Forthcoming Journal of Empirical Finance

Asymmetric Effects of the Limit Order Book on Price Dynamics

Tolga Cenesizoglu* Georges Dionne† Xiaozhou Zhou‡

HEC Montréal HEC Montréal Université du Québec à Montréal

November 9, 2021

Abstract

The objective of this paper is to analyze how the information embedded in different parts of the limit order book (LOB) affects price dynamics. We distinguish between slopes of lower and higher levels of the bid and ask sides and include them in a linear vector autoregressive system, along with the midquote return and trade direction, as in Hasbrouck (1991). The slopes of the higher levels have significantly greater (in terms of absolute value) immediate effects on prices than do the corresponding slopes of the lower levels. The immediate effects of the slope variable are reversed when we take the cumulative effects into consideration, suggesting that prices overreact to information embedded in the LOB. Furthermore, this reversal is stronger for the higher levels than for the lower levels, resulting in the cumulative effects of the higher levels being smaller than those of the lower levels. We link the asymmetries in the immediate effects of the lower and higher levels to the existence of more patient traders, in line with recent theoretical findings. Finally, we show that ignoring these asymmetries might have cost a high-frequency day trader 16 to 25 basis points in monthly profits.

Key words: Limit order book slope, Hasbrouck model, High-frequency trading, Asymmetric price impact, Patient traders.

JEL Classification: G10, G14, G19

*Address: 3000, chemin de la Côte-Sainte-Catherine, Montréal (Québec) Canada H3T 2A7; Telephone: +1-514-3405668; Email: tolga.cenesizoglu@hec.ca

†Corresponding Author. Address: 3000, chemin de la Côte-Sainte-Catherine, Montréal (Québec) Canada H3T 2A7; Telephone: +1-514-3406596; Email: georges.dionne@hec.ca

‡Address: 315, rue Sainte-Catherine Est, Montréal (Québec) Canada H2X 3X2; Telephone: +1-514-9873000 ext 0999; Email: zhou.xiaozhou@uqam.ca

1 Introduction

Regardless of their original trading mechanism, almost all of the world's major exchanges now feature electronic limit order books (LOBs). Most exchanges offer investors access to historical and real-time data on their LOBs at ever-increasing frequencies. Thus, there is an immense wealth of information in these high-frequency LOBs, which investors can use to gain insights into future price movements.

Hence, it is not surprising that there is a growing body of empirical literature analyzing whether the LOB affects future price movements.[1] However, most studies in the literature overlook the possibility that different levels of the LOB might affect future price dynamics differently.

In this paper, we contribute to the literature by filling this gap. Specifically, we examine whether different parts of the LOB affect future price dynamics in line with our predictions based on recent theoretical models. The result of this analysis is important because traders cannot observe underlying factors determining future prices, but they can make inferences about these factors based on the shape of the LOB. The recent theoretical literature shows that traders might strategically choose to place their orders in different levels depending on specific factors. Different levels might contain information about the underlying factors and have different effects on future price dynamics. We distinguish between the ask and bid sides, but also between their lower and higher levels, to better capture the shape of the LOB. To summarize the information in these different parts of the LOB, we use their slope, which is defined as the ratio of the change in price to the change in the cumulative quantity (depth) available between two given levels of a given side. To understand how the shape of the LOB might affect future price dynamics, two major underlying factors, based on the recent theoretical literature, are proposed. One is asymmetric information. For example, Goettler, Parlour, and

[1]Biais, Hillion, and Spatt (1995) are among the first to analyze the dynamics of limit order markets and present many interesting facts. Specifically, they show that price revisions tend to move in the direction of previous limit order flows, suggesting that the limit order book contains information relevant to future price paths. In contrast, Griffiths, Smith, Turnbull, and White (2000) find that limit orders tend to have a negative impact on prices in the Toronto Stock Exchange, because limit orders can be "picked off" by better-informed investors. This result, in turn, suggests that limit orders are placed by less-informed investors and thus do not convey much relevant information about prices. In contrast, Cao, Hansch, and Wang (2009) provide empirical evidence based on data from the Australian Stock Exchange that the limit order book is somewhat informative, contributing approximately 22% to price discovery. They also show that order imbalances between the demand and supply schedules along the book are significantly related to future short-term returns, even after controlling for autocorrelations in returns, inside spread, and trade imbalance. Similarly, using data from NYSE's Trades, Orders, Reports, and Quotes, Kaniel and Liu (2006) argue that informed traders prefer limit orders to market orders and limit orders are therefore more informative than market orders. More recently, Beltran-Lopez, Giot, and Grammig (2009) also demonstrate that factors extracted from the limit order book have non-negligible information relevant to the long-run evolution of prices in the German Stock Exchange. Specifically, they find that shifts and rotations of the order book can explain between 5% to 10% of the long-run evolution of prices, depending on the liquidity of the asset. Kozhan and Salmon (2012) provide empirical evidence that variables summarizing the information in the limit order book have statistically significant power in predicting future price movements. However, they argue that this statistical relation cannot be exploited to provide economic value in a simple trading exercise.

Rajan (2009) develop a model in which informed traders place orders at different levels of the book to profit from their superior information. In this context, the shape of the LOB affects future price dynamics due to the underlying asymmetries in traders' information sets. The second proposed underlying factor is traders' level of patience. For example, Rosu (2009) proposes a model where traders' waiting costs determine their level of patience and how they place their orders in the LOB. Rosu (2009) shows that patient traders place their orders in the higher levels of the ask side when they expect to take advantage of large market orders that are more likely to arrive. In this paper, we focus on Rosu's (2009) model because it is closest, in terms of its modeling assumptions, to the market structure of the Xetra trading system at the Frankfurt Stock Exchange, where our data are from. Specifically, in Xetra, liquidity is provided by anonymous traders who place orders in the LOB. These traders may trade immediately or wait in the order book, and they may change their orders strategically at any time. They can change or cancel their limit orders dynamically at no cost. Rosu (2009) considers a dynamic model of an order-driven market where traders trade via limit orders that they are free to modify or cancel, as is the case in Xetra.

Motivated by Rosu (2009), we first develop three sets of testable hypotheses on the effects of LOB slopes on returns, and then we analyze the relation between these effects and traders' patience. To test our hypotheses, we begin by reconstructing the first 20 levels of the historical LOB at every millisecond, for all stocks in the DAX30 index in July 2010 and in May and June 2011, based on data from the Xetra system. We follow Hasbrouck (1991) and consider data in transaction periods rather than in regularly spaced time periods. We compute midquote returns between two trades, and various slope measures right after each trade. We then consider stock-specific vector autoregressive (VAR) models that include midquote return, trade direction, and four slope measures, i.e., the bid- and ask-side slopes based on the lower and higher levels. The coefficient estimates in this framework can be interpreted as the effect of the slope variables on prices, conditional on a transaction, in the sense of Goettler et al. (2009). In this framework, we distinguish between the immediate effect of a slope variable, defined as the coefficient estimate on its first lag, and the slope variable's cumulative effect, defined as the sum of the coefficient estimates on all of its lags.

Our first set of hypotheses concerns the asymmetries in the immediate effects on returns of the lower and higher levels of the same side. We expect the immediate effects on returns of the lower and higher levels to differ depending on the proportion of patient traders. Given that the existence of more patient traders has a greater effect on the slope of the higher levels than the lower levels, any evidence that the immediate effect of the higher levels is greater in magnitude than that of the lower levels might indicate the existence of more patient traders. Our second set of hypotheses concerns the asymmetries between the overall effects of the lower and higher levels based on coefficient estimates on all lags. We consider these asymmetries because it might be the case that the higher and lower levels affect returns differently, not only instantaneously but also with a delay. More importantly, an intraday trader can potentially exploit these asymmetries to obtain economic benefits, which is indeed the case, as we show later, in Section 8. Our third and final set of hypotheses concerns the relation between the immediate and cumulative effects of a given slope measure. Rosu (2009) predicts that a trade's subsequent price impact is smaller than its immediate impact, which he refers to as price overshooting. Given that the price movements are jointly determined by market orders and the shape of the LOB, this prediction suggests that the cumulative effect of a slope variable might also be different than its immediate effect. Any evidence that the immediate effect of a slope variable is larger (smaller) in magnitude than its cumulative effect indicates that prices overreact (underreact) to information in the LOB.

We estimate the VAR model for each stock separately and find supporting empirical evidence in line with our expectations discussed above. First of all, not surprisingly, the immediate and cumulative effects of different ask-side slopes are, on average, positive, which is in line with the basic law of supply and demand, as in Kalay and Wohl (2009). In other words, an increase in the ask-side slope indicates a decrease in the selling pressure in the LOB, which in turn results in an increase in the prices and higher returns. Similar results hold for the bid side but with opposite signs, as one would expect. More importantly, the immediate effects of the slopes of the higher levels exhibit more heterogeneity, and they are, on average, greater (in magnitude) than those of the lower levels. That said, there is a reversal in the immediate effects when we take the cumulative effects into consideration. This in turn suggests that the prices overreact to information embedded in the LOB. Finally, this reversal is stronger for the higher levels than the lower levels. In other words, the cumulative effects of the higher levels are smaller (in magnitude) than those of the lower levels. This stronger reversal in the effects of the higher levels in turn points to asymmetries between the effects of the lower and higher levels, which is in line with our second set of hypotheses. These results suggest that the observed differences in the magnitudes of the effects might be due to the existence of patient traders in our sample. We next turn our attention to the relation between these effects and traders' patience. To do this, we consider an intraday analysis. We begin by analyzing how the effects of slope measures change as the traders' patience changes over the trading day. Foucault, Kadan, and Kandel (2005) hypothesize that traders become less patient over the trading day because they might want, or need, to liquidate their positions before the end of the continuous trading session. Hence, there should be, on average, more patient traders in the morning and fewer patient trade rs at the end of the day. We use the convexity of the LOB as our proxy for traders' patience, motivated by Rosu's finding relating the two.

In line with these expectations, the hourly convexity measure, averaged over all the stocks and trading days in our sample, is indeed low initially and increases almost monotonically over the trading day, suggesting that there are indeed more patient traders at the beginning of the trading day and fewer patient traders at the end. We thus expect that the ask- and bid-side slopes of the higher levels should have significantly greater (in absolute value) immediate effects on prices than the corresponding slopes of the lower levels in the morning, whereas the opposite should hold at the end of the trading day. Consequently, we estimate separate VAR models for each stock, trading day, and hour, and obtain a panel data set of the immediate effects of the slope variables. We find that the immediate effects of the slope variables of the higher levels are, on average, greater in magnitude than the corresponding coefficients on the slope variables of the lower levels at the beginning of the trading day, and that they are smaller in magnitude at the end of the trading day. We then regress the differences between the coefficient estimates of the higher and lower levels (in absolute values for the bid side) on dummy variables, for the first and last hour of trading. We find that the dummy variable for the first hour of trading has significantly positive coefficient estimates, whereas the dummy variable for the last hour of trading has significantly negative coefficient estimates for both the ask and bid sides. This implies that the slope variables of the higher levels do indeed have significantly greater (in absolute value) immediate effects on prices than the corresponding slopes of the lower levels, at the beginning of the trading day, when there are more patient traders; and the slope variables have significantly smaller immediate effects at the end of the trading day, when there are fewer patient traders. We thus argue that these findings provide statistically significant evidence linking traders' patience to the immediate effects of slope measures on price dynamics, in line with our predictions.

Having found that different slope measures affect prices differently, we then show that these asymmetric effects can also be economically important. We do this by observing the performance of high-frequency day-trading strategies that ignore these asymmetries in forecasting midquote returns at each transaction, and we compare these with a strategy that does use them. Our main trading strategy, which we refer to as the unrestricted strategy, uses an unrestricted VAR model to forecast midquote returns at each transaction period. The competing strategies, which we refer to as restricted strategies, employ restricted versions of the VAR model, such that a chosen pair of slope variables has symmetric effects on price dynamics. In all trading strategies, similarly to Kozhan and Salmon (2012), we consider a forecast that is greater (smaller) than a given threshold to be a buy (sell) signal, and we revise our existing position in the stock accordingly. Comparing the average monthly cumulative returns of the unrestricted and restricted trading strategies, we find that ignoring these asymmetric effects might have cost a trader between 16 and 25 basis points in monthly profits in June 2011, corresponding to about 2% and 3% in annual terms.

Our paper contributes to the literature in three important empirical dimensions. First of all, to the best of our knowledge, we are among the first to analyze asymmetries in the effects of information embedded in different parts of the LOB. Several interesting previously undocumented results emerge from our analysis. The ask- and bid-side slopes of the higher levels have significantly greater immediate effects on prices than do the corresponding slopes of the lower levels. We also find that there is a reversal in the immediate effects when we take cumulative effects into consideration, suggesting that the prices overreact to information embedded in the LOB. This reversal is stronger for the higher levels than it is for the lower ones, resulting in smaller cumulative effects for the higher levels than for the lower. Second, in line with recent theoretical findings, we link the asymmetries in the effects of the lower and higher levels to the existence of more patient traders. Specifically, the slope variables of the higher levels indeed have significantly greater (in absolute value) immediate effects on prices than do the corresponding slopes of the lower levels, at the beginning of the trading day when there are more patient traders; and they have significantly smaller immediate effects at the end of the trading day, when there are fewer patient traders. Finally, we show that these asymmetric effects can be economically important.

The rest of the paper is organized as follows. Section 2 presents the details of our data set and defines the LOB slope. Section 3 introduces the empirical model along with the empirical choices made to obtain the main set of results. Section 4 discusses the theoretical motivation behind our empirical analysis and presents our testable hypotheses based on the theoretical discussion and empirical model. In Section 5, the effects of slope variables on price dynamics are given. Section 6 reports our main empirical results on the asymmetries in the short-run and cumulative effects of different slope variables on price dynamics. Section 7 relates the asymmetries in the immediate effects of slope variables on price dynamics to traders' patience. Section 8 shows that the asymmetries in the effect of slope measures on price dynamics are also economically important. Section 9 concludes the paper. Additional results and robustness checks are provided in an online appendix.

2 Data

Our data are from the automated order-driven trading system Xetra, operated by the Deutsche Börse Group at the Frankfurt Stock Exchange. It is the main German trading platform, accounting for more than 90% of total transactions at all German exchanges. In Xetra, there are no dedicated market makers for blue chip and other liquid stocks. Thus, all liquidity in Xetra is provided by market participants submitting limit orders.

The raw data set contains all events that are tracked and sent through the data streams. We first process the raw data set using XetraParser software, developed by Bilodeau (2013).[2] We then reconstruct the first 20 levels of the limit order book in millisecond time intervals between the normal trading hours of 9:00 a.m. and 5:30 p.m.[3] The limit order book can change when either a trade is executed or a limit order is placed, modified, or canceled. In the unlikely event that these two types of events have the same millisecond time stamp, we need to make an assumption on the sequence of events, given that we cannot observe which one arrived earlier. We assume that a trade is always executed before any other change to the limit order book with the same millisecond time stamp. Thus, we first modify the limit order book to reflect the trade execution before taking its snapshot. In other words, if a trade is executed at a given millisecond, then the snapshot of the limit order book for that millisecond already reflects the executed trade.

Our data cover all stocks in the DAX30 index and all trading days in July 2010, and May and June 2011. We focus on three months of data simply due to the sheer size of ultra high-frequency limit order books. Table 1 presents the list of stocks in the DAX30 index as of June 2011, along with some daily summary statistics over the whole sample from the Security Daily files in Compustat Global.

[2] We thank Yann Bilodeau for his comments and help in constructing the data set.

[3] During normal trading hours, there are two types of trading mechanisms: call auctions and continuous auctions. For stocks listed on the DAX 30, there are three call auctions during a trading day: the open, mid-day, and closing auctions. The prices during call auctions are not determined by trading activity but rather are based on a set of rules determined by the exchange. Between the call auctions, the market is organized as a continuous auction in which traders can submit round lot-sized limit and/or market orders only. The prices from the call auctions serve as the opening prices for the following continuous auctions. To avoid any bias due to the peculiar structure of the call auctions, we ignore all data corresponding to the three call auctions for a DAX 30 stock.

