross-sectional averages of estimated cumulative effects of one-minute mid-quote returns due to positive one-standard-deviation shocks in hidden (blue) and displayed (red) bid-ask order imbalances. The effects correspond to impulse response functions based on a vector autoregressive model for one-minute returns, depth imbalances, bid-ask spreads, volatility, and order activity, estimated for 10 randomly picked stock traded at NASDAQ, November to December 2008. The dashed lines show the approximate 95% confidence intervals of the averaged impulse response functions.



In this paper, we show theoretically that these price pressures arise from mis-coordination between the supply and demand side of liquidity, when the former do not sufficiently display their trade intentions. Our results, therefore, suggest that mis-coordination can be a major source of non-informational trading frictions. We substantiate our conclusions by analyzing the microstructure foundations of hidden order submissions and their causal effect on prices in a structural equilibrium framework. Our model considers strategic liquidity suppliers and demanders. Liquidity suppliers trade in the public limit order book, face liquidity competition, and are impatient and pre-committed to trade (cf. Foucault et al. (2005)). Liquidity demanders passively monitor the order book market for trading opportunities and make strategic decisions about the timing and the venue of their trades (see, e.g., Grossmann (1992), Harris (1997), Hasbrouck and Saar (2009) and Bessembinder et al. (2009)). They only enter the limit order book market whenever the current market situation makes trading beneficial. These strategic liquidity demanders therefore provide a reservoir of *latent* liquidity demand. While such so-called 'latent traders' can monitor the public limit order book, liquidity suppliers only have a probabilistic belief about their presence.

Such a setting incorporates two important aspects determining the optimal (equilibrium) strategies of liquidity suppliers and demanders: i) competition in liquidity supply and ii) competition for latent order flow. The presence of liquidity competition confronts liquidity suppliers with the decision whether to quote more aggressively or to avoid competition by strategically hiding their trade intentions. On the other hand, competition for (latent) order flow incentivizes liquidity suppliers to signal their trade interest in order to attract such trade demand. Accounting for both aspects results into a trade-off between the benefits and costs of order display. Order display fuels liquidity competition and therefore increases liquidity suppliers' trade execution costs. On the other hand, order display increases the probability to attract trade counterparties which in turn reduces execution costs.

In the equilibrium, therefore, the optimal amount of displayed orders (the so-called 'display size') balances this trade-off. We show that the equilibrium display size depends on liquidity suppliers' beliefs about the presence of latent counterparty demand. If this presence is unlikely, they limit their display size to prevent other liquidity suppliers from overbidding. Conversely, if the probability for the presence of latent liquidity demand is high, it is optimal to fully reveal trading intentions to attract such counterparties. In this case, the costs arising from liquidity competition are overcompensated by the benefit of attracting counterparty liquidity demand.

Although off-exchange brokerage services charge a brokerage fee for their counterparty search service, they are generally still more cost-efficient for large traders than public limit order book markets as they provide more liquidity. Only if liquidity suppliers in the public limit order book provide sufficient *displayed* liquidity, it appears to be beneficial for the latent trader to trade in this market. Consistent with this reasoning, we show that there is a critical order display size for attracting latent liquidity demand. This display size can be interpreted as a liquidity premium required by the latent trader in order to enter the limit order book market. Liquidity suppliers offering this display size then maximize the probability of executing their limit orders against latent trade demand. A central feature of the equilibrium is, therefore, that liquidity supply and (latent) liquidity demand is optimally synchronized, minimizing the costs of trading for all counterparties and avoiding frictions arising from sub-optimal coordination.

If liquidity suppliers, however, undercut this optimal display size, i.e., hiding too much of their orders, they face a higher risk of *not* (fully) executing their order. Whenever they are pre-committed to trade and thus face time constraints to liquidate their positions, they need to enforce trade execution by increasing order aggressiveness. Consequently, non-executed hidden orders are canceled and traded as market orders. This ultimately generates price pressures, which are unrelated to information and would be completely absorbed in case of perfect liquidity synchronization. Our theory produces therefore a set of testable predictions on the causal effects of excessive hidden order submissions. The predictions build on the assumption that liquidity suppliers face liquidation constraints and thus need to re-position their orders in case of non-execution. Consequently, our model predicts changes in the order flow composition whenever there is too much hidden liquidity on one side of the market. In these situations, we predict lower order execution rates on the same side of the market, higher cancellation rates and an increase of market order submissions on the same side of the market.

To empirically validate the testable predictions, we employ a unique data set on hidden orders and high-frequency order messages from NASDAQ. The data set yields high-frequency snapshots of the entire order book, including the hidden part. It therefore allows us to relate changes in the hidden and displayed order book imbalance on the order flow. To appropriately capture the dynamics in high-frequency order flow data, we specify a high-frequency vector autoregressive process of hidden and displayed depth imbalances, order flow, mid-quote returns and volatility. The short-run and long-run effects of hidden order submissions are quantified by estimated impulse response functions. Beyond a distinct effect of hidden orders on subsequent returns, as depicted by Figure 1, we also find strong empirical support for the predicted effects on changes in the order flow decomposition.

Our model moreover allows us to derive conclusions on the driving forces affecting the propensity for hidden liquidity supply. For instance, we theoretically and empirically show that wider spreads increase the provision of hidden liquidity. Markets with wider spreads are thus subject to higher return volatility. Our model therefore complements existing theory based on asymmetric information by establishing a non-informational causal link between the bid-ask spread and the tick size as drivers of liquidity mis-coordination to asset return volatility. While in classical information-based models, as, e.g., Glosten and Milgrom (1985) or Kyle (1985), the bid-ask spread arises as a compensation for market maker's adverse selection costs and, thus, is triggered by volatility, our setting explains why causality might run in the opposite direction as well. We argue that wider spreads force traders to limit their display size due to increased liquidity competition. Excess volatility then arises from an extensive use of hidden orders, creating matching frictions and ultimately non-expected liquidity demand. In fact, we empirically show that the cross-sectional variation in volatility is well explained by the variation in hidden liquidity and in bid-ask spreads.

Our paper contributes to the extant theoretical literature on limit order book markets, focusing on order bidding strategies and their determinants in limit order book markets.[4] Only few models, however, address the impact of hidden orders on trading. Boulatov and George (2013) and Moinas (2010) study the impact of hidden liquidity when some investors have private information. They analyze the role of informed and non-informed trading and the informational efficiency of prices. Generally, they find that hidden liquidity improves market quality as it generates a deeper book and more intense competition in liquidity supply. Buti and Rindi (2013) study the use of hidden orders in a limit order book market, where traders make endogenous trade decisions. Similar to our setting, there is no information asymmetry in their model. They show that hidden orders help traders to compete for liquidity provision and to reduce order display costs.

A major distinction to these approaches, however, is that our model does not only focus on liquidity competition but captures an important additional source of liquidity externality: liquidity demanders strategically choose between different venues. A major effect is that competition for liquidity demand incentivizes liquidity suppliers to expose trading intentions even when they face liquidity competition. In this sense, our paper contributes to the recent line of research that simultaneously analyzes the role of liquidity competition and liquidity externalities (see, e.g., Hendershott and Mendelson (2002) and Foucault et al. (2005)). Our findings complement this literature and demonstrate that fundamental empirical microstructure relationships in limit order book markets (as analyzed, e.g., in Biais (1993), Ranaldo (2004), and Hall and Hautsch (2006)) can be explained by the interplay between liquidity competition and the disclosure of trading intentions.

A further major difference to the existing literature is that our model focuses on frictions that purely arise from trade execution without endogenizing trading motifs. This is in line with the view of Foucault et al. (2005), who stress that a considerable amount of trading frictions in

[4] See Glosten (1994), Chakravarty and Holden (1995), Rock (1996), Seppi (1997), Parlour (1998) Biais et al. (2000), Parlour and Seppi (2003), Foucault (1999), Foucault et al. (2005), Goettler et al. (2005) and Rosu (2009).

practice is not caused by information asymmetry but by the process of optimal trade execution. In practice, the decision to trade is typically separated from the process of trade *execution* as most market participants delegate this task to specialized brokerage firms. At this stage, however, trade execution follows generic principles of transaction cost minimization irrespective of the investor's underlying trading motif. One of our aims is, therefore, to gain a deeper understanding to which extent fundamental microstructure relationships originate by frictions arising from the trading process alone. Therefore, we do not consider a general equilibrium model, where trading motifs are endogenized. In our setting, traders trade due to exogenous reasons. Strategic order choice decisions are entirely due to the objective of (expected) transaction cost minimization.

Our results are important for both market regulators and exchange operators. Public exchanges compete for order flow in an increasingly fragmented market. If they loose too much order flow to competitors, the public price formation process may be harmed. Extant literature suggests that these liquidity externalities are closely related to market transparency. In this work, we show that transparency on primary exchanges can enhance market quality in terms of lower transaction costs, mitigate fragmentation and attract large order flow from latent investors. To increase market share and improve price formation on public exchanges, our analysis suggests that market operators should broaden their network with other liquidity pools, enhance their order routing infrastructure and provide large institutional investors direct market access and real-time monitoring capabilities. Then, liquidity opportunities can be seized instantly as they arise.

The remainder of this paper is structured as follows. Section 2 presents the data, the underlying econometric framework, and empirical evidence on the main finding of this paper: effects of hidden order imbalances on the subsequent price process. In Section 3, we present the theoretical model and derive testable predictions. These predictions are empirically validated in Section 4. Section 5 concludes.

2 The price impact of hidden orders – empirical evidence

In this section, we substantiate the empirical evidence shown in Figure 1 which constitutes the main empirical result of this paper. In this context, we present the data and the underlying econometric framework. This section therefore builds the empirical basis for the theoretical model presented in Section 3.

2.1 Data

Our empirical analysis uses a combination of two data sets based on NASDAQ trading. Information on consolidated hidden and displayed depth for each price level on a minute-by-minute basis for all NASDAQ traded stocks originates from the NASDAQ ModelView data set. The initial sample covers the constituents of the S&P500 universe through the period from November to December 2008. To reduce the impact of very illiquid stocks, we restrict the analysis to stocks that have a quoted spread below 25 cents on average. This leaves us with a sample of $N = 468$ stocks.

Table 1 reports averages across stock groups and time for mid-quote levels, spreads, visible and hidden depth, and limit order activities. We group the stocks into quintiles based on the average daily trading volumes (ADV). Distinct variations in trading activities are reflected by inter-trade durations ranging from 2.65 seconds for the least actively traded stocks to 0.35 seconds for stocks in the largest liquidity quintile. Similar monotonic relationships across the liquidity quintiles are found for trade sizes (increasing in ADV), bid-ask spreads (decreasing in ADV), price levels (decreasing in ADV), first-level order book depth (increasing in ADV), and daily volatility, measured based on the daily high-low range relative to the daily average mid-quote (increasing in ADV). Hence, the highest trading activity (in terms of both the number of transactions and the size of shares) is observed for stocks with small spreads, low price levels, and high depth.

We observe that the proportion of hidden shares in total posted shares is declining for less liquid stocks, amounting to approximately 17% on average. The relative amount of shares executed against standing hidden orders, however, is decreasing with overall daily trading volume. While on average 26% of trading volume is executed against hidden orders in the smallest liquidity quintile, this number declines to on average 7% for the most actively traded stocks. Hence, hidden liquidity is more prevalent for less liquid stocks with wider spreads and lower displayed depth. As shown in Section 3, this relationship is strongly supported by economic reasoning.

To utilize information on order flow *between* the minute-by-minute snapshots, we augment the NASDAQ ModelView data by TotalView message-level data, which is processed via the data service LOBSTER[5]. The data contains information on any (visible) order activity and the corresponding fully reconstructed (displayed) NASDAQ limit order book at each instant. We aggregate order executions, cancelations, and submissions stemming from NASDAQ TotalView for each minute and merge this information with the minute-by-minute snapshots on hidden depth from NASDAQ ModelView. The merged data set then consists of 390 daily minute-by-minute information observed over 40 trading days, resulting into $15,600$ observations per stock.

To limit the computational burden, we conduct this analysis for 10 randomly selected stocks reflecting an arbitrary cross-section of differently liquid S&P500 constituents. The tickers are APC, AZO, CAH, GAS, GOOG, EMR, LEG, PAYX, STJ, and TDC. Table 3 in the Appendix reports time-series averages of mid-quotes, bid-ask spreads, visible and hidden depth as well as order activities based on one-minute aggregates for these 10 stocks. The statistics indicate that order submission behavior and market dynamics are obviously strongly driven by liquidity competition and order activity. In line with Hautsch and Huang (2012), we observe that most order activity originates from order submission and order cancellation activity: on average, approx-

[5]See http://lobsterdata.com.

Table 1: Averages across stocks and time for daily trading volume (ADV), inter-trade durations (DUR), daily high-low ranges standardized by corresponding daily average mid-quotes (HL), and trade sizes (TS) as well as averages of minute-by-minute snapshots of bid-ask spreads (SPR), mid-quotes (MQ), visible depth on top (first level) of the book ($D1$), and total hidden depth on the first 10 levels ($HD10$). Moreover, we report the average ratios of hidden to total depth on the first 10 levels (evaluated based on minute-by-minute snapshots) ($RHD10$), the average number of shares traded against hidden volume (THD), and the corresponding ratio of executed hidden shares to average trading volumes ($RTHD := THD/ADV$). The amount of traded hidden volume, THD, is computed as the average daily trade volume executed on the best quotes. The averages are computed within liquidity quintiles based on ADV. The stock universe consists of all S&P500 constituents that are traded on NASDAQ, excluding stocks with an average spread below 25 cents. The sample ultimately includes 468 stocks for the period between November and December 2008.

Liquidity Quintile	Observable Stock Properties							Hidden Liquidity posted		Hidden Liquidity traded	
	ADV ($10^6 sh.$)	DUR ($sec.$)	HL ($ratio$)	TS ($sh.$)	SPR ($ticks$)	MQ (\$)	$D1$ ($sh.$)	$HD10$ ($sh.$)	$RHD10$ ($ratio$)	THD ($10^6 sh.$)	$RTHD$ ($ratio$)
q_{20}	1.39	2.65	0.07	147	4.91	36.46	308	656	0.19	0.37	0.26
q_{40}	2.72	1.38	0.08	158	3.39	32.84	576	1318	0.20	0.57	0.20
q_{60}	4.23	0.94	0.09	165	2.40	27.41	800	1671	0.17	0.69	0.15
q_{80}	7.13	0.61	0.10	178	1.87	24.59	1278	2292	0.16	0.83	0.11
q_{100}	16.98	0.35	0.11	219	1.38	23.32	3490	6202	0.13	1.10	0.07
Total	6.57	1.19	0.09	174	2.79	28.91	1305	2440	0.17	0.71	0.16

imately 47% of the order flow volume is caused by order submissions, 49% by cancellations, and only approximately 4% by trades.

