SFB 649 Discussion Paper 2012-023

Hidden Liquidity: Determinants and Impact

Gökhan Cebiroglu*
Ulrich Horst*

* Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

SFB 649 ECONOMIC RISK BERLIN

Hidden Liquidity: Determinants and Impact *

Gökhan Cebiroğlu
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin
goekhan.cebiroglu@math.hu-berlin.de

Ulrich Horst
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin
horst@math.hu-berlin.de

February 29, 2012

Abstract

We cross-sectionally analyze the presence of aggregated hidden depth and trade volume in the S&P 500 and identify its key determinants. We find that the spread is the main predictor for a stock's hidden dimension, both in terms of traded and posted liquidity. Our findings moreover suggest that large hidden orders are associated with larger transaction costs, higher price impact and increased volatility. In particular, as large hidden orders fail to attract (latent) liquidity to the market, hidden liquidity provision gives rise to negative liquidity externalities.

JEL classification: G10;G11;G12;G14;G24

Keywords: Hidden Liquidity; Pretrade Transparency, Iceberg Orders, Informed Trading, Market Impact; Market Quality, Liquidity Externalities; Upstairs Markets, Trade Negotiation

*We thank Emmanuel Bacry, Mark DiBattista, Boris Drovetsky, Nikolas Hautsch, Ruihong Huang and Daniel Nehren and Matheus Rosenbaum for their support, helpful comments and suggestions. We'd also like to thank the participants and attendees of the 6th World Congress of the Bachelier Finance Society at the Fields Institute (2010), the Princeton-Humboldt Finance Workshop at Princeton University (2009), the Conference on Modeling and Managing Financial Risks at Ecole Polytechnique in Paris (2011) and seminar participants at various institutions for their helpful discussions and comments. Financial support and data provision from Deutsche Bank is gratefully acknowledged. Horst gratefully acknowledges financial support through the SFB 649 Economic Risk.

1 Introduction

Hidden liquidity is an indispensable feature of today's electronic exchanges. Recent studies indicate that hidden liquidity comprises a substantial and growing proportion of overall liquidity. Understanding the motifs of hidden order submission, the determinants of hidden liquidity, as well as its impact on market dynamics is of growing importance to investors, regulators, policy-makers, and exchanges alike. We identify which observable stock characteristics most strongly relate to hidden liquidity provision in order-driven markets, and analyze the impact of hidden liquidity on different dimensions of the market. Our results suggest a negative liquidity externality associated with large hidden orders: unlike large displayed orders they fail to attract additional (latent) liquidity. We also find that unexecuted hidden orders are canceled and resubmitted as a series of market orders thereby increasing post-submission returns. As a result, large hidden orders often carry information about future price movements and tend to increase realized volatility.

1.1 Hidden Liquidity in Modern Exchanges

Most of today's order-driven electronic-exchanges provide hidden liquidity in the form of specific order types. The most prominent one is the so-called "Iceberg Order". An iceberg order is a passive order that has been split into smaller parts of which just a small proportion is visible to the public (if at all). Hiding the actual order quantity reduces information leakage as well as price movements and order flows caused by substantial changes in a stock's supply (*market impact*). Order-splitting is a standard trade strategy of institutional investors (Foster and Vishwanathan, 1990; Keim and Madhavan, 1995), often automated through order/execution management systems.

A series of empirical studies confirms the growing and substantial use of hidden orders among most major stock exchanges. For instance, Pascual Gasco and Veredas (2008) report that 26% of all trades on the Spanish Stock Exchanges involve hidden volume. Frey and Sandas (2009) find that 9.3% of submitted and 15.9% of executed shares contain Iceberg orders on the German Xetra Stock Exchange. De Winne and D'Hondt (2004; 2007; 2009) report that 27.2% (20.4) of the total liquidity in the book is hidden for the French CAC40 (Belgian BEL20) exchanges and moreover that the hidden ratios can even reach 50% at the best limit prices. Tuttle (2003) finds that around 25% of liquidity for all NASDAQ National Market quotes are hidden. Further studies confirm that hidden liquidity is particularly prevalent among large investors: D'Hondt et al. (2004) report that 81% of orders with total sizes in the largest quartile are Iceberg orders or (partly) hidden orders. Further supplementing these findings, Frey and Sandas (2009) find that Iceberg orders are on average 12-20 times larger than limit orders. Bessembinder et al. (2009); Aitken et al. (2001); De Winne and D'Hondt (2007) report similar findings, among others.

Various sources of risk associated with order-exposure – and hence motives to use hidden liquidity – have been identified in the literature. Copeland and Galai (1983) propose adverse-selection risk as the prime concern behind hidden liquidity origination. Harris (2003) attributes exposure risk to the presence of so-called *parasitic traders*. He argues (p.) that, at the expense of limit order traders, these "rogue" traders exploit the free trading option of limit orders by "front-running" them. In particular, the limit order trader will suffer higher liquidity competition and eventually higher execution-risk.[1] Among more theoretical works, Moinas (2010) suggests that *informed* traders "scare-away" counter-party market order flow by exposing too much of their intentions. Buti and Rindi (2008) study exposure-impact on the same-side (limit) order flow. In contrast to Moinas and more in line with Harris' view, they suggest that *uninformed liquidity traders* have an interest in hiding their position as exposure can increase same-side liquidity competition. Cebiroglu and Horst (2011) model both dimensions simultaneously, same-side liquidity competition as well as opposite-side liquidity demand. Their results suggest that market impact is primarily (though not exclusively) felt through same-side liquidity competition and that at least partially hiding passive orders can substantially decrease trading costs.

Although this and further empirical findings (Bessembinder et al., 2009; Frey and Sandas, 2009) suggest that hidden liquidity provision can be beneficial to individual investors in certain cases, it is still an ongoing debate whether or not it actually benefits market quality in general. For instance, while findings in Aitken et al. (2001),Anand and Weaver (2004),Tuttle (2003) and Frey and Sandas (2009) indicate that hidden liquidity provision attracts additional liquidity to the market, Hendershott and Jones (2005) show that overall market quality in the Island electronic communication network (ECN) deteriorated after responding to a September 2002 regulatory enforcement to stop "displaying its limit order book in the three most active exchange-traded funds (ETFs) where it was the dominant venue". Edwards et al. (2004) and Bessembinder and Maxwell (2008) report similar results for certain bond markets.

1.2 Our Contributions

The contribution of our paper is two-fold. First, we study the impact of hidden liquidity on the stock market. Recent studies on the information content of hidden liquidity are to some degree conflicting. For instance, Pascual and Tornero (2003), studying data from the Spanish Stock Exchange, suggest that hidden order traders trade for liquidity and not for information reasons. On the other hand, Frey and Sandas (2009), using data from the Frankfurt Stock Exchange,

[1] Note also, that the exposure impact is related to the field of studies that analyze how the visible (exposed) portion of the order book affects the market. See Biais et al. (1995); Ranaldo (2004); Harris and Panchapagesan (2005); Bessembinder et al. (2009); Griffiths et al. (2000); Chordia et al. (2002); De Winne and D'Hondt (2007); Cao et al. (2009); Beber and Caglio (2005); Ellul et al. (2003); Pascual Gasco and Veredas (2008); Hautsch and Huang (2009).

report that undetected Iceberg orders carry information about future price movement, while detected Icebergs do not. Finally, Tuttle (2003), analyzing NASDAQ National Market quotes, finds that while displayed depth carries little information, aggregated hidden depth is indicative of price changes even beyond 30 minutes post-trade. Our work further extends this line of research. Using NASDAQ's Modelview data set, we find that aggregated hidden liquidity clusters around few price quotes and happens to enter the market sporadically. Displayed depth on the other hand is more regular and more evenly distributed along both dimensions. For instance, we report that while 80% of total *displayed* depth at the ten best price quotes is concentrated on five prices quotes on average, hidden depth concentrates on average on only 2-3 price quotes. This suggest that the presence of hidden depth is associated with single large orders (traders).

1.2.1 Impact of Hidden Liquidity

We test whether these single spikes of hidden liquidity are related to informed trading. Using an event-study framework, we analyze the exante and expost impact of hidden (and displayed) imbalances on different dimensions of the market, including spread, volatility, depth, and abnormal returns. We find strong evidence that large hidden imbalances (orders) are associated with significant expost returns indeed, whereas displayed liquidity exerts almost no impact on price returns. In comparing previous results on the return impact of Reuters news-items (Gross-Klussmann and Hautsch, 2010), we find that the impact of hidden orders can even exceed the impact of statistically significant earnings announcement news. Moreover, our results suggest that (large) hidden orders are more likely to be submitted by traders who follow a trend, and when volatility is low and spreads are narrow.

The Costs of Hidden Liquidity and Grossmann's Conjecture

A growing number of empirical studies suggests that hidden liquidity is used by liquidity traders (see Aitken et al. (2001); Pascual and Tornero (2003); Gozluklu (2010)). At the same time our study suggests that hidden orders carry substantial information with respect to future prices. To reconcile both facts, we use an analogy to Grossmann's conjecture about markets without *upstairs* negotiation; see Grossmann (1992). He argues that in pure downstairs markets, trading must be costlier as investors cannot signal their interest-to-trade to *latent* liquidity (by this we mean liquidity demand of traders that act primarily on observable liquidity supply). In particular, as *non-discretionary*[2] liquidity demand typically exceeds the displayed liquidity supply in primary markets, attracting latent (discretionary) liquidity is necessary to tap into additional sources of liquidity. However, according to Admati and Pfleiderer (1988), discretionary liquid-

[2]We use non-discretionary in the sense of Admati and Pfleiderer (1988), i.e. traders that do not have discretion over when to trade.

ity traders[3] prefer to trade in "thick" markets, in order to reduce their trading costs. Hence, when investors hide their liquidity demand, discretionary traders will not be attracted to the market. Consequently, faced with a reduced likelihood to execute their hidden orders, these investors will eventually "cross the spread" and use market orders instead, effectively bypassing "unexpressed" liquidity.

We report empirical evidence that large displayed orders do in fact attract latent liquidity, that large hidden orders have a much lesser chance of being executed and that unexecuted hidden orders are eventually cancel and resubmitted as a series of market orders. Large hidden orders therefore tend to incur higher transaction costs and increase post-submission returns. It is in this sense that they carry information about future price movements. This is often accompanied by an increase in realized volatility. The very same stylized facts have previously been verified for pure downstairs markets.[4] Moreover, the fact that large traders use hidden orders, that are less likely to get executed as compared to visible orders, indicates that hidden orders are used by unaware and unsophisticated investors. These investors face higher transaction costs and increase stock price volatility.

Volatility and Hidden Liquidity

The link between hidden liquidity and volatility has recently been studied in the literature, albeit not extensively. The results are inconclusive. For instance, while Aitken et al. (2001) report that volatility increases hidden order proportions (or vice versa), results of Bessembinder et al. (2009) suggest that greater return volatility is associated with a lower likelihood of hidden order usage.[5] Our study indicates that these distinct observations are not necessarily inconsistent but may rather refer to different aspects and dimensions of the interplay between hidden liquidity and volatility. While the first study analyzes the cross-sectional aspect, the latter looks at its intertemporal dimension. Our impact study shows that this discrepancy between the cross-sectional and intertemporal dimension can be resolved when looking at the post- and presubmission impact of hidden liquidity, where two distinct effects are observed. Namely, (large) hidden order submission takes place when volatility is low. However, the postsubmission impact involves a persistent and substantial increase in volatility, indicating that hidden orders overall increase volatility.

[3]Discretionary traders can also be considered as *opportunistic* traders, in that they actively monitor the market for liquidity opportunities. In particular, they have discretion over when to trade. In agency-brokerage firms, this is typically facilitated by the DMA (Direct Market Access) desk.

[4]Bessembinder and Venkatamaran (2004); Smith et al. (2001); Booth et al. (2002).

[5]Indeed theoretical literature indicates that the role of volatility is ambiguous; see Moinas (2010).

1.2.2 Determinants of Hidden Liquidity

Our second contribution is to identify visible stock characteristics that correlate most strongly with hidden liquidity submission. We focus on determinants of hidden liquidity in general, both in terms of traded as well as posted volumes. In particular, we study cross-sectionally how the presence and magnitude of hidden depth relates to observable and readily available stock characteristics. Earlier studies by, e.g., De Winne and D'Hondt (2007); Bessembinder et al. (2009) and more recently Hautsch and Huang (2011), study the predictability of single hidden orders. However, due to data limitations they are not able to detect all hidden orders. Our approach is different. We are trying to understand cross-sectional variations instead of temporal predictability and instead of taking a focus on predicting single orders, we look at the broader picture of cross-sectional, aggregated hidden depth variation. Our work extends the line of literature initiated in De Winne and D'Hondt (2009). They study how a stock's tick size, volatility, market capitalization, and trading activity relate to the posted hidden fraction on the best limit. Their regression analysis reports a significant r^2-goodness-of-fit of 0.292 points and indicates that the ratio of posted hidden liquidity decreases with tick sizes and liquidity and increases with volatility. We extend their approach by incorporating additional stock characteristics, like spread, price, overall depth, inter-trade time interval, trade size and traded volume. Moreover, instead of using depth at the best limit price only, we incorporate the full depth of the best ten price quotes. Our sample of $N = 448$ stocks is larger than their's using only 88 stocks. We not only look at posted liquidity, but also on traded liquidity and estimate both hidden liquidity ratios and volumes separately.