[Insert Table 1 here]

Figure 1 presents two snapshots of the limit order book for ALV on June 1, 2011. As can be easily seen in Figure 1, both the bid and ask sides of the limit order book can take on different shapes at different times. We summarize different shapes that the limit order book can exhibit by distinguishing between its lower and higher levels and measuring their slopes separately. More precisely, let $P^B_{l_1,t}$ and $P^A_{l_1,t}$ denote l_1^{th} best bid and ask prices, respectively, in period t. Similarly, let $D^B_{l_1+1,l_2,t}$ and $D^A_{l_1+1,l_2,t}$ denote the cumulative quantity available between levels $l_1 + 1$ and l_2 (both levels inclusive and $l_2 > l_1$) in the bid and ask sides of the limit order book, respectively. The slopes of the bid and ask sides between levels l_1 and l_2 in period t, $S^B_{l_1,l_2,t}$ and $S^A_{l_1,l_2,t}$, are defined as the change in the price relative to the cumulative quantity available between levels l_1 and l_2:

$$S^B_{l_1,l_2,t} = \frac{P^B_{l_2,t} - P^B_{l_1,t}}{D^B_{l_1+1,l_2,t}}, \tag{1}$$

$$S^A_{l_1,l_2,t} = \frac{P^A_{l_2,t} - P^A_{l_1,t}}{D^A_{l_1+1,l_2,t}} \tag{2}$$

for $l_1 = 1, \ldots, 19$ and $l_2 > l_1$. The slope of the bid side is a measure of price sensitivity to changes in quantity demanded and is always negative. A high (in absolute value) bid slope implies that the price between two levels of the bid side will decrease more, on average, for a given change in quantity demanded. In other words, an increase in the bid slope suggests that investors are willing to buy the same total quantity only at lower prices. Similarly, the slope of the ask side is a measure of price sensitivity to changes in quantity supplied and is always positive. A high ask slope implies that the price between two levels of the ask side will increase more, on average, for a given change in quantity supplied, which, in turn, suggests that investors are willing to sell the same total quantity only at higher prices. We consider the log of slope variables to diminish the effect of outliers on our results.

[Insert Figure 1 here]

3 The Empirical Model

In this section, we present the empirical model and the empirical choices to obtain the main set of results. Our empirical model is based on that of Hasbrouck (1991), who suggests using the following VAR model to analyze the effects of trades on prices:

$$r_t = \sum_{\tau=1}^{\infty} \alpha_{r,\tau} r_{t-\tau} + \sum_{\tau=1}^{\infty} \alpha_{x,\tau} x_{t-\tau} + \varepsilon_{r,t}, \tag{3a}$$

$$x_t = \sum_{\tau=1}^{\infty} \beta_{r,\tau} r_{t-\tau} + \sum_{\tau=1}^{\infty} \beta_{x,\tau} x_{t-\tau} + \varepsilon_{x,t}, \tag{3b}$$

where t indexes trades; x_t is the sign of the trade in period t (+1 for a trade initiated by a buyer and -1 for a trade initiated by a seller based on Lee and Ready (1991)'s algorithm); r_t is the log midquote return defined as the change in log midquote price (i.e. the average of the best bid and ask quotes) between periods $t-1$ and t, that is, $r_t = \Delta q_t = q_t - q_{t-1}$; and q_t is the log midquote price at the end of period t. This is a very general and flexible model that nests many of the standard microstructure models as special cases. The disturbances in this framework, $\varepsilon_{r,t}$ and $\varepsilon_{x,t}$, are generally modeled as white noise processes and can be interpreted as public information embedded in unexpected returns and private information embedded in unexpected trades, respectively.

We assume that the dynamics of a limit order market can also be approximated by a linear VAR system similar to that proposed by Hasbrouck (1991). Specifically, we include slope measures based on different levels of the ask and bid sides as additional state variables in the above VAR, which yields:[4]

$$r_t = c_r + \sum_{\tau=1}^{\infty} \alpha_{r,\tau} r_{t-\tau} + \sum_{\tau=1}^{\infty} \alpha_{x,\tau} x_{t-\tau} + \sum_{\tau=1}^{\infty} \boldsymbol{\alpha}'_{z,\tau} \mathbf{z}_{t-\tau} + \varepsilon_{r,t}, \tag{4a}$$

$$x_t = c_x + \sum_{\tau=1}^{\infty} \beta_{r,\tau} r_{t-\tau} + \sum_{\tau=1}^{\infty} \beta_{x,\tau} x_{t-\tau} + \sum_{\tau=1}^{\infty} \boldsymbol{\beta}'_{z,\tau} \mathbf{z}_{t-\tau} + \varepsilon_{x,t}, \tag{4b}$$

$$\mathbf{z}_t = \mathbf{c}_z + \sum_{\tau=1}^{\infty} \boldsymbol{\gamma}_{r,\tau} r_{t-\tau} + \sum_{\tau=1}^{\infty} \boldsymbol{\gamma}_{x,\tau} x_{t-\tau} + \sum_{\tau=1}^{\infty} \boldsymbol{\gamma}'_{z,\tau} \mathbf{z}_{t-\tau} + \boldsymbol{\varepsilon}_{z,t}, \tag{4c}$$

where $\mathbf{z}_t$ is a vector that includes the slope measures of interest.

To implement this model empirically, we need to make some empirical choices regarding the sampling approach,

[4] We use bold letters to distinguish vectors and matrices from scalars.

the slope measures and the truncation point for the infinite sums. Regarding the sampling frequency, we use tick-by-tick data, i.e. sampling every time there is a trade, for our main sets of results following Hasbrouck (1991) and Engle and Dufour (2001). In the online appendix, we estimate a version of the model in Equation (4) using data sampled at a regular frequency of 100 milliseconds. With tick-by-tick sampling, we need to make a choice about whether to take the limit order book snapshot right before or after a trade. We can measure limit order book variables, including the best bid and ask prices, every millisecond. However, a trade can only be matched to a given millisecond and thus one needs to decide whether to take a snapshot of the limit order book right before or right after a trade. The theory does not provide much guidance on this issue. We measure the slope variables right after a trade.[5] Our main sampling approach implies that the midquote return and slope variables in period t are observed right after (less than a millisecond after) the trade in period t and its direction. Hence, we include the trade direction in period t to control for its contemporaneous effect on returns and limit order book variables in the estimated version of Equation (4).

Regarding the truncation issues, we truncate the infinite sums in Equation (4) at five lags. We choose five lags following the previous literature. For example, both Hasbrouck (1991) and Dufour and Engle (2000) also truncate the sums at five lags. This allows us to compare some of our results to theirs. Furthermore, as we will discuss below, we use the sum of the coefficients on these lags as a proxy for the cumulative effect of slope variables on returns. Given that the stocks in our sample are traded very frequently, we believe that five transactions are more than enough to capture the cumulative effect of slope variable on returns. That said, we also estimated the model with different numbers of lags. The results based on different number of lags are presented in the online appendix and are similar to those based on the model with five lags.

The timing convention discussed above is reflected in the starting points of the summations in the estimated version of Equation (4). The summations for trade direction in the equations for the returns and limit order book variables

[5] We also considered the alternative sampling approach of measuring the limit order book variables right before a trade. Our results remain qualitatively similar.

start at zero instead of one in the first sampling approach. The estimated version of the model is then as follows:

$$r_t = c_r + \sum_{\tau=1}^{5} \alpha_{r,\tau} r_{t-\tau} + \sum_{\tau=0}^{5} \alpha_{x,\tau} x_{t-\tau} + \sum_{\tau=1}^{5} \boldsymbol{\alpha}'_{z,\tau} \mathbf{z}_{t-\tau} + \varepsilon_{r,t}, \quad (5a)$$

$$x_t = c_x + \sum_{\tau=1}^{5} \beta_{r,\tau} r_{t-\tau} + \sum_{\tau=1}^{5} \beta_{x,\tau} x_{t-\tau} + \sum_{\tau=1}^{5} \boldsymbol{\beta}'_{z,\tau} \mathbf{z}_{t-\tau} + \varepsilon_{x,t}, \quad (5b)$$

$$\mathbf{z}_t = \mathbf{c}_z + \sum_{\tau=1}^{5} \boldsymbol{\gamma}_{r,\tau} r_{t-\tau} + \sum_{\tau=0}^{5} \boldsymbol{\gamma}_{x,\tau} x_{t-\tau} + \sum_{\tau=1}^{5} \boldsymbol{\gamma}'_{z,\tau} \mathbf{z}_{t-\tau} + \boldsymbol{\varepsilon}_{z,t}. \quad (5c)$$

We include four slope measures: the first two are the bid- and ask-side slopes between their corresponding first and fifth levels, $S^B_{1,5,t}$ and $S^A_{1,5,t}$, which we use to capture the slopes of lower levels, and the other two are the bid- and ask-side slopes between their corresponding fifth and twentieth levels, $S^B_{5,20,t}$ and $S^A_{5,20,t}$, which we use to capture the slopes of higher levels. This empirical choice is motivated by two factors. The first level is undoubtedly the most frequently updated one, and we thus want to include this information in our definition of the slope of the lower level. Second, levels of the limit order book higher than 10 are less frequently updated and might have stale information. We want to minimize the effect of this stale information by including levels between five and ten, which are still updated quite frequently, in our definition of higher levels.[6] We estimate the empirical specification in Equation 5 via ordinary least squares (OLS) with heteroskedasticity and autocorrelation consistent standard errors a la Newey and West (1987).

This specification has several advantages compared with a simple regression framework. First, it can be considered as a reduced-form linear approximation that is designed to capture the dynamics of limit order market models discussed in the introduction. It is also used by Brogaard, Hendershott, and Riordan (2016) in a similar fashion to understand the role of limit orders submitted by high-frequency traders in price discovery. The coefficient estimates on the slope variables can be interpreted as the effects of these variables on prices conditional on a transaction in the sense of Goettler et al. (2009). In addition, this empirical specification allows us to analyze not only the short-run effects of slope variables on prices but also their cumulative effects after several transactions. To be more precise on the latter point, the coefficient estimates on the first lags of slope variables can be interpreted as their immediate effects on prices, while the coefficient estimates on further lags provide an idea about their short-run dynamic effects on prices. Furthermore, Dufour and Engle (2000) argue that the sum of the coefficient estimates on all lags of a given slope

[6] We consider using alternative definitions and cutoff points for lower and higher levels in the online appendix. The results based on these alternative definitions and cutoff points are presented in the online appendix and are similar to our main set of empirical results.

variable can be considered as a first raw approximation of its impact on prices after several transactions.

4 Hypothesis Development

In this section, we first briefly summarize the theoretical literature before focusing on a specific model to motivate our hypotheses. We then discuss how we can test our hypotheses using the empirical model in Equation 5.

Whether the state of the limit order book can predict future price movements is a theoretical question. Earlier microstructure models, such as those of Glosten and Milgrom (1985), Kyle (1985), Glosten (1994), and Rock (1996), treat limit orders as free options provided by uninformed investors because they tend to be picked off by better-informed investors. They thus implicitly assume that the limit order book cannot be informative for future price movements. However, recent theoretical models allow informed investors to strategically choose between limit and market orders, and show that investors use not only market orders, as assumed in the previous literature, but also limit orders in a rational expectations equilibrium.[7] Regardless of their underlying assumptions, the common prediction of these recent theoretical models is that limit orders can predict future price movements.

As mentioned in Section 2, our data is from the automated order-driven trading system Xetra. In Xetra, liquidity is provided by anonymous traders who place orders in the limit order book. These traders may trade immediately or wait in the order book, and may change their orders strategically at any time. They can change or cancel their limit orders dynamically at no cost. Due to the complexity of modeling such markets, there are only a few theoretical papers modelling the dynamics of an order-driven market. To the best of our knowledge, Rosu (2009) is closest to the market structure of Xetra in terms of its modeling assumptions. Specifically, Rosu (2009) presents a dynamic model of an order-driven market where traders trade via limit orders that they are free to modify or cancel, as in Xetra. In his model, traders' patience is the main driving factor. He argues that symmetrically informed patient sellers (buyers) with a low waiting cost will place their order at a higher level of the ask (bid) to get a better expected selling (buying)

[7] For example, informed investors could use limit orders to avoid detection, as in Kumar and Seppi (1994), to insure themselves against the price they could obtain for their market orders, as in Chakravarty and Holden (1995), or to take advantage of their sufficiently persistent private information, as in the studies by Kaniel and Liu (2006) and Kalay and Wohl (2009). There is also a more recent literature on dynamic limit order markets with strategic traders, such as the works of Foucault et al. (2005), Goettler et al. (2009), and Rosu (2009). Foucault et al. (2005) show that patient traders tend to submit limit orders, while impatient ones submit market orders in equilibrium. Rosu (2009) shows that fully strategic, symmetrically informed liquidity traders can choose between market and limit orders based on their trade-off between execution prices and waiting costs. Goettler et al. (2009) find that limit orders tend to be submitted mostly by informed traders, and competition among them results in their private information being reflected in the limit order book. Bhattacharya and Saar (2016) find that informed traders tend to provide liquidity in illiquid markets and demand liquidity from more liquid markets.

price, whereas a very impatient seller (buyer) will immediately place a market order.

Rosu (2009) has many predictions regarding the relation between traders' patience, the shape of the limit order book and the effect of a transaction on returns. Regarding the shape of the limit order book, he shows that it depends on the relative arrival rates of limit orders from different types of traders. Specifically, if large market orders are more likely, as in a more volatile market, the limit order book will be less convex, indicating that patient traders have placed their orders at the higher levels of the ask.[8] If large market orders are unlikely, the limit order book will be more convex, and the limit orders will cluster at the best ask level or at intermediate levels in a mixed environment.

Regarding the price impact of transactions, he shows that it will be smaller in markets with higher trading activity (measured by the sum of arrival rates of all agents) and with higher competition (measured by the ratio of the arrival rates for patient and impatient traders). He also distinguishes between instantaneous and subsequent price impacts and finds that the subsequent price impact is smaller than the instantaneous price impact, which he refers to as price overshooting.

In this paper, we are mostly interested in the effect of the limit order book shape on returns while controlling for the arrival of market orders. Hence, we first use Rosu's theory to motivate our hypotheses on the effect of limit order book slopes on returns before linking them to traders' patience.

We start with our hypotheses on the effect of limit order book slopes on returns. First of all, an increase in the ask-side slope indicates a decrease in the selling pressure in the limit order book, which in turn results in an increase in the prices and higher returns. Similarly, an increase in the bid-side slope indicates a decrease in the buying pressure in the limit order book, which in turn results in a decrease in the prices and lower returns. In other words, these basic laws of supply and demand indicate that the slope of the ask (bid) side as measured in this paper is related inversely to the supply (demand) for the stock. We thus expect the slope of the ask (bid) side to have a positive (negative) effect on returns. Furthermore, one can also analyze the similarities and differences between the effects of bid- and ask-side slopes. Although we present our estimation results and briefly discuss the similarities and differences between the effects of bid- and ask-side slopes, we do not put too much emphasis on these results given that they have been

[8] Rosu (2009, p.4624) states that: "If large market orders are more likely, as in the quadratic case (i), then the price impact function $I_m(i)$ is concave, which indicates that patient sellers indeed place their limit orders at higher levels above the ask." We prefer to use the term convexity rather than concavity because the limit order book is convex on average, as will be shown in Section 7. That said, there are, of course, concave limit order books in our sample. Given that we are interested in relative convexity, we refer to a concave limit order book as being less convex than a convex limit order book.

extensively studied in the previous literature (see for example Kalay and Wohl (2009)). Instead, we focus on the similarities and differences between the effects of the slopes of the higher and lower levels of the same side, which have not been studied in the previous literature.