Figure 3 in the Appendix provides evidence on the autocorrelation properties of fundamental order book characteristics. It shows across-stock averages of autocorrelations of one-minute returns, 10-min volatilities, and one-minute snapshots of depth and displayed depth imbalances, defined as standing buy volume in excess of sell volume. Moreover, we report one-minute aggregates of limit order submissions (SUB and SUS), cancellations (CAB and CAS), and executions (EXB and EXS). Nearly all variables are strongly autocorrelated and reveal a pretty strong persistence (i.e., slowly decaying autocorrelation functions). Interestingly, the liquidity supply (reflected by standing limit orders) is more persistent than the liquidity demand (reflected by order executions). In line with the finding that liquidity competition is a substantial driver of market dynamics, this finding suggests that traders actively micro-manage, modify, and cancel limit orders when they feel that orders become mis-priced or have a low chance of execution. The presence of strong serial dependence in execution volumes (i.e., market orders) is in line with the fact that traders do not execute their position by means of a single market order, but tend to slice larger orders into smaller orders and feed them sequentially into the market. This is in line with the literature on optimal liquidation (e.g., Obizhaeva and Wang (2013)).

2.2 Estimating the price impact of hidden orders

Variations in liquidity supply are expected to affect future price directions only if they occur on *one* side of the market. We therefore consider *buy-sell imbalances*, defined as the difference between standing volume on the buy side and the sell side. According to Figure 3 (Appendix), depth imbalances are significantly autocorrelated. We moreover find that *hidden* order imbalances are more persistent than displayed imbalances. This result yields some evidence that displayed imbalances tend to be absorbed by counterparties' market order flow. Conversely, hidden order imbalances "survive" longer, as their presence cannot easily be detected by counterparties.

Modelling depth imbalances therefore requires an autoregressive model. This is also true for midquote returns, which are clearly less persistent, but still reveal significant autocorrelations on one-minute frequencies. We moreover find significant cross-autocorrelations between depth imbalances, returns, volatility, liquidity demand, and liquidity supply.[6] To properly quantify the effect of imbalances in (hidden and displayed) market depth on subsequent returns, it is therefore required to utilize an econometric framework which explicitly accounts for these dynamic interdependencies. The equi-distant nature of the underlying ModelView data, allows us to employ classical techniques of multivariate time series analysis. A natural and computationally tractable framework is the class of vector autoregressive (VAR) models, as applied, e.g., by Hasbrouck (1991) and Hautsch and Huang (2012), among others.

To capture all major variables which are dynamically interrelated with midquote returns and liquidity characteristics, we specify a VAR process of the following variables: one-minute mid-quote returns (RET), minute-by-minute snapshots of bid-ask spreads (SPR), minute-by-minute rolling window estimates of 10-min realized volatilities (corresponding to the sum of squared one-minute mid-quote returns; RV), minute-by-minute snapshots of hidden and displayed order imbalances ($HI10$ and $DI10$, respectively) and total depth (sum of hidden and displayed depth on the first 10 levels; $TD10$), and per-minute numbers of submitted, executed, and canceled buy and sell limit orders (SUB, SUS, EXB, EXS, CAB, and CAS).[7]

The state of the order book at t is thus represented by the 12-dimensional vector y_t, consisting of the variables RET, SPR, $VOLA$, $HI10$, $DI10$, TD, SUB, SUS, EXB, EXS, CAB, and CAS. We then model y_t in terms of a VAR(p) process of the form

$$y_t = \sum_{j=1}^{p} A_j y_{t-j} + u_t, \tag{2.1}$$

[6]For sake of brevity, these statistics are not reported here.

[7]We log-transform total depth $TD10$, realized volatility RV, and spread SPR in order to reduce the impact of outliers and to make a normal distribution more appropriate.

with A_j denoting (12×12) coefficient matrices for $j = 1, ..., p$ and u_t denoting a vector of zero mean white noise error terms with $E[u_t u_t'] = \Sigma_u$.

Given the paper's objective, we are interested in the effect of depth imbalances ($HI10$ and $DI10$) at minute t on the midquote return process (RET) over the next trading minutes $t + h$, $h > 1$. Changes in depth imbalances, however, might affect future returns not only directly but also indirectly through other variables in the system. To measure the long-run effect of depth imbalances on returns, while accounting for possible dynamic feedback between all variables, it is necessary to analyze how a change in depth imbalances is dynamically traced through the system. Such analysis is performed by means of an *impulse response* analysis. In a VAR system, the impulse response function is derived based on the underlying moving average representation,

$$y_t = \Phi_0 u_t + \Phi_1 u_{t-1} + \Phi_2 u_{t-2} + \Phi_3 u_{t-3} + \dots, \tag{2.2}$$

with $\Phi_0 = I_K$ and $\Phi_s = \sum_{j=1}^{p} \Phi_{s-j} A_j$ for $s > 0$. We compute *generalized* impulse response functions according to Pesaran and Shin (1998), defined as the difference between the (conditionally) expected value of y_{t+n} (i.e., RET) given a shock in the innovation of variable j ($HI10$ or $DI10$) and the corresponding (conditional) expectation if this shock would not occur. Define Ω_t as the information set up to time t, then, the generalized impulse response is given by

$$\Theta_j(n) := E[y_{t+n} \mid u_{jt} = \delta_j, \Omega_{t-1}] - E[y_{t+n}, \Omega_{t-1}], \tag{2.3}$$

where δ_j is the size of the shock. Assuming multivariate normality for u_t, the conditional expectation given a shock $\delta_j := \sqrt{\sigma_{jj}}$ in one variable yields $E[u_t \mid u_{jt} = \delta_j] = \Sigma_u e_j \sigma_{jj}^{-1} \delta_j$, with e_j denoting the unit vector. By setting the shock to one standard deviation, i.e., $\delta_j = \sqrt{\sigma_{jj}}$, the generalized impulse on all variables in $t + n$, induced by a shock in variable j, is given by

$$\Theta_j(n) = \frac{\Phi_n \Sigma_u e_j}{\sqrt{\sigma_{jj}}}, \quad j = 1, \dots, K. \tag{2.4}$$

By summing over all periods $k = 1, \dots, n$, we obtain the *cumulative* (generalized) impulse response given by

$$\Xi_j(n) := \sum_{k=1}^{n} \Theta_j(k) = \sum_{k=1}^{n} \Phi_k \frac{\Sigma_u e_j}{\sqrt{\sigma_{jj}}}, \quad j = 1, \dots, K, \tag{2.5}$$

which is consistently estimated by

$$\hat{\Xi}_j(n) = \sum_{k=1}^{n} \hat{\Phi}_k \frac{\hat{\Sigma}_u e_j}{\sqrt{\hat{\sigma}_{jj}}}. \tag{2.6}$$

The main advantage of this approach is that generalized impulse response functions are invariant to the re-ordering of the endogenous variables. As shown by Pesaran and Shin (1998), orthogonalized impulse responses coincide with orthogonalized impulse responses (based on a Cholesky decomposition of Σ_u) if the respective variable is the first one in the ordering. Pesaran and Shin (1998) derive the asymptotic properties of the impulse response functions based on a co-integrated VAR model. We adapt these derivations and provide the asymptotic distributions of the generalized impulse response functions in the Appendix.

The high persistence of the underlying order book process requires using a VAR process with high lag order. Information criteria and residual diagnostic suggest a lag order of 30. We check the robustness of the resulting impulse response functions with respect to the choice of the lag order and estimate alternative specifications that are parameterized more parsimoniously, particularly VAR(5) and VAR(15) processes. In line with the results of Jorda (2005), showing that impulse-response estimates are relatively stable regarding the choice of the underlying lag order (given that a dominant part of the serial dependence is sufficiently captured), we find that our results are not qualitatively affected and are remarkably quantitatively stable with respect to the model choice.

We refrain from reporting individual VAR estimates, which are hardly interpretable for such a highly parameterized process and rather focus on the resulting impulse response function. The plots shown in Figure 1 represent the cumulative impulse response of mid-quote returns triggered by a positive one-standard-error shock in hidden and displayed net buy order imbalances ($HI10$ and $DI10$). The reported impulse response functions are cross-sectional averages (across the analyzed 10 stocks). Since the effects are symmetric in the sign of the shock, we restrict our analysis to positive shocks only and refrain from showing the opposite case. Given a fixed time interval $t = n$ after the shock and variable j, the corresponding asymptotic variance of the *averaged* impulse response function is approximated by $M^{-2} \sum_{l=1}^{M} \Lambda_{jn}^{l}$, with Λ_{jn}^{l} denoting the asymptotic covariance of the generalized impulse response (see Appendix).[8]

We thus find strong evidence for significantly positive price reactions caused by submissions of hidden orders on the buy side. This effect is strikingly different to the reaction of *displayed* orders causing slightly negative price reactions. We therefore observe that one-sided liquidity supply shielded from the market yields a fundamentally different price impact than visible liquidity supply. Two competing explanations for the price impact of hidden liquidity arise: First, the presence of hidden liquidity on one side of the market indicates the presence of information which then materializes in the subsequent price process. Second, these price reactions simply originate from trading frictions. In the remainder of the paper we provide theoretical reasoning and additional empirical support for the latter hypothesis.

[8]This approximation obviously ignores potential cross-equation correlations between the estimated asset-specific impulse response functions of stock l. Given the high parameterization, the latter, however, is not straightforwardly taken into account. We therefore use this approximation as a convenient but still sufficiently precise way to assess and compactly illustrate the overall significance of our estimates. The latter – and consequently – our conclusions regarding the empirical validity of our hypotheses are not affected by this approximation and is confirmed by individual (asset-specific) estimates, which are not reported here.

3 Hidden liquidity in a sequential trade game

3.1 Structure of the game

In this game we model the fundamental trade-off in the decision to hide when liquidity suppliers face both, benefits and punishment for displaying their trade intentions. For instance, order display and thus the revelation of trading intentions fosters the attraction of possible trading counterparties which otherwise would not partake in trading. As a result, order-display can reduce liquidity supplier's trading costs by attracting additional liquidity demand. On the other hand, displayed orders are more likely to get overbid and, therefore, can lead to higher trading costs by increasing liquidity competition on the same side of the market (see Harris (1997)). Hence, it is conceivable that order-display affects the supply and demand side of liquidity differently and that the equilibrium display size marks a trade-off.

Therefore, to properly capture this fundamental trade-off, we propose a minimalistic game-theoretical framework that features both channels of interactions. This requires the presence of three strategically interacting traders. The so-called *hidden trader* (H) aims at executing a (buy) order with minimal costs within a given time and has the discretion over the extent of order display. He faces liquidity competition from the so-called *liquidity competitor* (C) who trades on the same side of the market. While trader C competes with trader H in order-execution, the third trader, the *latent trader* L, acts as a potential counterparty. The latent trader has the option to trade in the public limit order book against the standing orders of H and C or entirely skip the exchange order book market and conduct trades in an off-exchange trading platform. Hence, we think of L representing a large institutional investor, who actively monitors the public primary market for trading opportunities, but otherwise trades on an off-exchange trading venue (therefore we call him *latent*), such as, e.g., a dark pool, broker-dealer network or a classical upstairs market. While these off-exchange markets generally provide more liquidity, they also charge additional commission fees for the counterparty search service involved. Therefore, institutional traders typically balance their positions between different types of platforms depending on their liquidity needs and implied (expected) transaction costs.

In such a setup, liquidity competition between H and C is driven by the costs of non-execution (or delayed execution) and the costs of overbidding (i.e., paying a worse price). Costs of non-execution particularly arise whenever traders face liquidation time constraints, i.e., are required to liquidate their positions over a given time horizon. For nowadays trading, this is a rather realistic assumption and requires traders to enforce order execution if time elapses. In such a situation, non-executed (buy) limit orders need to get canceled and re-submitted as market (buy) orders at the end of the trading period. The higher expected transaction costs due to market order executions correspond to the (opportunity) costs of non-execution. In case of H, these costs particularly arise if he hides his order and thus reduces the execution probability.

The timing of events is as follows. The hidden trader H arrives at the initial time point t_0, submits his order and decides about order display. Then, at t_1 trader C arrives and submits his order in reaction to trader $H's$ display decision at t_0. His expected action therefore is taken into account by H. After the liquidity suppliers have settled their quotes at t_1, liquidity demanders arrive. First, at t_2, the latent trader L arrives with a given probability. Depending on the (visible) liquidity in the limit order book market, he decides whether to enter the market and to trade against the standing liquidity supply. At t_3, a noise liquidity demander arises and submits random sizes of market orders which are matched by the standing (buy) limit orders posted by H and C. Finally, liquidity supply by H and C which remains non-executed, needs to get executed via market orders at time point t_4.

The timing and structure of the game yields a generic setting to study the trade-off between order liquidity competition and counterparty attraction under the assumption of liquidation time constraints. There are some simplified assumptions. We argue that these simplifications are justified because our focus is to address the partial aspect of the origination and effects of hidden liquidity. Therefore, we may abstract from a range of more general aspects of the trading process. In particular, we abstract from general endogenous order setting strategies. Instead, we assume that the three strategic players H, C and L react to each other in the above sequential structure, which, however, makes the trade-off and the game between all traders non-trivial.

Specific details about the order placement choices and traders' optimal strategies are given in the next section.

3.2 Order placement choices

Let B_{t_i} denote the best bid price at time t_i. At time t_0, the hidden trader H submits a so-called Iceberg order at the best bid price B_{t_0}. Without loss of generality we assume that H establishes a new best bid price level.[9] We denote the size of the Iceberg order by N_H. In contrast to standard limit orders, Iceberg orders allow traders to specify how much of the posted order volume should be visible to the public. We refer this to as the 'display size'. Accordingly, H can choose to openly display any number from $[0, N_H]$ in the order book; the remaining shares are hidden and shielded from public view. They can be matched with incoming orders but have lower execution probability than any displayed orders. We denote a generic action of H by

$$a_H \in [0, N_H].$$

H knows that he faces competition from the liquidity competitor C, who will arrive at time t_1 and wants to buy N_C shares. The assumption that C arrives with certainty, however, is made for convenience only. Our results easily carry over to a more general setting where H has a belief about the distribution of N_C, thus reflecting the degree of liquidity competition.

At time t_1, C acts based on the *visible* order book as determined by the hidden trader's display size a_H set in t_0. If C posts his limit order on the same price level, it is queued behind H's order, looses priority and thus faces a higher execution risk. This effect increases in the display size a_H. Therefore, to increase execution priority, C may overbid H's submission price

[9]Our implications can be, however, straightforwardly generalized to a setting where H posts at an existing order book level.

level. Consequently, he faces a trade-off between overbidding H's order by a tick Δ (and therefore increasing his execution costs) and submitting at the same submission price level (and thus being subject to higher execution risk). Formally, we denote C's order placement choices by

$$a_C \in \{0, 1\},$$

where $a_C = 0$ is associated with a submission at B_{t_0} and $a_C = 1$ is associated with a submission at $B_{t_0} + \Delta$. If the competitor overbids the hidden trader's best bid price B_{t_0}, the new best bid at t_1 is $B_{t_1} = B_{t_0} + \Delta$ and $B_{t_1} = B_{t_0}$ otherwise.

We assume that C fully displays his trade intentions and his only discretion lies in the choice of the limit order submission price level. This does not pose a restriction of our model. As shown below, overbidding is the main threat and the reason for other liquidity suppliers to hide their trade intentions. Hence, liquidity competitors who do not face competition by others (i.e., arriving last in our setting) always fully display their trade intention as they can not get overbid. In our setting, this assumption is generic and does not restrict the generality of the setup.