Our explanatory models reach significantly higher r^2-goodness-of-fit, 0.457 (0.515) for the posted hidden volume (ratio) and 0.657 (0.849) for the traded hidden volume (ratio). Using suitable nonlinear transformations, we are able to bring the numbers up to 0.657 (0.695) and to 0.803 (0.939) for traded hidden liquidity. In order to identify the hierarchy of determinants for hidden liquidity, we additionally apply a LARS-procedure. First, we obtain that total hidden volumes are best predicted by liquidity measures (with a less significant fit). Hidden ratios, on the other hand, are governed mostly by the stock's spread. For instance, we obtain 0.92 (0.29) points of r^2 for traded (posted) hidden ratios if we only consider the stock's spread.

1.2.3 Outline

The remainder of this paper is structured as follows. Section 2 describes the data set and presents descriptive statistics on traded and posted hidden liquidity for the S&P 500. Section 3 reports analysis on the determinants of hidden liquidity. In section 4 we use LARS regression to provide a parsimonious model for posted and traded hidden liquidity in the S&P 500. Section 5 presents the impact-study of hidden liquidity. Finally we conclude.

2 Data

2.1 Sample Selection

Our analysis will takes a view at both, traded as well as posted hidden liquidity for all stocks from the S&P 500 during October 2008 and March 2009.We obtain posted hidden liquidity by using the NASDAQ Modelview data set. This data set contains minute-by-minute snapshots of the full aggregated order book depth, including visible and hidden depth for all NASDAQ-, NYSE- and AMEX- listed stocks during the time period. The order book data is presented in aggregated form, that is displayed and hidden volumes are aggregated to their total depths. Our initial sample consists of all stocks that were continuously listed in the S&P 500 index through the whole time period. To reduce the impact of *outliers*, we constrain our analysis to stocks that show an average daily traded volume (ADV) of less then 50 million shares, 0.2 million average number of trades, average spread of less than 25 cents and average price of less then 100$ (trade volumes are extracted from the Trade and Quote Database (TAQ) data set provided by Deutsche Bank). We finally obtain a sample size of $N = 448$ shares. To reduce the impact of opening and closing auctions we constrain our analysis to daily periods between 09:15 and 15:45. Thus, the daily sample size for each stock counts 390 minute-by-minute snapshots. We also consider depth up to the best ten price levels and not beyond.

The Modelview data set provides information about posted (displayed and hidden) depth only. To study traded/executed hidden liquidity we use TAQ data. However, it does not provide additional information as to whether the traded volume was hidden. We therefore approximate the traded hidden volumes by identifying them with trades that executed within the spread, as every order executed within the spread, by definition, must have been hidden. We point out that this assumption crucially depends on the reporting mechanism and the latency of reporting. However, assuming that latency is far lower than the average inter-trade-arrival time and that best bid and ask prices are adjusted after each trade immediately, it is reasonable to assume that hidden order can only be executed within the spread.

Besides extracting the unobservable hidden volumes from the two data sets, we further consider the stock's main observable characteristic statistics: The average daily traded volume (ADV), the average inter-trade time interval ($Time$), average volatility ($Vola$), average trade size ($TrSize$), average spread ($Spread$), average price ($Price$), and average top of the book depth (Top).

2.2 Summary Statistics

Based on the ADV we sort stocks into liquidity quintiles q_1, q_2, q_3, q_4, q_5, starting with the least liquid quintile q_1 and q_5 representing the most liquid one. Table 1 reports cross-sectional sample statistics on posted and displayed liquidity as taken from the NASDAQ Modelview data set.

Liquidity Quantile	Mean (in 1000)					Median (in 1000)					Max (in 1000)					Min (in 1000)					SD (in 1000)				
	H_B	H_S	H	D	$Total$	H_B	H_S	H	d	$Total$	H_B	H_S	H	D	$Total$	H_B	H_S	H	D	$Total$	H_B	H_S	H	D	$Total$
least liquid	0.78	0.11	0.89	4.2	5.1	0.2	0.01	0.28	3.7	4.3	39	15	41	29	53	0	0	0	0.01	0.02	2.14	0.49	2.25	2.65	3.8
q_2	1.24	0.12	1.36	6.5	7.8	0.45	0.01	0.52	5.9	6.9	60	25	63	43	79	0	0	0	0.03	0.04	3.2	0.68	3.4	3.7	5.3
q_3	1.8	0.11	1.9	12	14	0.59	0.01	0.65	11	12	83	27	85	73	120	0	0	0	0.09	0.12	4.8	0.67	4.9	6.4	8.8
q_4	2.29	0.12	2.41	14	17	0.55	0.01	0.62	14	15	143	33	145	84	173	0	0	0	0.04	0.07	7	0.82	7.1	7.7	11
most liquid	5.8	0.1	5.8	45	51	1.53	0	1.61	43	47	229	50	231	273	367	0	0	0	0.33	0.5	15	1.03	15	24	30
total	2.37	0.11	2.48	16	19	0.66	0.01	0.74	15	17	111	30	113	100	159	0	0	0	0.1	0.15	6.3	0.74	6.5	8.8	12

Table 1: Sample descriptive statistics sorted by liquidity quintile. The table reports the cross-sectional mean, maximum (Max), minimum (Min) and standard deviation (SD) for the posted hidden volume without the spread, i.e. "deep" in the book (H_B), in the spread (H_S), the total posted hidden liquidity (H), the total posted displayed liquidity (D) and the total posted liquidity ($Total$). Numbers reported in shares.

Generally, more liquid stocks report more volume no matter whether displayed or hidden. Note that for more liquid stocks the hidden volumes posted in the spread decrease. This is due to the fact that liquid stocks typically trade at around one tick of spread. For instance Microsoft has an average spread of 1.06 cent, i.e. in most cases the spread trades at one cent. Therefore, opportunities to post hidden liquidity within the spreads are few.

Based on liquidity quintiles Table 2 reports cross-sectional averages of ADV, $Time$, $Vola$, $TrSize$, $Spread$, $Price$ and Top and the corresponding posted and traded hidden liquidity volumes by ratio and total volume. $Ratio$ denotes the hidden proportion of the total liquidity. Generally, more liquid stocks show higher ADV and Top and smaller $Time$, $Spread$, $Price$ and $TrSize$. $Vola$ and $TrSize$ exhibit a weakly pronounced increase for more liquid stocks. Generally, less liquid stocks trade more hidden. The ratio of executed hidden liquidity drops from 26 %, for the least liquid quintile, to 7 %, for the most liquid quintile. The decrease is weaker but nevertheless significant for posted liquidity dropping respectively from 19 % to 13 %. All in all, hidden liquidity accounts for a substantial fraction of trading activity at 16 % (17 %) of posted (traded) liquidity. The findings are in line with earlier studies, for instance see Bessembinder et al. (2009). Table 3 reports cross-correlations between the observable stock characteristics. Our results suggest that more liquid stocks are characterized by narrower spreads, smaller inter-trade time, higher traded volume, larger trade sizes, larger top-of-book depth and smaller prices. $Spread$ and $Price$ as well as $TrSize$ and Top show strongest correlation. The substantial $Spread - Price$ correlation is consistent with tight competition among market-makers. This can bee seen as follows. Assume that the spread would be the same or larger for lower-priced stocks. As it is reasonable to assume that the market maker's average per-share-profit is a monotonically increasing function of $\frac{Spread}{Price}$, in competitive markets market makers would immediately switch to lower priced stocks, therefore increasing the competition

Liquidity Quantile	Average Stock Properties							Posted Hidden Liquidity		Traded Hidden Liquidity	
	ADV	$Time$	$Vola$	$TrSize$	$Spread$	$Price$	Top	$Volume$	$Ratio$	$Volume$	$Ratio$
least liquid	1.39	2.65	0.07	147	4.91	36.46	308	656	0.19	0.37	0.26
q_2	2.72	1.38	0.08	158	3.39	32.84	576	1318	0.2	0.57	0.2
q_3	4.23	0.94	0.09	165	2.4	27.41	800	1671	0.17	0.69	0.15
q_4	7.13	0.61	0.1	178	1.87	24.59	1278	2292	0.16	0.83	0.11
most liquid	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

Table 2: Cross-sectional averages for the ADV, $Time$, $Vola$, $TrSize$, $Spread$, $Price$, Top and the hidden $Volume$ and $Ratio$ for traded and posted liquidity based on liquidity quintiles. ADV and traded hidden $volume$ is given in millions of shares. Top, $TrSize$ and $volume$ of posted hidden depth are given in single shares. $Time$ is reported in seconds, $Price$ is reported in dollar and $Spread$ is reported in ticks (i.e. cents).

and ultimately resulting in lower spreads. Using this kind of Bertrand-competition argument, we expect that in equilibrium average spread and average price are strongly correlated. Moreover, $TrSize$ and Top correlate, since more posted liquidity is going to attract more liquidity demand and vice versa. Notice that $Vola$ and $Spread$ show least correlation.

	ADV	$Time$	$Vola$	$TrSize$	$Spread$	$Price$	Top
ADV	$*$	−0.570	0.248	0.584	−0.382	−0.213	0.529
$Time$		$*$	−0.267	−0.298	0.529	0.176	−0.218
$Vola$			$*$	0.463	−0.146	−0.403	0.287
$TrSize$				$*$	−0.351	−0.490	0.941
$Spread$					$*$	0.698	−0.285
$Price$						$*$	−0.42
Top							$*$

Table 3: Reporting cross correlation-coefficients for the mean average daily traded volume (ADV), volatility ($Vola$), the spread ($Spread$), price ($Price$) and the top-of-book depth (Top).

3 Determinants of Hidden Liquidity

In this section, we examine how the observable stock characteristics (i.e. ADV, $Time$, $Vola$, $TrSize$, $Spread$, $Price$, and Top) relate to the stock's hidden properties, precisely the mean posted and traded hidden volumes (ratios), i.e. H_{ps}, H_{tr} (H^R_{ps}, H^R_{tr}). Our main purpose is to identify the key determinants of the stock's hidden dimension. Our choice of the explanatory

variables reflects the holistic approach by covering the observable, standard mean order book properties (i.e. $Spread$, $Price$, Top), liquidity quantities ($Time$, $TrSize$, ADV) as well as its volatility ($Vola$). We provide two models, a standard (linear) model and to account for possible nonlinear relationships, a partly log-transformed model. The respective standard models read

$$\begin{aligned}
H_{ps} &= \alpha + \alpha_A ADV + \alpha_{Ti} Time + \alpha_V Vola + \alpha_{Tr} TrSize \\
&\quad + \alpha_S Spread + \alpha_P Price + \alpha_T Top + \epsilon_1, \\
H^R_{ps} &= \alpha^R + \alpha^R_A ADV + \alpha^R_{Ti} Time + \alpha^R_V Vola + \alpha^R_{Tr} TrSize \\
&\quad + \alpha^R_S Spread + \alpha^R_P Price + \alpha^R_{To} Top + \epsilon_2, \\
H_{tr} &= \beta + \beta_A ADV + \beta_{Ti} Time + \beta_V Vola + \beta_{Tr} TrSize \\
&\quad + \beta_S Spread + \beta_P Price + \beta_{To} Top + \epsilon_3, \\
H^R_{tr} &= \beta^R + \beta^R_A ADV + \beta^R_{Ti} Time + \beta^R_V Vola + \beta^R_{Tr} TrSize \\
&\quad + \beta^R_S Spread + \beta^R_P Price + \beta^R_{To} Top + \epsilon_4.
\end{aligned} \tag{3.1}$$

For the transformed model, we simultaneously apply log-transforms on selected predictor variables (mainly the $Spread$) as well as on some response variables (i.e. H_{ps}, H^R_{ps} and H_{tr}).

$$\begin{aligned}
\log(H_{ps}) &= a + a_A ADV + a_{Ti} Time + a_V Vola + a_{Tr} TrSize \\
&\quad + a_S Spread + a_P Price + a_{To} Top + \epsilon_5, \\
\log(H^R_{ps}) &= a^R + a^R_A ADV + a^R_{Ti} Time + a^r_V Vola + a^R_{Tr} TrSize \\
&\quad + a^R_S \log(Spread) + a^R_P Price + a^R_{To} Top + \epsilon_6, \\
\log(H_{tr}) &= b + b_A ADV + b_{Ti} Time + b_V Vola + b_{Tr} TrSize \\
&\quad + b_S \log(Spread) + b_P Price + b_{To} Top + \epsilon_7, \\
H^R_{tr} &= b^R + b^R_A ADV + b^R_{Ti} Time + b^R_V Vola + b^R_{Tr} TrSize \\
&\quad + b^R_S \log(Spread) + b^R_P Price + b^R_{To} Top + \epsilon_8,
\end{aligned} \tag{3.2}$$

We apply standard assumptions for the models (3.1) and (3.2). In particular, we assume that the error terms $(\epsilon_i)_{i=1,2,...,8}$ $(i = 1, 2, ..., 8)$ are iid and normally distributed.

3.1 Estimation Results

Our results can be summarized in three points. First, according to table 4, both, the linear models as in (3.1) as well as the transformed ansatz according to (3.2) show a significant goodness-of-fit in terms of adjusted-r^2 as well as f-statistics. For all models, we report r^2-values in a range starting from 0.457 points to up to substantial 0.939 points, depending on whether hidden ratios or volumes and whether linear or transformed models are considered. In particular, the model significantly improve earlier models[6]. Moreover, with 440 degrees of

[6] De Winne and D'Hondt (2007) report r^2 of 0.295 for the correct hidden fraction at the top of the book.