Our first set of hypotheses concerns the asymmetries in the immediate effects on returns of the higher and lower levels of the same side. Results in Rosu (2009) discussed above suggest that the relative slopes of the higher levels depend closely on the relative proportion of patient traders in the market. The relative effect of the higher levels on returns for a given trade direction and/or size might also depend on the relative proportion of patient traders. We thus expect the immediate effects of the lower and higher levels on returns to differ from each other depending on the proportion of patient traders. Denoting our hypotheses for the ask side with a and those for the bid side with b, we can express these hypotheses as restrictions on the immediate effects of slopes on returns as follows:

$$H1a \quad : \quad \boldsymbol{\alpha}_{z,1}(S^A_{1,5,t}) = \boldsymbol{\alpha}_{z,1}(S^A_{5,20,t});$$

$$H1b \quad : \quad \boldsymbol{\alpha}_{z,1}(S^B_{1,5,t}) = \boldsymbol{\alpha}_{z,1}(S^B_{5,20,t});$$

where $\boldsymbol{\alpha}_{z,1}(\cdot)$ denote the element of $\boldsymbol{\alpha}_{z,1}$ corresponding to the slope measure of interest in parentheses. Given that the existence of more patient traders affects the slope of the higher levels more than the lower levels, rejection of these null hypotheses in favor of the alternative where the immediate effect of the higher levels is bigger in magnitude than that of the lower levels, i.e. $\boldsymbol{\alpha}_{z,1}(S^A_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S^A_{5,20,t})$ for the ask side and $\boldsymbol{\alpha}_{z,1}(S^B_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S^B_{5,20,t})$ for the bid side, might indicate the existence of more patient traders.

We first test these hypotheses in the overall sample before focusing on different periods with potentially varying degrees of trader patience. Given that we cannot directly observe the degree of traders' patience, we follow Foucault et al. (2005) by assuming that their relative number changes during a trading day. Namely, traders are more patient in the morning and less patient at the end of the day because they want to liquidate their positions before the end of the continuous trading session. So, we expect the slope of the higher levels of the ask side should have a greater impact on stock prices than that of the lower levels in the morning, when there are potentially more patient traders. We test this conjecture by estimating the model using data within each hour of the trading day. We then compare the instantaneous and cumulative effects across different hours of the trading day. We expect these effects to be bigger in magnitude in the morning when there are more patient traders.

Our second set of hypotheses concerns the asymmetries between the overall, and not only the immediate, effects of the lower and higher levels. We consider these asymmetries because it might be the case that the higher and lower levels affect returns differently when we consider not their immediate effects but rather their delayed effects. Furthermore, it is also possible that higher and lower levels affect returns differently when we consider all lags jointly rather than considering only the first lag. We thus test the equality of the coefficients on all lags of slope variables based on different levels jointly. Our second set of testable hypotheses can be written:

$$H2a \quad : \quad \boldsymbol{\alpha}_{z,\tau}(S^{A}_{1,5,t}) = \boldsymbol{\alpha}_{z,\tau}(S^{A}_{5,20,t}) \quad \text{for } \tau = 1, 2 \ldots, 5;$$

$$H2b \quad : \quad \boldsymbol{\alpha}_{z,\tau}(S^{B}_{1,5,t}) = \boldsymbol{\alpha}_{z,\tau}(S^{B}_{5,20,t}) \quad \text{for } \tau = 1, 2 \ldots, 5.$$

where $\boldsymbol{\alpha}_{z,\tau}(\cdot)$ for $\tau = 1, 2 \ldots, 5$ denote the element of $\boldsymbol{\alpha}_{z,\tau}$ corresponding to the slope measure of interest in parentheses. Rejection of this null hypothesis suggests that at least one of the five lags of the higher and lower levels affects returns differently. This in turn points to the importance of distinguishing between lower and higher levels of the limit order book and, more importantly, implies that the slopes of the higher and lower levels have statistically different predictive powers for returns. An intraday trader can potentially exploit these differences in predictive powers to obtain economic benefits, which is what we analyze in Section 8.

Our third set of hypotheses concerns the relation between the immediate and cumulative effects of a given slope measure. As discussed above, Rosu (2009) also predicts that subsequent price impact of a trade is smaller than its immediate impact, which he refers to as price overshooting. Given that the price movements are jointly determined by market orders and the shape of the limit order book, this prediction suggests that the cumulative effect of a slope variable might also be different than its immediate effect. We can test this hypothesis by approximating the cumulative effect of a slope variable by the sum of the coefficient estimates on all five lags, as discussed in Section 3. Thus, our third and final set of hypotheses tests the equality of the immediate and cumulative effects of a given slope measure, and can be expressed as follows:

$$\begin{aligned}
H3a &: \quad \boldsymbol{\alpha}_{z,\tau}(S^A_{1,5,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{1,5,t}); \\
H3a^* &: \quad \boldsymbol{\alpha}_{z,\tau}(S^A_{5,20,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{5,20,t}); \\
H3b &: \quad \boldsymbol{\alpha}_{z,\tau}(S^B_{1,5,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^B_{1,5,t}); \\
H3b^* &: \quad \boldsymbol{\alpha}_{z,\tau}(S^B_{5,20,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^B_{5,20,t}).
\end{aligned}$$

where * denotes our hypothesis for the higher levels of the limit order book for both ask and bid sides. Rejection of these hypotheses in favor of the alternative, where the immediate effect of a slope variable is bigger (smaller) in magnitude than its cumulative effect, indicates that prices overreact (under-react) to information in the limit order book.

5 Estimation Results

Table 2 presents some summary statistics across stocks from the estimation of the empirical model. We start our discussion with the effects of the ask-side slopes on prices. The immediate effects (the coefficient estimates on the first lag) of the lower levels are significantly positive for each stock separately. These immediate effects are also jointly significant across all stocks based on Bonferroni p-values. The signs of these observed effects are consistent with the basic law of supply and demand, as in Kalay and Wohl (2009). Indeed, an increase in the ask-side slope indicates a decrease in the selling pressure in the limit order book which in turn results in an increase in the prices and higher returns. These effects are significant not only statistically but also economically. To understand the economic magnitude of these coefficient estimates, one needs to take into account the fact that we consider the logarithm of slope measures in the VAR. For example, the log stock price increases by 0.227 basis points ($0.227\ b.p. = 0.346 \times 0.656$) at the next trade, following a one standard deviation[9] increase in the slope of the ask-side between the first and fifth levels. This effect is economically important given that the mean log return between two transaction periods is practically

[9] We compute the standard deviation of ask-side slope based on the first five levels using limit order book snapshots one millisecond before each transaction. We use the average of these standard deviations across 30 stocks.

zero. Turning our attention to the immediate effects of the higher levels reveals that they are also significantly positive for almost all stocks (28 out of 30) separately and for all stocks jointly based on Bonferroni p-values. However, they also exhibit more variation across stocks than the immediate effects of the lower levels. More importantly, the immediate effects of the higher levels of the ask side are on average stronger than those of the lower levels.

[Insert Table 2 here]

The cumulative effects (the sum of the coefficient estimates on all lags) of the lower levels are all positive, and significantly so, for almost all stocks (29 out of 30) in our sample. The cumulative effects of the higher levels are also positive for almost all stocks (27 out of 30), but significantly so for around half (16 out of 30) of the stocks. More importantly, the slopes of both the lower and higher levels have cumulative effects smaller in magnitude than their immediate effects. This in turn implies that the effects of further lags on ask side slopes are on average negative. To put it differently, there is a reversal of the positive immediate effects of ask side slope variables. This reversal is stronger for the slopes of the higher levels than the lower levels. As a result, the slopes of the lower levels have stronger cumulative effects that the slopes of the higher levels.

The bid-side slopes, not surprisingly, have negative effects on returns because an increase in the bid-side slope indicates a decrease in the buying pressure in the limit order book which in turn results in an increase in the prices and higher returns. Furthermore, the immediate effects of bid-side slopes are slightly bigger in magnitude and exhibit more heterogeneity across stocks than those of the corresponding ask-side slopes. Otherwise, we observe patterns in terms of the relations between the immediate and cumulative effects of lower and higher levels of the bid side very similar to those discussed above for the ask side.

These results already provide preliminary evidence in line with our expectations discussed in Section 4. First, the immediate effects of the slopes of the higher levels exhibit more heterogeneity and are on average bigger (in magnitude) than those of the lower levels. As discussed in Section 4, this suggests the existence of patient traders who place their limit order in the higher levels of the book, as mentioned in the discussion of our first set of testable hypotheses. Second, there seems to be a reversal of the immediate effects when we consider cumulative effects, providing preliminary evidence of overreaction to information embedded in the limit order book, as discussed in the motivation of our third set of testable hypotheses. Finally, this reversal effect is stronger for the higher levels than the lower levels, which in turn points to asymmetries in the effects of the lower and higher levels, as discussed while motivating our second set of hypotheses. In the next section, we analyze whether these preliminary results in line with our expectations are indeed statistically significant by testing the null hypotheses presented in Section 4.

6 Hypothesis Tests

As discussed in Section 4, in addition to our predictions on the signs of these effects, we also predict differences in the magnitude of these effects, if asymmetries in traders' level of patience is the main driving factor. To reiterate, we expect the slope of the higher levels to have a stronger effect on prices when patient traders place their orders in the higher levels of the limit order book, making it less convex. In this section, before establishing the link between traders' patience and the magnitudes of these effects, we first test whether the effects of the lower and higher levels are significantly different from each other based on the above estimation .

Table 3 presents some summary statistics for our first set of hypotheses on the equality of the immediate effects of the lower and higher levels ($H1a$ and $H1b$). We start our discussion with our results for the ask side. The null hypothesis is rejected at the 5% significance level for 27 out of 30 stocks. The corresponding Bonferroni p-value is practically zero, suggesting that the lower and higher levels of the ask side have significantly different immediate effects when we consider all stocks jointly. These results provide strong empirical evidence that the lower and higher levels of the ask side have immediate effects on returns that are significantly different from each other. More importantly, we find that the lower levels of the ask side have smaller immediate effects for 23 of 30 stocks in our sample and significantly so at the 5% level for 22 of these 23 stocks. As discussed in Section 4, these results indicate that patient traders are more present for these 22 stocks in our sample period. That said, there are also 7 stocks for which the lower levels of the ask side have bigger immediate effects than the higher levels, and significantly so for 5 of these 7 stocks. For these stocks, impatient traders seem to be more present in our sample period.

Our findings are very similar for the bid side. Briefly, the higher levels have significantly more negative immediate effects on returns than the lower levels for most of the stocks (24 out of 30) while the opposite is true for only 6 stocks and significantly so for only 4. These results also point to the existence of more patient traders for most of the stocks in our sample during the period considered.

[Insert Table 3 here]

Table 4 presents the results for H2a and H2b testing whether the overall effects based on all lags of the slopes of the lower and higher levels are equal. For both bid and ask side, the overall effects of the slopes of the lower and higher levels are significantly different from each other for all stocks in our sample, regardless of whether we consider the empirical evidence for each stock separately or jointly across stocks based on Bonferonni p-values. In other words, the evidence for the asymmetric effects of lower and higher levels becomes even stronger when we take into account the effects of lags further than the first one. This result is important because it shows that one cannot compute a single slope variable to capture the state of the limit order book, meaning that the shape of the limit order book is not linear. More importantly, it is useful to distinguish between the slopes of the different levels as they might have different effects on returns.

[Insert Table 4 here]

Finally, Table 5 presents the results for our third set of hypotheses on the relation between the immediate and cumulative effects of a given slope variable ($H3a$ and $H3b$). As shown above, the cumulative effects of all slope variables considered are smaller in magnitude than their immediate effects. The results in Table 5 show that these differences between the cumulative and immediate effects are statistically significant. These results in turn suggest that prices following a transaction overreact to information embedded in limit order book slopes, which is followed by a significant reversal after several transactions. Furthermore, this reversal is much more pronounced for the higher levels of both ask and bid sides. This in turn explains why the slopes of the higher levels have smaller (in magnitude) effects after several transactions than the slopes of the lower levels, which is the opposite of what we observe in terms of the relationship between their immediate effects.

[Insert Table 5 here]

Overall, our results in this section show that the ask- and bid-side slopes of the higher levels have significantly greater (in absolute value) immediate effects on prices than the corresponding slopes of the lower levels. There is a reversal of this pattern, due to price overreaction, when we consider the effects of slope variables after several transactions. These results suggest that the observed differences in the magnitudes of the effects might be due to the existence of patient traders in our sample. That said, these results do not provide direct empirical evidence linking traders' patience to the effects of the slope variables on prices. In the next section, we provide that empirical evidence.

7 Traders' Patience and the Asymmetries in the Effects of Slope Variables on Price Dynamics

To establish the link between traders' patience and the asymmetries in the effects of slope variables on prices, we perform an intraday analysis of how the effects of slope variables vary as traders' patience changes over the trading day. We first provide empirical evidence that there are more patient traders in the morning and less patient traders at the end of the trading day, as argued by Foucault et al. (2005). We then show that the ask- and bid-side slopes of the higher levels have significantly greater (in absolute value) immediate effects on prices than the corresponding slopes of the lower levels in the morning, while the opposite holds at the end of the trading day as discussed in Section 4.

Although traders' patience cannot be directly observed, Rosu (2009) shows that the limit order book should be less convex when patient traders have placed their orders at the higher levels of the ask. Hence, we first verify how the shape of the limit order book changes over the trading day. To do this, we first standardize the prices and (cumulative) quantities in each level of a given limit order book snapshot to numbers between zero and one so that we can average them over different stocks and time periods. We standardize the price at a given level by dividing the difference between this price and the midquote price by the difference between the price of the 20th level and the midquote price.[10] Similarly, we standardize the quantities in each level of a given limit order book snapshot by dividing them by the total quantity available in all 20 levels. Figure 2 presents the shape of the limit order between 9:00 and 10:00, 12:00 and 13:00, and 16:30 and 17:30, averaged over all stocks and trading days. Both the bid and ask sides of the limit order book exhibit, on average, a convex shape throughout the trading day. More importantly, we find that the limit order book exhibits a much less convex shape at the beginning of the trading day compared with the rest of the trading day. This in turn suggests that traders are more patient at the beginning of the trading day, especially during the opening hour, and they become less patient throughout the day. This is in line with our expectations based on the

[10] We use the negative of the standardized prices for the bid side for presentational purposes.

arguments in Foucault et al. (2005).

[Insert Figure 2 here]

To demonstrate this point further, we measure the convexity of the limit order book based on the ratio of the slopes of the higher and lower levels and present its evolution over the trading day. We consider level five of the limit order book as our breakpoint for our main convexity measure. In other words, we measure convexity as $S^A_{5,20,t}/S^A_{1,5,t} - 1$ for the ask side and $|S^B_{5,20,t}|/|S^B_{1,5,t}| - 1$ for the bid side. We also considered other breakpoints to measure convexity and our results remain similar given that the limit order book exhibits, on average, a convex shape as shown in Figure 2. The interpretation of this convexity measure is straightforward: A positive (negative) convexity implies that the slope of the higher levels is greater (smaller) than the slopes of the lower levels, while a zero convexity implies a linear limit order book. For a given stock, we first compute the convexity of the ask and bid sides of its limit order book for each snapshot and then average them over each hour of the trading day.[11] Figure 3 presents the convexity measures of bid and ask sides averaged over stocks and trading days in our sample. The average convexity is low at the beginning of the trading day, and increases almost monotonically over the trading day. This is in line with our expectations that there are more patient traders in the morning and less patient traders at the end of the trading day.

[Insert Figure 3 here]

To provide further evidence that our measure is a good proxy for limit order book convexity, we analyze whether it changes as a function of order size and volatility in line with Rosu's predictions. Specifically, Rosu (2009) argues that the convexity of the limit order book is lower (1) when large market orders are likely or (2) when the market is more volatile. To test the first prediction, we compute the convexity based on the snapshot of the limit order book one millisecond before each trade. We classify a given trade for a given stock into deciles based on breakpoints for those specific stocks. In other words, the definition of large and small market orders is stock-specific. Figure 4 presents the 25th, 50th and 75th percentiles of limit order book convexity for each trade decile. It shows that the shape of the limit order book is always convex, which corresponds to the second case suggested by Rosu where large market orders are unlikely. Also, according to the theoretical model of Rosu (2009), if patient traders can correctly anticipate the

[11] The trading hours on Xetra are between 9:00 and 17:30. We average over 30 minute interval between 16:00 and 16:30 so that the first and last periods in a trading day, which are our main interests, are both an hour long.

probability of large market orders, the convexity should be monotonically decreasing along with the size of the trade. However, our results show that the relationship between convexity (both bid and ask side) and trade size is U-shaped. That is, the limit order book is relatively more convex right before small and large market orders. As mentioned by Rosu, the relation between the shape of the limit order book and size of the trade is often ambiguous in the empirical literature. Our results appear to support the explanation that patient traders can correctly anticipate the probability of small and medium size market order arrivals, but not the probability of large market order arrivals.