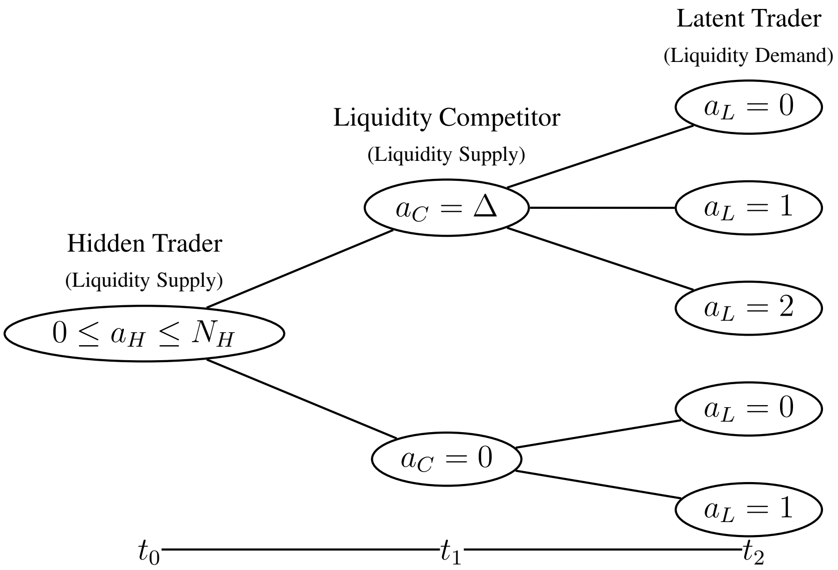
Liquidity (buy) suppliers set their quotes in anticipation of future incoming (sell) liquidity demand. We assume that liquidity demand stems from a strategic latent trader L and an exogenous noise trader. L arrives at t_2 with probability μ and has liquidity demand (i.e., sell volume) N_L. The (sell) noise trader arrives (with certainty) at t_3. His trade size is random and exogenous, i.e., independent of the previous traders' actions.

Both liquidity suppliers H and C share a consistent belief μ about the presence of the latent trader L. The latter is strategic in his decision about where to trade. He can either trade in the public limit order book ("downstairs market") using market orders, on an off-exchange trading platform such as a broker-dealer network, or through private brokers ("upstairs market"). The noise trader only trades at the public limit order book market.

Thus, L's decision about where to trade constitutes a central element of liquidity externality in our model. We assume that L is an institutional trader facing large trading (sell) demand. If he decides to trade in the upstairs market, he benefits from guaranteed order execution as such upstairs markets provide sufficient liquidity even for large block trades. As such brokerage mechanisms, however, require non-trivial commission fees, he may consider entering the downstairs (limit order book) market as an alternative. In this market, he avoids paying commission fees but faces (implied) transaction costs if there is not sufficient liquidity pending on the highest order book price level. In this case, part of his order need to be matched with liquidity standing on lower price levels (increasing transaction costs) or he may decide to execute the remaining part on the upstairs market. As discussed below in more detail, the latter alternative becomes more beneficial the higher the liquidity supply in the downstairs market and thus the lower the non-executed volume. Hence, a central element of our model is that the attraction of outside (external) liquidity demand becomes more likely if the (visible) liquidity supply in the limit order book market increases.

Without loss of generality and to keep the model tractable, we assume that L only trades in

Figure 2: *Structure of the game. At t_0, the hidden trader arrives and submits an Iceberg order with size N_H and display size a_H with $0 \leq a_H \leq N_H$. Other traders only observe a_H shares submitted. At t_1, the liquidity competitor arrives and decides whether to maintain ($a_C = 0$) or to overbid ($a_C = \Delta$) the hidden trader's price. At t_2, the latent trader arrives and trades a_L ticks deeply into the book.*



public limit order book markets via (sell) *market* orders. We moreover assume that L controls his (expected) transaction costs and thus has discretion about how deeply he wants to trade into the book (if there is insufficient liquidity supply on the best price level). A generic action of L is therefore denoted by

$$a_L \in \{0, 1, 2, ...\}.$$

Here, $a_L = 0$ means that L only trades in the upstairs market, while $a_L = i$. $i = 1, 2, ...$, implies that he trades exactly i ticks into the limit order book beyond the prevailing best bid price B_{t_1}. Trading a_L ticks into the book, implies that the new best bid price at t_2 is

$$B_{t_2} = B_{t_1} - a_L, \tag{3.1}$$

with $B_{t_1} = B_{t_0} + \mathbb{1}_{\{a_C=1\}}$, where $\mathbb{1}$ denotes the indicator function. An illustration of the trading game of the strategic players H, C and L is shown in Figure 2.

An important feature of the trading game is that the price setting in the upstairs market is coupled with the quote setting in the downstairs market. In line with practice, we assume the price in the upstairs market to follow a linear pricing rule, which takes the prevailing best bid price as a reference price and charges an additional brokerage commission fee $\gamma > 0$,

$$B_{t_2} - \gamma, \tag{3.2}$$

where B_{t_2} denotes the best bid price at time t_2, after L's arrival. This pricing rule influences the L's decision whether or not to enter the downstairs market. In particular, if L decides to trade a_L ticks into the book (see (3.1)), he implicitly makes trading on the upstairs market more expensive. Hence, the price impact of his own order on the downstairs market will increase the execution costs of any non-executed part on the upstairs market. As discussed in below in more detail, this effect is a central aspect in L's trading strategy.

Finally, the noise liquidity demander (seller) enters the game at t_3, enabling H and C to execute yet non-executed limit order shares. We assume that the noise trading demand arises from exogenous reasons, independent of the limit order book's liquidity provision, and is stochastic. Hence, the noise market sell order size is distributed according to some probability density distribution f.

An important assumption underlying this setting is that all strategic traders face commitments to liquidate their positions within a given time. This forces them to increase order aggressiveness in case of non-execution:

Assumption 1 (Pre-commitment to trade)**.** We assume that all strategic traders are pre-committed to trade. Accordingly,

i) trader H and C *cancel* outstanding non-executed shares and *re-submit* them as a corresponding buy *market order* at the end of the trading period, t_4,

ii) trader L trades non-executed shares via market orders in the upstairs market after t_2.

Trading pre-commitment is an important element in many trading strategies nowadays.[10] Many algorithmic trade execution services implement liquidation constraints, reflecting trader's trading horizons, irrespective of trader's underlying motifs. For instance, many so-called 'child orders' that are executed on a trajectory of a larger 'meta order' are often subject to time liquidation constraints ensuring that the underlying trading strategy meets certain execution requirements. In these strategies, market orders often serve as an action of last resort, when execution with limit orders does not provide the desired outcome.

Throughout the paper, we assume the common priority rules for matching market orders against standing limit orders: orders submitted at more competitive price levels have priority over orders submitted at less competitive price levels. Openly displayed orders have priority of hidden orders at the same price level. Finally, orders with the same display status submitted at the same price level are executed on a first-come-first-serve basis. The cost functions as well as the timing and structure of the game are assumed to be common knowledge among all strategic players.

3.3 Trader payoffs and strategies

In our model, traders coordinate their trade decisions in order to minimize expected trading costs. Since this is a major (albeit not the only) driving force of nowadays trading activity (particularly on high frequencies), our aim is to exclusively focus on frictions arising from the

[10] Foucault et al. (2005) shows that trader impatience plays an important role in the origination of non-informational sources of frictions.

mechanism of trading and the underlying market structure. We denote these frictions as 'non-informational' frictions. We are therefore not concerned with aspects of price discovery or frictions arising from information asymmetry. We neither assume that some traders have prior knowledge about the fundamental value of the asset. Our objective is rather to gain insights into the effect of these non-informational frictions on traders' strategic decisions and the resulting order book dynamics.

Likewise, we assume that traders are risk neutral and that the displayed shares in the public limit order book are the only common information source. Moreover, we assume that traders are aware about other traders' intentions and payoffs. Beliefs about possible hidden liquidity is not formed by any participant. We argue that this is a costly endeavor for most market participants as real-time information about hidden liquidity is not available and historical data on hidden liquidity is not extensively disseminated and expensive. It is thus reasonable to assume that trader decisions are mostly affected by the *visible* portion of the order book.

3.3.1 The latent trader's strategy

In principle, the latent trader L can decide to trade arbitrarily deeply into the limit order book. We restrict the analysis, however, to the most relevant cases $a_L \in \{0, 1, 2\}$. Considering the case $a_L > 2$ would not gain any additional insights but would just require to assume the presence of more liquidity suppliers and corresponding competitors. In this case, liquidity competition would spread over more than two price levels of the order book, making the model significantly more complex without changing the underlying rationale. In our model, the case $a_L > 2$ can be ruled out by assuming that the size of the latent trader L exceeds a critical threshold. Then, given his trading demand it will never pay off for him to shift prices in the downstairs market too much and thus to make trading in the upstairs market significantly more expensive.

On the other hand, we rule out the trivial case where L is so large (in terms of trade demand) that he will never benefit from trading in the downstairs market (because the non-executed part of his order will be prohibitively high in any case). In such a case, the game would reduce to a game without the latent trader thus concentrating on liquidity competition. To rule out such a case, we assume that there is also an upper bound on the latent trader's size, i.e., $N_L^- \leq N_L \leq N_L^+$. Given these bounds, it can be shown that L's order placement decisions are restricted to the set

$$a_L^* = \{0, 1, 2\}. \tag{3.3}$$

For a formal proof see Lemma 2 in the Appendix.

We start the equilibrium analysis by first deriving the latent trader's payoff and his best response given the competitor's and hidden trader's actions a_H and a_C. Denote $D := N_C + a_H$ as the displayed liquidity, provided jointly by H and C. To simplify notation, we set the initial price level set by H in t_0 to $B_{t_0} = 0$. L's payoff function $\Pi_L(a_H, a_C, a_L)$ then depends on H's display size a_H, C's decision on potential overbidding, a_C, and L's decision a_L whether to enter the downstairs market and – if yes – how deeply to trade into the book. The corresponding payoff can be presented in terms of the following matrix:

$$\begin{array}{c|cc} & \multicolumn{2}{c}{\Pi_L(a_H, a_C, a_L)} \\ \hline & a_C = 0 & a_C = \Delta \\ \hline a_L = 0 & -N_L\gamma & N_L(\Delta - \gamma) \\ a_L = 1 & -(N_L - D)(\Delta + \gamma) & N_C\Delta - (N_L - N_C)\gamma \\ a_L = 2 & n/a & N_C\Delta - (N_L - D)(\Delta + \gamma) \end{array} \tag{3.4}$$

Hence, if L does not enter the downstairs order book market ($a_L = 0$) and C does not improve the best bid price $a_C = 0$, then L's net transaction costs arise solely from the commission fee. If C, however, improves the best bid price ($a_C = \Delta$), L's net transaction costs reduce by the amount $N_L\Delta$.

Moreover, when L trades one tick into the order book ($a_L = 1$), he trades all visible liquidity at the best bid B_{t_1}. If, in this case, C posts his order at the same price level as H ($a_C = 0$), then L executes all D shares, but needs to trade the remaining $N_L - D$ shares in the upstairs market, confronting him with a commission fee γ. His net marginal trading costs for the remaining $N_L - D$ shares are therefore $\Delta + \gamma$ since his trade in the downstairs market has reduced the public (downstairs) bid price by one tick. Following the same logic, the transaction costs of L in the remaining cases follow similarly.

Since L is a seller and the trading costs enter $\Pi_L()$ negatively, his aim is to *maximize* Π_L. From (3.4), we can derive the best response of L, given H's and C's order placement decisions,

$$\begin{aligned} \underset{a_L \in \{0,1,2\}}{\operatorname{argmax}}\ \Pi_L(a_H, 0, a_L) &= \begin{cases} 0 & \text{if } a_H \leq \phi(a_C, N_L, N_C, \lambda), \\ 1 & \text{otherwise,} \end{cases} \\ \underset{a_L \in \{0,1,2\}}{\operatorname{argmax}}\ \Pi_L(a_H, 1, a_L) &= \begin{cases} \mathbb{1}_{N_C \geq N_L\lambda} & \text{if } a_H \leq \phi(a_C, N_L, N_C, \lambda), \\ 2 & \text{otherwise,} \end{cases} \end{aligned} \tag{3.5}$$

where $\phi(a_C, N_L, N_C, \lambda)$ is a critical threshold for the display size a_H, depending on the size of C's order (N_C) and his order aggressiveness a_C, the latent liquidity demand N_L and the parameter λ. The latter corresponds to the relative costs of price improvement in the order book (by one tick Δ) relative to the per-share commission costs prevailing in the upstairs market. The upper part of (3.5) refers to the case, where C does not overbid, i.e., $a_C = 0$. In this case, L enters the order book market only if H's display size a_H is sufficiently high. In this case, trading in the downstairs market becomes beneficial as L can execute a sufficiently large proportion of his order without commission fees. As long as $a_H > \phi()$, this cost benefit overcompensates the increasing trading costs of the non-executed part due to price impairment in the upstairs market. The critical threshold $\phi()$ gives the display size a_H, making the marginal costs of trading in the downstairs market equal to the costs of trading in the upstairs market.

A similar logic applies to the second case in (3.5), refering to the case of price improvement through the liquidity competitor C. Note that in this case the action space of L expands from $\{0, 1\}$ to $\{0, 1, 2\}$. The important difference to the case above is that L may enter the order book market even if the display size a_H is below $\phi()$. This is only possible, however, if the liquidity competitor C offers sufficient liquidity to attract the latent trader. In particular, C's liquidity supply must be sufficiently high to make L benefiting from C's price improvement in the downstairs market (recall that L is a seller) without making subsequent trading of any non-executed volume in the upstairs market (due to L's price impact in the order book market) too expensive. This is guaranteed by the size condition $N_C \geq N_L \lambda$. If, H's display size is too small, however, it is not beneficial for L, to trade against H's liquidity supply. In this case, only C's liquidity supply gets executed as L only trades against the first order book level. Only if C's display size is sufficiently high, it pays off for L to execute two order book levels and thus to execute liquidity of both H and C.

Trading in the downstairs market therefore implies a fundamental trade-off between benefitting from lower transaction costs (i.e., avoiding commission fees) and causing price impact which in turn makes subsequent trading in the upstairs market more expensive. Thus, trading in the downstairs market is only beneficial for L, if the (displayed) liquidity is sufficiently high. The critical threshold $\phi()$ can therefore be interpreted as a premium for liquidity display to attract outside trading demand.

The liquidity premium $\phi()$ can be explicitly computed as

$$\phi = \phi_1 \mathbb{1}_{\{a_C=0\}} + \mathbb{1}_{\{a_C=1\}} \Big(\phi_2 \mathbb{1}_{\{N_C \geq N_L \lambda\}} + \phi_3 \mathbb{1}_{\{N_C < N_L \lambda\}} \Big), \tag{3.6}$$

with ϕ_1, ϕ_2 and ϕ_3 given by

$$\phi_1 := N_L \lambda - N_C, \qquad \phi_2 := (N_L - N_C)\lambda, \qquad \phi_3 := 2N_L \lambda - N_C(\lambda + 1). \tag{3.7}$$

It therefore increases with liquidity demand N_L and decrease with the competitor's volume N_C. Hence, competition in liquidity supply reduces the display premium for hidden order traders.