Variable	Posted Hidden Liquidity				Traded Hidden Liquidity			
	Volume		Ratio		Volume		Ratio	
	standard	transformed	standard	transformed	standard	transformed	standard	transformed
$Intercept$ ($t-statistic$)	$-2.51e+03$ (-1.1)	$6.06e+00$ (10.7)	$-7.37e-01$ (-3.17)	$1.62e+00$ (3.2)	$-2.89e+06$ (-1.56)	$1.63e+01$ (48.37)	$2.23e-02$ (1.06)	$6.22e-01$ (24.6)
ADV ($t-statistic$)	$5.95e-05$ (1.41)	$1.92e-08$ (3.34)	$-1.18e-08$ (-2.75)	$-4.43e-09$ (-0.87)	$3.59e-01$ (10.49)	$4.02e-08$ (11.77)	$-2.65e-09$ (-6.82)	$-9.86e-10$ (-3.84)
$Time$ ($t-statistic$)	$-3.78e+02$ (-1.19)	$-2.38e-01$ (-5.47)	$-1.98e-01$ (-6.09)	$-2.68e-01$ (-6.93)	$-7.08e+05$ (-2.75)	$-4.18e-01$ (-16.17)	$2.42e-02$ (8.23)	$4.79e-03$ (2.47)
$Vola$ ($t-statistic$)	$-1.92e+04$ (-2.59)	$-2.00e+00$ (-2)	$1.06e+00$ (1.41)	$1.53e+00$ (1.72)	$5.60e+07$ (9.33)	$2.95e+00$ (4.95)	$2.91e-01$ (4.26)	$-5.41e-02$ (-1.21)
$TrSize$ ($t-statistic$)	$3.33e+01$ (2.12)	$7.68e-03$ (3.7)	$5.04e-03$ (3.15)	$1.02e-02$ (5.52)	$-8.57e+03$ (-0.67)	$2.62e-03$ (2.12)	$-5.56e-05$ (-0.38)	$1.59e-04$ (1.72)
$Spread$ ($t-statistic$)	$2.66e+04$ (1.58)	$-9.88e-02$ (-0.98)	$1.65e+01$ (9.62)	$1.13e+00$ (12.54)	$-7.96e+07$ (-5.84)	$3.97e-01$ (6.62)	$1.31e+00$ (8.43)	$1.29e-01$ (28.62)
$Price$ ($t-statistic$)	$-2.79e+01$ (-1.36)	$-9.38e-03$ (-2.93)	$9.39e-03$ (4.48)	$1.11e-02$ (3.89)	$2.59e+05$ (15.57)	$1.39e-02$ (7.31)	$2.61e-03$ (13.79)	$4.06e-04$ (2.84)
Top ($t-statistic$)	$7.85e+01$ (3.07)	$2.52e-03$ (0.75)	$-7.14e-03$ (-2.73)	$-1.53e-02$ (-5.13)	$-5.84e+04$ (-2.81)	$-1.72e-02$ (-8.63)	$-2.72e-04$ (-1.15)	$-5.01e-04$ (-3.34)
$f-statistic$ ($Gain$)	55 ($*$)	124 (+126%)	69 ($*$)	147 (+113%)	124 ($*$)	262 (+112%)	360 ($*$)	978 (+172%)
r^2 ($Gain$)	0.457 ($*$)	0.657 (+44%)	0.515 ($*$)	0.695 (+35%)	0.657 ($*$)	0.803 (+22%)	0.849 ($*$)	0.939 (+11%)

Table 4: Parameter estimates of the cross-sectional regression models for traded and posted hidden liquidity volume/ratio according to the *standard* linear models in (3.1) and the *transformed* nonlinear models in (3.2). T-statistics are reported below the parameter estimates in round brackets. We finally report the *adjusted* r^2, f-statistic and their corresponding gains in percentages of the standard models (3.1).

freedom, the f-statistics further substantiate model significance beyond the 0.01% significance level.

Secondly, considering hidden ratios instead of volumes[7] and applying the transformation rules according to (3.2) significantly enhances the predictability of hidden liquidity. For instance, for posted hidden liquidity, the r^2 value increases from 0.495 to 0.695 points, while for traded hidden liquidity, r^2 improves from 0.657 to significant 0.939 points. Similarly, we obtain significant improvement in model significance according to f-statistics. F-statistics improve from 55 to 147 points for posted and 124 to 978 points for traded hidden liquidity.

Thirdly, generally, the spread is a highly influential determinant of hidden liquidity. In fact, we observe that hidden ratios are mainly affected by the spread. For instance, for the transformed models of traded and posted hidden ratios, the spread by far dominates the other determinants in terms of t-statistics, with t-values of 28.62 and 12.54 respectively. Hidden volumes on the

[7]Notice, the fact that hidden ratios are better predicted than volumes is of importance, as one can easily recover the actual hidden volumes from the knowledge of the hidden ratios and the easily accessible displayed depth.

other hand, tend to be more susceptible towards liquidity quantities, i.e. inter-trade time, trade size and ADV .

Finally, our results indicate that stocks trade more hidden, in volume as well as in proportion, when spreads are large. Price and Top relate to hidden liquidity according to how they correlate to the spread 3, i.e. generally, hidden liquidity usage increases for higher-priced stocks and decreases with Top . Liquidity quantities affect hidden liquidity differently. In terms of ADV , the more liquid a stock is, the less is the proportion of traded and posted hidden liquidity, although it increases in absolute volumes. On the other hand, the usage of hidden liquidity tends to increases with the trade size and with smaller inter-trade times. The impact of volatility is less substantial is less conclusive.

3.2 A Minimal Model For Hidden Liquidity

The results attest that the order book's visible dimension can have significant explanatory power over its hidden properties. Our findings indicate that there is a hierarchy among the predictor variables. In particular, we found that hidden volumes are better explained by liquidity quantities while normalized volumes, i.e. hidden ratios, are mainly explained by the spread. In this section, we make use of a framework, the so called *Least Angle Regression* (LARS), to adequately capture this hierarchy. Our aim is to identify a *minimal* model for hidden liquidity with the most informative predictors, ranked them according to their explanatory power. Moreover, as the set of determinants partly exhibits strong cross-correlations according to table 3, LARS may help to adequately account for effects of *spurious regression*.

3.2.1 LARS Methodology

Typically, model selection algorithms such as *Lars*, *Lasso*, *All Subsets*, *Forward* and *Backward Elimination* are used and designed to reduce the number of covariates and to identify an efficient and parsimonious set of predictor variables.[8] According to Efron et al. (2004), " [the Lasso and Forwards Stagewise Regression] are variants of a basic procedure called Least Angle Regression. ... [Lars] can be viewed as moderately greedy forward stepwise procedure whose forward progress is determined by compromising among the currently most correlated covariates. LARS moves along the most obvious compromise direction, the *equiangular* vector, while the Lasso and Stagewise procedure put some restrictions on the equiangular strategy".

We briefly scratch the Lars methodology as explained in Efron et al. (2004) using their notation. Lars regression consists of multiple steps. At each step the algorithm builds up successive models and estimates $\widehat{\mu} = X\,\widehat{\beta}$, so that after k steps the model comprises only k parameters that are

[8] This is particularly useful in high-dimensional statistics as simpler models enhance the scientific insights for models with high degrees of freedom. See Gelper and Croux (2008) for an application in time series forecasting. Lars regression is related to the aforementioned model selection algorithms.

nonzero. More precisely, assume we have m linearly independent covariates $x_1, x_2, ..., x_m$. And denote by A some subset of the index set $\{1, 2, ..., m\}$ with cardinality $|\mathcal{A}| = a$ and denote $\mathbf{1}_{\mathcal{A}}$ the vector of all ones with length equaling a. Then Efron et al. define the following matrices

$$X_{\mathcal{A}} = (\cdots s_j X_j \cdots)_{j \in \mathcal{A}}, \qquad \mathcal{G}_{\mathcal{A}} = X_{\mathcal{A}}{}' X_{\mathcal{A}}, \tag{3.3}$$

$$A_{\mathcal{A}} = (\mathbf{1}_{\mathcal{A}}{}' \mathcal{G}^{-1} \mathbf{1}_{\mathcal{A}})^{-\frac{1}{2}}, \qquad w_{\mathcal{A}} = A_{\mathcal{A}} \mathcal{G}_{\mathcal{A}}^{-1} \mathbf{1}_{\mathcal{A}}, \tag{3.4}$$

where $s_j = \pm 1$ and $w_{\mathcal{A}}$ is the unit vector making *equal angles*, less than 90 degrees, with the columns of $X_{\mathcal{A}}$. The so called *equiangular vector* $\mathbf{u}_{\mathcal{A}}$ then reads as follows:

$$\mathbf{u}_{\mathcal{A}} = X_{\mathcal{A}} w_{\mathcal{A}} \qquad \text{with} \qquad ||\mathbf{u}_{\mathcal{A}}||^2 = 1. \tag{3.5}$$

The LARS algorithm can now be described as follows. Starting with the estimate $\widehat{\mu}_0 = 0$ one successively builds up $\widehat{\mu}$ in steps. Therefore, assume $\widehat{\mu}_{\mathcal{A}}$ to be the current LARS estimate. Then the current correlation reads

$$\widehat{\mathbf{c}} = X' (y - \widehat{\mu}_A), \tag{3.6}$$

where we define the *active* set $\mathcal{A}$ to be the set of indices corresponding to the covariates with the greatest absolute current correlations, i.e.

$$\widehat{C} = \max_j \{|\widehat{c}_j|\} \qquad \text{and} \qquad \mathcal{A} = \{j : |\widehat{c}_j| = \widehat{C}\}. \tag{3.7}$$

Now we define

$$s_j = sign\{\widehat{c}_j\} \qquad j \in \mathcal{A} \tag{3.8}$$

and we again compute $X_{\mathcal{A}}, A_{\mathcal{A}}$ and $\mathbf{u}_{\mathcal{A}}$ as in (3.3)-(3.5). Finally, the updated estimate $\widehat{\mu}_{\mathcal{A}+}$ reads

$$\widehat{\mu}_{\mathcal{A}+} = \widehat{\mu}_{\mathcal{A}} + \widehat{\gamma}\, \mathbf{u}_{\mathcal{A}}, \qquad \text{with} \qquad \widehat{\gamma} = \min_{j \in \mathcal{A}^c}{}^{+} \left\{ \frac{\widehat{C} - \widehat{c}_j}{A_{\mathcal{A}} - a_j}, \frac{\widehat{C}\widehat{c}_j}{A_{\mathcal{A}} + a_j} \right\}, \tag{3.9}$$

where $\min^+$ indicates that the minimum is taken only over positive components for each choice of j in (3.9). Once can easily show that the maximum absolute correlation declines with each step. In other words, the followings holds:

$$\widehat{C}_+ = \widehat{C} - \widehat{\gamma}\, A_{\mathcal{A}}. \tag{3.10}$$

3.2.2 The Akaike Information Criterion

The Lars procedure as outlined in the previous subsection, provides k model estimates for $\widehat{\mu}$ in k steps. However, one wants to know and choose only the *best* of these models. By *best*, one generally means the model that most effectively balances *goodness-of-fit* and model-parsimony. Let $\mathbf{y}$ denote some dependent variable we want to explain and

$$\mathbf{y} \propto (\mu, \sigma^2 \mathbf{I}), \tag{3.11}$$

indicating that the y_i are uncorrelated, with mean μ_i and variance σ^2. Then one can write

$$(\widehat{\mu}_i - \mu_i)^2 = (y_i - \widehat{\mu}_i)^2 - (y_i - \mu_i)^2 + 2(\widehat{\mu}_i - \mu_i)(y_i - \mu_i). \tag{3.12}$$

Summing over i and taking expectation yields

$$E\left[\frac{||\widehat{\mu} - \mu||^2}{\sigma^2}\right] = E\left[\frac{||\mathbf{y} - \widehat{\mu}||^2}{\sigma^2} - n\right] + 2\sum_{i=1}^{n} \frac{cov(\widehat{\mu}_i, y_i)}{\sigma^2}, \tag{3.13}$$

where the last term is defined as the model's *degree of freedom*, i.e.

$$df = \sum_{i=1}^{n} \frac{cov(\widehat{\mu}_i, y_i)}{\sigma^2}. \tag{3.14}$$

Now the Akaike information criterion reads

$$C_p = \frac{||\mathbf{y} - \widehat{\mu}||^2}{\sigma^2} - n + df. \tag{3.15}$$

When σ^2 and df are known, C_p is an unbiased estimator of the *true* risk $E\left[\frac{||\widehat{\mu} - \mu||^2}{\sigma^2}\right]$. In order to select the *best* model among a set of models, one chooses the one with the lowest C_p value. This model should represent the best trade-off between bias ("accuracy") and variance ("complexity").

3.2.3 LARS Results

Table 5 and 6 report results of the LARS procedure for posted and traded hidden liquidity according to the models (3.1) and (3.2). We report for each lars-step the selected variable (*action*), its resulting r^2-goodness-of-fit (r^2_{lars}) and its ratio with the adjusted-r^2 of the respective full-models (r^2_{ols}). Besides r^2-estimates, we also report estimates for C_p.

The LARS results generally confirm the hierarchy among predictor variables, that hidden volume is mostly affected by liquidity determinants, while hidden ratios are more affected by the stock's spread. In particular, for hidden posted and traded volume, we obtain that $TrSize$, ADV, $Time$ and Top generally rank higher than non-liquidity quantities like $Spread$. The picture drastically reverses for hidden ratios, where mainly the $Spread$ and less pronounced $Price$ dominate among the predictor variables. In these cases specifically, the spread ranks first and observe for instance that in the case of the hidden traded ratio, the spread alone explains already 98% of the predictive power of the full model with all predictor variables (see colum 4 in table 6).