We consider a similar approach to analyze the second prediction. We first compute the daily realized volatility for each stock-day pair as the sum of squared 5-minute midquote returns. We classify a given trading day for a given stock into deciles based on breakpoints for that specific stock. We then compute the 25th, 50th and 75th percentiles of the average limit order book convexity for stock-day pairs that are in a given volatility decile. Figure 5 presents these results and shows that there is a slight negative relation between limit order book convexity and volatility as predicted by Rosu (2009). This negative relation is more evident for the bid side than the ask side and for higher levels of convexity (75% percentile) than median or low levels of convexity (50% and 25% percentiles). These results provide some supporting evidence that the limit order book convexity is lower on days with higher volatility than days with lower volatility.

[Insert Figures 4 and 5 here]

Rosu (2009) also compares his model based on waiting costs of patient traders with the model of Glosten (1994), which is based on asymmetric information and perfect competition. He argues that one can test his model by checking whether the expected increase in the bid-ask spread conditional on the arrival of the next order is positive. The intuition is that the bid-ask spread should increase on average in order to compensate patient traders who wait in the order book. In other words, the bid-ask spread should be higher not only when there are more patient traders but also when there are more trades. A high volume combined with a low bid-ask spread might indicate a lack of patient traders in the market, while a high volume combined with a high bid-ask spread might point to the existence of patient traders. Thus, another way to understand how traders' patience changes over the trading day is to jointly analyze the evolution of bid-ask spread and volume over the trading day. To this end, we first present in panel (a) of Figure 6 the number of transactions (averaged over stocks and trading days) over the trading day.[12] The number of transactions in our data set exhibits the well-known U-shape pattern over the trading day where it is high both at the beginning and end of the trading day. We then present the evolution of the bid-ask spread over the trading day. To do this, we use the book bid-ask spread, defined as the depth weighted ask price minus the depth weighted bid price over all levels, instead of the difference between best ask and bid prices, which might be tick constrained. Furthermore, given that the Rosu's model involves the whole book and not just the top of the book, the book spread measure is a better fit and is not limited by the minimum tick in the market. We compute the book spread based on snapshots of the limit order book right after a transaction.[13] For a given hour of the trading day, we then average the book spread over all the stocks and trading days and present its evolution over the trading day in panel (b) of Figure 6. The book bid-ask spread is high at the beginning of the day and decreases almost monotonically throughout the trading day. These intraday patterns for volume and bid-ask spread point to traders' patience rather than their information as the driving force. More importantly, the patterns also suggest that there might be more patient traders at the beginning of the trading day and less at the end of the day, in line with our other sets of results.

[Insert Figure 6 here]

So far, we have provided empirical evidence in support of several predictions of Foucault et al. (2005) that point to the existence of more patient traders at the beginning of the trading day and less towards the end of the trading day, which is in line with our expectations. We now turn our attention to how the effect of slope variables changes over the trading day. We expect the slope variables of the higher levels to have greater (in absolute value) immediate effects on prices than the corresponding slopes of the lower levels at the beginning of the trading day. To test this, we first estimate the empirical model in Equation 4 for each stock, trading hour and day separately and obtain a panel data set of coefficient estimates.[14] Figure 7 presents these coefficient estimates for different trading hours averaged over stocks and trading days. The positive lines in Figure 7 are for the ask side, suggesting that the coefficient estimates are on average positive for the ask side, while the negative ones are for the bid side, suggesting that they are on average negative. This is consistent with our findings based on the estimation using the whole sample. Regardless of the side

[12] We also considered volume in Euros rather than the number of transactions and observed the similar U-shape pattern over the trading day.

[13] We also considered snapshots of the limit order book right before a transaction and our results are very similar to those presented.

[14] We estimate the empirical model in Equation 4 only if there are at least 100 observations for a given stock, trading hour and day.

considered, the coefficient estimates on all slope variables, on average, decrease in magnitude over the trading day. This pattern is more pronounced for the coefficients on the slope variables of the higher levels. More importantly, the coefficients on the slope variables of the higher levels are, on average, greater in magnitude than the corresponding coefficients on the slope variables of the lower levels at the beginning of the trading day and smaller in magnitude at the end of the trading day. In other words, the slope variables of the higher levels have greater (in absolute value for the bid side) immediate effects on prices than the corresponding slopes of the lower levels for an average stock in the DAX30 index in our sample period. This in turn provides empirical evidence that slope variables of the higher levels should have greater (in absolute value) immediate effects on prices than the corresponding slopes of the lower levels when there are more patient traders at the beginning of the trading day, in line with our expectations.

[Insert Figure 7 here]

We now test whether this relationship between traders' patience and the effects of slope variables on prices is statistically significant. To do this, we regress the differences between the coefficient estimates of the higher and lower levels on dummy variables for the first and last hour of trading. These results are presented in Table 6. For the ask side, the dummy variable for the first hour of trading has significantly positive coefficient estimates, whereas the dummy variable for the last hour of trading has significantly negative coefficient estimates. The results are very similar for the bid side with opposite signs. This suggests that slope variables of the higher levels should have significantly greater (in absolute value) immediate effects on prices than the corresponding slopes of the lower levels at the beginning of the trading day when there are more patient traders and significantly smaller immediate effects at the end of the trading day when there are less patient traders. We also considered dummy variables based on half-hour and two-hour periods at the beginning and end of the trading day. Our findings are very similar to those presented in Table 6.

[Insert Table 6 here]

Finally, we analyze how the differences between the coefficient estimates of the higher and lower levels change directly as a function of traders' patience, proxied by the convexity of the limit order book. Given that the convexity of the limit order book is a relatively noisy proxy for traders' patience, we use a non-parametric approach to the effect of this noise on our results. Specifically, we first sort the coefficient estimates for the ask (bid) side into deciles based on the convexity of the ask (bid) side and compute the average difference between the coefficient estimates of the higher and lower levels. Figure 8 presents these average differences for ask and bid sides along with their 95% confidence intervals. For the ask side, the difference is significantly positive when the convexity is low. This suggests that the effect of the higher levels on prices is significantly greater than that of the lower levels. As the convexity increases, this difference decreases and it becomes significantly negative for high levels of convexity. The results are very similar for the bid side with opposite signs. These results provide additional empirical evidence in line with our expectations. More precisely, the slopes of the higher levels have significantly greater (in absolute value) immediate effects on prices than the slopes of the lower levels when there are more patient traders, as proxied by convexity, whereas the opposite holds when there are less patient traders in the market.

[Insert Figure 8 here]

8 Economic Value of Asymmetric Effects

8.1 Basic Trading Strategy

Our empirical results discussed so far provide statistically significant evidence in support of certain asymmetries between the effects of different slope measures on price dynamics. In this section, we show that these asymmetries can also be economically significant. We do this by comparing the performances of a high-frequency day-trading strategy that ignores these asymmetries and a strategy that uses them. Our unrestricted strategy employs the return equation of the VAR system, Equation (5a), to forecast midquote returns at each transaction period, whereas the competing strategies employ restricted versions of this equation such that a chosen pair of slope variables has symmetric effects on price dynamics.[15]

To be consistent with the idea of high-frequency trading, we consider the possibility of trading at every transaction period and, thus, focus on the economic value of the asymmetries in the overall effects of different slope variables, rather than the asymmetries in their cumulative effects . To this end, we impose the restrictions implied by hypotheses

[15] In all these forecasting models, we choose to exclude the contemporaneous effect of trade direction on returns. As discussed above, any contemporaneous effect of the trade direction on returns is not a causal relationship, even if the trade direction is observed right before the return is calculated. Nevertheless, we also considered trading strategies based on forecasting models that include the contemporaneous effect of trade direction on returns. Our results based on this VAR model are similar to those based on the restricted VAR model and, thus, are not presented here in detail but are available upon request.

H1a, H1b, H2a, and H2b one at a time on Equation (5a) and use these versions as our forecasting models in these so-called restricted strategies. The restricted strategies thus ignore a certain type of asymmetry in forecasting price movements. Comparing the performances of the restricted strategies and the unrestricted strategy allows us to evaluate the economic value of different asymmetries. To give a more concrete example, in the restricted strategy based on H1a, we restrict the coefficients on the first lags on ask-side slopes based on higher and lower levels to be the same. This forces the ask-side slope measures to have symmetric immediate effects on prices while allowing for all other potential asymmetries in their effects on prices. If the performance of the unrestricted strategy is higher than that of the corresponding restricted strategy, we argue that the asymmetry between the immediate effects of these two slope measures is economically important.

The trading strategy we consider is similar to that discussed in Kozhan and Salmon (2012) and can be summarized as follows: At each transaction period t for a given stock, we take a snapshot of the limit order book immediately (less than a millisecond) after observing the transaction, in line with our empirical model discussed in Section 3. We then compute the forecast of the midquote return in the next transaction period $t + 1$, $\hat{r}_{t+1}$, based on this snapshot and a given forecasting model, and then reevaluate our existing position if the return forecast is either greater than a nonnegative threshold, κ, or less than $-\kappa$. We consider a forecast greater than κ, i.e. $\hat{r}_{t+1} > \kappa$, to be a buy signal and do one of the following actions depending on our existing position in the stock: (1) buy one share of the stock if we do not already have an existing position in the stock; (2) buy two shares of the stock if we have an existing short position in the stock, i.e. close the short position and take a long position of one share; (3) do nothing if we already have a long position. Similarly, we consider a forecast less than $-\kappa$, i.e. $\hat{r}_{t+1} < -\kappa$, to be a sell signal and do one of the following actions depending on our existing position in the stock: (1) short-sell one share of the stock if we do not already have an existing position in the stock; (2) short-sell two shares of the stock if we have an existing long position in the stock, i.e. close the long position and take a short position of one share; (3) do nothing if we already have a short position. We implicitly assume that our trading does not alter the dynamics of the relationship between returns, trade directions, and limit order book. We believe that this is a reasonable assumption because we consider trading at most two shares at a time and holding a short or long position of one share at any point in time. The effect of our trading and position should be negligible given that the limit order book depth of stocks in our sample is large.

We also close any existing position at the end of each trading day and, thus, do not hold an overnight position, in line with the idea of day trading. We evaluate the performance of each strategy based on its average cumulative monthly return.

In this paper, we focus on in-sample results based on the model estimated using tick-by-tick data from June 2011 due to the sheer size of the data set as well as the time it takes to run the trading algorithms.[16] Table 7 presents the differences between the monthly cumulative returns of the unrestricted and restricted trading strategies. A positive number implies that the unrestricted strategy provides, on average, higher monthly cumulative returns than the trading strategy that uses the restricted forecasting model implied by hypotheses in the corresponding column heading. In other words, a positive number suggests that a given asymmetry is economically important. Compared with the strategies imposing the restrictions implied by H1a and H1b on the empirical model, the unrestricted strategy provides a higher cumulative return for 25 and 24 out of 30 stocks and statistically so for 9 and 11 stocks, respectively. Averaged over all stocks, the unrestricted strategy provides cumulative returns that are, respectively, 16.2 and 16.9 basis points higher than the strategies imposing the restrictions implied by H1a and H1b, with Bonferroni p-values lower than 5%. The results are even stronger for the economic importance of asymmetries in the overall effects of slope measures. Indeed, the unrestricted strategy provides cumulative returns that are higher than the strategies imposing the restrictions implied by H2a and H2b for 26 and 25 out of 30 stocks respectively, and statistically so for 14 and 10 stocks respectively, with an average outperformance of 24.6 and 24.2 basis points, respectively. Overall, these results suggest that asymmetries in the immediate and overall effects of both ask- and bid-side slopes are economically important; ignoring them might have cost a trader between 16 and 25 basis points in return in June 2011.

[Insert Table 7 here]

8.2 Alternative Trading Settings

In this section, we analyze the robustness of our results on the economic importance of asymmetries to alternative empirical choices. We start with several remarks regarding our main set of results presented in Table 7. First, as in Kozhan and Salmon (2012), κ can be considered a parameter to filter out potentially weak signals, and trading

[16] Our results are in-sample results in the sense that we use the same data period to estimate parameters of the forecasting models and to test the relative performances of trading strategies based on these forecasting models.

frequency decreases as higher values of κ are considered. In our main set of results, we consider a κ of zero. In other words, we only consider the sign, not the strength, of our forecasts. This allows us to analyze the economic value of different asymmetries in forecasting the direction of prices. In robustness checks, we also consider alternative values for κ and, thus, take the strength of our forecasts into account in our trading strategies. Second, our main empirical results are based on the assumption that we can trade at midquote prices, instead of the more realistic assumption of trading at the best bid and ask prices. We made this assumption mainly to be consistent with our empirical models, which forecast midquote returns. This assumption also allows us to analyze the economic value of the asymmetries without having to worry about transaction costs. Nevertheless, we test the robustness of our results to trading at the best bid and ask, rather than midquote, prices. Third, we calculate our trading signals based on the snapshots of the limit order book immediately (less than a millisecond) after a transaction and assume that we can trade at the prices in that snapshot of the limit order book. However, in reality we can neither observe a transaction (and the snapshot of the limit order book right after it) instantaneously nor trade at prices observed in this snapshot of the limit order book. This is due to latency in different legs of the trading process where information first flows from the exchange to the trader's system, and is then processed by the trader, and an order is finally sent from the trader's system to the exchange. Thus, a trader will trade based on information and prices different from those in the observed snapshot of the limit order book.[17] Ideally, the effects of the three legs of the information latency on the economic values of asymmetries should be examined separately, yet this is difficult because we need to know the time duration of each leg. Instead, we take a very simplistic approach and assume that we can only trade at prices observed in the snapshot of the limit order book 500 milliseconds after a transaction.

Trading at Different Thresholds

We first show that asymmetries continue to be economically important when we take into account not only the sign but also the strength of forecasts by considering alternative values of the threshold parameter κ. For each stock, we consider values of κ that correspond to 10%, 20%, 30%, 40% and 50% of the standard deviation of its tick-by-tick returns in June 2011. As mentioned above, κ can be considered a parameter to filter out potentially weak signals, and trading frequency decreases as higher values of κ are considered. In other words, we can trade potentially at each

[17] Hasbrouck (2018) discusses the costs associated with traders' latencies.

observed transaction period if we completely ignore the trading signal. For example, in our main set of results where we set κ to zero, we trade, on average, approximately 30% to 35% of the time, i.e. 30% to 35% of all transaction periods, for a given stock in a given day. On the other hand, when we set κ to 50% of the corresponding standard deviation, we trade less than 1% of the time. Thus, choosing higher values of κ also allows us to analyze the effect of the trading frequency on our results in addition to the effect of jointly considering the sign and strength of our forecasts. We find that the number of trades indeed decreases on average when we consider higher values of κ, as expected. Furthermore, the absolute performances of all trading strategies, also not presented, decrease as we trade less and less frequently at higher values of κ.

[Insert Table 8 here]

Table 8 presents the difference between the cumulative returns of the unrestricted strategy and the trading strategy that uses the restricted forecasting model implied by hypotheses in the corresponding column heading, averaged over all stocks and trading days in June 2011. All return differences are positive and statistically significant at 5%, suggesting that the asymmetries in the immediate and overall effects of slope measures continue to be economically important when we consider alternative values of κ. There is a small tendency for the return differences to increase as we consider higher values of κ, suggesting that taking both the sign and strength of the signal provides the day trader with further economic profits. The asymmetries in the overall effects of slope measures continue to be slightly more important economically compared with the asymmetries in the immediate effects for different values of κ.

Trading at the Best Bid and Ask Prices

We now consider the effect of trading at the best bid and ask prices corresponding to the direction of the required position. We should note that monthly cumulative returns of the unrestricted and restricted strategies are, on average, negative when we assume that we trade at the best bid and ask prices. This is in contrast to mostly positive average returns when we assume that we can trade at midquote prices. In other words, positive profits of all strategies are negated by transaction costs, in line with the findings of Kozhan and Salmon (2012). Nevertheless, Table 9 shows that the return differences between the unrestricted and restricted strategies are positive not only on average but also for most firms in our sample when we trade at the best bid and ask prices with a κ of zero. This is significantly so for all asymmetries at the 5% significance level when the empirical evidence for all stocks is considered jointly and for eight to twelve stocks depending on the asymmetry when the empirical evidence for each stock is considered separately. We also find that these results mostly hold when we consider higher values of κ up to 25% of the standard deviation of tick-by-tick returns for a given stock.