The analysis is simplified if we restrict ourselves to the case of a *large* latent trader, i.e., assuming $N_L \lambda \geq N_C$.In this case, it can be shown that $\phi_3 - \phi_2 = \phi_1 > 0$, and

$$\phi_1 < \phi_2 < \phi_3, \tag{3.8}$$

in which case L's best strategy (3.5) simplifies to

$$\begin{aligned} \underset{a_L \in \{0,1,2\}}{\operatorname{argmax}} \Pi_L(a_H, 0, a_L) &= \begin{cases} 0 & \text{if } a_H \leq \phi_1, \\ 1 & \text{otherwise,} \end{cases} \\ \underset{a_L \in \{0,1,2\}}{\operatorname{argmax}} \Pi_L(a_H, 1, a_L) &= \begin{cases} 0 & \text{if } a_H \leq \phi_3, \\ 2 & \text{otherwise.} \end{cases} \end{aligned} \tag{3.9}$$

3.3.2 Trading costs of liquidity suppliers

In our model, trading costs of liquidity suppliers arise whenever they are not able to execute their trading volume via the posted limit order. In this case, they are forced to cancel the non-executed limit orders and to re-submit them as market orders at the end of the trading period

(t_4). In this case, the trading costs depend on prevailing liquidity on the opposite side of the market. If market depth on top of the book is sufficiently high, the market order can be fully executed on the first level and the trader only pays the bid-ask spread. Conversely, if the market is illiquid, the order needs to get executed against liquidity on higher price levels, incrementally increasing marginal execution costs.

The latter feature is captured by a so-called impact function $c(m)$, representing the trading costs of a market order of size m which is executed against a standing limit order book.[11] We assume that $c(m)$ and thus market (il-)liquidity on the sell-side book is exogenous for the trading decisions of all strategic players. This simplification is made to abstract from interactions between liquidity supply on both sides of the book. The latter would make the model significantly more complex without gaining additional insights in the given context. For simplicity and without loss of generality we assume that $c(m)$ depends on an individual trader's order.

For ease of illustration, we normalize the costs of both liquidity suppliers H and C with respect to the initial best bid price $B_{t_0} = 0$. Accordingly, executions at B_{t_0} do not cause any costs. To provide a concrete example, denote the expected spread at t_4 by $s > 0$ and assume that the sell-side limit order book is block-shaped with the (deterministic) inverse depth given by $\beta > 0$. Then, trading costs $c(m)$ are quadratic in the trade size m and are given by

$$c(m) = s \cdot m + \frac{\beta}{2} m^2. \tag{3.10}$$

To quantify the expected execution costs of a generic (potentially partly hidden) limit order of size n, we assume that a shares are displayed and $n - a$ shares are hidden. We moreover assume that there exist d additional displayed shares, which have been submitted by a competitor on the same price level but are queued behind. Then, the a displayed shares have highest priority, the d displayed shares have second highest priority and finally the Iceberg's $n - a$ hidden shares have least priority. In our setting, this may represent the current state of a limit order (initially submitted by H or C) *after* the decision of the latent trader L who may potentially has executed parts of the order. In such a situation, standing liquidity supply can be matched only with random trading demand from noise traders. Any remaining portions need to be executed via market orders under the cost function m.

The expected costs to fully execute the limit orders in dependence on n, a, and d are thus given by

$$Q(a, d, n-a) = \sum_{j=0}^{a} c(n-j)p_j + \sum_{j=a+1}^{a+d} c(n-a)p_j + \sum_{j=a+d+1}^{n+d} c(n+d-j)p_j, \tag{3.11}$$

where $p_j := \mathbb{P}[x = j]$ defines the probability for the arrival of sell market orders from noise traders of the size j.

[11]To exclude mathematically pathological cases, we impose some (weak) restriction on the shape of $c(m)$, guaranteeing that the proportional cost growth due to increases in m converges to zero if m gets very large. For the exact condition, which is valid for typical order book shapes, see (5.1) in the Appendix.

The first term is associated with the costs arising from trading the displayed shares $n - a$ as market orders. The amount of shares is reduced whenever noise sell orders of size $j \leq a$ arrive. If the noise trading demand exceeds a, but is smaller than $a + d$, all displayed orders a and part of C's displayed d orders, but none of the $n - a$ hidden shares are automatically executed. Correspondingly, exactly $n - a$ shares need to be executed subsequently via market orders inducing costs of the magnitude $c(n - a)$. The last term captures the trading costs whenever the noise market orders exceed the size $a + d$. In this case, the hidden $n - h$ shares get partially executed, causing costs $c(n + d - j)$.

In the next section, we will extend the cost function to account for the latent trader's arrival probability and his strategic actions.

3.3.3 The liquidity competitor's best strategy

Denote $p_C(a_H, a_C)$ as the probability that C's order is *not* executed against L's trading demand given H's action a_H and L's best response (3.5). Then, for a given action $a_C \in \{0, \Delta\}$, $p_C(a_H, a_C)$ is

$$p_C(a_H, 0) = \begin{cases} 1 - \mu & \text{if } a_H \geq \phi_1, \\ 1 & \text{else,} \end{cases}$$
$$p_C(a_H, \Delta) = \begin{cases} 1 - \mu & \text{if } a_H \geq \phi_3, \\ 1 & \text{else.} \end{cases} \tag{3.12}$$

Recall that we benchmark all trading costs against the best bid price at the time of order submission. Hence, if C does not improve the price set by H, then C's trading costs are the expected costs of liquidating the volume N_C, which, however, has lower execution priority than the displayed volume of H, a_H. The resulting expected execution costs are equivalent to those of an Iceberg order with hidden size $n - a = N_C$, display size $a = 0$ and $d = a_H$ shares having higher priority than the N_C shares. Accordingly, the expected trading costs for C are given by $p_C(a_H, 0)Q(0, a_h, N_C)$.

Likewise, if he improves the price ($a_C = 1$), he gains priority over all other orders in the queue and the expected liquidation costs are $p_C(a_H, \Delta)Q(0, 0, N_C)$ plus the costs of the price improvement, ΔN_C. In summary, C's cost function is therefore given by

$$\Pi_C(a_H, a_C) = \mathbf{1}_{\{a_C=0\}} p_C(a_H, 0)Q(0, a_H, N_C) + \mathbf{1}_{\{a_C=\Delta\}}\, p_C(a_H, \Delta)Q(0, 0, N_C) + \Delta N_C \tag{3.13}$$

.

Using (3.13), the competitor's best response is given by the following lemma.

Lemma 1. *C's best response is given by*

$$a_C^* = \begin{cases} 0, & \textit{if } \ a_H \leq \psi(a_H), \\ \Delta, & \textit{else,} \end{cases}$$

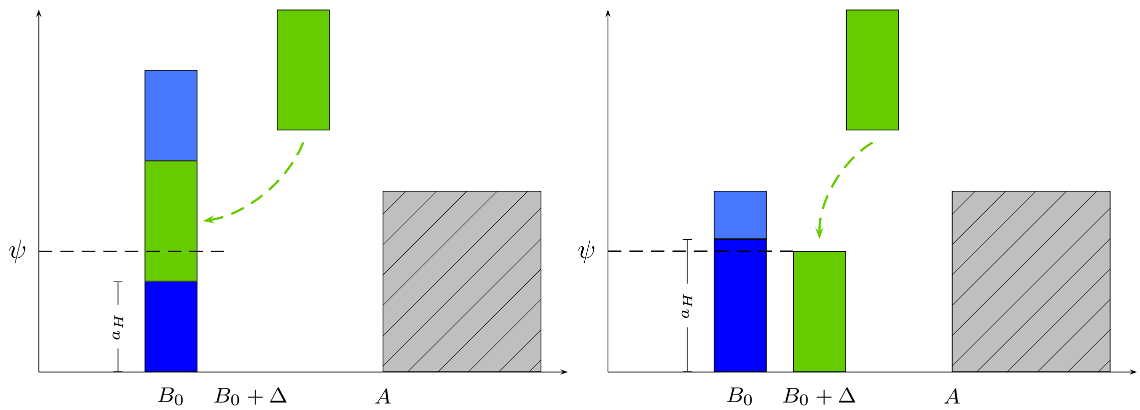
with $\psi(a_H) := \psi_1 \mathbf{1}_{\{a_H < \phi_1\}} + \psi_2 \mathbf{1}_{\{a_H \geq \phi_1\}}$ *and* $0 \leq \psi_1 \leq \phi_1 \leq \psi_2 \leq N_H$.

PROOF: See Appendix. □

Here, $\psi(a_H)$ gives the critical display size of H, making the liquidity competitor C being indifferent between overbidding or not. The higher the displayed volume by H, the lower the execution probability and thus the higher the expected execution costs of C, if he places his order on the same price level (and thus looses execution priority). If $a_H > \psi(a_H)$, these costs exceed the costs induced by the price improvement (by Δ), which forces him to overbid.

The threshold $\psi(a_H)$ depends on whether or not L partakes in trading. This decision, however, depends on the display size a_H. Therefore, $\psi()$ depends on a_H and one can readily verify that $\psi()$ increases in a_H. Hence, the threshold, making C indifferent between overbidding or not, becomes higher if the display size is higher as it becomes more likely that L enters the market. In this case, C's execution risk declines as he can expect to get executed against L (even if he is queued behind the displayed part of H). Figure 3 shows an illustration of the overbidding mechanics.

Figure 3: Illustration of C's overbidding decision a_C^*. At t_0, H submits an order of size N_H with display size a_H (blue bar). The remaining $N_H - a_H$ are not visible to C (light-blue bar). Left picture: When H does not over-display, i.e., $a_H \leq \psi_1$, C does not overbid ($a_C^* = 0$) at t_1. Right picture: When H displays too much, for instance, $a_H > \psi_2$, C overbids ($a_C^* = \Delta$).



3.4 Equilibrium strategies

3.4.1 The hidden trader's equilibrium strategy

In this section, we derive the equilibrium by obtaining the hidden trader's best response conditional on the other trader's best strategies. To derive the expected execution costs for H, we consider the two possible outcomes where H's volume is *not* executed against trading demand by L. First, C overbids ($a_C = 1$), in which case the total N_C shares have priority over the N_H hidden shares, causing execution costs of $Q(0, N_C, N_H)$. Second, C does not overbid ($a_C = 0$), in which case H's displayed part has priority over N_C shares, while the remaining $N_H - a_H$ have less priority. In this case, the costs are $Q(a_H, N_C, N_H - a_H)$.

Finally, in the case, where the latent trader L enters the downstairs market, all displayed orders get executed and H's execution costs for remaining shares equals $Q(N_H - a_H, 0, 0)$. By denoting $p_H(a_H, a_C)$ as the probability for H's volume being *not* executed against trading demand by L, the expected execution costs of H are given by

$$\begin{aligned}\Pi_H(a_H, a_C) =& p_H(a_H, a_C)\Big(\mathbf{1}_{\{a_C=0\}}Q(a_H, N_C, N_H - a_H) + \mathbf{1}_{\{a_C=\Delta\}}Q(0, N_C, N_H)\Big) \\ &+ \Big(1 - p_H(a_H, a_C)\Big)Q(N_H - a_H, 0, 0).\end{aligned} \tag{3.14}$$

Denote $a^*_C(a)$ as C's best response given H's display size $a_H \geq 0$, and denote the corresponding expected execution costs according to (3.14) by $\Pi^*_H()$. The equilibrium actions of H and C, (a^*_H, a^*_C), are then given by

$$a^*_H \in \arg\max_a \Pi^*_H(a, a^*_C); \quad a^*_C = a^*_C(a^*_H).$$

The following theorem characterizes the optimal strategies of H and C in the equilibrium:

Theorem 1. *Let $\psi()$ denote the overbidding threshold according to Lemma 1. Then, there exists a threshold for the size of H's Iceberg order $N_0 \in [\phi_3, \infty]$, such that the unique equilibrium strategy is given by*

$$(a^*_H, a^*_C) = \begin{cases} (\tilde{a}_H, 0) & \text{if } N_H \leq N_0, \\ (N_H, \Delta) & \text{else}, \end{cases}$$

with $\tilde{a}_H := \min(N_H, \psi(a_H = N_H))$.

PROOF: See Appendix. □

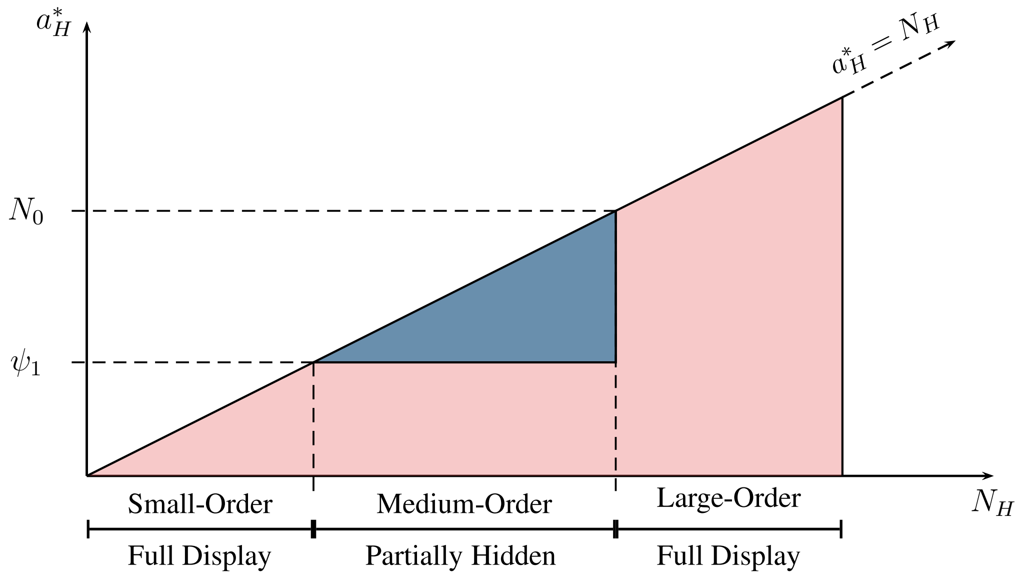
When the order size of H is below the critical threshold N_0, H limits his display size to the 'overbidding threshold' $\psi()$, which (just) prevents the competitor C from overbidding. Hence, consistent with the conjecture by Harris (1997), the presence of liquidity competitors restricts the extent by which trade intentions can be made public.

If H's order volume, however, is too small ($N_H < \min(\psi_1, \psi_2) \leq \psi(N_H)$), it never triggers the overbidding threshold. Consequently, small hidden traders do not face any liquidity competition and thus can afford fully displaying their trade intentions.

A third case occurs if the order size of H is larger than N_0. Such orders are sufficiently large to attract (large) latent counterparties if they are fully disclosed. In this case, the expected benefits arising from attracting the latent trader outweigh the expected losses due to overbidding. Therefore, traders that are sufficiently large, fully display their trade intentions.

Accordingly, hiding orders is only a rational decision for *medium-size* traders with $\psi_1 < N_H < N_0$. These traders are neither sufficiently small to avoid liquidity competition, nor they are sufficiently large to attract latent liquidity demanders. Figure 4 illustrates the different regimes of optimal order display in the equilibrium.

Figure 4: The figure shows the different equilibrium display regimes. There are two regimes. If H is large ($N_H > N_0$) or small ($N_H < \psi$), he fully displays his order. Otherwise, if the order is medium-sized ($\psi < N_H < N_0$), he partially hides his order and shows at most ψ shares.