We report some deviations from this hierarchy, particularly with respect to $TrSize$ and $Price$. However, we can attribute these to spurious-regression-effects and inferior fitting. For instance, $TrSize$ ranks least for traded hidden volume, although it is a liquidity quantity. However, as $TrSize$ is highly correlated with ADV and $Time$ according to 3, the addition of ADV or $Time$ in early steps of the LARS procedure already incorporates its (full) predictive power into the model. Hence, the residual explanatory power of $TrSize$ may be neglected in later steps.

Secondly, we observe that $Price$, although a non-liquidity quantity, ranks highest for hidden traded volume. However, its r^2 value reveals that its statistical contribution is only marginal. This regression-effect, that the LARS procedure occasionally selects inferior variables has been observed earlier and tends to be a general problem of model selection procedures. See also Weisberg (2004) for a more detailed discussion.

Step	Posted Hidden Volume								Posted Hidden Ratio							
	Volume				Volume (Transformed)				Ratio				Ratio (Transformed)			
	action	r^2_{lars}	$\frac{r^2_{lars}}{r^2_{ols}}$	C_p	*action*	r^2_{lars}	$\frac{r^2_{lars}}{r^2_{ols}}$	C_p	*action*	r^2_{lars}	$\frac{r^2_{lars}}{r^2_{ols}}$	C_p	*action*	r^2_{lars}	$\frac{r^2_{lars}}{r^2_{ols}}$	C_p
1	*Top*	0.34	0.73	99	*TrSize*	0.3	0.46	464	*Spread*	0.18	0.34	316	*Spread*	0.29	0.41	601
2	*TrSize*	0.44	0.94	22	*ADV*	0.37	0.56	376	*Price*	0.4	0.77	108	*Price*	0.55	0.78	219
3	*ADV*	0.45	0.98	9	*Time*	0.47	0.71	249	*Time*	0.41	0.78	105	*Vola*	0.59	0.85	159
4	*Vola*	0.46	0.99	6	*Price*	0.6	0.9	84	*Vola*	0.5	0.96	23	*Time*	0.67	0.96	42
5	*Time*	0.46	0.99	7	*Top*	0.65	0.97	27	*ADV*	0.51	0.97	18	*TrSize*	0.68	0.97	35
6	*Spread*	0.46	0.99	9	*Vola*	0.65	0.98	20	*TrSize*	0.51	0.98	16	*Top*	0.7	0.99	12
7	*Price*	0.47	1	8	*Spread*	0.66	1	8	*Top*	0.52	1	8	*ADV*	0.7	1	8

Table 5: Lars procedure for posted hidden liquidity.

Step	Executed Hidden Volume								Executed Hidden Ratio							
	Volume				Volume (Transformed)				Ratio				Ratio (Transformed)			
	action	r^2_{lars}	$\frac{r^2_{lars}}{r^2_{ols}}$	C_p	*action*	r^2_{lars}	$\frac{r^2_{lars}}{r^2_{ols}}$	C_p	*action*	r^2_{lars}	$\frac{r^2_{lars}}{r^2_{ols}}$	C_p	*action*	r^2_{lars}	$\frac{r^2_{lars}}{r^2_{ols}}$	C_p
1	*Price*	0.05	0.08	796	*Price*	0.08	0.09	1657	*Spread*	0.35	0.41	1476	*Spread*	0.92	0.98	176
2	*Time*	0.21	0.32	585	*Time*	0.33	0.41	1072	*Price*	0.56	0.66	849	*TrSize*	0.92	0.98	167
3	*ADV*	0.4	0.6	343	*Top*	0.57	0.7	544	*ADV*	0.69	0.81	468	*ADV*	0.93	0.99	78
4	*Top*	0.42	0.63	322	*ADV*	0.59	0.73	503	*Time*	0.72	0.84	393	*Top*	0.94	0.99	74
5	*Vola*	0.61	0.92	70	*Vola*	0.73	0.9	189	*TrSize*	0.79	0.92	199	*Price*	0.94	0.99	70
6	*Spread*	0.66	1	8	*Spread*	0.8	1	11	*Top*	0.84	0.99	32	*Time*	0.94	1	23
7	*TrSize*	0.66	1	8	*TrSize*	0.81	1	8	*Vola*	0.85	1	8	*Vola*	0.94	1	8

Table 6: Reporting results of the LARS procedure for traded hidden liquidity. For each LARS-step we report the choice of the selected variable ("'action"'), the current goodness-of-fit in terms of pure (r^2_{lars} and relative to the full model r^2_0 as reported in table 4 and finally the Akaike information Criteria C_p.

4 Impact and Information Content of Hidden Orders

Apart from looking at the cross-sectional determinants of hidden liquidity, our second contribution considers its impact on the stock market. In particular, in this section we shift our attention to temporal and dynamic aspects of hidden order submission, rather than its cross-sectional properties. The temporal dimension of liquidity is directly affected by investor's trading behavior. Hence, detecting distinct temporal features between displayed and hidden liquidity may help to identify the true motifs of (hidden) order submission. In fact, the question who the originators of hidden liquidity and what their intentions are is a highly active and ongoing area of market microstructure research as it is intimately related with aspects of market quality, fair trading procedures and market regulation. We will in particular analyze whether or not the presence of hidden liquidity is associated with informed trading. To motivate our program, we recycle an earlier result. To this end, recall table 1, were descriptive sample statistics for displayed and hidden depth. From the given numbers, we can construct the so called *coefficient-of-variation*, $\frac{\sigma}{\mu}$ (see table 7). As usual σ denoted the standard-deviation and μ the standard mean of the sample. Observe that the coefficient-of-variation for hidden liquidity

	hidden	*displayed*	*total*
$\frac{\sigma}{\mu}$	2.62	0.55	0.63

Table 7: Coefficients of Variation for hidden and displayed liquidity.

exceeds the one for displayed liquidity by a factor of roughly 5. In particular, this indicates a substantially higher degree of statistical dispersion and variance for hidden as compared to displayed liquidity. This assertion is confirmed by observing the time evolution of the total (displayed and hidden) depth in the limit order book for a random stock, *Health Care REIT* (figure 1). Indeed, the hidden depth evolution (blue curve) appears to exhibit a more irregular and variable pattern as compared to the displayed time evolution and is characterized by large spikes accompanied by extended period of low-depth activity. Displayed depth on the other hand exhibits a less *spiky* and more evenly distributed temporal pattern. The hidden spikes are significant as in this case they can roughly account for more than 15 times the average total displayed depth in the book.

The very fact that these arguably large spikes emerge and vanish in the course of only a few minutes provides a strong indication that they were originated by individual large traders, rather than a collective crowd of investors who coincidentally happened to place orders at the same time.[9] Two questions naturally arise out of this observation: 1) What is true motif of these large hidden traders? Do these shocks of hidden liquidity represent liquidity or informed

[9]This reasoning would be in line with prior empirical findings that suggest that hidden orders are mainly used by large investors, see Bessembinder et al. (2009) for instance.

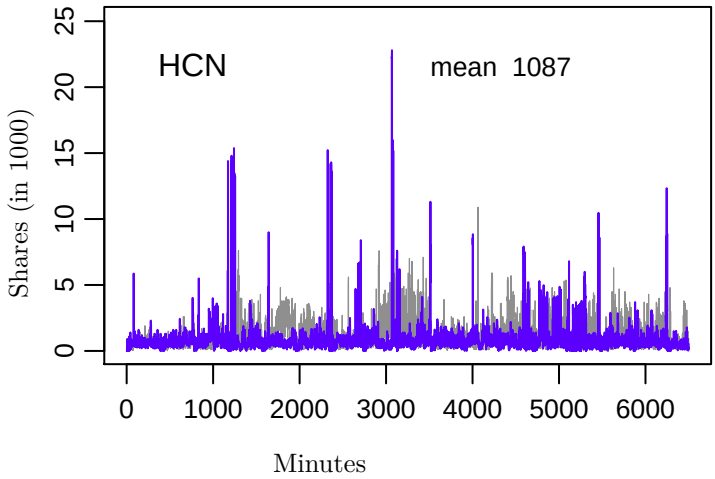

Figure 1: Time Series evolution of hidden and displayed depth in the order book for 6500 consecutive minutes for a selected stock of the S&P 500.

traders? 2) What is the impact of hidden liquidity on the stock market? In particular, how do large spikes of hidden orders affect the properties of the market?

4.1 Concentration and Dispersion of Hidden Liquidity

First, we will try to confirm the above observation on more solid empirical grounds. To this end, we introduce a range of measures to estimate the degree of *dispersion* or *localization* of hidden and displayed liquidity and check if we can quantify any notable difference between the two. We will estimate the dispersion along two dimensions: price and time. For the purpose of defining the measures we first introduce some notation. Denote x^h_{ij} (x^d_{ij}) the hidden (displayed) depth at time t_i ($i = 1, ..., n$) at price quote p_j ($0 \leq j \leq m$) for some stock. The hidden and displayed liquidity distribution at time t_i is then given by $x^h_i = (x^h_{i1}, x^h_{i2}, x^h_{i3}, ..., x^h_{im})$ and $x^h_i = (x^h_{i1}, x^h_{i2}, x^h_{i3}, ..., x^h_{im})$ respectively. We assume that the list of prices is finite and m denotes the maximum number of price quotes. Moreover, we denote by y^h_i (y^d_i) the total hidden (displayed) depth at time t_i, i.e. $y^h_i = \sum_{j=1}^m x^h_{ij}$ ($y^d_i = \sum_{j=1}^m x^h_{ij}$). Their time evolution is denoted by y^h (y^d), that is to say $y^h = (y^h_1, y^h_2, y^h_3, ..., y^h_m)$ and $y^d = (y^d_1, y^d_2, y^d_3, ..., y^d_m)$ hold.

4.1.1 Measures of Liquidity Concentration/Dispersion

To reduce the risk in measure-specification and to achieve robust estimation, we conduct our analysis by applying four measures. Our first measure is motivated by the concept of entropy.[10]

[10]The entropy measure has historical roots. It was established as early as the pioneering works of Boltzmann and Gibbs in the field of classical statistical mechanics and in statistical quantum mechanics. In thermodynamics,

In line with our desire to capture dispersion along both, the time and the price domain, we define two measures of entropy: The *temporal* ϕ_{time} and the *spatial* or *order-book entropy* ϕ_{book}:

$$\phi^{time} = -\frac{1}{\log(n)} \sum_{i=1}^{n} g_i^q \log\big(g_i^q\big) \tag{4.1}$$

$$\phi^{book}(i) = -\frac{1}{\log(m)} \sum_{j=1}^{m} h_{ij}^q \log\big(h_{ij}^q\big) \tag{4.2}$$

g_i^q and h_{ij}^q denote the corresponding empirical distribution functions, i.e. $g_i^q := \frac{y_i^q}{\sum_{i=1}^m y_i^q}$ and $h_{ij}^q := \frac{x_{ij}^q}{\sum_{j=1}^m x_{ij}^q}$ with $q = d, h$ and $i = 1, ..., n$. The choice of the normalization factors, $\frac{1}{\log(n)}$ and $\frac{1}{\log(m)}$, where n and m denote the respective sample (state) sizes, ensures that entropy is normalized and values range between 0 and 1, which facilitates cross-sectional comparison and comparisons across different sample sizes.[11] It is well-known that the entropy measure is non-negative and that it takes on its maximum value for equidistributed weights (i.e. state of highest dispersion) and its minimum when all but one weight is non-zero (i.e. state of highest of localization). As a second measure of dispersion, we consider the coefficient-of-variation (C), i.e.

$$C^{time} = \frac{\sigma(y^q)}{\mu(y^q)} \qquad\qquad C^{book}(i) = \frac{\sigma(x_i^q)}{\mu(x_i^q)} \tag{4.3}$$

where σ denotes the respective sample's standard-deviation and μ its the standard mean. We prefer the coefficient of variation over the more standard *variance-to-mean ratio* ($\frac{\sigma^2}{\mu}$) as it is normalized and thus facilitates cross-sample comparisons.

Entropy (4.2) and the coefficient-of-variation (4.3) are constructed to scale with the degree of dispersion. However, the numbers hardly allow for an illustrative understanding of the degree of dispersion. To account for this deficit, in the last step, we provide an additional measure, namely the minimum fraction of elements of a vector that account for at least a fraction s of the total sum of the vector. Or more formally, consider the values of y^q and x_i^q in descending order and denote the respective vectors $\widetilde{x}_i^q$ and $\widetilde{y}^q$. We define their partial sums according to $\overline{y}_j^q = \sum_{i=1}^j y_{l_i}^q$, $j = 1, ..., m$ and $\overline{x}_{ij}^q = \sum_{r=1}^j x_{ik_r}$. Then we can define the measures of *concentration* for both the price and the time domain according to

$$L_s^{time} = \frac{1}{n} \underset{\substack{1 \le j \le n \\ \overline{y}_j^q \ge s \cdot \overline{y}_n^q}}{\arg\min}\, \overline{y}_j^q \qquad\qquad L_s^{book}(i) = \frac{1}{m} \underset{\substack{1 \le j \le m \\ \overline{x}_{ij}^q \ge s \cdot \overline{x}_m^q}}{\arg\min}\, \overline{x}_{ij}^q \tag{4.4}$$

entropy is understood to represent the degree of dispersion (or disorder) in the thermodynamical system's micro state-space. To be more precise, according to the famous Gibbs formula the Entropy S is defined according to $S = -k_B \sum_i p_i \log p_i$, where p_i represents the probability of a finite system to reside in the state i, where k_B denotes the Boltzmann-constant. Nowadays, entropy finds more and more use in social and economic sciences. For instance, see Hart (1971)

[11] Observe that without the chosen normalization, entropy would increase with the sample size and thus the notion of *concentration* would consequently depend on the sample size n.