[Insert Table 9 here]

Trading at a Slower Speed

As mentioned above, in our main sets of results we assume that we can observe a transaction and the associated snapshot of the limit order book and trade at prices in this snapshot without any delay. However, in reality any trader who wants to implement such a trading strategy faces several delays. First, the trader observes the transaction and the associated snapshot of the limit order book with a delay, which corresponds to the time it takes for this information to travel from the exchange to the trader's system. Second, whether the trader is a human or a computer, it takes time to process this information. Finally, if the trader decides to trade based on this information, it takes time for the order to arrive at and be executed by the exchange's system.

With each delay, the market conditions can undoubtedly change and the trader's information can become stale quite rapidly in today's extremely fast markets. As a result, the trader would make decisions based on stale information. Ideally, one should distinguish between the first leg of the trading process during which the information travels from the exchange's systems to the trader's system (and is processed by the trader) and the second leg, during which the information flows in the other direction. It would be intriguing to analyze the effects of these legs of the information delay on the economic values of asymmetries separately, yet this is relatively difficult because these delays can be random and can change significantly from trader to trader depending on the physical proximity of the trader to the exchange and the information processing speeds.

We thus adopt an approach similar to the one discussed in Kozhan and Tham (2012), which can be considered as an approximation for the effect of these delays on the economic value of asymmetries. However, contrary to Kozhan and Tham (2012), we assume there is a positive delay in the first leg of the trading process and that the trader can observe a transaction (and the associated snapshot of the limit order book) and react to it after having considered the state of the entire book, which is an important consideration in our model. The violation of this assumption in reality might make the trader change the forecasting model due to the associated delay. In the second leg of the trading process, we assume that there is still a positive delay. In other words, once a trader places an order, it takes time to be executed by the exchange's systems, which is a reasonable assumption for the year 2011. We consider a total latency of 500 milliseconds for all legs of the trading process. This implies that the orders will be executed at prices in the snapshot of the limit order book observed exactly 500 milliseconds after a transaction. The average latency on the Xetra system, i.e. the average time required for an order to travel from a trader's system across the network to its backend and for confirmation of its receipt to be sent back to the trader, is about 13 milliseconds but we do not know which information is treated by the average trader. A delay of 500 milliseconds including the treatment of the information in the entire order book can be considered as relatively long compared with the average latency in the system. Nevertheless, we also considered a delay of 100 milliseconds, and our results remain similar to those presented in Table 10.

[Insert Table 10 here]

Our results can be summarized as follows: We find that the absolute performance of all trading strategies when we trade with a latency of 500 milliseconds is much lower than that based on a latency of zero milliseconds, in line with our expectations. More importantly, Table 10 presents the differences between the monthly cumulative returns of the unrestricted and restricted trading strategies when we trade at midquote prices with a κ of zero and a latency of 500 milliseconds, and shows that our main set of results is mostly robust to trading slower speeds.[18] For example, the number of stocks for which asymmetries are both economically important and statistically significant is similar to our main set of results. However, as expected, the return differences are, on average, smaller than those in our main set of results, which is also reflected in the Bonferroni p-values of positive return differences.

9 Conclusion

In this paper, we analyze whether different parts of the LOB affect immediate and cumulative price dynamics in line with our predictions based on recent theoretical models. To do this, we use high-frequency LOB data from the

[18] We also considered trading at the best bid and ask prices as well as at higher values of κ, and with a latency of 500 milliseconds. Our results remain qualitatively similar.

Xetra electronic trading system for all 30 stocks in the DAX30 index. We distinguish between not only the bid and ask sides, but also their lower and higher levels, since the recent theoretical literature shows that traders might strategically choose to place their orders in different levels of the book depending on various factors. We consider one such factor, namely, traders' patience, and we analyze how different levels of the LOB affect future price dynamics based on this factor. If traders' patience is present in the data, we expect the ask-side slope to have a positive effect on prices and the bid-side slope to have a negative one, regardless of the levels used to measure them. Furthermore, we also expect the slope of the higher levels to have a stronger effect on prices when the LOB is less convex, which would indicate the presence of patient traders in the higher levels.

We analyze these predictions by estimating a separate linear vector autoregression that includes the midquote return, trade direction, and four slope measures, i.e., the bid- and ask-side slopes based on the lower and higher levels, for each stock. We find that ask-side slopes have mostly positive effects on prices, while those of the bid-side slopes are mostly negative, regardless of the levels used to measure them, suggesting that traders' patience might be present in our sample. We then turn our attention to the relative magnitudes of these effects. We find that the ask- and bid-side slopes of the higher levels have significantly greater (in absolute value) immediate effects on prices than do the corresponding slopes of the lower levels, for most of the stocks based on the estimation of our model using the whole sample period. This finding suggests that the differences observed between the effect magnitudes might be due to the existence of patient traders in our sample, but no formal link can be established between these two phenomena. To do this, we consider an intraday analysis of how traders' patience and the effects of slope measures on prices change over the trading day. In line with our expectations, we find that the effects of the slope variables of the higher levels are indeed greater in magnitude than those of the lower levels, at the beginning of the trading day, when traders are more patient, as proxied by the convexity of the LOB.

Having found statistically significant evidence in support of asymmetries between the effects of different slope measures on price dynamics, we also demonstrate that these asymmetries can be economically significant. We do this through a simple high-frequency day-trading exercise comparing the performance of strategies that ignore the asymmetries and a strategy that uses them. In this framework, we show that the unrestricted strategy provides monthly profits that are, on average, 16 to 25 basis points higher than each restricted strategy.

We would like to thank to an anonymous referee, Yann Bilodeau, Michael Brennan, Dennis Chung, Sophie Moinas, Maureen O'Hara, Stephen Sapp, Gabriel Yergeau and participants at the Northern Finance Association Meetings and Financial Management Association Asia/Pacific Meetings for comments on earlier versions of this paper. We would also like to thank Yann Bilodeau for his help in constructing the data set. This research received financial support from SSHRC Canada [435-2019-1183], the Canada Foundation for Innovation [36738] and Fonds de Recherche sur la Société et la Culture de Québec [2019-NP-253336].

References

Beltran-Lopez, H., Giot, P., Grammig, J., 2009. Commonalities in the Order Book. Financial Markets and Portfolio Management 23, 209–242.

Bhattacharya, A., Saar, G., 2016. Limit Order Markets under Asymmetric Information. Available at SSRN: https://ssrn.com/abstract=3688473 .

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

Bilodeau, Y., 2013. Xetra parser. [computer software], HEC Montreal .

Brogaard, J., Hendershott, T., Riordan, R., 2016. Price Discovery without Trading: Evidence from Limit Orders. Available at SSRN: https://ssrn.com/abstract=2655927 .

Cao, C., Hansch, O., Wang, X., 2009. The Information Content of an Open Limit-Order Book. Journal of Futures Markets 29, 16–41.

Chakravarty, S., Holden, C. W., 1995. An Integrated Model Of Market And Limit Orders. Journal of Financial Intermediation 4, 213 – 241.

Dufour, A., Engle, R. F., 2000. Time and the Price Impact of a Trade. Journal of Finance 55, 2467–2498.

Foucault, T., Kadan, O., Kandel, E., 2005. Limit Order Book as a Market for Liquidity. Review of Financial Studies 18, 1171–1217.

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

Glosten, L. R., Milgrom, P. R., 1985. Bid, Ask and Transaction Prices In a Specialist Market with Heterogeneously Informed Traders. Journal of Financial Economics 14, 71–100.

Goettler, R. L., Parlour, C. A., Rajan, U., 2009. Informed Traders and Limit Order Markets. Journal of Financial Economics 93, 67 – 87.

Griffiths, M. D., Smith, B. F., Turnbull, D. A. S., White, R. W., 2000. The Costs and Determinants of Order Aggressiveness. Journal of Financial Economics 56, 65 – 88.

Hasbrouck, J., 1991. Measuring the Information Content of Stock Trades. Journal of Finance 46, 179–207.

Hasbrouck, J., 2018. High-Frequency Quoting: Short-Term Volatility in Bids and Offers. Journal of Financial and Quantitative Analysis 53, 613–641.

Kalay, A., Wohl, A., 2009. Detecting Liquidity Traders. Journal of Financial and Quantitative Analysis 44, 29–54.

Kaniel, R., Liu, H., 2006. So What Orders Do Informed Traders Use? Journal of Business 79, 1867–1913.

Kozhan, R., Salmon, M., 2012. The Information Content of A Limit Order Book: The Case of an FX Market. Journal of Financial Markets 15, 1–28.

Kozhan, R., Tham, W. W., 2012. Execution Risk in High-Frequency Arbitrage. Management Science 58, 2131 – 2149.

Kumar, P., Seppi, D. J., 1994. Limit and Market Orders with Optimizing Traders. Working paper Carnegie Mellon University .

Kyle, A. S., 1985. Continuous Auctions and Insider Trading. Econometrica 53, 1315–1335.

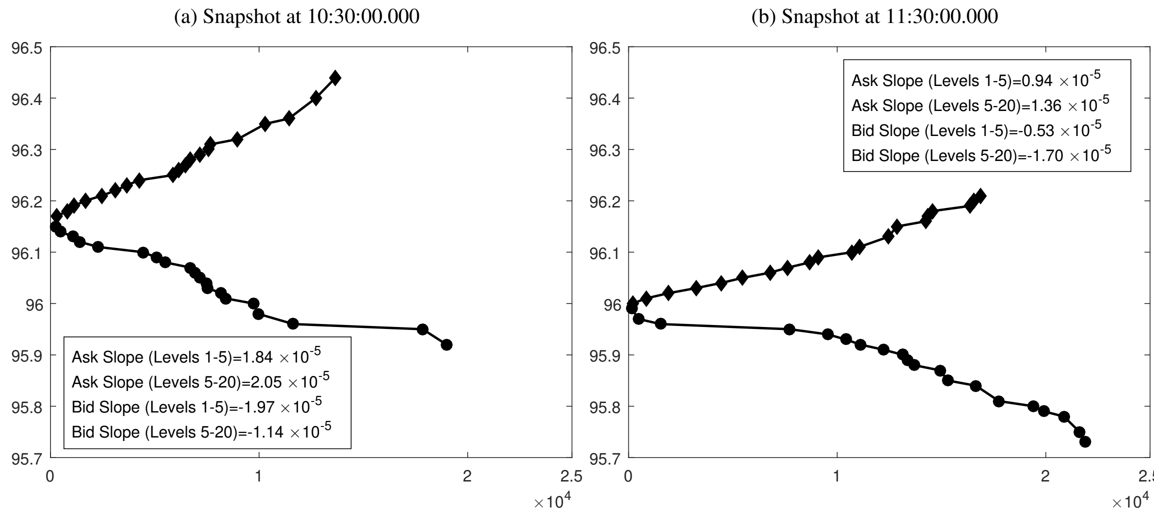
Lee, C., Ready, M., 1991. Inferring Trade Direction from Intraday Data. Journal of Finance 46, 733–746.

Newey, W. K., West, K. D., 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 55, 703–708.

Rock, K., 1996. The Specialist's Order Book and Price Anomalies. Working Paper, Harvard University .

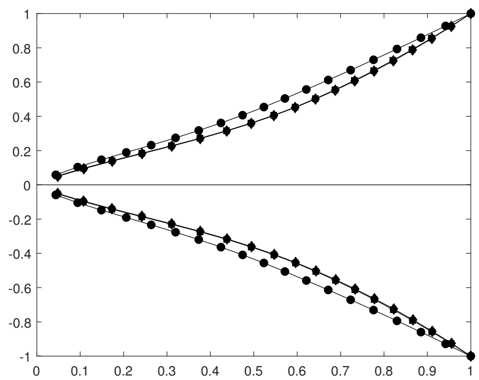
Rosu, I., 2009. A Dynamic Model of the Limit Order Book. Review of Financial Studies 22, 4601.

Figure 1: Snapshots of the Limit Order Book for ALV on June 1, 2011



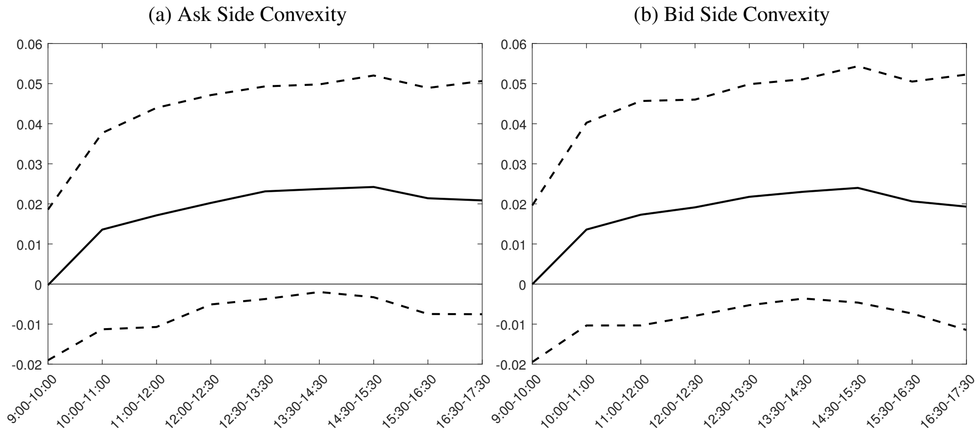
Note: This figure presents two snapshots of the limit order book for ALV on 1 June 2011. Panel (a) and (b) present the snapshot at 10:30:00.000 and 11:30:00.000, respectively. The x axis presents the cumulative quantity available and the y axis presents the price in the first 20 levels of the limit order book. The diamonds and squares represent different levels of the ask and bid sides, respectively. The annotation displays the corresponding ask- and bid-side slope measures between the first and fifth levels and between the fifth and twentieth levels.

Figure 2: Average Shape of the Limit Order Book at the Beginning, Middle and End of the Trading Day



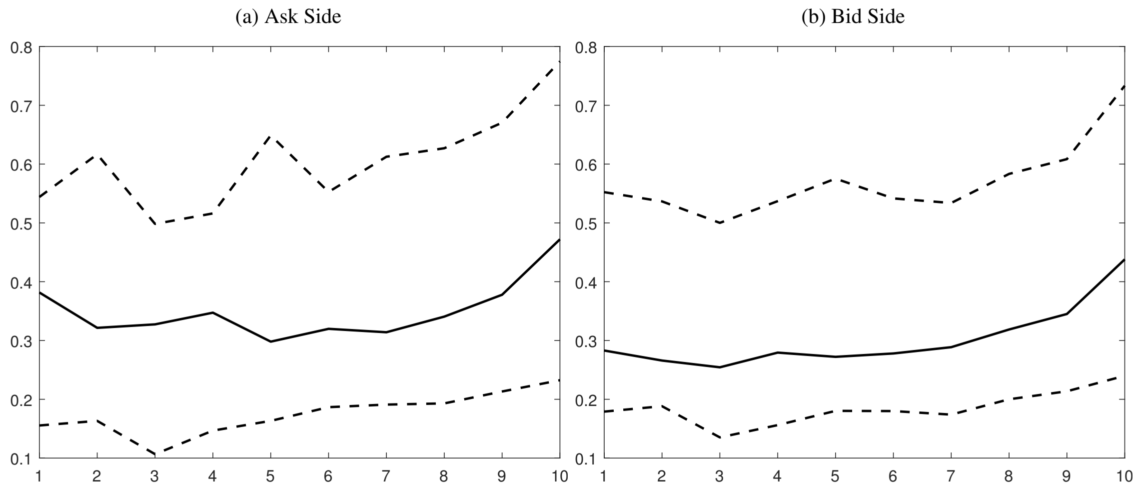
Note: This figure presents the shape of the normalized limit order book during a given hour at the beginning (9:00-10:00 (circles)), middle (12:00 and 13:00 (squares)) and end (16:30-17:30 (diamonds)) of the trading day averaged over all snapshots, stocks and trading days. We normalize the prices and quantities at each level so that we can average them over different stocks and trading days. The price at a given level for a given snapshot is normalized by dividing the difference between that price and the midquote price by the difference between price in the 20th level and the midquote price for that snapshot. Mathematically, the normalized prices for the ask and bid sides at time t are respectively given by $(P_l, t^A - M_t)/(P_20, t^A - M_t)$ and $(M_t - P_l, t^B)/(M_t - P_20, t^B)$ where M_t is the midquote given by $(P_1, t^A + P_2, t^B)/2$ and P_l, t^A and P_l, t^B are, respectively, the price at level l of the ask and bid sides for $= l = 1, 2, \ldots, 20$. To make the graph easier to read, we use the negative of the normalized prices for the bid side. The cumulative quantity at a given level for a given snapshot is normalized by dividing it by the total cumulative quantity available in the first 20 levels of the corresponding side. The normalized price and cumulative quantity are by definition between zero and one for all limit order book snapshots and thus can be averaged across stocks and time periods. For a given level of the limit order book, the figure presents the normalized price and cumulative quantity averaged over all snapshots for all stocks and trading days during the same hour.