Thus, order-display can attract additional latent demand that would otherwise not partake in trading. This is beneficial for all market participants. Liquidity suppliers benefit from increased execution probability and a reduction in the usage of costly market orders, while large liquidity demanders can meet their liquidity demand without conducting costly counterparty search in the upstairs market. Hence, the welfare benefits from the disclosure of trade intentions ultimately arise from synchronizing the timing and locations of trades through order display signaling. The absence of such trade coordination, due to an excessive use of hidden orders, does not only affect welfare and trading costs, but does also help explaining our main finding in Figure 1, suggesting that hidden orders induce significant price adjustments while displayed orders do not.

3.5 Testable predictions

3.5.1 Effects of hidden order submissions

According to the equilibrium derived in the previous section, it is optimal for sufficiently large liquidity suppliers to fully display their orders as long as they expect the presence of latent trade demand. Then, liquidity supply and demand is optimally coordinated and there is no need to confront the market with unexpected trade demand arising from non-executed hidden orders. In this scenario, any submission of hidden orders is sub-optimal and implies a deviation from equilibrium. As a result, liquidity supply is not matched by (latent) trading counterparts, which induces liquidity mis-coordination and forces traders facing liquidation constraints to enforce trade execution. It therefore becomes necessary to cancel any non-executed hidden orders and to trade them as market orders, which, however, causes price impact. A major implication of our model therefore is that an excessive use of hidden orders causes a temporal mis-match between liquidity supply and demand which in turn causes price pressure. This price pressure emerges from arising trade demand, which, however, is unexpected as it has been shielded from the market.

According to our theory, full order revelation is optimal for large and small traders but not necessarily for medium-size traders. In the latter case, the increase of hidden volume does not necessarily induce trading frictions. Empirically, however, we cannot identify individual hidden orders and thus cannot separate between small, medium or large traders. Effects of hidden order submissions, however, should be also valid even if they are tested unconditionally (i.e., without controlling for the size of individual orders), though they might be watered down in case of a dominance of medium-size orders. Empirical evidence will tell us to which extent such causalities can be identified. This yields the following hypotheses:

Testable Hypothesis 1 (Effect on returns)**.** Increases in hidden buy order volume cause positive returns.

Testable Hypothesis 2 (Effect on cancellation activity)**.** Increases in hidden buy order volume amplify cancellations of buy volume.

Testable Hypothesis 3 (Effect on order aggressiveness)**.** Increases in hidden buy order volume amplify the submission of market order volume on the buy side.

Note that Hypothesis 1 solely is also consistent with the hypothesis of informed trading: informed traders may use hidden orders to conceal their trade intentions. Therefore, one might expect that (buy) hidden orders are associated with (positive) price reactions. Therefore, just testing Hypothesis 1 does not allow us to empirically separate it from the hypothesis of informed trading.

Under the hypothesis that coordination frictions – as discussed above – play a role, also Hypotheses 2 and 3 should be true. The latter are hardly in line with informed trading and thus provide additional evidence in favor of our theory. Hence, if the observed price adjustments originate from mis-coordination frictions resulting from an excessive hiding of orders, not only Hypothesis 1, but also Hypotheses 2 and 3 should hold jointly.

Finally, note that the same predictions hold for hidden *sell* orders, but with reversed sign. Therefore, assuming that buy hidden and sell hidden traders arrive randomly in time, we expect to see price fluctuations in both directions, which increase with the extent of hidden order submissions. Consequently, we predict that markets with a higher proportion of hidden liquidity should face a higher level of volatility due to a generally higher level of liquidity mis-coordination:

Testable Hypothesis 4 (Effect on volatility)**.** Markets with a higher level of hidden liquidity exhibit more volatility.

3.5.2 Determinants of hidden order submissions

The previous section addressed the effects of hidden order submissions on subsequent prices and order submission activities. In this section, we analyze the implications of our theoretical model for the *determinants* of hidden order submissions. According to our theory, the main reason for the submission of hidden orders is to avoid liquidity competition by limiting the display size. The display size is triggered by the overbidding threshold $\psi(a_H)$ as of (3.12) and is determined by liquidity suppliers' expectations of the cost of trading. The latter are driven by the bid-ask spread, the minimum tick size, and the order book depth on the opposite side of the market.

Hence, in our framework, the driving forces of hidden order submissions correspond to the determinants of the threshold $\psi()$. Note that according to (3.13), $\psi()$ depends on thresholds ψ_1 and ψ_2, for which, however, a closed-form solution is generally not available. We nevertheless consider a special case which illustrates how the thresholds depends on underlying liquidity characteristics: Assume that noise trading volume is exponentially distributed with mean $\lambda > 0$. Moreover, assume that the impact function $c(m)$ follows (3.10) with $1/\beta$ denoting the role of the (constant) order book density and s being the quoted bid-ask spread. Under these simplifying assumptions, the sub-threshold in (3.13) ψ_1 is given by

$$\psi = \psi_1 := \begin{cases} -\lambda \log\left(1 - \frac{N_C}{\lambda}\frac{\Delta}{\tilde{s}}\right) & \text{if } \frac{N_C}{\lambda} < \frac{\tilde{s}}{\Delta}, \\ +\infty & \text{else,} \end{cases} \tag{3.15}$$

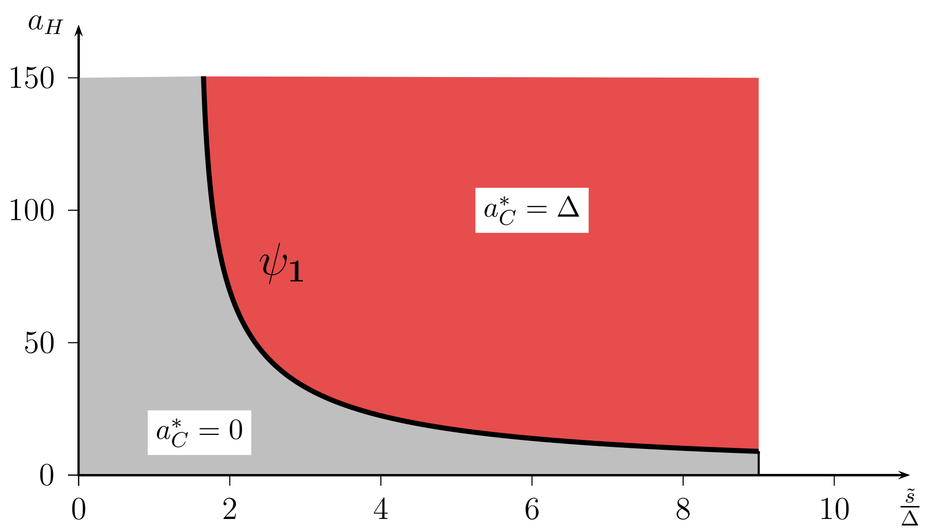
where

$$\tilde{s} := s\Big(1 - e^{-\frac{N_C}{\lambda}}\Big) + \lambda\beta\Big(\frac{N_C}{\lambda} - (1 - e^{-\frac{N_C}{\lambda}})\Big) > 0. \tag{3.16}$$

A similar calculation can be done for ψ_2. For sake of brevity, we concentrate our discussion on ψ_1. Notice that $\tilde{s}$ can be interpreted as the C's expected *effective spread*, corresponding to the effective per share costs when market (sell) orders arrive with exponentially distributed size x, and the remaining non-executed $N_C - x$ shares are traded via limit buy market orders with linear price impact $x\beta$.

The relationship between ψ and the effective spread $\tilde{s}$ can be graphically easily illustrated when $\tilde{s}$ is expressed relatively to the tick size Δ. Then, as shown in Figure 5, the overbidding threshold monotonously declines in $\frac{\tilde{s}}{\Delta}$.

Figure 5: The figure illustrates the overbidding threshold ψ_1 (3.15) as a function of the relative effective spread, $\frac{\tilde{s}}{\Delta}$. C overbids H whenever $a_H > \psi_1$. The illustration is based on a setting with $\mu = 0$, $\lambda = N_C = 100$, and $\beta = 0$, corresponding to stocks with deep books implying small price impacts.



Hence, as higher effective spreads make market orders more expensive, liquidity competitors have a higher interest in executing their position via limit orders. Consequently, they are more willing to overbid standing orders in order to increase the execution probability of their order. As a result, the overbidding threshold becomes smaller. Therefore, ceteris paribus, effective spreads increase liquidity competitors' order aggressiveness. Following the same logic, it is intuitive that the tick size acts in exactly the opposite way: A high tick size makes it more costly to overbid which in turn reduces liquidity competition and yields a higher overbidding threshold. Consequently, hidden traders are more willing to display their trade intentions. This broadly confirms the conjecture by Harris (1996) that high tick sizes prevent displayed orders from being undercut.

As an example, consider highly liquid stocks such as Cisco and Microsoft, which mostly have a quoted spread of one tick. Moreover, their order book is thick, suggesting that average trade sizes do not have much price impact, i.e., $\beta\lambda \approx 0$. It is furthermore known that the level of incoming liquidity supply clearly exceeds the amount of liquidity demand, i.e., $N_C/\lambda \gg 1$. Hence, in these cases, our model predicts $\psi_1 \to \infty$, suggesting that there is no (or only little) hidden volume in the order book.

Note that liquidity characteristics such as the quoted bid-ask spread s, the (opposite side) order book depth β, and the market order arrival rate λ affect $\psi()$ implicitly through the effective spread $\tilde{s}$. Hence, ceteris paribus, smaller spreads s, higher depth $1/\beta$ and more market order arrivals λ reduce the effective spread and therefore increase the overbidding threshold. Consequently, the display size increases and the proportion of hidden liquidity shrinks. We formulate these implications as testable hypotheses:

Testable Hypothesis 5 (Bid-ask spreads and hidden liquidity)**.** Markets with smaller bid-ask spreads reveal a higher proportion of hidden liquidity.

Testable Hypothesis 6 (Relative tick size and hidden liquidity)**.** Markets with wider (relative) tick sizes reveal a lower proportion of hidden liquidity.

4 Testing the model implications

4.1 Dynamic implications

Given the data availability, the theoretical model presented in Chapter 3 cannot be tested in structural form. Nevertheless, using the econometric framework presented in Chapter 2, we can test Hypotheses 1, 2 and 3 in reduced form. Hence, employing the VAR model for the 12-dimensional vector

$$y_t = (RET, SPR, VOLA, HI10, DI10, TD, SUB, SUS, EXB, EXS, CAB, CAS)',$$

the effect of a shock in the order imbalance of hidden orders ($HI10$) and displayed orders ($DI10$) on subsequent mid-quote returns cancellation activities and order aggressiveness, while accounting for dynamic (cross-)dependencies, is quantified by (cumulative) impulse response functions as formulated in (2.6). Accordingly, Hypothesis 1 is strongly confirmed by the cumulative impulse response function depicted in Figure 1 in Chapter 1. Hence, in line with our theory, imbalances in hidden volume induce a significantly stronger price impact than imbalances in displayed liquidity.

As discussed in Section (3.5.1), our theory is additionally supported by corresponding evidence for cancellation and order placement activities. Figures 6 through 8 show the estimates of the cumulative impulse responses of one-minute buy and sell limit order cancellation volumes ($\hat{\Xi}_{CAB}$ and $\hat{\Xi}_{CAS}$), execution volumes ($\hat{\Xi}_{EXB}$ and $\hat{\Xi}_{EXS}$), and submission volumes ($\hat{\Xi}_{SUB}$ and $\hat{\Xi}_{SUS}$), triggered by a positive one-standard-error shock in hidden and displayed order imbalances ($HI10$ and $DI10$).

As NASDAQ ModelView data does not contain the cancellation of hidden orders, we cannot verify whether hidden volume gets canceled in case of high depth imbalances and thus cannot fully validate Hypothesis 2. Recorded cancellations of *displayed* orders, however, can (at least partly) correspond to displayed parts of larger (partially) hidden orders. In this case, the cancellation of partly hidden orders might also trigger a fraction of displayed cancellations. Hence, the impulse response of hidden and displayed order imbalances on cancellations, as shown in Figure 6, might provide at least some evidence in favor of Hypothesis 2. In fact, we observe that shocks in non-displayed net buy depth tends to increase the cancellation rate of buy limit orders. The fact that this relationship is only borderline significant (and insignificant for net sell volumes) might be, however, due to the fact that this might be only partly driven by cancellations of hidden orders.

Figure 6: Estimates of cross-sectional averages of the cumulative generalized impulse response of one-minute aggregated limit order buy and sell **cancellation volumes** ($\hat{\Xi}_{CAB}$ and $\hat{\Xi}_{CAS}$) due to positive one-standard-deviation shocks in hidden (blue) and displayed (red) order imbalances. The dashed lines show the approximate 95% confidence intervals of the averaged impulse response functions. Based on one-minute aggregates of NASDAQ ITCH data and one-minute snapshots of NASDAQ ModelView data for the stocks APC, AZO, CAH, GAS, GOOG, EMR, LEG, PAYX, STJ, and TDC, for November to December 2008.

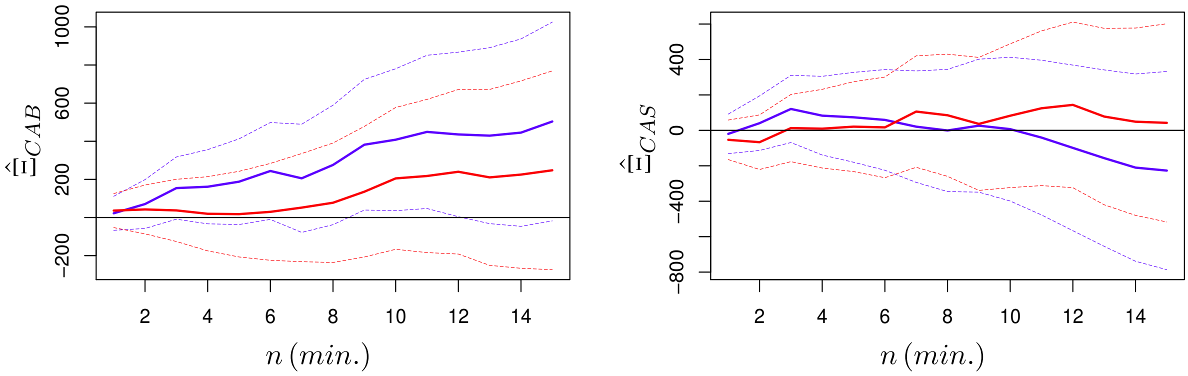
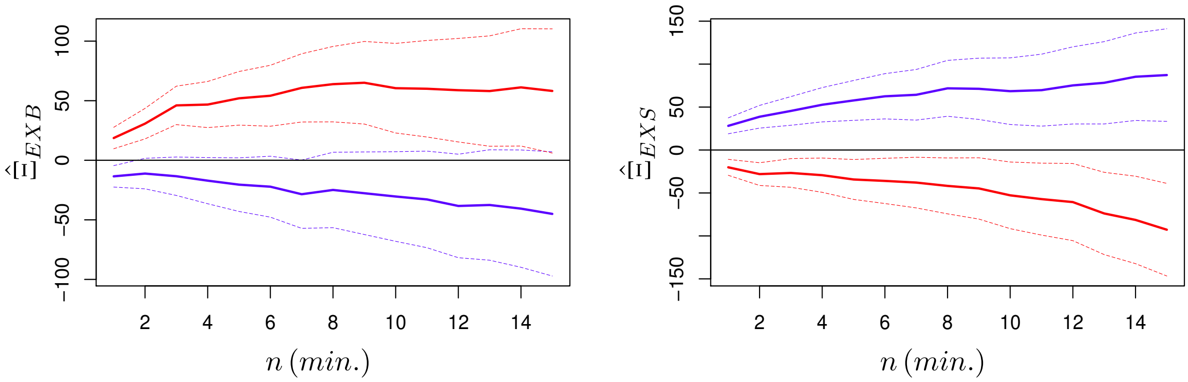
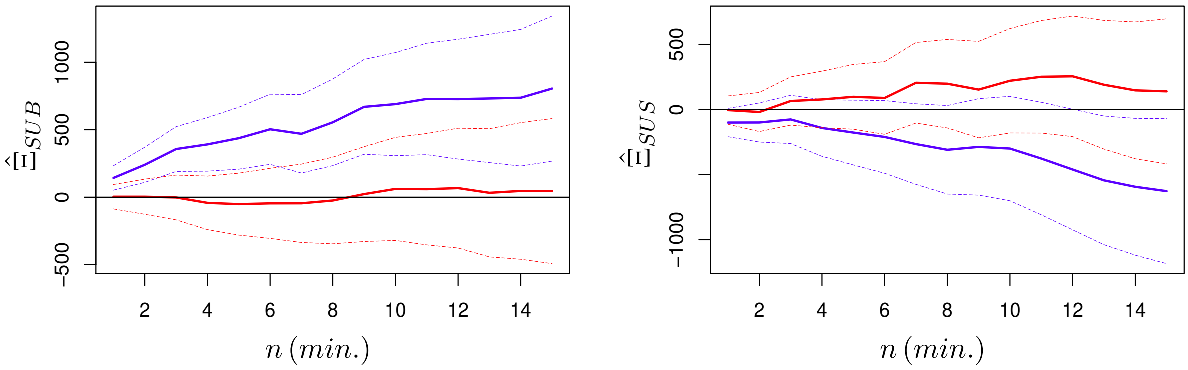