The smaller this number, the fewer states occupy s percent of the overall depth. For example check the case $y^q = (0, 0, 0, 0, Q)$ with $Q > 0$. We have $\overline{y}_j = Q$ for all $j \geq 1$. Hence,

$$L_s^{time} = \frac{1}{5} \underbrace{\arg\min_{\substack{1 \leq j \leq 5 \\ \overline{y}_j^q \geq s \cdot Q}} \overline{y}_j^q}_{=1} = \frac{1}{5} \cdot 1 = \frac{1}{5} = 0.20 \tag{4.5}$$

We check, that indeed 20 percent of the state $j = 5$ occupy more than the fraction s of the total depth.

We aggregate time dispersion estimates across stocks in the usual way. As an example, consider the k the estimates for the time entropy, i.e. ϕ_k^{time}. We construct the cross-sectional aggregate as follows

$$\phi^{time} = \frac{1}{N} \sum_{k=1}^{N} \phi_k^{time}. \tag{4.6}$$

For the price dispersion estimates correspond to each time t_i. Hence we need to pre-average on a per stock basis before cross-aggregating the results. To this end, consider again for each stock k its average price dispersion estimates, i.e. $\phi_k^{book} = \frac{1}{n} \sum_{i=1}^{n}$ And denote $\sigma^2(\phi_k^{book})$ the respective standard variance. Then we define the cross-sectional aggregate according to

$$\phi^{book} = \frac{\sum_{k=1}^{n} \sigma^{-2}(\phi_k^{book}) \phi_k^{book}}{\sum_{k=1}^{n} \sigma^{-2}(\phi_k^{book})}. \tag{4.7}$$

4.1.2 Estimation Results and Discussion

Table 8 and 9 report estimation results grouped in liquidity quintiles as in the previous fashion. Estimates for the entropy (ϕ) are grouped in the first column, the coefficient of variation (C) in the second and the localization measure (L) for $s = 0.50, 0.80$ are shown in the third and forth column of each table. We show estimates for each, hidden, displayed as well as total posted liquidity. Table 8 reports estimation of the average *price-dispersion* of liquidity, while table 9 reports estimates for the *time-dispersion* of liquidity. The results of table 8 and 9 can be broadly summarized in five points. First, hidden liquidity is significantly more localized (i.e. less dispersed) in the time, as well as in the price domain. In particular, consistent with the observations in figure 1 hidden liquidity tends to happen more sporadically in time and is more concentrated on fewer price quotes than displayed liquidity. For instance, according to the L^{book}in table 8, 26% of price quotes already contain more than 80% of the hidden volume. While 80% of displayed liquidity is only contained in 56% of the quotes. In the same way, table 9 shows, that hidden liquidity is concentrated on few time points, whereas the time evolution of displayed liquidity tends to be more spread out in time. Secondly, the difference in the degree of dispersion between hidden and displayed liquidity seems to be larger in the time domain than in the price domain. For instance, while the unconditional order book entropies for hidden and displayed liquidity report 0.86 and 0.97, we obtain entropies for the same pair of 0.19 and 0.71 in the time domain. Thirdly, by the third column, we observe that hidden liquidity varies

significantly more than displayed liquidity. Our unconditional coeficient of variation reports 2.72 for hidden and 0.73 for displayed liquidity in the price domain and 2.88 and 1.23 in the time domain. Fourthly, these results essentially don't depend on the stock's liquidity. We only report significant differences between liquidity quintiles with respect to the price-dispersion (i.e. table 8). Apart from a few exceptions, it appears that the concentration of hidden depth increases with a stock's liquidity. Fifth, the fact that hidden liquidity varies strongly, together with the fact that hidden liquidity is concentrated on fewer price quotes and submitted sporadically in time, suggests that hidden liquidity enters the market in few large blocks that are short-lived, whereas displayed liquidity enters the market more regularly and more continously and moderate in sizes. This is conistent with figure 1.

Liquidity Quintiles	ϕ			C			$L_{0.50}$			$L_{0.80}$		
	hidden	*displayed*	*total*	*hidden*	*displayed*	*total*	*hidden*	*displayed*	*total*	*hidden*	*displayed*	*total*
lowest	0.88	0.97	0.97	2.45	0.82	0.92	0.09	0.24	0.24	0.3	0.54	0.54
2	0.87	0.97	0.97	2.61	0.77	0.87	0.09	0.26	0.25	0.3	0.55	0.56
3	0.86	0.97	0.97	2.69	0.72	0.84	0.08	0.26	0.26	0.26	0.56	0.56
4	0.84	0.98	0.97	3.04	0.69	0.88	0.06	0.28	0.26	0.23	0.57	0.57
highest	0.85	0.98	0.98	2.81	0.66	0.78	0.06	0.29	0.27	0.22	0.59	0.59
total	0.86	0.97	0.97	2.72	0.73	0.86	0.08	0.27	0.26	0.26	0.56	0.56

Table 8: Measures of Dispersion for the order book distribution of hidden, displayed and total depth. Estimates are shown for four measures, ϕ, C, $L_{0.50}$ and $L_{0.80}$

Liquidity Quintiles	ϕ			C			$L_{0.50}$			$L_{0.80}$		
	hidden	*displayed*	*total*	*hidden*	*displayed*	*total*	*hidden*	*displayed*	*total*	*hidden*	*displayed*	*total*
least	0.2	0.62	0.58	3.11	1.46	1.87	0.1	0.2	0.15	0.13	0.37	0.29
2	0.21	0.66	0.64	2.87	1.35	1.62	0.1	0.21	0.17	0.14	0.4	0.34
3	0.2	0.72	0.69	2.81	1.22	1.41	0.11	0.23	0.2	0.14	0.44	0.39
4	0.17	0.75	0.73	2.82	1.15	1.31	0.1	0.24	0.21	0.14	0.47	0.42
highest	0.16	0.82	0.78	2.8	0.95	1.1	0.11	0.28	0.25	0.14	0.53	0.49
total	0.19	0.71	0.68	2.88	1.23	1.46	0.1	0.23	0.2	0.14	0.44	0.38

Table 9: Measures of Dispersion for the temporal distribution of hidden, displayed and total depth. Estimates are shown for four measures, ϕ, C, $L_{0.50}$ and $L_{0.80}$

4.2 Hidden Liquidity Impact: Informed or Liquidity Traders?

Our previous results establish that indeed hidden liquidity is more concentrated on single price quotes and enter the market sporadically, suggesting that the presence of hidden liquidity is associated to single large orders, whereas we observe a more regular and continuous and steady provision of displayed liquidity. Naturally, the question arises what drives these hidden liquidity-spikes and whether these large chunks of hidden liquidity carry valuable information with respect to future returns. Are they even related to informed trading? What motivates traders to issue large, hidden short-lived orders? And, how does hidden liquidity, as compared to displayed liquidity affect the market in general? We will further explore these connections in the course of this section and try to derive clues towards the true motifs behind large hidden orders. In particular, we will analyze the effect of large hidden order imbalances on the different dimensions of the market, i.e. volatility, return, spread and depth. We will also compare the results with the corresponding impact of displayed imbalances, to discern the main feature of hidden orders. We will use an event-study-framework as outlined in Campbell et al. (1997) and recently applied in a work on *quantifying market reactions to real-time news sentiment announcements*, see Gross-Klussmann and Hautsch (2010).

For this purpose, we first define events of significant imbalance-skew, i.e. liquidity shocks. That is, markets that exhibit an order-imbalance of a minimum magnitude. Specifically, to capture *hidden peculiarities*, we will distinguish between displayed and hidden order-imbalances. Hence, let t_i denote the times at which we record the market and denote D_i^{bid}, H_i^{bid} and T_i^{bid} cumulated displayed, hidden and total depth at time t_i on the bid-side of the order book and D_i^{ask}, H_i^{ask} and T_i^{ask} respectively for the ask side. Note that $D_i^{bid} + H_i^{bid} = T_i^{bid}$ holds. Then we define order-imbalances as the differences of the respective bid and sell-side depths

$$I_i^D = D_i^{bid} - D_i^{ask} \qquad I_i^H = H_i^{bid} - H_i^{ask} \qquad I_i^T = T_i^{bid} - T_i^{ask}. \tag{4.8}$$

Let us now fix an imbalance-threshold, say I. We call a time-point t_j an *event* of large hidden (displayed) excess-imbalance, whenever the corresponding imbalance exceeds the critical value I, i.e. $I_i^H > I$ ($I_i^D > I$). As we will have to consider cross-sectional comparison between different stocks, we need to normalize the choice of I and make it independent of the specific stock at hand. With this in mind, given each stocks total imbalances I_i^T, we consider its p-quantile-function $F_T^{-1}(p)$.[12] Now we define the threshold values according to

$$I_p := F_T^{-1}(p). \tag{4.9}$$

Now the choice of the critical threshold imbalance I_p depends solely on $p \in [0, 1]$, which normalizes the threshold imbalance. By considering the total instead of the displayed or hidden imbalances for identifying the threshold imbalances, we reduce the hidden/displayed selection bias.

[12] The quantile function is defined as $F^{-1}(p) := \inf\{x \in \mathbb{R}_+ : F(x) \geq p\}$, where F denotes the cumulative distribution of x.

We now fix an imbalance quantile p. Let $X = (X_{t_1}, X_{t_2}, ..., X_{t_m})$ denote some market quantity. Then, for some fixed time interval δ the liquidity shock's (mean) market impact $\overline{X}$ is simply defined as

$$\overline{X}^D_{\delta,p} = E[X_{t_k+\delta}|I^D_{t_k} \geq I_p] \qquad \overline{X}^H_{\delta,p} = E[X_{t_k+\delta}|I^H_{t_k} \geq I_p] \quad . \tag{4.10}$$

Notice that for $\delta > 0$, we study the ex-post, while for $\delta < 0$ we study the respective ex-ante impact. In the usual fashion, we conduct cross-sectional aggregation by weighted averages, where the weights equal the inverse of the respective sample variances. In other words, for stock k denote $\overline{X}^{D,k}_{\delta,p}$ and $\overline{X}^{H,k}_{\delta,p}$ its impact estimates according to (4.10). Denote $\sigma^2(\overline{X}^{D,k}_{\delta,p})$ and $\sigma^2(\overline{X}^{H,k}_{\delta,p})$ its respective variances. Then cross-sectional mean are obtained as follows:

$$\overline{X}^D_{\delta,p} = \frac{\sum_{k=1}^N \sigma^{-2}(\overline{X}^{D,k}_{\delta,p})\overline{X}^{D,k}_{\delta,p}}{\sum_{k=1}^N \sigma^{-2}(\overline{X}^{D,k}_{\delta,p})} \qquad \overline{X}^H_{\delta,p} = \frac{\sum_{k=1}^N \sigma^{-2}(\overline{X}^{H,k}_{\delta,p})\overline{X}^{H,k}_{\delta,p}}{\sum_{k=1}^N \sigma^{-2}(\overline{X}^{H,k}_{\delta,p})}. \tag{4.11}$$

We will henceforth concentrate on the stock's realized volatility, spread, depth and the so called *cumulative abnormal return* CAR.

4.2.1 Cumulative Abnormal Returns

In defining abnormal returns, we follow closely Campbell et al. (1997). As a model for stock $k's$ "normal returns" we assume the following market model:

$$R^k_{t_i} = \alpha_k + \beta_k R^{market}_{t_i} + \epsilon^k_{t_i}. \tag{4.12}$$

As usual t_i denotes time and $\epsilon^k_{t_i}$ normal and iid, random variable. R^{market}_t is the so called *market return*, and stock i's actual return R^i_t.[13] We will henceforth choose the S&P 500 index as the market return. In order to derive CAR, first we estimate (4.12), based on the 1-minute-snapshot NASDAQ Modelview data. Estimation is done without including the event windows. Provided with the parameter estimates α_k and β_k, we may again use (4.12) to compute the single abnormal returns according to $\widehat{R}^k_{t_i} = R^k_{t_i} - \widehat{\alpha}_k - \widehat{\beta}_k R^{market}_{t_i}$. Now, starting at some time t_i, the kth stock *cumulative abnormal return* up to some later time $t_i + \delta$ reads

$$\widehat{CAR}^k_{i\delta} = \prod_{t_i \leq t_l \leq t_i+\delta} (\widehat{R}^k_{t_l} + 1) - 1. \tag{4.13}$$

4.3 Impact Estimates

Figures 4, 5, 3, and 2 provide estimates of realized volatility, spread, total depth, and cumulative abnormal returns conditional on excess hidden and displayed order-imbalances according to (4.10) and (4.11). To test for generical effects, we report estimates for varying degrees of

[13] Typically R^{market} is chosen to be some predictor for the stock's actual return. For instance, in the most simple case, one may choose the stock's expected return.

imbalance I_p, i.e. $p = 60\%, 70\%, 80\%, 90\%$. We may henceforth refer to the event when order-imbalance exceeds the critical imbalance threshold I_p a *large order imbalance shock* (LOS). Impact estimates have been conducted for $\delta = -60\ min$ to (one hour pre-LOS) to up to $\delta = 120\ min$ (two hours post-LOS). The realized conditional mean estimates are each drawn with respect to the time-lag δ. Spreads volatility and depth are normalized to the unconditional mean spread, to wit a value of one represents the unconditional mean. The red-colored line represents the estimated mean conditional on hidden excess-imbalance, while the blue-colored line represents the respective estimate for the case of displayed excess-imbalances. The light-red and light-blue colored lines correspond to the respective upper and lower confidence intervals, obtained by standard t-tests. To check for statistical significane, thick black lines provide the unconditional standard deviation of the spread itself.