Figure 3: Limit Order Book Convexity Over the Trading Day



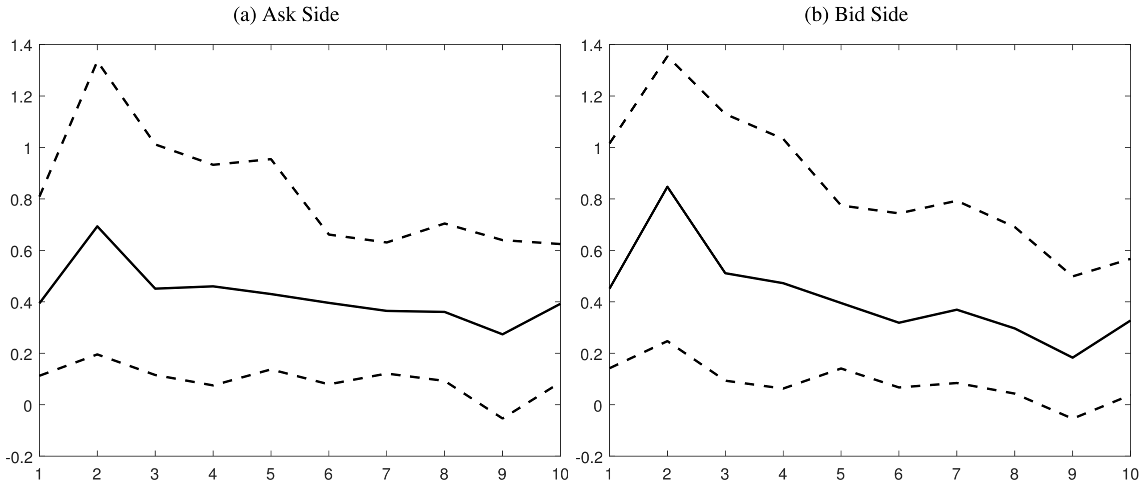
Note: This figure presents the average convexity measures of the ask (panel (a)) and bid (panel (b)) sides over the trading day. The convexity of the limit order book is defined as the ratio of the slopes of the higher and lower levels. Mathematically, the convexity is measured as $S^A_{5,20,t}/S^A_{1,5,t} - 1$ for the ask side and $|S^B_{5,20,t}|/|S^B_{1,5,t}| - 1$ for the bid side, where $S^i_{l_1,l_2,t}$ is the slope of a given side $i = A, B$ between levels l_1 and l_2 of the snapshot at time t as described in the text. For a given stock, we first compute the convexity of ask and bid sides of its limit order book for each snapshot and then average them over each hour of the trading day. For a given trading hour, we then compute the average convexity measure (solid line) and the 25^{th} and 75^{th} quantiles (dashed lines) over all stocks and trading days.

Figure 4: The Relation Between Limit Order Book Convexity and Trade Size



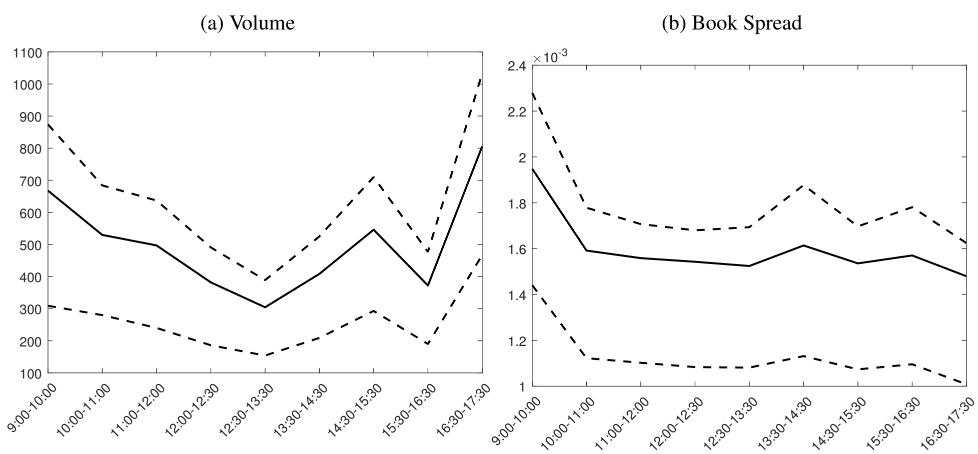
Note: This figure presents how the convexities of the ask (panel (a)) and bid (panel (b)) sides right before a transaction vary as a function of the number of shares traded in that transaction. The convexity of the limit order book is defined as the ratio of the slopes of the higher and lower levels. Mathematically, the convexity is measured as $S^A_{5,20,t}/S^A_{1,5,t} - 1$ for the ask side and $|S^B_{5,20,t}|/|S^B_{1,5,t}| - 1$ for the bid side, where $S^i_{l_1,l_2,t}$ is the slope of a given side $i = A, B$ between levels l_1 and l_2 of the snapshot at time t as described in the text. For a given transaction, we measure limit order book convexity based on the snapshot one millisecond prior to that transaction. We classify each trade for a given stock into deciles based on breakpoints for that specific stock. The x-axis presents the deciles of the trade size and the y-axis presents the 25^{th} (lower dashed lines), 50^{th} (middle solid lines) and 75^{th} (upper dashed lines) percentiles of convexity measures for that decile.

Figure 5: The Relation Between Limit Order Book Convexity and Daily Volatility



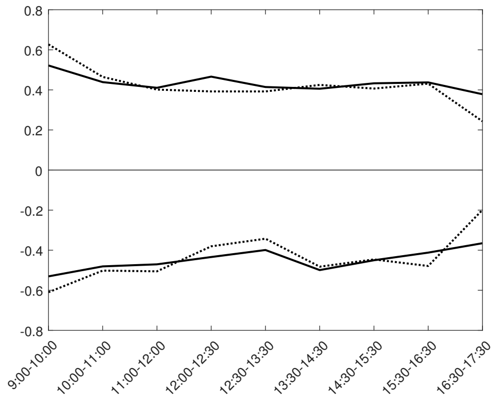
Note: This figure presents how the convexities of the ask (panel (a)) and bid (panel (b)) sides vary as a function of daily volatility. The convexity of the limit order book is defined as the ratio of the slopes of the higher and lower levels. Mathematically, the convexity is measured as $S^A_{5,20,t}/S^A_{1,5,t} - 1$ for the ask side and $|S^B_{5,20,t}|/|S^B_{1,5,t}| - 1$ for the bid side, where $S^i_{l_1,l_2,t}$ is the slope of a given side $i = A, B$ between levels l_1 and l_2 of the snapshot at time t as described in the text. For a given stock-day pair, we compute the daily realized volatility as the sum of squared 5-minute midquote returns for that stock on that day. We classify each trading day for a given stock into deciles based on breakpoints for that specific stock. The x-axis presents the deciles of daily realized volatility and the y-axis presents the 25^{th} (lower dashed lines), 50^{th} (middle solid lines) and 75^{th} (upper dashed lines) percentiles of convexity measures for that decile.

Figure 6: Book Spread and Volume Over the Trading Day



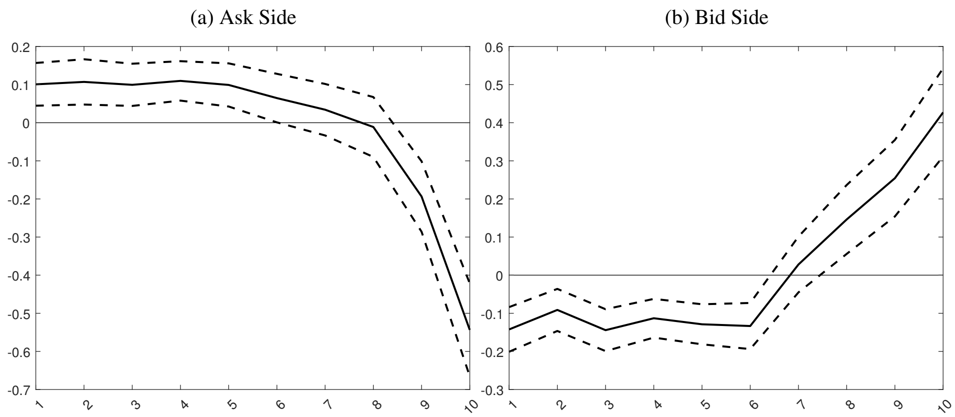
Note: This figure presents how the volume (panel (a)) and book spread (panel (b)) vary over the trading day. We compute the book bid-ask spread, defined as the depth weighted ask price minus the depth weighted bid price over the first 20 levels, based on the snapshot of the limit order book one millisecond after a transaction. For a given stock, we first compute book spread for each snapshot after a transaction and then average them over each hour of the trading day. For a given trading hour, we then compute the average book spread (solid line) and the 25^{th} and 75^{th} percentiles (dashed lines) over all stocks and trading days. The volume of a transaction is defined as the number of shares traded. For a given hour of the trading day, we compute the average volume (solid line) and the 25^{th} and 75^{th} percentiles (dashed lines) over all stocks and trading days. The x-axis presents the hours of the trading day, with the exception of the thirty minute period between 12:00-12:30. The y-axis presents the average, the 25^{th} and 75^{th} percentiles of the variable of interest.

Figure 7: Immediate Effects of Slope Variables Over the Trading Day



Note: This figure presents how the immediate effects of different slope variables on returns vary over the trading day. The immediate effects are the coefficient estimates on the first lags of a given slope variable from the estimation of the empirical model in Equation 4. We estimate the empirical model in Equation 4 for each stock, trading day and hour separately. For a given trading hour, we compute the average of the immediate effects over all stocks and trading days. The figure presents these averages over different hours of the trading day. The immediate effects of the slopes of both lower and higher levels of the ask side are on average positive and thus presented above the zero line in the graph. The solid line above the zero line presents the immediate effect of the slope of the lower levels of the ask side, while the dashed line above the zero line presents that of the higher levels. The immediate effects of the slopes of both the lower and higher levels of the bid side are on average negative and thus presented below the zero line in the graph. The solid line below the zero line presents the immediate effect of the slope of the lower levels of the ask side while the dashed line below the zero line presents that of the higher levels.

Figure 8: The Effect of Convexity on the Difference Between the Immediate Effects of Slopes of the Higher and Lower Levels



Note: This figure presents how the differences between the immediate effects of the slopes of the higher and lower level varies with limit order book convexity. We first estimate the empirical model in Equation 5 for each stock, trading hour and day with at least 100 observations separately and obtain a panel data set of coefficient estimates. We then compute the differences between the immediate effects of the higher and lower levels for ask and bid sides. We obtain the average ask and bid side convexities using limit order book snapshots for each stock, trading hour and day. We then group each stock, trading hour and day into one of the deciles based on the average convexity of the ask side in panel (a) and that of the bid side in panel (b). The x-axis presents these deciles and the y-axis presents the mean (solid line) and its 95% confidence intervals (dashed lines) of the difference between the immediate effects of the higher and lower levels that are in a given decile.

Table 1: List of Stocks and Their Characteristics

Ticker	Company Name	Avg. Market Capitalization (in billion Euros)	Avg. Daily Trading Volume (in million shares)	Avg. Daily Turnover (in percentage)	Avg. Daily Return (in percentage)	Std. Dev. of Daily Return (in percentage)
ADS	ADIDAS AG	10.372	1.084	0.518%	0.082%	1.758%
ALV	ALLIANZ SE	41.942	2.579	0.568%	-0.080%	1.485%
BAS	BASF SE	55.074	4.355	0.474%	-0.115%	1.658%
BAYN	BAYER AG	44.467	3.062	0.370%	-0.116%	1.419%
BEI	BEIERSDORF AG	10.199	0.476	0.210%	0.031%	0.913%
BMW	BAYER MOTOREN WERKE AG	35.468	2.969	0.493%	0.141%	1.844%
CBK	COMMERZBANK	11.422	47.425	1.390%	-0.342%	3.730%
DAI	DAIMLER AG	51.039	4.824	0.453%	-0.049%	1.562%
DB1	DEUTSCHE BOERSE AG	10.375	1.398	0.717%	-0.015%	1.520%
DBK	DEUTSCHE BANK AG	35.703	6.862	0.816%	0.003%	1.742%
DPW	DEUTSCHE POST AG	15.627	4.436	0.367%	0.026%	1.172%
DTE	DEUTSCHE TELEKOM	45.090	14.698	0.339%	0.031%	1.396%
EOAN	E.ON SE	40.915	10.682	0.534%	-0.174%	1.524%
FME	FRESENIUS MEDICAL CARE AG&CO	14.601	0.717	0.241%	-0.051%	1.170%
FRE	FRESENIUS SE & CO KGAA	9.837	0.342	0.219%	0.039%	1.269%
HEI	HEIDELBERGCEMENT AG	8.188	1.036	0.552%	-0.354%	1.889%
HEN3	HENKEL AG & CO KGAA	8.164	0.786	0.441%	-0.034%	1.161%
IFX	INFINEON TECHNOLOGIES AG	7.604	12.378	1.139%	-0.020%	1.886%
LHA	DEUTSCHE LUFTHANSA AG	6.435	3.707	0.810%	0.005%	1.576%
LIN	LINDE AG	18.788	0.518	0.306%	0.077%	1.238%
MAN	MAN SE	12.301	0.956	0.678%	-0.092%	1.564%
MEO	METRO AG	13.989	1.115	0.344%	-0.271%	1.384%
MRK	MERCK KGAA	4.670	0.483	0.748%	0.197%	1.413%
MUV2	MUNICH RE CO	19.884	0.868	0.461%	-0.058%	1.181%
RWE	RWE AG	22.476	2.708	0.517%	-0.205%	1.405%
SAP	SAP SE	50.418	3.847	0.314%	-0.049%	1.205%
SDF	K&S AG	9.614	1.185	0.619%	0.107%	1.367%
SIE	SIEMENS AG	80.435	3.146	0.344%	-0.097%	1.306%
TKA	THYSSENKRUPP AG	15.510	3.114	0.605%	0.121%	2.002%
VOW3	VOLKSWAGEN AG	20.302	1.186	0.697%	0.188%	1.803%
Mean		24.364	4.765	0.543%	-0.036%	1.551%
Median		15.568	2.643	0.505%	-0.027%	1.452%
Min		4.670	0.342	0.210%	-0.354%	0.913%
Max		80.435	47.425	1.390%	0.197%	3.730%

Note: This table presents the average market capitalization (in billion Euros), the average daily trading volume (in million shares), the average daily turnover (in percentage) defined as the trading volume divided by the number of shares outstanding, the average daily (log) return and its standard deviation (in percentage) for the 30 stocks in the DAX30 index over July 2010, May 2011, and June 2011. All data are from the Compustat Global Security Daily files and based on the primary issues.