Figure 7: Estimates of cross-sectional averages of the cumulative generalized impulse response of one-minute aggregated limit order buy and sell **execution volumes** ($\hat{\Xi}_{EXB}$ and $\hat{\Xi}_{EXS}$) due to positive one-standard-deviation shocks in hidden (blue) and displayed (red) order imbalances. The dashed lines show the approximate 95% confidence intervals of the averaged impulse response functions. Based on one-minute aggregates of NASDAQ ITCH data and one-minute snapshots of NASDAQ ModelView data for the stocks APC, AZO, CAH, GAS, GOOG, EMR, LEG, PAYX, STJ, and TDC, for November to December 2008.



Hypothesis 3 is tested based on estimated impulse response functions of order execution volumes. According to Figure 7, an increase in buy-side hidden orders does *not* generate a significant increase in buy order executions (EXB). The effect tends to be negative but is widely insignificant. In contrast, positive shocks in hidden buy-sell imbalances have a significantly positive impact on sell order order executions, i.e., buy market orders (EXS). These results confirm Hypothesis 3 and show that hidden liquidity influences the market in a very different way than displayed liquidity. Indeed, for the latter, we observe exactly opposite effects, which are naturally explained by the fact that displayed liquidity is likely to be absorbed by trading counterparts and thus induce executions on the *same* side of the market. The fact that hidden liquidity imbalances trigger executions on the *opposite* side is strongly in line with our reasoning that impatient hidden liquidity suppliers enforce execution by increasing their trading aggressiveness.

Figure 8: Estimates of cross-sectional averages of the cumulative generalized impulse response of one-minute aggregated limit buy and sell **order submission volumes** ($\hat{\Xi}_{SUB}$ and $\hat{\Xi}_{SUS}$) due to positive one-standard-deviation shocks in hidden (blue) and displayed (red) order imbalances. The dashed lines show the approximate 95% confidence intervals of the averaged impulse response functions. Based on one-minute aggregates of NASDAQ ITCH data and one-minute snapshots of NASDAQ ModelView data for the stocks APC, AZO, CAH, GAS, GOOG, EMR, LEG, PAYX, STJ, and TDC, for November to December 2008.



Additional supportive evidence is provided by Figure 8, showing the estimated cumulative impulse responses for submission volumes. We observe that hidden order imbalances increase the rate of *displayed* limit orders on the same side of the market and tend to decrease limit order activities on the opposite side. Hence, hidden order imblances trigger corresponding imbalances in displayed limit orders. These effects are in line with Hypothesis 3 as they indicate that excessive hidden liquidity generally increases order aggressiveness. However, this does not only result in higher market order activity but obviously also in more *displayed* limit order submissions. Hence, a possible strategy to accelerate the execution of non-executed (hidden) orders is to make them visible to the market. This strategy is not captured by our model but can be seen as a less aggressive alternative to market order trading.

Again we observe a striking difference between the effects of displayed and hidden order submissions. While imbalances in hidden orders cause significant reactions in limit order submissions, these effects are insignificant in case of imbalances in displayed orders. Hence, strong one-sided displayed liquidity supply does not trigger further effects in the same direction but is likely to be absorbed by trading demand on the opposite of the market. The re-positioning and revelation of volume which has been concealed before, however, causes fundamentally different effects as it is perceived as newly arriving liquidity supply or trading demand, respectively. As a result to this unexpected trading demand, prices significantly move, as documented in Figure 1, manifesting the key theoretical and empirical finding of this paper: hiding orders h liquidity sychnronization and makes prices inefficient as they only incorporate visible but not invisible trade demand. As a result, we observe price fluctuations which would not occur if liquidity demand and supply would be synchronized.

4.2 Cross-sectional implications

While Hypotheses 1 to 3 are associated with the causal temporal effects of hidden liquidity submission, Hypotheses 4 to 6 formulated in Chapter 3 postulate cross-sectional relationships between the extent of hidden liquidity, volatility, bid-ask spreades and the tick size. In order to test these hypotheses, we therefore define the relative bid-ask spread $RSPR$ as the ratio between the spread SPR and the mid-quote MQ, and the relative tick size $RTCK$ as the ratio between the tick size TCK and MQ. Volatility is estimated by the daily realized variance (RV), computed as the sum of squared 10-min returns.

Testing the role of relative bid-ask spreads and relative tick sizes as drivers of hidden liquidity (Hypothesis 5 and 6) yields the following simple regression

$$RHD10_i = \alpha_h + \beta_{h,2} RTCK_i + \beta_{h,3} RSPR_i + \varepsilon_{hi}. \tag{4.1}$$

Testing the effect of volatility on liquidity (Hypothesis 4) implies the model

$$RV_i = \alpha_v + \beta_{v,1} RHD10_i + \beta_{v,2} RTCK_i + \beta_{v,3} RSPR_i + \varepsilon_{vi}, \tag{4.2}$$

for $i = 1, \ldots, N$, and white noise error terms ε_{hi} and ε_{vi}. Although not predicted by our theory, equation (4.2) also incorporates $RTCK_i$ and $RSPR_i$ as additional control variables.All variables enter in logarithmic form as time averages across days and (in case of $RHD10_i$ and $RSPR_i$) one-minute snapshots within a day. This leaves us with $N = 468$ cross-sectional observations.[12]

Equations (4.1) and (4.2) constitute a triangular relationship, where causality runs from $RHD10$ to RV. Although not captured by our framework, this causality, however, might be reversed. For instance, Harris (1996) argues that liquidity suppliers use hidden orders to reduce the risk of being picked off. Since the picking-off risk is particularly high in volatile markets, causality may run from volatility to hidden liquidity. Moreover, in our setting, simultaneity between RV and $RHD10$ can simply arise because of the use of time averages of both variables. To account for this effect, we consider a second specification in which we explicitly include RV_i in the first equation, resulting into a bivariate simultaneous equations system:

$$RHD10_i = \tilde{\alpha}_h + \tilde{\beta}_{h,1} RV_i + \tilde{\beta}_{h,2} RTCK_i + \tilde{\beta}_{h,3} RSPR_i + \tilde{\varepsilon}_{hi}, \tag{4.3}$$

$$RV_i = \tilde{\alpha}_v + \tilde{\beta}_{v,1} RHD10_i + \tilde{\beta}_{v,2} RTCK_i + \tilde{\beta}_{v,3} RSPR_i + \tilde{\varepsilon}_{vi}. \tag{4.4}$$

As soon as both $\tilde{\beta}_{h,1}$ and $\tilde{\beta}_{v,1}$ are truly non-zero, $RHD10_i$ and RV_i are simultaneous, and

[12] An alternative to using averaged variables would be to estimate the model in a panel setting, which would allow us to exploit not only cross-sectional variation but also time variation. Properly capturing the strong serial (cross-)dependencies of most variables in a panel setting would be, however, quite challenging. Corresponding panel VAR approaches would be cumbersome, or even impossible, to estimate given the amount of underlying observations.

Table 2: Estimation results of cross-sectional regressions of (4.1), (4.2), (4.3) and (4.4). The first two columns give the OLS estimates of (4.1) and (4.2). The next columns give the 2SLS estimates of (4.3) and (4.4) with instruments $RET2_i$ and $D10_i$. Standard errors are shown in brackets. Below, we report the F-statistics based on the first-stage regressions as tests for weak instruments and the Sargan test for over-identification.

	Structural Model		*Simultaneous Equations Model*			
	$RHD10_i$	RV_i	$RHD10_i$	RV_i	$RHD10_i$	RV_i
	(4.1)	(4.2)	(4.3)	(4.4)	(4.3)	(4.4)
$RHD10_i$		3.356***		3.356***		1.560***
		(0.427)		(0.427)		(0.163)
RV_i			0.298***		0.294***	
			(0.038)		(0.038)	
$RTCK_i$	−0.464***	1.877***	−0.559***	1.877***	−0.558***	1.043***
	(0.027)	(0.217)	(0.029)	(0.217)	(0.029)	(0.090)
$RSPR_i$	0.172***	−0.234*	0.070**	−0.234*	0.071**	0.075
	(0.031)	(0.126)	(0.033)	(0.126)	(0.033)	(0.063)
Const.	−1.663***	−5.438***	1.620***	−5.438***	1.581***	−8.426***
	(0.106)	(0.791)	(0.431)	(0.791)	(0.430)	(0.332)
Instruments	–	–	$RET2_i$		$RET2_i, D_i$	
N	468	468	468	468	468	468
Weak-Instr.	–	–	1067.78***	68.08***	532.46***	86.52***
Sargan	–	–	–	–	71.75***	109.30***
Note:						*p<0.1; **p<0.05; ***p<0.01

the parameters cannot be consistently estimated by OLS. We therefore employ two-stage least squares (2SLS) to estimate the system equation-by-equation. We use the squared daily mid-quote return, $RET2_i$, as an obvious instrument for RV_i. As a second instrument, we utilize the displayed depth $D10_i$. Both $RET2_i$ and $D10_i$ are correlated with the endogenous variables. While the uncorrelatedness of $RET2_i$ and $\tilde{\varepsilon}_{hi}$ is easily justified (given that RV_i serves as a regressor in (4.3) and thus captures most volatility-associated variation in $\tilde{\varepsilon}_{hi}$), the uncorrelatedness of $RET2_i$ and $\tilde{\varepsilon}_{vi}$ is more critical and relies on the ability of the regressors that are included in (4.4) to sufficiently capture variations in RV_i. We conjecture, however, that cross-sectional variation in RV_i is particularly captured by the included regressors rather than by squared daily returns, diminishing the remaining explanatory power of $RET2_i$ for RV_i and making correlations between $RET2_i$ and $\tilde{\varepsilon}_{vi}$ unlikely. The uncorrelatedness between $D10_i$ and both $\tilde{\varepsilon}_{hi}$ and $\tilde{\varepsilon}_{vi}$ can be similarly justified, as $D10_i$ and $RTCK_i$ are strongly correlated[13], and thus, we expect the explanatory power of $RTCK_i$ to capture most of the variation in both equations (4.3) and (4.4), making correlations between $D10_i$ and both $\tilde{\varepsilon}_{vi}$ and $\tilde{\varepsilon}_{hi}$ less likely.

Table 2 presents the equation-by-equation OLS estimates of (4.1) and (4.2) and 2SLS estimates of (4.3) and (4.4). Without exception, the coefficient estimates based on all model specifications confirm Hypotheses 4 to 6 and are significant on the 5% level. Accordingly, there is significant evidence for hidden liquidity provision ($RHD10_i$) being higher for stocks that trade at wider spreads ($RSPR_i$) and smaller tick sizes ($RTCK_i$). Furthermore, markets that exhibit a higher proportion of hidden liquidity supply ($RHD10_i$) are more volatile (RV_i). Moreover, as expected, we find evidence for simultaneity between volatility and hidden liquidity provision. Although the test of over-identification does not fully support our choice of over-identifying moment conditions, the results are nevertheless qualitatively similar across the different specifications. The results are also qualitatively similar if additional or other instruments are employed (not shown here). We therefore conclude that simultaneity effects do not fundamentally influence the coefficient estimates of our variables of interest.

5 Conclusions

This paper shows that the excessive use of hidden orders can pose a possible source of trading frictions in limit order book markets. Shielding liquidation needs from public view distorts the synchronization of liquidity supply and demand and, as a result, cause excess price fluctuations which are unrelated to information and would not occur in case of full order display.

In contrast, displaying trade intention improves the time synchronization of time schedules between liquidity demander's trading needs so that price pressures resulting from trading demand and trading impatience are mutually absorbed. Consistent with this reasoning, we show that the excessive use of hidden liquidity can artificially increase market volatility and thus can harm the price discovery process and price efficiency.

[13]The estimated correlation coefficient between $D10_i$ and $RTCK_i$ equals 0.81.

An important conclusion arising from this study is that trading frictions due to hidden liquidity ultimately imply higher trading costs and a higher market fragmentation. The latter effect arises because (large) liquidity demanders refrain from trading in a public limit order book if the displayed liquidity is low and they face the risk of causing (too much) price impact. Consequently, public limit order book exchanges suffer from a shortage in liquidity demand which increases the costs for trading.

The key determining factor in the provision for hidden liquidity is liquidity suppliers' belief about the presence of latent liquidity demand and their belief about the intensity in liquidity competition. Therefore, anything that reduces competition in liquidity supply or increases their information about the presence of (large) latent liquidity demand will increase public supply of displayed liquidity. Our analysis, therefore, provides several profound implications for reducing trading frictions and increasing market participation on public exchanges.

First, in order to increase pre-trade transparency, exchange operators can enhance market makers' and liquidity suppliers' incentives to keep spreads narrow, reducing the costs of liquidity competition and thus increasing the incentives for order-exposure. Wide spreads increase order-aggressiveness of incoming orders and -thereby- force liquidity suppliers to hide a greater fraction of their order. Therefore, incentivizing market makers to provide narrower spreads can reduce hidden liquidity related trading frictions.

Second, exchange operators can introduce a rebate structure that rewards liquidity demand. Although this violates the general pricing rule that liquidity demanders have to pay a fee for consuming liquidity, attracting large liquidity demanders will reduce the amount of large hidden orders in the order book.

Third, increasing monitoring as well as *direct market access* capabilities will increase the chances of exposure but will also increase the rate at which latent investors observe the public market to seize liquidity opportunities. Knowing that latent traders monitor the market, liquidity suppliers are more willing to reveal their trade intention and thereby increasing market liquidity and reducing market volatility.