4.3.1 Return Impact

The impact on abnormal returns is shown in figure 2. The results can be summarized in two points. First, the expost impact of large hidden orders is significant and qualitatively follows expost a square-root law On the other hand, impact is hardly observed for displayed liquidity. For instance, after one hour of trading ($\delta = 60$), hidden liquidity generates about 35 basis points of cumulated abnormal returns for high imbalances, i.e. $p = 90\%$ (blue line). Even for lower excess imbalance, i.e. $p = 40\%$, hidden liquidity still manages to generate up to 13 basispoints impact on cumulated abnormal returns. In comparison, at high imbalances, i.e. $p = 90\%$, displayed LOS generates only minor 3.5 basis points impact on cumulated returns after 60 minutes of trading ($\delta = 60$). Hence, as the average market impact for market orders follows a *square-root-law*[14], our findings strongly suggest, that the large hidden orders, as they are less likely to get executed, get cancelled and feeded to the market via market orders. On the other hand, results indicate that large displayed depth gets consumed by other liquidity-seeking investors. We wil revive this argument in a more detailed discussiion in section in 4.4.

A comparison to a recent study on the information content of news related to earnings announcements illustrates how substantial the impact of hidden LOS is. Gross-Klussmann and Hautsch (2010) report that news arrivals exert substantial impact on cumulated abnormal returns. For instance, approximately 7 minutes after arrival of positive news, the cumulated abnormal return may reach about 7 basis points for significant news items. Our results suggest that this level of return impact can be reached for hidden LOS at an imbalance level of $p = 60\%$. For even larger hidden imbalances, i.e. $p = 90\%$, the cumulated abnormal return reaches up to 10 basis points. More than that, the impact on cumulated returns is shown to be persistent beyond 60 minutes. Our analysis thus suggests, that hidden orders can have a higher impact on prices than relevant news items.

Secondly, aside the expost impact ($\delta > 0$) we observe significant exante impact for both,

[14]See for instance Farmer and Lillo (2003).

displayed and hidden LOS: for $p = 70\%$, the prior increase in CAR reaches 29 basispoints for hidden and only about 20 basispoints for displayed imbalances. This difference reduces for smaller imbalances as for $p = 40\%$ we obtain 21 basispoints for hidden and 19 basispoints for displayed LOS at the 40%-level of imbalance. Moreover, the positive pre-event impact on abnormal returns suggests, that large hidden and displayed orders tend to be submitted in *trending markets*, that is hidden order traders tend to be momentum traders. This is in line with previous findings as reported by Pascual and Tornero (2003); Griffin et al. (2003). However, in contrast to Pascual and Tornero, we argue that this assertion is not necessarily due private information, as *momentum trading* is driven by revealed and public information.

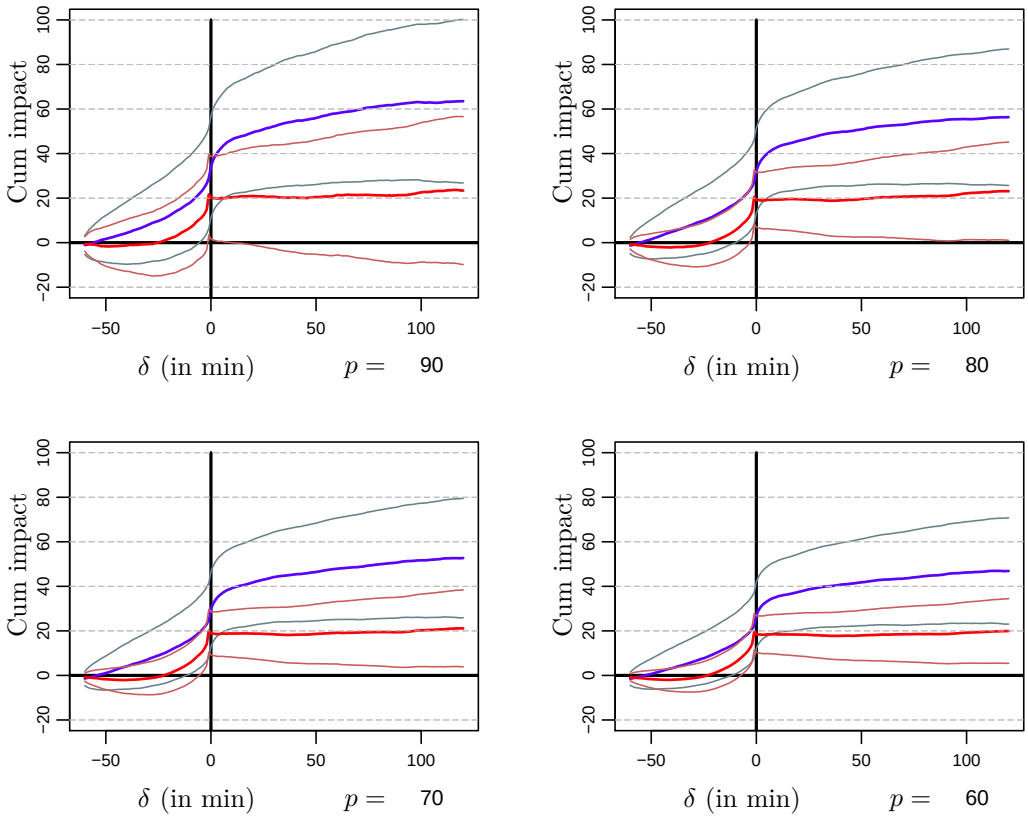

Figure 2: Impact of Hidden Order Imbalances on the cumulated abnormal return at around times of hidden order submissions. We report cumulated abnormal returns conditional on varying degrees of hidden (displayed) order imbalances p.

4.3.2 Volatility Impact

We now turn our attention to the volatility impact in figure 3. From its exante impact (i.e. $\delta < 0$) we deduce that hidden LOS happen when volatility is comparably low. In other words, large hiden order traders tend to submit their orders when volatility is low. For instance,

$p = 90\%$, volatility is approximately 27% lower than itshistorical mean, significantly below its lower standard deviation. On the other hand, at post-submission time ($\delta > 0$), volatility shows a substantial increase. The effect for displayed displayed LOS is less pronounced and occurs mostly within the band of historic standard deviation. In fact, we will show in later chapters that the increase of volatility for hidden LOS is a natural consequence of the increase in cumulative abnormal returns and the discrete sampling of realized volatility.

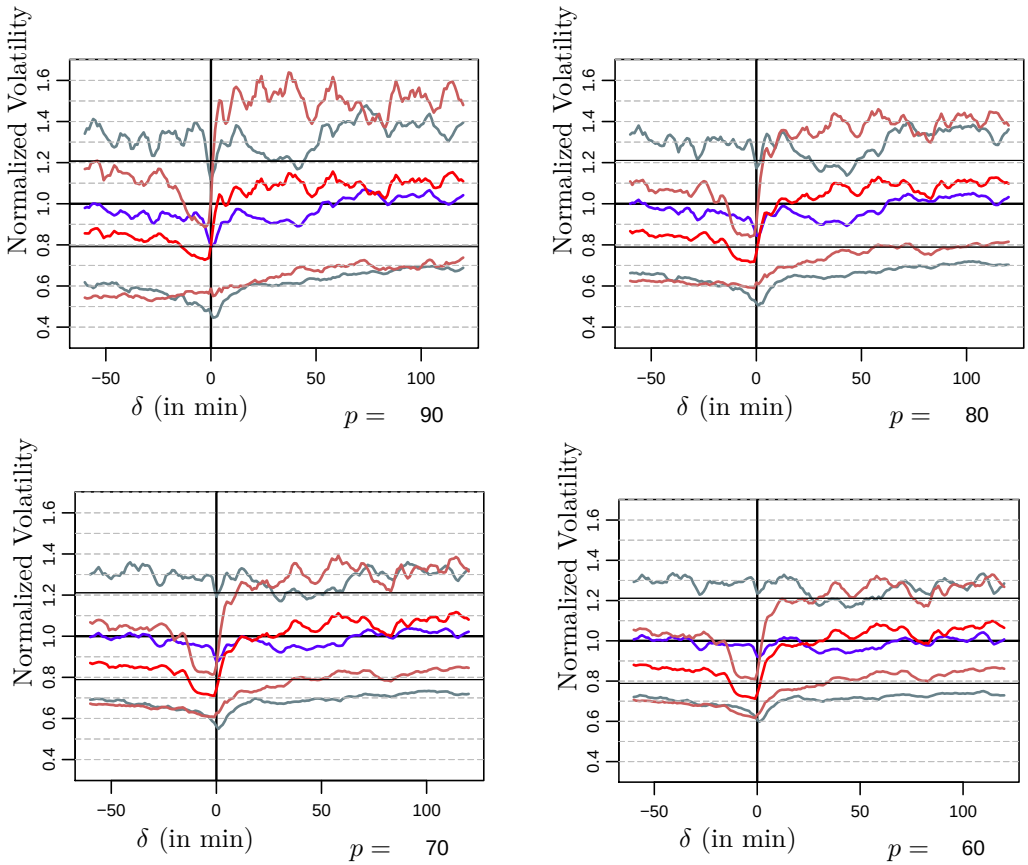

Figure 3: Impact of Hidden Order Imbalances on the realized 10-minute volatility at around times of hidden order submissions. We report the volatility conditional on varying degrees of hidden (displayed) order imbalances p. Volatility has been normalized by the historical volatility.

4.3.3 Spread Impact

Our results strongly suggest a significant decline in the spread up to 10-15 minutes prior to hidden LOS. In terms of standard deviation, the declines are significant reaching more than 10% as compared its mean spread. Likewise, spreads narrow down for displayed LOS at approximately the same amplitude around their peak level ($\delta \approx 0$) as well. However, they show pronounced reaction 10-15 minutes prior to the imbalance-shock. On the contrary, for hidden imbalances, spreads start to decline earlier. Moreover, it is interesting to note that for large imbalances, say

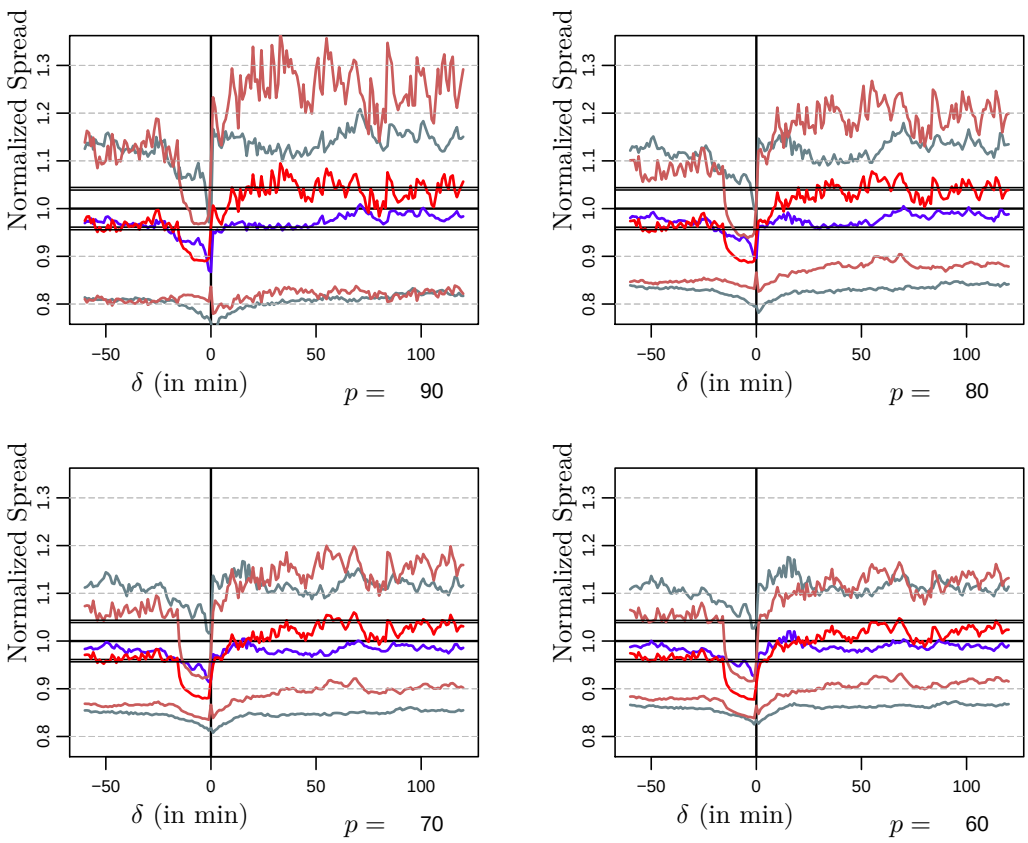

Figure 4: Impact of Hidden Order Imbalances on the realized spread at around times of hidden order submissions. We report realized spread conditional on varying degrees of hidden (displayed) order imbalances p. Spread has been normalized by its historical mean.