Table 2: Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask Slope (Levels 1-5)		Ask Slope (Levels 5-20)		Bid Slope (Levels 1-5)		Bid Slope (Levels 5-20)	
	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.
Mean	0.346	0.169	0.435	0.039	-0.365	-0.168	-0.464	-0.038
Median	0.328	0.167	0.440	0.031	-0.355	-0.170	-0.472	-0.030
Std. Dev.	0.125	0.044	0.166	0.043	0.141	0.041	0.200	0.039
Positive	30	30	29	27	0	0	1	4
Significantly Positive	30	28	28	19	0	0	1	1
Negative	0	0	1	3	30	30	29	26
Significantly Negative	0	0	0	3	30	28	29	18
Bonferonni p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the immediate and cumulative effects of limit order book slopes on returns from the estimation of the model in Equation 5. We estimate the VAR system for each stock separately using all tick-by-tick data available for that stock in our sample over July 2010, May 2011, and June 2011. The immediate effect is the coefficient estimate on the first lag of the corresponding slope variable in the return equation 5a ($\alpha_{z,1}$). The cumulative effect is the sum of coefficient estimates on all lags of the slope variable ($\sum_{\tau=1}^{5} \alpha_{z,\tau}$). The mean, median and standard deviations are computed across the 30 stocks in our sample. Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is positive (negative). Significantly Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is significantly positive (negative) at the 5% significance level.

Table 3: The Relation Between the Immediate Effects of the Slopes of the Higher and Lower Levels on Returns

Number of stocks for which	Ask Side	Bid Side
$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	23	6
$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	7	24
$H1 : \boldsymbol{\alpha}_{z,1}(S_{1,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5%	25	26
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	19	5
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	6	21
Bonferonni p-value for $H1$ across stocks	0.000	0.000

Note: This table presents the relation between the immediate effects of the slopes of the higher and lower levels on returns and results for our first set of hypotheses on the equality of the immediate effects of lower and higher levels ($H1a$ and $H1b$). All numbers, with the exception of the last row, are out of 30 stocks in our sample. The row "$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which the immediate effect of the lower levels is smaller (greater) than that of the higher levels is rejected at the 5% significance level. The row "$H1 : \boldsymbol{\alpha}_{z,1}(S_{1,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5%" presents the number of stocks for which our first hypothesis on the equality of the immediate effects of the slopes of the higher and lower levels. The row "$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which our hypothesis is rejected at the 5% significance level in favor of the alternative that the immediate effect of the slope of the lower levels is smaller (greater) than that of the higher levels. The row "Bonferonni p-value for $H1$ across stocks" presents the p-value for testing $H1$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is computed as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H1$ for the i^{th} stock.

Table 4: The Relation Between the Overall Effects of the Slopes of the Higher and Lower Levels on Returns

	Ask	Bid
Mean of F-statistic for $H2$	178.140	185.874
Median of F-statistic for $H2$	167.821	170.897
Std. Dev. of F-statistic for $H2$	78.368	80.115
Nb. of Stocks for which $H2$ is rejected at 5%	30	30
Bonferroni p-value	0.000	0.000

Note: This table presents results on whether the overall effects (based on all lags considered) of the slopes of the lower and higher levels of the same side are equal ($H2a$ and $H2b$). We test $H2$ for ask and bid sides and for each stock separately. The mean, median and standard deviation are computed over these individual F-statistics for $H2$. The row "Nb. of Stocks for which $H2$ is rejected at 5% " presents the number of stocks (out of 30) for which our second hypothesis that the overall effects of the slopes of the lower and higher levels of the same side are equal is rejected at the 5% significance level. The row "Bonferonni p-value for $H2$ across stocks" presents the p-value for testing $H2$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is computed as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H2$ for the i^{th} stock.

Table 5: The Difference Between the Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask		Bid	
	Low	High	Low	High
Mean	0.177	0.396	-0.197	-0.426
Median	0.156	0.407	-0.176	-0.440
Std. Dev.	0.101	0.141	0.120	0.178
Positive	30	29	0	1
Sign. Positive	29	29	0	1
Negative	0	1	30	29
Sign. Negative	0	0	29	29
Bonferonni pvalue	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the difference between the immediate and cumulative effects of a given slope variable on returns. The difference between the immediate and cumulative effects of a given slope variable on returns is computed as the difference between the coefficient estimate on the first lag of a slope variable in the return equation and the sum of the coefficients on all its five lags. It is also given as the sum of the coefficient estimates on the second to fifth lags of this slope variable. We compute these differences for each stock separately and the mean, median and standard deviation are computed across the 30 stocks in our sample. Positive (Negative) and Sign. Positive (Sign. Negative) present the number of stocks for which the difference is positive (negative) and significantly so at the 5% significance level, respectively. These differences correspond to our third hypothesis, related to the over- or under-reaction of returns to information embedded in the limit order book. The results under Ask-Low correspond to $H3a : \boldsymbol{\alpha}_{z,1}(S^A_{1,5,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{1,5,t})$ and Ask-High to $H3a^* : \boldsymbol{\alpha}_{z,1}(S^A_{5,20,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{5,20,t})$. The results under Bid-Low and Bid-High are for the corresponding hypotheses for the bid sides, i.e. $H3b$ and $H3b^*$. The row "Bonferonni p-value for $H3$ across stocks" presents the p-value for testing $H3$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is computed as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H3$ for the i^{th} stock.

Table 6: The Effect of Trading Hour on the Difference Between Immediate Effects of the Higher and Lower Level Slopes

	Ask	Bid
Constant	-0.0129	0.0004
First Hour Dummy	0.1181***	-0.0796**
Last Hour Dummy	-0.1224***	0.1657***

Note: This table presents the coefficient estimates from a regression of the difference between immediate effects of the higher and lower level slopes on a constant and on dummy variables for the first and last hour of trading. We estimate the empirical model in Equation 4 for each stock, trading hour and day separately and obtain a panel data set of the coefficient estimates on the first lag of different slope variables, i.e. their immediate effects. We regress the differences between the immediate effects of the higher and lower level slopes on a constant and on dummy variables for the first and last hour of trading. ***, ** and * denote significant coefficient estimates at the 1%, 5% and 10% levels, respectively.

Table 7: Relative Performances of the Restricted Trading Strategies

	Restrictions on the Immediate Effect		Restrictions on the Overall Effect	
	H1a	H1b	H2a	H2b
Mean	0.162%	0.169%	0.246%	0.242%
Median	0.126%	0.159%	0.252%	0.222%
nb positive	25	24	26	25
nb positive & signf	9	11	14	10
Bonferonni p-val	0.015	0.000	0.014	0.001

Note: This table presents summary statistics on the differences between the monthly cumulative returns of the unrestricted and restricted trading strategies. A positive number implies that the unrestricted strategy provides, on average, higher monthly cumulative returns than the trading strategy that uses the restricted forecasting model implied by the hypotheses in the corresponding column heading. Mean and median are calculated over all stocks in our sample. nb pos and nb pos & signf present the number of stocks out of 30 in our sample for which the return difference is positive and significantly positive, respectively. Bonf. p-val is the p-value based on the Bonferroni correction for stocks with positive return differences and is calculated as $\min(1, \min(p_1, \cdots, p_m) \times m)$ where the minimum is calculated only over stocks for which the return difference is positive and, thus, m is the number of stocks for which the return difference is positive. p_i is the p-value of the hypothesis testing whether monthly cumulative returns of the unrestricted and a given restricted trading strategy are equal.

Table 8: Average Relative Performances of the Restricted Trading Strategies at Different Thresholds

	Restrictions on the Immediate Effect		Restrictions on the Overall Effect	
κ	H1a	H1b	H2a	H2b
0	0.162%	0.169%	0.246%	0.242%
0.1	0.238%	0.259%	0.375%	0.412%
0.2	0.118%	0.085%	0.245%	0.294%
0.3	0.462%	0.416%	0.740%	0.620%
0.4	0.604%	0.583%	0.910%	0.953%
0.5	0.425%	0.450%	0.696%	0.688%

Note: This table presents the difference between the monthly cumulative returns of the unrestricted trading strategy and the trading strategy that uses the restricted forecasting model implied by the hypotheses in the corresponding column heading, averaged over all stocks and trading days in June 2011, for different values of the threshold parameter, κ. We trade less frequently as higher values of κ are considered, which filters out potentially weak signals. A positive number implies that the unrestricted strategy provides, on average, higher monthly cumulative returns than the trading strategy that uses the restricted forecasting model implied by the hypotheses in the corresponding column heading.

Table 9: Relative Performances of the Restricted Strategies When Trading at the Best Bid and Ask Prices

	Restrictions on the Immediate Effect		Restrictions on the Overall Effect	
	H1a	H1b	H2a	H2b
Mean	0.140%	0.139%	0.196%	0.141%
Median	0.172%	0.165%	0.267%	0.138%
nb positive	22	22	19	19
nb positive & signf	8	12	12	10
Bonferonni p-val	0.024	0.000	0.001	0.002

Note: This table presents the difference between the monthly cumulative returns of the unrestricted and the trading strategy that uses the restricted forecasting model implied by hypotheses in the corresponding column heading, when we trade at the best bid and ask prices instead of midquote prices and a κ of zero. A positive number implies that the unrestricted strategy provides, on average, higher monthly cumulative returns than the trading strategy that uses the restricted forecasting model implied by hypotheses in the corresponding column heading. Mean and median are calculated over all stocks in our sample. nb pos and nb pos & signf present the number of stocks out of 30 in our sample for which the return difference is positive and significantly positive, respectively. Bonf. p-val is the p-value based on the Bonferroni correction for stocks with positive return differences and is calculated as $\min(1, \min(p_1, \cdots, p_m) \times m)$ where the minimum is calculated only over stocks for which the return difference is positive and, thus, m is the number of stock for which the return difference is positive. p_i is the p-value of the hypothesis testing whether monthly cumulative returns of the unrestricted trading strategy and a given restricted trading strategy are equal.

Table 10: Relative Performances of the Restricted Strategies when Trading at a Slower Speed

	Restrictions on the Immediate Effect		Restrictions on the Overall Effect	
	H1a	H1b	H2a	H2b
Mean	0.057%	0.125%	0.106%	0.123%
Median	0.050%	0.131%	0.114%	0.130%
nb positive	21	26	23	24
nb positive & signf	14	15	13	15
Bonferonni p-val	0.208	0.247	0.228	0.225

Note: This table presents the differences between the monthly cumulative returns of the unrestricted and restricted trading strategies when we trade at a latency of 500 milliseconds and at midquote prices with a κ of zero. A positive number implies that the unrestricted strategy provides, on average, higher monthly cumulative returns than the trading strategy that uses the restricted forecasting model implied by hypotheses in the corresponding column heading. Mean and median are calculated over all stocks in our sample. nb pos and nb pos & signf present the number of stocks out of 30 in our sample for which the return difference is positive and significantly positive, respectively. Bonf. p-val is the p-value based on the Bonferroni correction for stocks with positive return differences and is calculated as $\min(1, \min(p_1, \cdots, p_m) \times m)$ where the minimum is calculated only over stocks for which the return difference is positive and, thus, m is the number of stocks for which the return difference is positive. p_i is the p-value of the hypothesis testing whether monthly cumulative returns of the unrestricted trading strategy and a given restricted trading strategy are equal.

Online Appendix

October 16, 2021

A.1 Robustness Checks

A.1.1 Sampling Frequency

In our main set of results, we follow the previous literature and estimate the Hasbrouck's VAR model using tick-by-tick data. This sampling approach has several advantages and allow one to compare some of our results with those in the literature. It also has few disadvantages. For example, as discussed above, one needs to choose whether to take limit order book snapshots right before or after a trade. Furthermore, this sampling approach allows one to analyze only the conditional (on the arrival of a market order) effects of a given slope variable on returns. To overcome these issues, we also consider sampling at a regular frequency of 100 milliseconds. As mentioned above, we can sample all variables at millisecond frequency. However, sampling every millisecond results in too many observations as well as too many periods with zero trades. On the other hand, sampling at a very low frequency results in fewer number of observations. Sampling at a 100 millisecond frequency represents a fair compromise. We take snapshots of the limit order book every 100 milliseconds and also compute the midquote returns based on these snapshots. We use signed volume in number of shares for x_t instead of the trade direction. This sampling approach has several advantages compared with the first sampling approach. First, this approach allows us to avoid choosing whether to take limit order book snapshots right before or after a trade. Second, it also allows us to analyze the effects of slope variables on returns on an ex-ante basis without having to condition on the arrival of a trade. Finally, it allows us to analyze the effects of order size on the limit order book slopes at different times over the trading day.

Here, we mostly focus on the last point and analyze the relation between order size and limit order book slopes over the trading day. We nevertheless first summarize the estimation and hypothesis test results based on this sampling approach, which are presented in Tables A.1-A.4. The immediate and cumulative effects of slope variables based on this sampling approach are very similar to those in our main set of results based on tick-by-tick sampling, with two relatively minor differences. First, the effects are smaller in magnitude although still statistically significant with the similar signs. Second, the immediate effects of the higher levels are bigger (in magnitude) than those of the lower levels only when we consider their median across stocks but not their mean, unlike our main set of results. This difference can be explained by the overreaction of prices to information in limit order book slopes observed in our main set of results. Recall that in our main set of results the coefficient estimates on further lags of slope variables have on average the opposite sign of the coefficient estimates on the first lag. We also find in our main results that these opposite effects are much stronger for slopes of the higher levels than that of the lower levels. These opposite effects of further lags observed in tick-by-tick data are captured by the immediate effects and more strongly so for slopes of higher levels when we consider sampling at 100 millisecond intervals, which is on average longer than the duration between two transactions. This difference is reflected in the results for the first hypothesis. Specifically, we find in Tables A.1-A.4 that the immediate effects of higher levels of the ask side are bigger than those of the lower levels for around half of the stocks in our sample, which is in contrast to our main results where this is true for most of the stocks. The results for the other two hypotheses based on 100 millisecond sampling are very similar to those based on tick-by-tick sampling.

We now turn our attention to the relation between order size and limit order book slopes over the trading day. Our analysis resembles the one discussed in Section 7. Specifically, we consider data for each trading hour of the day separately and estimate the empirical specification in Equation 5 using data sampled every 100 millisecond from that hour. We then compute the immediate effect of signed volume on a given slope measure as the coefficient estimate on its first lag in the equation for that slope measure in the VAR system. Similarly, we compute its cumulative effect as the sum of the coefficients on all of its five lags. Figure A.1 presents how the median immediate and cumulative effects of signed volume on slope measures change over the trading day. Several interesting facts emerge. The immediate effect of signed volume on slopes of the lower levels is practically zero regardless of the side considered. This suggests that the signed volume does not really affect liquidity supply in the limit order book. This changes when we consider higher levels of the limit order book. An increase in the signed volume increases slopes of the higher levels of the ask side and much more so of the bid side. In other words, a smaller sell volume or a bigger buy volume decreases the liquidity supply in the higher levels of the limit order book. The fact that this effect is stronger for the slope of the higher levels of the bid side suggests that the buyers in the limit order book become even more impatient following a decrease in the selling pressure or an increase in the buying pressure observed in market orders. This is an interesting finding that has not been previously reported in the literature. Turning our attention to the cumulative effects reveals similar results. One important difference is that the cumulative effects of signed volume on slopes of the higher levels are smaller in magnitude. For example, the cumulative effects of signed volume on the slope of the higher levels of the ask side become practically zero. These findings suggest that there is a certain degree of reversal in the immediate effects of signed volume on the higher levels. Finally, in terms of intraday trends in these effects, we find that the effects decrease over the trading day but only slightly so. This in turn suggests that the limit order traders are less sensitive to changes in buying or selling pressure from the market orders towards the end of the trading day than at the beginning where there are more patient traders.