Fourth, we propose a novel order type, the *mutually binding Indication Order of Interest* that mitigate the downsides of hidden orders but maintain the signaling capacity of openly displayed orders. When a buyer submits a bIOI, they instantly match with pre-existing sell bIOI's at a pre-specified price. If there is no sell bIOI, then it is not actively displayed. Hence, mutually binding IOIs effectively operate as a dark pool on top of the pre-existing public exchange. Therefore, signals are only received by those counterparties who ar pre-committed to trade and completely bypasses those traders who act at the same side of the market. Because orders are matched at a pre-specified price, the matching of bIOIs can not have any price-effects. We therefore argue that the inception of these order types can reduce market volatility, attract non-exchange order flow and improve trade execution for large traders.

Appendix

A Proofs

As discussed in Chapter 3, the cost function $c(m)$ needs to satisfy a 'growth condition' of the form

$$\lim_{n\to\infty} \frac{c(n-m)}{c(n)} = 1 \quad \text{for all } m \geq 0. \tag{5.1}$$

There are several reasons why this condition does not pose a substantial restriction to our model. First, it is satisfied, whenever $c()$ is of polynomial or power-law growth. This is in line with much empirical evidence on power-law scalings in financial markets. Second, the growth condition ensures that for sufficiently large trade sizes n and a fix size m, the marginal transaction costs c of a market order of size $n-m$ and market order of size n are asymptotically equal.[14] Through the remainder, we assume that (5.1) is satisfied.

Before we proof Lemma 1 and Theorem 1, we present one preparatory lemma:

Lemma 2. *Assume that the buy-side order book depth behind H's submission price level is block-shaped, i.e., $V^j = V = 1$, and assume that L's order size N_L satisfies $N_L^- \leq N_L \leq N_L^+$ with $N_L^- = N_H + N_C + 2\frac{\gamma}{\Delta} - 1$ and $N_L^+ = (N_C + N_H)\left(1 + \frac{\gamma}{\Delta}\right)$. Then, L executes at most the visibly displayed orders of H and C, i.e.,*

$$a_L^* \in \{0, 1, 2\}. \tag{2.2}$$

PROOF OF LEMMA 2: Assume that L has consumed already the (visible) liquidity provided by H and C and denote by $N := N_L - a_H - N_C$ the number of remaining shares. Under our assumption of a block-shaped order book, i.e., $V^j = V$, for all j, the cost of trading is minimal. In this case, if L trades i ticks deeper into the book, the costs (in ticks) of trading downstairs, respectively upstairs, when benchmarked against H's submission price level, equal

$$\sum_{j=1}^{i} j = \frac{i(i-1)}{2}, \qquad \text{respectively,} \qquad \left(i + \frac{\gamma}{\Delta}\right)(N-i).$$

The resulting cost function

$$C : \{0, 1, ..., N\} \to \mathbb{R}; \quad i \mapsto \frac{i(i-1)}{2} + \left(i + \frac{\gamma}{\Delta}\right)(N-i)$$

attains its minimum at $i^* \in \{0, N\}$. By direct computation we verify that the minimum is indeed attained at $i = 0$ because $N_L > N_H + N_C + 2\frac{\gamma}{\Delta} - 1$. □

PROOF OF LEMMA 1: There are three regimes where the structure of C's cost function changes as a function of H's and his own strategy. There are, however, only two thresholds

[14] For instance, there should be no much difference in the cost function if a trader submits an order of say 10000 shares or 10001 shares.

where C's optimal action changes from 0 to Δ. For $a_H \in [0, \phi_1)$, the total liquidity offered by H and C is too small to trigger a trade execution by L, hence, $p_C \equiv 1$. In this case, $a_C = 0$ is optimal if and only if

$$Q(0, 0, N_C) + \Delta N_C \geq Q(0, a_H, N_C).$$

Since the mapping $a_H \mapsto Q(0, a_H, N_C)$ is strictly increasing, we obtain the first possible threshold.

On $[\phi_1, \phi_3)$, the best response of L yields $p_C \equiv 1-\mu$ for $a_C = 0$ and $p_C \equiv 1$ for $a_C = \Delta$. As a result, the costs associated with $(N_H, 0)$ decrease discontinuously at ϕ_1 for both players, and $a_C = 0$ might be optimal for the competitor at ϕ_1 in which case a second threshold $\psi_2 \in [\phi_1, \phi_3)$ might exist, where the best response switches back to $a_C = \Delta$. If such a threshold exists, i.e., if $a_C = \Delta$ is optimal for some $a_H < \phi_3$, then $a_C = \Delta$ must be optimal for all $a_H \geq \phi_3$ as the costs associated with price overbidding decreases at ϕ_3 with p_C declining from $p_C = 1$ to $p_C = 1 - \mu$.

Finally, if no threshold exists in $[\phi_1, \phi_3)$, then a threshold might exist in $[\phi_3, N_H]$, where the probability of latent trader execution is independent of C's action and $a_C = 0$ is optimal if and only if

$$(1 - \mu)Q(0, 0, N_C) + \Delta N_C \geq (1 - \mu)Q(0, a_H, N_C).$$

This proves the structural result on C's best response function. □

PROOF OF THEOREM 1:

For the proof, we have to show that the equilibrium display size a_H^* must satisfy $\Pi_H(a_H^*) \leq \Pi_H(a)$ for all $a \in [0, N_H]$. We will show this for the case $\phi_3 > \psi_2$. The opposite case does not add much further insights and follows by analogy and slight adjustments.

Below we discuss different scenarios depending on the size of N_H:

- If $N_H \in [0, \phi_1)$, then $p_H(a, a_C(a)) = 1$ for all display sizes $a \leq N_H$. As a result,

  $$\Pi^*(a) = \begin{cases} Q(a, N_C, N_H - a) & \text{if } a < \psi_2, \\ Q(0, N_C, N_H) & \text{else.} \end{cases}$$

  Since the mapping $a \mapsto Q(a, N_C, N_H - a)$ is decreasing, this yields

  $$a_H^* = \min\{N_H, \psi_1\}.$$

- If $N_H \in [\phi_1, \psi_2)$, then $a_C^*(a) = 0$ for all $a \in [\phi_1, \psi_2)$ as of Lemma 1. Thus, since the mapping $a \mapsto Q(a, N_C, N_H - a)$ is decreasing, $a_H = N_H$ is the dominating display size for all $a \in [\phi_1, \psi_2)$. For the same reason, it is also the dominating display size for $a < \psi_1$. It remains to be shown that $a_H = N_H$ also dominates any choice for $a \in [\psi_1, \phi_1)$. This is true, because in this case $a_C^*(a) = \Delta$ holds according to Lemma 1, while the probability for L entering the market is zero (see Lemma 2). Thus, in equilibrium $a_H^* = N_H$.

- If $N_H \in [\psi_2, \phi_3)$, then according to the previous case, $a_H^* \geq \psi_2$ must hold. However, for $\psi_2 < a < \phi_3$, we have $a_C^*(a) = \Delta$, while L does not trade downstairs, i.e., $a_L = 0$ and $p_H(a, a_C(a)) = 1$. Hence, there is no benefit in displaying more than ψ_2, therefore $a_H^* = \psi_2$.

- If $N_H \in [\phi_3, \infty)$, then $p_H(a, a_C(a)) = 1 - \mu$ for all $a \in [\phi_3, N_H]$,

$$\Pi^*(a) = \begin{cases} \mu Q(N_H - a, 0, 0) + (1-\mu)Q(a, N_C, N_H - a) & \text{if } a < \psi_2, \\ \mu Q(N_H - a, 0, 0) + (1-\mu)Q(0, N_C, N_H) & \text{else.} \end{cases}$$

Then, the costs resulting from full display ($a = N_H$) are

$$\mu \underbrace{Q(0,0,0)}_{=0} + (1-\mu)Q(0, N_C, N_H), \tag{2.3}$$

while that resulting from partial hiding ($a = \psi_2$) is

$$\mu Q(0, 0, N_H - \psi_2) + (1-\mu)Q(\psi_2, N_C, N_H - \psi_2). \tag{2.4}$$

Under the growth condition (5.1) for the market impact function, the costs of missing out the opportunity of being executed by L outweighs the cost of price undercutting for large orders. Indeed, for all $n, h, d \in \mathbb{N}_0$ we have

$$Q(a, N_C, n-a) = \sum_{i=0}^{a} c(n-i)p_i + c(n-a)\Big(F(a+N_C) - F(a)\Big) + \sum_{j=N_C+a+1}^{N_C+n} c(N_C+n-j)p_j.$$

Now, (5.1) yields:

$$\begin{aligned} \sum_{i=0}^{a} \frac{c(n-i)}{c(n)} p_i &\overset{n\to\infty}{\longrightarrow} F(a), \\ \frac{c(n-a)}{c(n)}\Big(F(a+N_C) - F(a)\Big) &\overset{n\to\infty}{\longrightarrow} F(a+N_C) - F(a), \\ \sum_{j=N_C+a+1}^{N_C+n} \frac{c(N_C+n)}{c(n)} p_j &\overset{n\to\infty}{\longrightarrow} 1 - F(a+N_C). \end{aligned} \tag{2.5}$$

Thus, for all $\delta > 0$ there exists $N_0 \in \mathbb{N}$ such that for all $N_H \geq N_0$ that satisfy $N_H \geq \phi_3$:

$$\begin{aligned} (1-\delta)c(N_H) \leq Q(a, N_C, N_H - a) \leq (1+\delta)c(N_H), \quad a \in \{0, \psi_2\} \\ (1-\delta)c(N_H) \leq Q(0, 0, N_H - \psi_2) \leq (1+\delta)c(N_H). \end{aligned}$$

Hence the assertion follows from (2.3) and (2.4) as $Q(0,0,0) = 0$.

□

B Asymptotic theory of generalized impulse response functions

To derive the asymptotic properties of the cumulative impulse response functions Ξ, we follow Lütkepohl (2007). Therefore, consider the $K-$dimensional VAR(p) process

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t, \tag{2.6}$$

with $y_t = (y_{1t}, \ldots, y_{Kt})'$ and the $(K \times K)$ coefficients matrices A_i and $K-$dimensional white noise with $E(u_t) = 0$ and

$$E(u_t u_s') = \begin{cases} \Sigma_u, & \text{if} \quad t = s, \\ 0, & \text{otherwise}, \end{cases} \tag{2.7}$$

with

$$\Sigma_u = \begin{bmatrix} \sigma_{11} & \ldots & \sigma_{1K} \\ \vdots & \ddots & \vdots \\ \sigma_{K1} & \ldots & \sigma_{KK} \end{bmatrix}. \tag{2.8}$$

Let vec denote the column stacking operator and $vech$ the corresponding operator that stacks only the elements on and below the diagonal. Then, the duplication operator D_k is such that for any $(K \times K)$-matrix T, $D_K vech(T) = vec(T)$ holds. Then, we define the following matrices

$$D_K^+ = (D_K' D_K)^{-1} D_K', \qquad \sigma = vech(\Sigma_u), \tag{2.9}$$

$$J = \begin{bmatrix} I_k & \ldots & 0 & 0 \end{bmatrix}, \qquad \alpha = vec(A_1, A_2, \ldots, A_p), \tag{2.10}$$

and

$$\Gamma = \left(\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ t_{t-p+1} \end{bmatrix} \begin{bmatrix} y_t', & \ldots & y_{t-p+1} \end{bmatrix} \right), \qquad \overline{\Sigma}_u = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \ldots & 0 \\ 0 & \sqrt{\sigma_{22}} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & \sqrt{\sigma_{KK}} \end{bmatrix}, \tag{2.11}$$

$$\mathbf{A} = \begin{bmatrix} A_1 & A_2 & \ldots & A_{p-1} & A_p \\ I_K & 0 & \ldots & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & I_K & 0 \end{bmatrix}. \tag{2.12}$$

Let $\hat{\alpha}$ and $\hat{\sigma}$ denote the least squares estimators with respect to (2.6). According to Lütkepohl (2007) these estimators have asymptotic covariances Σ_α and Σ_σ with

$$\Sigma_\alpha = \Gamma^{-1} \otimes \Sigma_u \qquad \Sigma_\sigma = 2D_K^+ \left(\Sigma_u \otimes \Sigma_u \right) D_K^{+\prime}. \tag{2.13}$$

The orthogonalized impulse response functions, Θ^o and Ξ^o, arise from diagonalizing the residual covariance matrix Σ_u, such that $\Sigma_u = PP'$ holds. With the definition $w_t = P^{-1}u_t$, w_t obeys $\Sigma_w = E[w_t w_t'] = I_K$. The corresponding MA representation of y_t can be written as $y_t = \sum_{i=0}^{\infty} \Theta^o(i) w_{t-i}$ with the *orthogonalized impulse response function* $\Theta^o(i)$, i.e.,

$$\Theta^o(i) = \Phi_i P \qquad \Xi^o(n) = \sum_{i=0}^{n} \Theta^o(i). \tag{2.14}$$

Comparing with the corresponding *generalized* impulse responses in (2.4) and (2.5), we have

$$\Theta^g(n) = \Phi_n Q \qquad \Xi^g(n) = \sum_{i=0}^{n} \Phi_i Q, \tag{2.15}$$

$$\Theta^o(n) = \Phi_n P \qquad \Xi^o(n) = \sum_{i=0}^{n} \Phi_i P, \tag{2.16}$$

with $Q = \Sigma_u \left(\overline{\Sigma}_u\right)^{-1}$. Observe that the only difference between orthogonalized and generalized impulse response functions lies in the right-multiplication of the matrices Q and P. Thus, we can use the analogy of the asymptotic properties of the orthogonalized impulse response as of Lütkepohl (2007) to derive the asymptotic properties for the generalized impulse response.

Theorem 2 (Asymptotic Theory of Generalized Impulse Response Functions)**.** *Suppose*

$$\sqrt{T}\begin{bmatrix}\hat{\alpha}-\alpha\\ \hat{\sigma}-\sigma\end{bmatrix} \xrightarrow{d} N\left(0, \begin{bmatrix}\Sigma_\alpha & 0\\ 0 & \Sigma_\sigma\end{bmatrix}\right). \tag{2.17}$$

Then

$$\sqrt{T}vec\left(\hat{\Xi}^o(n) - \Xi^o(n)\right) \xrightarrow{d} N\left(0, B_n\Sigma_\alpha B_n' + \overline{B}_n\Sigma_\sigma\overline{B}_n'\right), \qquad n = 1, 2\ldots, \tag{2.18}$$

where the matrices B_n and $\overline{B}_n$ obey

$$B_n = \left(P' \otimes I_K\right)F_n, \qquad \overline{B}_n = (I_K \otimes \Psi_n)H, \tag{2.19}$$

$$\tag{2.20}$$

with the F_n matrices obeying $F_n = \sum_{i=1}^{n} G_i$ and $G_i = \sum_{m=0}^{i-1} J(A')^{i-1-m} \otimes \Phi_m$ and the H matrix being defined as $H = \partial vec(P)/\partial\sigma'$. Moreover, $\Psi_j = \sum_{i=1}^{j}\Phi_i$.