$p = 90\%$, the maximum decline is stronger (peaked at $\delta \approx 0$) for displayed LOS than for hidden. Concerning the post-LOS ($\delta > 0$) impact, we observe that spreads recover its mean level after approximately 70 minutes of trading when imbalances are large, i.e. $p = 90, 80$. They retain normal levels already after 20 minutes of trading for smaller imbalances ($p = 60\%$). On the other hand, in case a hidden LOS, the post-event impact is significantly stronger and spreads widen up more. After an hour of trading ($\delta \approx 60\ min$), spreads still reside at the upper level of standard-deviation, whereas spreads never reach that level in the case of displayed imbalance shocks.

4.3.4 Depth Impact

Figure 5 reports results for the total depth impact. Similar to the previous findings, total depth shows its peak at event-time ($\delta = 0$). Results are in line with naive expectation. Note that total depth is larger in case of displayed imbalances than hidden imbalances. For instance, at event-time ($\delta = 0$) and for large imbalances, say $p = 90\%$, displayed imbalances lead to total

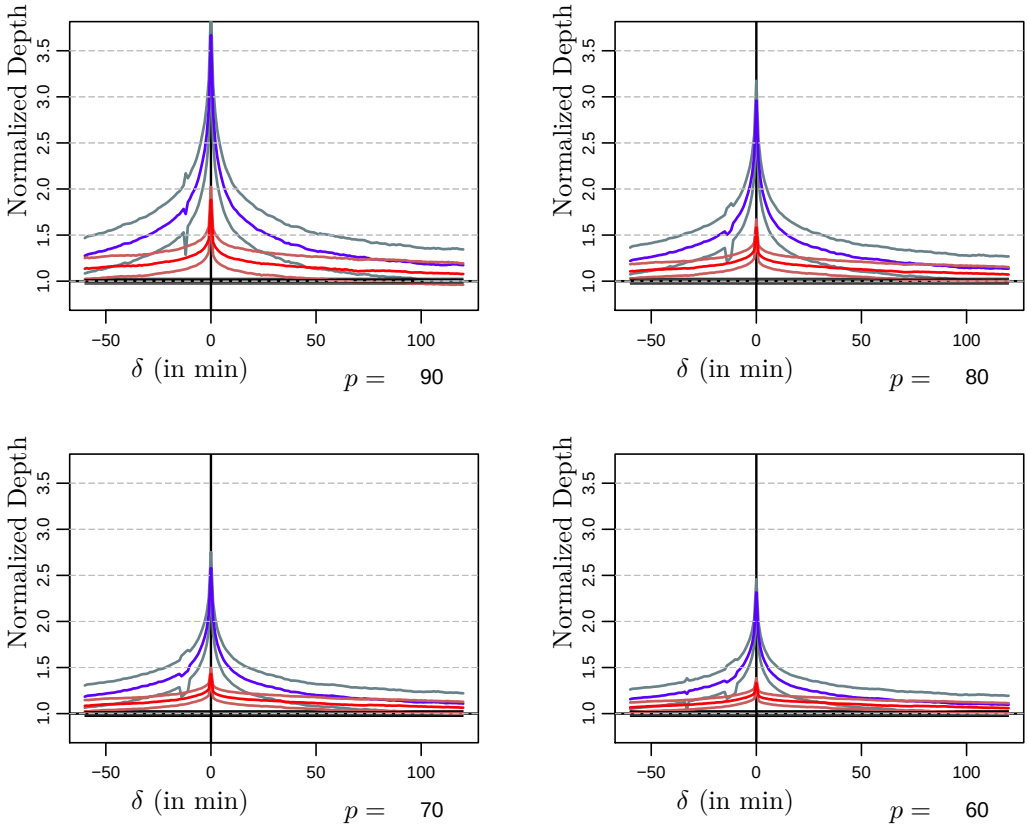

Figure 5: Impact of Hidden Order Imbalances on the total depth in the book at around times of hidden order submissions. We report the total order depth mean conditional on varying degrees of hidden (displayed) order imbalances p. Total depth is normalized by the historical total depth.

depth equaling 3.5 the average total depth in the book. Hidden imbalances on the other hand only increase the depth by a factor of roughly 1.8. We can easily explain this observation. To this end, assume that the depths of both sides of the book can be written as

$$H^{bid} = \frac{H}{2}(1+\epsilon^{bid}) \qquad H^{ask} = \frac{H}{2}(1+\epsilon^{ask}) \tag{4.14}$$

$$D^{bid} = \frac{D}{2}(1+\epsilon^{bid}) \qquad D^{ask} = \frac{D}{2}(1+\epsilon^{ask}), \tag{4.15}$$

where H and D are fixed and assume ϵ^{bid} and ϵ^{ask}, D_1^{bid} to be non-negative, iid, random variables. The respective imbalances and total depth then read

$$I^H = \frac{H}{2}(\epsilon^{bid} - \epsilon^{ask}) \qquad I^D = \frac{D}{2}(\epsilon^{bid} - \epsilon^{ask}) \qquad . \tag{4.16}$$

Then the total depths reads $T = H^{bid} + H^{ask} + D^{bid} + D^{ask}$ and from table 2 in section 2, we know that generally total displayed depth is larger than total hidden depth, i.e. $D \geq H$, we

can write their conditional expectations as follows

$$
\begin{aligned}
E[T|I^H > I_p] &= E[D^{bid} + D^{ask}|I^H > I_p] + E[H^{bid} + H^{ask}|I^H > I_p] \\
&= D + H + \frac{H}{2} E\left[\epsilon^{bid} + \epsilon^{ask}|\frac{H}{2}(\epsilon^{bid} - \epsilon^{ask}) > I_p\right] \\
&\overset{*}{\leq} D + H + \frac{D}{2} E\left[\epsilon^{bid} + \epsilon^{ask}|\frac{H}{2}(\epsilon^{bid} - \epsilon^{ask}) > I_p\right] \\
&\overset{**}{\leq} D + H + \frac{D}{2} E\left[\epsilon^{bid} + \epsilon^{ask}|\frac{D}{2}(\epsilon^{bid} - \epsilon^{ask}) > I_p\right] \\
&= E[T|I^D > I_p].
\end{aligned}
\tag{4.17}
$$

In step $*$ and $**$ we only used the fact that $D \geq H$ holds. This confirms, that total liquidity conditional on displayed imbalances must be larger than total liquidity conditional on hidden imbalances, which is solely a result of the fact that on average there is more displayed depth D in the market than hidden depth H.

4.4 Further Discussion

We report that hidden order submission is associated with larger postsubmission returns and volatility. On the other hand, the impact of displayed depth is only marginal. In the following sections, we will show why this is consistent with the view that hidden orders are used by unsophisticated liquidity traders, not informed traders. Moreover, we will elaborate on the consequences for the hidden order trader as well as for overall market quality.

Our results suggests a square-root-law like expost market impact structure for hidden liquidity, while it is literally "flat" for displayed liquidity. This suggests that hidden order traders trade their residual hidden size via market orders, while as displayed traders do not. This suggests that displayed depth gets consumed by other liquidity-seeking investors upon observation. Because hidden depth is kept hidden, these traders can not trade against these orders. Eventually, the hidden order trader has to use market orders, with the said effect on market impact.

4.4.1 Unsophisticated Liquidity Traders, why Hidden Orders Cause Price Impact and Grossmann's Conjecture

Our results suggests a square-root-like law for the market impact of hidden liquidity. Square-root market impact laws have been associated with market order submissions in recent literature Farmer and Lillo (2003). Moreover, we find that the expost impact of large hidden orders critically depends on the magnitude of imbalance (i.e. order size). These results strongly suggest that the hidden order trader is using market orders to fulfill his trading target, as his hidden orders are less likely to get executed.[15] On the other hand, as we report no significant expost market impact for displayed depth, our results suggest that (large) hidden depth is

[15] As markets generally penalize the use of hidden orders through priority arrangements, hidden orders by nature are less likely to get executed. For empirical evidence, see Bessembinder et al. (2009).

not succeeded by market order submission, indicating that displayed depth gets consumed by other "opportunistic" liquidity-seeking investors (i.e. latent liquidity).[16] This also suggests that large hidden order traders in their majority are *irrational*, as their strategy clearly increases transaction costs.

We want to point out that this reasoning is essentially in analogy of Grossmann's view about secondary, so called upstairs markets. Namely, he predicts that as the trader taps into *unexpressed* additional liquidity sources (i.e. latent liquidity) in the upstairs market, trading costs and price impact are expected to be lower as compared to the *anonymous* downstairs market (Grossmann, 1992). Indeed, Bessembinder and Venkatamaran (2004); Booth et al. (2002) could empirically prove that markets with an additional upstairs markets are associated with less transaction costs and smaller price impact. This is absolutely in line with our previous assertion about hidden traders. Namely, by not displaying his intentions, the hidden order trader unwittingly bypasses a potentially large pool of uncommitted and unexpressed liquidity. In the end, he has to pay higher trading costs and generates higher price impact. Hence, This reasoning of course, strongly relies on the assumption, that not only there is *latent liquidity* but that their net liquidity supply/demand is large enough to compensate for the downside risk faced by *parasitic traders*, see Harris (1997).

Thus together with the previous results and the assumption, that there is latent liquidity, several implications follow. First large hidden order traders are not only not informed but they are truly unsophisticated. In other words, large institutional investors are misled into believing that hidden orders are advantageous. As they will likely cancel their orders and subsequently use market orders, trading with displayed limit orders in the presence of opposite-side hidden orders is beneficial. Especially, we conjecture that in markets with a high degree of hidden liquidity provision the rate of market orders and order cancellations increases.

4.4.2 Does Hidden Liquidity increase Volatility?

Let us draw attention to the second aspect of the consequences of (large) liquidity trader's using hidden orders, namely its effect on the stock price volatility. To this end note that it is safe to say that realized volatility will be larger in high-return markets as return positively affects volatility. This follows from the fact, that realized volatility is measured over discrete time intervals. To see this, assume the standard price process as a geometric Brownian motion at discrete times t_i, i.e. $S_{t_{i+1}} - S_{t_i} = S_{t_i}\big(\mu(t_{i+1} - t_i) + \sigma\left(W(t_{i+1}) - W(t_i)\right)\big)$. Where as usual $W(t)$ denotes the standard Wiener process. Assume equidistant partition of the time line, i.e. $\Delta t = t_{i+1} - t_i$, such that $T = \Delta t N$. Now consider the realized volatility as the sum of squared

[16] In the sense of Admati and Pfleiderer (1988), we consider latent liquidity to be comprised as the set of passive traders that have discretion over when to trade (*discretionary traders*) and prefer to trade when they observe "thick" market – that is, when they believe trading has little effect on prices.

returns.

$$
\begin{aligned}
Vola_{\Delta t} &= \sum_{i=0}^{N} \left(\frac{S_{t_{i+1}} - S_{t_i}}{S_{t_i}} \right)^2 \\
&= \sum_{i=0}^{N} \left(\mu(t_{i+1} - t_i) + \sigma \left(W(t_{i+1}) - W(t_i) \right) \right)^2 \\
&= \sum_{i=0}^{N} \left(\mu \Delta t + \sigma \left(W(t_{i+1}) - W(t_i) \right) \right)^2 \\
&= \sum_{i=0}^{N} \left(\mu^2 \Delta t^2 + \sigma^2 \left(W(t_{i+1}) - W(t_i) \right)^2 + 2\mu \Delta t \sigma \left(W(t_{i+1}) - W(t_i) \right) \right) \\
&= N\mu^2 \Delta t^2 + \sigma^2 \sum_{i=1}^{N} \left(W(t_{i+1}) - W(t_i) \right)^2 + 2\mu\sigma \left(W(t_{i+1}) - W(t_i) \right)
\end{aligned}
\tag{4.18}
$$

Now taking expectations, we obtain

$$
\begin{aligned}
E\left[Vola_{\Delta t}\right] &= E\left[N\mu^2 \Delta t^2\right] + \sigma^2 \sum_{i=0}^{N} \underbrace{E\left[\left(W(t_{i+1}) - W(t_i) \right)^2 \right]}_{=t_{i+1}-t_i} + 2\mu\sigma \sum_{i=0}^{N} \underbrace{E\left[\left(W(t_{i+1}) - W(t_i) \right) \right]}_{=0} \\
&= N\mu^2 \Delta t^2 + \sigma^2 \sum_{i=1}^{N} (t_{i+1} - t_i) \\
&= \mu^2 T \Delta t + \sigma^2 T
\end{aligned}
\tag{4.19}
$$

It follows that as realized volatility is measured at discrete times, i.e. $\Delta t > 0$, any non-zero drift (i.e. return) is going to increase volatility itself. Of course, in the limit $\Delta t \to 0$ we have the well known result

$$
\lim_{\Delta t \to 0} Vola_{\Delta t} = \sigma^2 T, \tag{4.20}
$$

that volatility does only depend on the risk source of the underlying price process, namely σ. In particular, together with the previous findings, that hidden orders induce high returns, we can conclude that markets with hidden order provision have higher volatility. This assertion is also consistent with our empirical volatility-impact estimates in figure 3, that shows that the submission of (large) hidden orders (i.e. hidden imbalances) increases intertemporally the post-submission volatility as well as earlier empirical studies that attribute hidden liquidity to increased volatility (Tuttle, 2003; Aitken et al., 2001). This conclusion substantiates in particular our earlier correspondence between primary markets that provide hidden orders and markets without upstairs trade-negotiation. Remember that on both trading-mechanisms, trading costs and induced price impact are large because traders misleadingly do not or structurally can not - using Grossmann's formulation - *tap into additional unexpressed sources of liquidity.* In particular, by the arguments in (4.19), we expect that not only markets with hidden order provision exhibit larger volatility, but markets without upstairs markets do so as well. Indeed, Smith et al. (2001) show empirically that this is exactly the case.

A second point, Moinas (2010) suggests that the role of volatility is ambiguous. Indeed for instance, while Aitken et al. (2001) reports that volatility increases hidden order proportions (or vice versa), results in Bessembinder et al. (2009) rather suggest that greater return volatility is associated with a lower likelihood of hidden order usage. However, these two empirical findings are not conflicting at all. While Bessembinder et al. study the affect of volatility on traders' exposure decision, Aitken et al. on the other hand study how the overall use of hidden orders relate the stock's unconditional volatility. Although intimately related, these two views are not identical. The first study analyzes the cross-sectional link between volatility and hidden liquidity, whereas the latter studies its intertemporal dimension. Indeed, according to the estimates of the volatility impact in figure 3, we report two distinct effects. One is the presubmission (or ex-ante) impact, the other is the post-submission (ex-post) impact. From the ex-ante impact of hidden LOS, we clearly observe that the submission of large hidden orders (i.e. hidden imbalances) takes place when volatility is low compared to its historical level. However, after submission, the volatility of price quickly recovers and even may significantly exceed beyond its historical level. Hence, low volatility indeed attracts hidden traders as predicted by Bessembinder et al.. However, ex-post, volatility will increase (in line with Aitken et al.). We conclude, that hidden orders are not used to mitigate adverse-selection risks.

5 Conclusion

Using NASDAQ *Modelview*-data, we cross-sectionally analyze and identify the main determinants of posted and traded hidden liquidity for the S&P500. Observable stock quantities suffice to explain the presence of hidden liquidity as we report compelling goodness-of-fit. Using a model-selection approach, we identify the stock's average spread as the main determinant for the the proportion of traded and posted hidden liquidity. Hidden traded and posted volumes are better explained by quantities that represent the liquidity dimension of a stock.

Secondly, we find that hidden liquidity localized around few price quotes and few points in time. That is, hidden liquidity tends to enter the market in individual large chunks, sporadically in time, resembling spike-dynamics. These observations motivate the question whether hidden orders are related to single *informed trader* or more generally and whether hidden orders carry informational content with respect to future price evolution. We conduct an event-study analysis to check how the significant skews in the hidden and displayed order-imbalance affect (ex-post and ex-ante) the different dimensions of the markets. Our findings reveal a discrepancy between displayed and hidden liquidity, in that the presence of (large) hiden orders carry substantial information, whereas the informational content of displayed liquidity is negligeable. In particular, we find that the information content of hidden orders is persistent and can exceed the information content of news related to important earnings-announcements. Besides increased abnormal returns, we report an ex-post increase in volatility that is associated with hidden

order submission. Moreover, our ex-ante analyisis suggest that hidden orders are submitted in trending markets and when spreads and volatility are comparatively low.

By observing a simple analogy to Grossmann's conjecture about stocks that don't trade in upstairs market, we are able to tie together these different and to some degree seemingly, contradictory observations. According to Grossmann, taping into additional, unexpressed (latent) liquidity sources reduces transaction costs, price impact and consequently volatility in markets. Exactly these stylized facts are observed when institutional investors submit hidden orders instead of displayed orders. Obiously, institutional investors are misled into believing that hiding their intentions to trade improves their execution performance. Contrary, as we claim, as institutional investors are not likely to match their total liquidity demand with the markets endogenous liquidity flow, they need to attract additional, unexpressed liquidity to the market by exposing their full interest-to-trade a-priori. In failing to do so, the likelihood of cancelling the hidden order and subsequently issuing a large market order increases, in effect *bypassing* uncommitted latent liquidity and thus causing price impact, large transaction costs and thereby higher volatility. In other words, as the non-discretionary, institutional investor chooses to hide his intention-to-trade, the discretionary opportunity trader (latent liquidity) will less likely detect the counterpart, and therefore mutual commitment of liquidity and trades will be less synchronized. Hence, our results suggest, that not only is using hidden orders suboptimal for institutional investors but hidden orders in primary market can create negative liquidity externalities.

References

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

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.

Anand, A. and D. Weaver (2004): "Can Order Exposure be Mandated," *Journal of Financial Markets*, 4, 405–426.

Beber, A. and C. Caglio (2005): "Order Submission Strategies and Information: Empirical Evidence from the NYSE," .

Bessembinder, H. and W. Maxwell (2008): "Transparency and the Corporate Bond Market," *Journal of Economics Perspectives*, 22, 217–234.

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.

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

Biais, B., P. Hillion, and C. Spatt (1995): "An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse," *The Journal of Finance*, 50, 1655–1689.

Booth, G. G., J.-C. Lin, T. Martikainen, and T. Yiuman (2002): "Trading and Pricing in Upstairs and Downstairs Stock Markets," *Review of Financial Studies*, 15, 1111–1135.

Buti, S. and B. Rindi (2008): "Hidden Orders and Optimal Submission Strategies in a Dynamic Limit Order Market," Working Paper Series.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997): *The Econometrics of Financial Markets*, Princeton University Press.

Cao, C., O. Hansch, and X. Wang (2009): "The information content of an open limit-order book," *Journal of Futures Markets*, 29, 16–41.

Cebiroglu, G. and U. Horst (2011): "Optimal Display of Iceberg Orders," Working paper series, SFB 649 Economic Risk.

Chordia, T., R. Roll, and A. Subrahmanyam (2002): "Order imbalance, liquidity, and market returns," *Journal of Financial Economics*, 65, 111 – 130.

Copeland, T. E. and D. Galai (1983): "Information Effects on the Bid-Ask Spread," *The Journal of Finance*, 38, 1457–1469.

De Winne, R. and C. D'Hondt (2007): "Hide-and-Seek in the Market: Placing and Detecting Hidden Orders," *Review of Finance*, 11, 663–692.

——— (2009): "Hidden Liquidity in a Pure Order-Driven Market," *Finance Letters, forthcoming.*

D'Hondt, C., R. De Winne, and A. Francois-Heude (2004): "Hidden Orders on Euronext: Nothing is Quote as it Seems," Working paper series, FUCaM-Catholic University of Mons, Belgium.

Edwards, A. K., L. Harris, and M. S. Piwowar (2004): "Corporate Bond Market. Transparency and Transaction Costs," Unpublished Working Paper.

Efron, B., T. Hastie, J. Iain, and T. Robert (2004): "Least Angle Regression," *The Annals of Statistics*, 32, 407–499.

Ellul, A., C. W. Holden, P. Jain, and R. Jennings (2003): "Determinants of Order Choice on the New York Stock Exchange," Working Paper Series.

Farmer, J. D. and F. Lillo (2003): "On the origin of power law tails in price fluctuations," Tech. rep., Santa Fe Institute.

Foster, D. and S. Vishwanathan (1990): "A Theory of Intraday Variations in VOlume, Variance, and Trading Costs in Securities Markets," *Review of Financial Studies*, 3, 593–624.

Frey, S. and P. Sandas (2009): "The Impact of Iceberg Orders in Limit Order Books," Cfr working paper, Centre for Financial Research.

Gelper, S. and C. Croux (2008): "Least angle regression for time series forecasting with many predictors," Tech. rep., Katholieke Universiteit Leuven, Belgium.

Gozluklu, A. E. (2010): "Pre-Trade Transparency and Informed Trading. An Experimental Approach to Hidden Liquidity," Working Paper Series.

Griffin, J. M., J. H. Harris, and S. Topaloglu (2003): "The Dynamics of Institutional Trading," *Journal of Finance*, 58, 2285–2320.

Griffiths, M. D., B. F. Smith, D. Alasdair, S. Turnbull, and R. W. White (2000): "The costs and determinants of order aggressiveness," *Journal of Financial Economics*, 6, 65 – 88.

Gross-Klussmann, A. and N. Hautsch (2010): "When machines Read the News: Using Automated Text Analytics to Quantify High Frequency News-Implied Market Reactions," *Journal of Empirical Finance*, 18, 321–340.

Grossmann, S. (1992): "The Informational Role of upstairs and downstairs markets," *Journal of Business*, 65, 509–529.

Harris, L. (1997): "Order Exposure and Parasitic Traders," Working paper series, University of Southern California.

——— (2003): *Trading & Exchanges. Market Microstructure for Practitioners*, Oxford University Press.

Harris, L. E. and V. Panchapagesan (2005): "The information content of the limit order book: evidence from NYSE specialist trading decisions," *Journal of Financial Markets*, 8, 25 – 67.

Hart, P. (1971): "Entropy and Other Measures of Concentration," *Journals of the Royal Statistcal Society. Series A*, 131, 73–85.

Hautsch, N. and R. Huang (2009): "The Market Impact of a Limit Order," Working paper, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany.

——— (2011): "On the dark side of the market: identifying and analyzing hidden order placements," Tech. rep., Quantitative Finance Laboratory, Berlin.

Hendershott, T. and C. M. Jones (2005): "Island goes dark: Transparency, Fragmentation, Liquidity Externalities, and Multimarket Regulation," *Review of Financial Studies*, 18, 743–793.

Keim, D. B. and A. Madhavan (1995): "Anatomy of the trading process empirical evidence on the behavior of institutional traders," *Journal of Financial Economics*, 37, 371–398.

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

Pascual, R. and A. P. Tornero (2003): "On the Hidden Side of Liquidity," Working Paper Series.

Pascual Gasco, R. and D. Veredas (2008): "What Pieces of Limit Order Book Information Matter in Explaining Order Choice by Patient and Impatient Traders?" Working Paper Series.

Ranaldo, A. (2004): "Order aggressiveness in limit order book markets," *Journal of Financial Markets*, 7, 53 – 74.

Smith, B. F., D. A. Turnbull, and R. W. White (2001): "Upstairs Market for Principal and Agency Trades: Analysis of Adverse Information and Price Effects," *Journal of Finance*, 56, 1723–1746.

Tuttle, L. A. (2003): "Hidden Orders, Trading Costs and Information," Working Paper Series.

Weisberg, S. (2004): "[Least Angle Regression]: Discussion," *The Annals of Statistics*, 32, 490–494.

SFB 649 Discussion Paper Series 2012

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "HMM in dynamic HAC models" by Wolfgang Karl Härdle, Ostap Okhrin and Weining Wang, January 2012.
002 "Dynamic Activity Analysis Model Based Win-Win Development Forecasting Under the Environmental Regulation in China" by Shiyi Chen and Wolfgang Karl Härdle, January 2012.
003 "A Donsker Theorem for Lévy Measures" by Richard Nickl and Markus Reiß, January 2012.
004 "Computational Statistics (Journal)" by Wolfgang Karl Härdle, Yuichi Mori and Jürgen Symanzik, January 2012.
005 "Implementing quotas in university admissions: An experimental analysis" by Sebastian Braun, Nadja Dwenger, Dorothea Kübler and Alexander Westkamp, January 2012.
006 "Quantile Regression in Risk Calibration" by Shih-Kang Chao, Wolfgang Karl Härdle and Weining Wang, January 2012.
007 "Total Work and Gender: Facts and Possible Explanations" by Michael Burda, Daniel S. Hamermesh and Philippe Weil, February 2012.
008 "Does Basel II Pillar 3 Risk Exposure Data help to Identify Risky Banks?" by Ralf Sabiwalsky, February 2012.
009 "Comparability Effects of Mandatory IFRS Adoption" by Stefano Cascino and Joachim Gassen, February 2012.
010 "Fair Value Reclassifications of Financial Assets during the Financial Crisis" by Jannis Bischof, Ulf Brüggemann and Holger Daske, February 2012.
011 "Intended and unintended consequences of mandatory IFRS adoption: A review of extant evidence and suggestions for future research" by Ulf Brüggemann, Jörg-Markus Hitz and Thorsten Sellhorn, February 2012.
012 "Confidence sets in nonparametric calibration of exponential Lévy models" by Jakob Söhl, February 2012.
013 "The Polarization of Employment in German Local Labor Markets" by Charlotte Senftleben and Hanna Wielandt, February 2012.
014 "On the Dark Side of the Market: Identifying and Analyzing Hidden Order Placements" by Nikolaus Hautsch and Ruihong Huang, February 2012.
015 "Existence and Uniqueness of Perturbation Solutions to DSGE Models" by Hong Lan and Alexander Meyer-Gohde, February 2012.
016 "Nonparametric adaptive estimation of linear functionals for low frequency observed Lévy processes" by Johanna Kappus, February 2012.
017 "Option calibration of exponential Lévy models: Implementation and empirical results" by Jakob Söhl und Mathias Trabs, February 2012.
018 "Managerial Overconfidence and Corporate Risk Management" by Tim R. Adam, Chitru S. Fernando and Evgenia Golubeva, February 2012.
019 "Why Do Firms Engage in Selective Hedging?" by Tim R. Adam, Chitru S. Fernando and Jesus M. Salas, February 2012.
020 "A Slab in the Face: Building Quality and Neighborhood Effects" by Rainer Schulz and Martin Wersing, February 2012.
021 "A Strategy Perspective on the Performance Relevance of the CFO" by Andreas Venus and Andreas Engelen, February 2012.
022 "Assessing the Anchoring of Inflation Expectations" by Till Strohsal and Lars Winkelmann, February 2012.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2012

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

023 "Hidden Liquidity: Determinants and Impact" by Gökhan Cebiroglu and Ulrich Horst, March 2012.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".