With (2.15) and (2.16) in mind, it is easy to check that the asymptotic property of the corresponding cumulative generalized impulse response Ξ^g are derived similarly by replacing the matrix P with Q. Together with (2.13) and Theorem 2 we finally obtain:

Corollary 1 (Asymptotic Distribution of Generalized Impulse Response)**.**

$$\sqrt{T}vec\left(\hat{\Xi}^g(n) - \Xi^g(n)\right) \xrightarrow{d} N\left(0, B_n^g\Sigma_\alpha B_n^{g\prime} + \overline{B}_n^g\Sigma_\sigma\overline{B}_n^{g\prime}\right), \qquad n = 1, 2\ldots \tag{2.21}$$

with F_n, G_n Ψ_n as in Theorem 2 and

$$B_n^g = \Big(Q' \otimes I_K\Big) F_n, \qquad \overline{B}_n^g = (I_K \otimes \Psi_n) H^g, \qquad H^g = \partial vec(Q)/\partial\sigma', \tag{2.22}$$

and

$$\Sigma_\alpha = \Gamma^{-1} \otimes \Sigma_u, \qquad \Sigma_\sigma = 2D_K^+ \Big(\Sigma_u \otimes \Sigma_u\Big) D_K^{+\prime}. \tag{2.23}$$

Due to Corollary 1, it is straightforward to show that the cumulative impulse response $\Xi_j^g(n)$ of the j-th endogenous variable at time n after the shock is obtained by right-multiplying Ξ_n^g with the column-vector e_i, which consists of zeros except at the jth entry. Thus, we have

$$\sqrt{T} vec\Big(\hat{\Xi}_j^g(n) - \Xi_i^g(n)\Big) \xrightarrow{d} N\Big(0, \Lambda_{jn}\Big), \tag{2.24}$$

and

$$\Lambda_{jn} = e_j' \Big(B_n^g \Sigma_\alpha B_n^{g\prime} + \overline{B}_n^g \Sigma_\sigma \overline{B}_n^{g\prime}\Big) e_j. \tag{2.25}$$

C Descriptive statistics

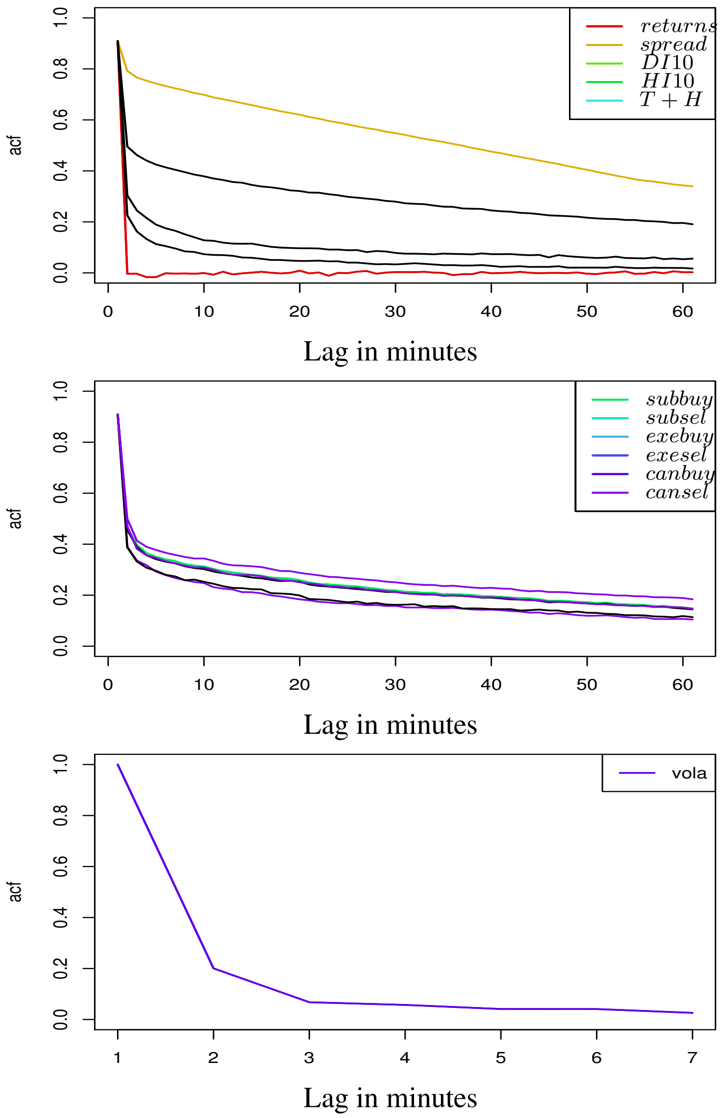
Table 3: Time-series averages of mid-quotes, bid-ask spreads, visible and hidden depth, and order activities based on one-minute aggregates for the stocks APC, AZO, CAH, GAS, GOOG, EMR, LEG, PAYX, STJ, and TDC. Reported variables: Averages of one-minute mid-quotes (MQ), the bid-ask spread (SPR), visible depth on the first 10 levels of the book ($D10$), and total hidden depth on the first 10 levels ($HD10$), and averages of minute-by-minute aggregated volume of limit buy order submissions (SUB), limit sell order submissions (SUS), limit buy order cancellations (CAB), limit sell order cancellations (CAS), buy limit order executions (EXB), and sell limit order executions (EXS). Data are based on one-minute aggregates of NASDAQ ITCH data and one-minute snapshots of NASDAQ ModelView data. The sample period is from November to December 2008, corresponding to $15,600$ one-minute intervals.

	MQ (in \$)	SPR (in \$)	$D10$ (10^3 sh.)	$HD10$ (10^3 sh.)	SUB (10^3 sh.)	SUS (10^3 sh.)	CAB (10^3 sh.)	CAS (10^3 sh.)	EXB (10^3 sh.)	EXS (10^3 sh.)
APC	36.42	0.07	1.41	0.89	22.20	18.99	21.47	21.29	1.63	1.68
AZO	117.14	0.53	0.24	0.51	7.36	4.80	7.18	6.93	0.28	0.31
CAH	33.87	0.05	2.24	0.40	8.08	7.58	8.43	8.41	0.60	0.62
EMR	33.08	0.04	3.41	0.45	18.81	19.50	18.23	19.10	1.82	1.54
GAS	37.84	0.24	1.11	0.26	3.94	4.01	4.05	4.27	0.14	0.13
GOOG	301.67	0.30	0.65	1.48	11.04	9.19	9.07	10.19	2.11	2.11
LEG	14.75	0.03	9.28	0.26	11.48	11.45	11.48	12.11	0.55	0.54
PAYX	26.31	0.02	9.88	0.28	31.51	26.39	29.50	30.02	2.09	2.10
STJ	31.52	0.08	2.23	0.27	13.40	12.36	13.44	13.43	0.87	0.85
TDC	13.87	0.03	6.06	0.45	4.96	4.80	5.23	4.97	0.25	0.26
Average	64.65	0.14	3.65	0.53	13.28	11.91	12.81	13.07	1.03	1.01

Table 4: Summary statistics (mean, standard deviation and 10%, 25%, 75% and 90% quantiles) of mid-quotes, bid-ask spreads, visible and hidden depth and order activities on one-minute aggregates for the stock AZO. Reported variables: one-minute mid-quotes (MQ), minute-by-minute snapshots of bid-ask spreads (SPR), visible depth on the first 10 levels of the book ($D10$), total hidden depth on the first 10 levels ($HD10$), and of total depth and displayed depth imbalances, defined as standing buy volume in excess of sell volume ($DI10$ and $HI10$). Moreover, we report statistics of the minute-by-minute aggregated volume of limit buy order submissions (SUB) and limit sell order submissions (SUS), minute-by-minute aggregated volume of buy limit order cancellations (CAB) and sell limit order cancellations (CAS), and minute-by-minute aggregated volume of buy limit order executions (EXB) and sell limit order executions (EXS). Order flow minute-by-minute aggregation is based on NASDAQ ITCH data and one-minute snapshots are based on NASDAQ ModelView data. Sample period November to December 2008 corresponding to $15,600$ one-minute intervals.

Variable	Mean	St. Dev.	$q10$	$q25$	$q75$	$q90$
MQ (in $)	117.14	12.92	100.27	106.48	129.02	132.45
SPR (in $)	0.53	3.01	0.10	0.14	0.27	0.38
$DI10$ (in $)	−0.02	0.16	−0.23	−0.12	0.08	0.18
$HI10$ (in $)	0.01	0.60	−0.46	−0.16	0.19	0.49
$D10$ (in 1000 sh.)	0.24	0.16	0.04	0.11	0.33	0.46
$HD10$ (in 1000 sh.)	0.51	0.60	0.03	0.15	0.67	1.14
SUB (in 1000 sh.)	7.36	23.56	0.70	1.82	8.20	15.33
SUS (in 1000 sh.)	4.80	14.77	0.00	0.00	5.90	11.47
CAB (in 1000 sh.)	7.18	23.40	0.60	1.70	7.90	14.90
CAS (in 1000 sh.)	6.93	16.54	0.62	1.77	8.00	14.56
EXB (in 1000 sh.)	0.28	0.53	0.00	0.00	0.36	0.76
EXS (in 1000 sh.)	0.31	0.53	0.00	0.00	0.40	0.80

Figure 3: Estimated average unconditional autocorrelation functions of buy- and sell-side order flow volume variables (EXB, EXS, CAB,CAS, SUB, SUS) in quantities that refer to the state of the order book, including the spread ($SPRD$), the hidden and displayed order imbalances ($HI10$, $DI10$), the total order depth ($TD10 = H10 + D10$), the midpoint return RET, and return volatility $VOLA$. $VOLA$ is computed as the sum of squared 1-min returns over a 10-min window. Accordingly, the ACF of $VOLA$ is computed based on 10-minute intervals. Autocorrelation estimates for the other variables are based on one-minute aggregates over snapshots. In particular, order flow volumes are aggregated on a minute-by-minute basis, while order-book quantities originate from one-minute snapshots of the order book. Based on one-minute aggregates of NASDAQ ITCH data and one-minute snapshots of NASDAQ ModelView data for the stocks APC, AZO, CAH, GAS, GOOG, EMR, LEG, PAYX, STJ, and TDC, November to December 2008.



References

AITKEN, M. J., H. BERKMAN, AND D. MAK (2001): "The Use of Undisclosed Limit Orders on the Australian Stock Exchange," *Journal of Banking & Finance*, 25, 1589–1603. 2

BESSEMBINDER, H., M. PANAYIDES, AND K. VENKATAMARAN (2009): "Hidden Liquidity: An Analysis of Order Exposure Strategies in Electronic Stock Markets," *Journal of Financial Economics*, 94, 361–383. 2, 3

BIAIS, B. (1993): "Price Formation and Equilibrium Liquidity in Fragmented and Centralized Markets," *Journal of Finance*, 48, 157–184. 5

BIAIS, B., D. MARTIMORT, AND J.-C. ROCHET (2000): "Competing mechanisms in a common value environment," *Econometrica*, 68, 799–837. 5

BOULATOV, A. AND T. J. GEORGE (2013): "Hidden and displayed liquidity in securities markets with informed liquidity providers," *Review of Financial Studies*, hhs123. 5

BUTI, S. AND B. RINDI (2013): "Undisclosed Orders and Optimal Submission Strategies in a Limit Order Market," *Journal of Financial Economics*, 109, 797 – 812. 5

CHAKRAVARTY, S. AND C. W. HOLDEN (1995): "An integrated model of market and limit orders," *Journal of Financial Intermediation*, 4, 213–241. 5

FOUCAULT, T. (1999): "Order Flow Composition and Trading Costs in a Dynamic Limit Order Market," *Journal of Financial Markets*, 2, 99 – 134. 5

FOUCAULT, T., O. KADAN, AND E. KANDEL (2005): "Limit Order Book as a Market for Liquidity," *Review of Financial Studies*, 18, 1171–1217. 3, 5, 16

FREY, S. AND P. SANDAS (2009): "The Impact of Iceberg Orders in Limit Order Books," Working paper, Center for Financial Research. 2

GLOSTEN, L. R. (1994): "Is the Electronic Open Limit Order Book Inevitable?" *The Journal of Finance*, 49, 1127 – 1161. 5

GLOSTEN, L. R. AND P. R. MILGROM (1985): "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 14, 71 – 100. 5

GOETTLER, R. L., C. A. PARLOUR, AND U. RAJAN (2005): "Equilibrium in a Dynamic Limit Order Market," *The Journal of Finance*, 60, 2149–2192. 5

GROSSMANN, S. (1992): "The Informational Role of Upstairs and Downstairs Markets," *Journal of Business*, 65, 509–529. 3

HALL, A. D. AND N. HAUTSCH (2006): "Order Aggressiveness and Order Book Dynamics," *Empirical Economics*, 30, 973–1005. 5

HARRIS, L. (1996): *Does a Large Minimum Price Variation Encourage Order Exposure?*, vol. 96, New York Stock Exchange. 27, 31

——— (1997): "Order Exposure and Parasitic Traders," Working paper series, University of Southern California. 3, 12, 23

HASBROUCK, J. (1991): "Measuring the Information Content of Stock Trades," *The Journal of Finance*, 46, 179–207. 9

HASBROUCK, J. AND G. SAAR (2009): "Technology and Liquidity Provision: The Blurring of Traditional Definitions," *Journal of Financial Markets*, 12, 143 – 172. 3

HAUTSCH, N. AND R. HUANG (2012): "The Market Impact of a Limit Order," *Journal of Economic Dynamics and Control*, 36, 501–522. 7, 9

HENDERSHOTT, T. AND H. MENDELSON (2002): "Crossing Networks and Dealer Markets: Competition and Performance," *The Journal of Finance*, 55, 2071–2115. 5

JORDA, O. (2005): "Estimation and Inference of Impulse Responses by Local Projections," *The American Economic Review*, 95, 161–182. 11

KYLE, R. (1985): "Continous Auctions and Insider-Trading," *Econometrica*, 53, 1315–1335. 5

LÜTKEPOHL, H. (2007): *New Introduction to Multiple Time Series Analysis*, Springer. 38, 39

MOINAS, S. (2010): "Hidden Limit Orders and Liquidity in Order Driven Markets," IDEI Working Papers 600, Institut d'onomie Industrielle (IDEI), Toulouse. 5

OBIZHAEVA, A. AND J. WANG (2013): "Optimal Trading Strategy and Supply/Demand Dynamics," *Journal of Financial Markets*, 16, 1–32. 9

PARLOUR, C. A. (1998): "Price Dynamics in Limit Order Markets," *Review of Financial Studies*, 11, 789–816. 5

PARLOUR, C. A. AND D. J. SEPPI (2003): "Liquidity Based Competition for Order Flow," *Review of Financial Studies*, 16, 301–343. 5

PESARAN, H. H. AND Y. SHIN (1998): "Generalized Impulse Response Analysis in Linear Multivariate Models," *Economics Letters*, 17–29. 10

RANALDO, A. (2004): "Order Aggressiveness in Limit Order Book Markets," *Journal of Financial Markets*, 7, 53 – 74. 5

ROCK, K. (1996): "The Specialist's Order Book and Price Anomalies," *Review of Financial Studies*, 9, 1–20. 5

ROSU, I. (2009): "A Dynamic Model of the Limit Order Book," *Review of Financial Studies*, 22, 4601–4641. 5

SEPPI, D. J. (1997): "Liquidity Provision with Limit Orders and a Strategic Specialist," *Review of Financial Studies*, 10, 103–150. 5