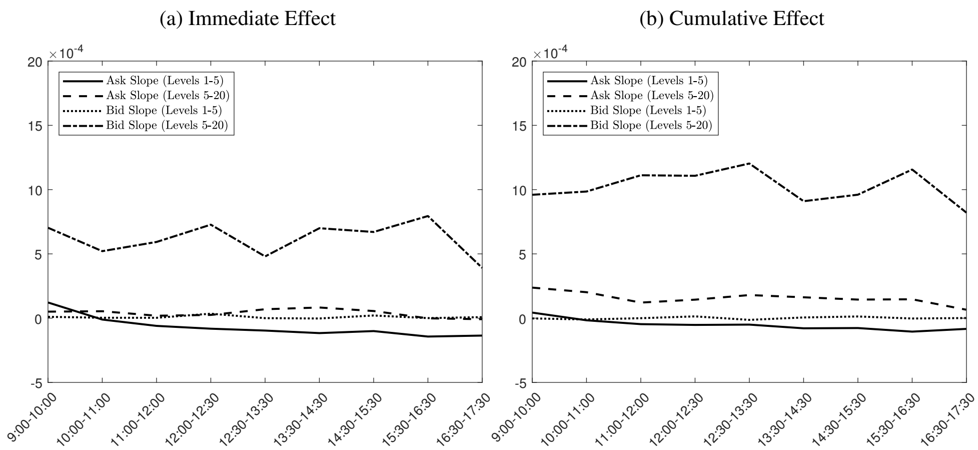
A.1.2 Additional Control Variables

In our main empirical model, we include three sets of variables, namely returns, trade direction and slope variables. As discussed in Section 3, we consider this model because it can be seen as a natural extension of Hasbrouck's model, which is theoretically motivated. In the original Hasbrouck model, the trade direction is the main variable whose effect on returns is of interest. In our extension of this model, the trade direction serves as a control variable when analyzing the asymmetries in the effects of slope variables on returns. From an empirical standpoint, we can also include other variables that control for differences in liquidity and trading activity across stocks. To this end, we include the bid-ask spread, the volume (in euros) of a market order, the duration in millisecond between two market orders as state variables in our main empirical specification. The bid-ask spread is measured as the difference between the best bid and ask prices of the same limit order book snapshot used for measuring the slope variables right before a trade. The volume of a market order is measured as the number of shares transacted multiplied by the transaction price. The duration is simply the time in milliseconds between two transactions. The estimation and hypothesis test results from this model are presented in Tables A.5-A.8. They are very similar to our main set of empirical results. This in turn suggests that controlling for differences in liquidity and trading activity across stocks does not significantly alter our main conclusions on the asymmetries in the effects of slope variables on returns.

A.1.3 Other Robustness Checks

In this subsection, we list the additional robustness checks that we performed. First, we consider alternative truncation points for the infinite sum. Specifically, we consider three and eight lags instead of the five lags considered in our main set of results. These results are presented in Tables A.9-A.16. Second, we consider alternative definitions and cutoff points for the lower and higher levels: (1) we change the cutoff level to two and define lower and higher levels based on the first two and the second to twentieth levels, respectively; (2) we exclude the first level from the definition of lower levels and define it to be levels between the second and fifth, and we define higher levels as levels between the fifth and the twentieth. These results are presented in Tables A.17-A.24. Finally, we estimate our main empirical specification separately for each of the three months in our sample. These results are presented in Tables A.25-A.36. Overall, the estimation and hypothesis test results based on these alternative empirical choices are similar to our main sets of results with some minor differences and this suggests that our results are robust to a battery of additional checks.

Figure A.1: The Effect of Signed Volume on Limit Order Book Slopes



Note: This figure presents how the immediate (panel (a)) and cumulative (panel (b)) effects of signed volume on the slopes of the higher and lower levels of the ask and bid sides vary over the trading day. For a given trading hour of the day, we estimate the model in Equation 5 using data sampled every 100 millisecond for all stocks and trading days. The immediate effect of signed volume on a given slope measure is the coefficient estimate on its first lag in the equation for that slope measure in the VAR system. The cumulative effect of signed volume on a given slope measure is the sum of the coefficients on all of its five lags. The x-axis presents the trading hours and the y-axis presents the median of the immediate and cumulative effects of signed volume on slope measures.

The following four tables (Tables A.1-A.4) present the estimation and hypothesis test results from the estimation of the empirical model using data sampled at regular frequency of 100 milliseconds as discussed in Section A.1.1. Tables A.1-A.4 can be directly compared to Tables 2-5 of the main text.

Table A.1: Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask Slope (Levels 1-5)		Ask Slope (Levels 5-20)		Bid Slope (Levels 1-5)		Bid Slope (Levels 5-20)	
	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.
Mean	0.105	0.009	0.090	0.002	-0.102	-0.009	-0.087	-0.002
Median	0.089	0.008	0.090	0.002	-0.094	-0.008	-0.095	-0.001
Std. Dev.	0.042	0.002	0.050	0.002	0.043	0.002	0.037	0.002
Positive	30	30	29	27	0	0	1	6
Significantly Positive	30	28	26	21	0	0	0	4
Negative	0	0	1	3	30	30	29	24
Significantly Negative	0	0	0	3	29	28	28	23
Bonferonni p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the immediate and cumulative effects of limit order book slopes on returns from the estimation of the model in Equation 5. We estimate the VAR system for each stock separately using data sampled at regular frequency of 100 milliseconds for that stock in our sample over July 2010, May 2011, and June 2011. The immediate effect is the coefficient estimate on the first lag of the corresponding slope variable in the return equation 5a ($\alpha_{z,1}$). The cumulative effect is the sum of coefficient estimates on all lags of the slope variable ($\sum_{\tau=1}^{5} \alpha_{z,\tau}$). The mean, median and standard deviations are computed across the 30 stocks in our sample. Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is positive (negative). Significantly Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is significantly positive (negative) at the 5% significance level.

Table A.2: The Relation between the Immediate Effects of the Slopes of the Higher and Lower Levels on Returns

Number of stocks for which	Ask Side	Bid Side
$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	14	18
$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	16	12
$H1 : \boldsymbol{\alpha}_{z,1}(S_{1,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5%	24	27
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	12	15
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	12	12
Bonferonni p-value for $H1$ across stocks	0.000	0.000

Note: This table presents the relation between the immediate effects of the slopes of the higher and lower levels on returns and results for our first set of hypotheses on the equality of the immediate effects of lower and higher levels ($H1a$ and $H1b$). All numbers with the exception of the last row are out of 30 stocks in our sample. The row "$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which the immediate effect of the lower levels is smaller (greater) than that of the higher levels. The row "$H1 : \boldsymbol{\alpha}_{z,1}(S_{1,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5%" presents the number of stocks for which our first hypothesis on the equality of the immediate effects of the slopes of the higher and lower levels. The row "$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which our hypothesis is rejected at the 5% significance level in favor of the alternative that the immediate effect of the slope of the lower levels is smaller (greater) than that of the higher levels. The row "Bonferonni p-value for $H1$ across stocks" presents the p-value for testing $H1$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H1$ for the i^{th} stock.

Table A.3: The Relation between the Overall Effects of the Slopes of the Higher and Lower Levels on Returns

	Ask	Bid
Mean of F-statistic for $H2$	1773.841	1757.622
Median of F-statistic for $H2$	1665.655	1960.088
Std. Dev. of F-statistic for $H2$	674.646	686.745
Nb. of Stocks for which $H2$ is rejected at 5%	30	30
Bonferroni p-value	0.000	0.000

Note: This table presents results on whether the overall effects (based on all lags considered) of the slopes of the lower and higher levels of the same side are equal ($H2a$ and $H2b$). We test $H2$ for ask and bid sides and for each stocks separately. The mean, median and standard deviation are computed over these individual F-statistics for $H2$. The row "Nb. of Stocks for which $H2$ is rejected at 5% " presents the number of stocks (out of 30) for which our second hypothesis that the overall effects of the slopes of the lower and higher levels of the same side are equal. The row "Bonferonni p-value for $H2$ across stocks" presents the p-value for testing $H2$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H2$ for the i^{th} stock.

Table A.4: The Difference between the Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask		Bid	
	Low	High	Low	High
Mean	0.096	0.088	-0.094	-0.086
Median	0.079	0.087	-0.085	-0.092
Std. Dev.	0.042	0.050	0.042	0.036
Positive	30	29	0	1
Sign. Positive	30	28	0	1
Negative	0	1	30	29
Sign. Negative	0	1	30	29
Bonferonni pvalue	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the difference between the immediate and cumulative effects of a given slope variable on returns. The difference between the immediate and cumulative effects of a given slope variable on returns is computed as the difference between the coefficient estimate on the first lag of a slope variable in the return equation and the sum of the coefficients on all its five lags. It is also given as the sum of the coefficient estimates on the second to fifth lags of this slope variable. We compute these differences for each stock separately and the mean, median and standard deviation are computed across the 30 stocks in our sample. Positive (Negative) and Sign. Positive (Sign. Negative) present the number of stocks for which the difference is positive (negative) and significantly so at the 5% significance level, respectively. These differences correspond to our third hypothesis, related to the over- or under-reaction of returns to information embedded in the limit order book. The results under Ask-Low correspond to $H3a : \boldsymbol{\alpha}_{z,1}(S^A_{1,5,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{1,5,t})$ and Ask-High to $H3a^* : \boldsymbol{\alpha}_{z,1}(S^A_{5,20,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{5,20,t})$. The results under Bid-Low and Bid-High are for the corresponding hypotheses for the bid sides, i.e. $H3b$ and $H3b^*$. The row "Bonferonni p-value for $H3$ across stocks" presents the p-value for testing $H3$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H3$ for the i^{th} stock.

The following four tables (Tables A.5-A.8) present the estimation and hypothesis test results where we include the bid-ask spread, the volume (in euros) of a market order, the duration in millisecond between two market orders as state variables in our main empirical specification. Tables A.5-A.8 can be directly compared to Tables 2-5 of the main text.

Table A.5: Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask Slope (Levels 1-5)		Ask Slope (Levels 5-20)		Bid Slope (Levels 1-5)		Bid Slope (Levels 5-20)	
	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.
Mean	0.345	0.169	0.436	0.037	-0.366	-0.169	-0.463	-0.040
Median	0.325	0.164	0.458	0.031	-0.352	-0.172	-0.447	-0.030
Std. Dev.	0.129	0.043	0.162	0.043	0.138	0.042	0.209	0.041
Positive	30	30	29	27	0	0	1	4
Significantly Positive	30	28	28	17	0	0	1	1
Negative	0	0	1	3	30	30	29	26
Significantly Negative	0	0	0	3	30	28	29	18
Bonferonni p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the immediate and cumulative effects of limit order book slopes on returns from the estimation of the model in Equation 5 that also includes the bid-ask spread, the volume (in euros) of a market order, the duration in millisecond between two market orders as state variables. We estimate the VAR system for each stock separately using all tick-by-tick data available for that stock in our sample over July 2010, May 2011, and June 2011. The immediate effect is the coefficient estimate on the first lag of the corresponding slope variable in the return equation 5a ($\alpha_{z,1}$). The cumulative effect is the sum of coefficient estimates on all lags of the slope variable ($\sum_{\tau=1}^{5} \alpha_{z,\tau}$). The mean, median and standard deviations are computed across the 30 stocks in our sample. Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is positive (negative). Significantly Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is significantly positive (negative) at the 5% significance level.

Table A.6: The Relation between the Immediate Effects of the Slopes of the Higher and Lower Levels on Returns

Number of stocks for which	Ask Side	Bid Side
$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	23	6
$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	7	24
$H1 : \boldsymbol{\alpha}_{z,1}(S_{1,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5%	27	26
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	21	4
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	6	22
Bonferonni p-value for $H1$ across stocks	0.000	0.000

Note: This table presents the relation between the immediate effects of the slopes of the higher and lower levels on returns and results for our first set of hypotheses on the equality of the immediate effects of lower and higher levels ($H1a$ and $H1b$). All numbers with the exception of the last row are out of 30 stocks in our sample. The row "$\boldsymbol{\alpha}_{z,1}(S_{2,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$\boldsymbol{\alpha}_{z,1}(S_{2,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which the immediate effect of the lower levels is smaller (greater) than that of the higher levels. The row "$H1 : \boldsymbol{\alpha}_{z,1}(S_{2,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5%" presents the number of stocks for which our first hypothesis on the equality of the immediate effects of the slopes of the higher and lower levels. The row "$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{2,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{2,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which our hypothesis is rejected at the 5% significance level in favor of the alternative that the immediate effect of the slope of the lower levels is smaller (greater) than that of the higher levels. The row "Bonferonni p-value for $H1$ across stocks" presents the p-value for testing $H1$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H1$ for the i^{th} stock.

Table A.7: The Relation between the Overall Effects of the Slopes of the Higher and Lower Levels on Returns

	Ask	Bid
Mean of F-statistic for $H2$	177.638	180.073
Median of F-statistic for $H2$	169.501	160.602
Std. Dev. of F-statistic for $H2$	76.562	83.045
Nb. of Stocks for which $H2$ is rejected at 5%	30	30
Bonferroni p-value	0.000	0.000

Note: This table presents results on whether the overall effects (based on all lags considered) of the slopes of the lower and higher levels of the same side are equal ($H2a$ and $H2b$). We test $H2$ for ask and bid sides and for each stocks separately. The mean, median and standard deviation are computed over these individual F-statistics for $H2$. The row "Nb. of Stocks for which $H2$ is rejected at 5% " presents the number of stocks (out of 30) for which our second hypothesis that the overall effects of the slopes of the lower and higher levels of the same side are equal. The row "Bonferonni p-value for $H2$ across stocks" presents the p-value for testing $H2$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H2$ for the i^{th} stock.

Table A.8: The Difference between the Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask		Bid	
	Low	High	Low	High
Mean	0.176	0.399	-0.197	-0.423
Median	0.154	0.418	-0.173	-0.417
Std. Dev.	0.105	0.137	0.116	0.186
Positive	30	29	0	1
Sign. Positive	29	29	0	1
Negative	0	1	30	29
Sign. Negative	0	0	29	29
Bonferonni pvalue	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the difference between the immediate and cumulative effects of a given slope variable on returns. The difference between the immediate and cumulative effects of a given slope variable on returns is computed as the difference between the coefficient estimate on the first lag of a slope variable in the return equation and the sum of the coefficients on all its five lags. It is also given as the sum of the coefficient estimates on the second to fifth lags of this slope variable. We compute these differences for each stock separately and the mean, median and standard deviation are computed across the 30 stocks in our sample. Positive (Negative) and Sign. Positive (Sign. Negative) present the number of stocks for which the difference is positive (negative) and significantly so at the 5% significance level, respectively. These differences correspond to our third hypothesis, related to the over- or under-reaction of returns to information embedded in the limit order book. The results under Ask-Low correspond to $H3a : \boldsymbol{\alpha}_{z,1}(S^A_{2,5,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{2,5,t})$ and Ask-High to $H3a^* : \boldsymbol{\alpha}_{z,1}(S^A_{5,20,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{5,20,t})$. The results under Bid-Low and Bid-High are for the corresponding hypotheses for the bid sides, i.e. $H3b$ and $H3b^*$. The row "Bonferonni p-value for $H3$ across stocks" presents the p-value for testing $H3$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H3$ for the i^{th} stock.

The following four tables (Tables A.9-A.12) present the estimation and hypothesis test results from the estimation of the empirical model with 3 lags. Tables A.9-A.12 can be directly compared to Tables 2-5 of the main text.

Table A.9: Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask Slope (Levels 1-5)		Ask Slope (Levels 5-20)		Bid Slope (Levels 1-5)		Bid Slope (Levels 5-20)	
	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.
Mean	0.109	0.111	0.231	0.048	-0.114	-0.115	-0.256	-0.045
Median	0.108	0.114	0.228	0.043	-0.113	-0.117	-0.222	-0.037
Std. Dev.	0.035	0.027	0.123	0.037	0.039	0.031	0.132	0.038
Positive	30	30	28	28	0	0	1	2
Significantly Positive	30	28	26	23	0	0	0	2
Negative	0	0	2	2	30	30	29	28
Significantly Negative	0	0	0	2	30	28	28	20
Bonferonni p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the immediate and cumulative effects of limit order book slopes on returns from the estimation of the model in Equation 5. We estimate the VAR system for each stock separately using all tick-by-tick data available for that stock in our sample over July 2010, May 2011, and June 2011. The immediate effect is the coefficient estimate on the first lag of the corresponding slope variable in the return equation 5a ($\alpha_{z,1}$). The cumulative effect is the sum of coefficient estimates on all lags of the slope variable ($\sum_{\tau=1}^{5} \alpha_{z,\tau}$). The mean, median and standard deviations are computed across the 30 stocks in our sample. Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is positive (negative). Significantly Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is significantly positive (negative) at the 5% significance level.

Table A.10: The Relation between the Immediate Effects of the Slopes of the Higher and Lower Levels on Returns

Number of stocks for which	Ask Side	Bid Side
$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	26	4
$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	4	26
$H1 : \boldsymbol{\alpha}_{z,1}(S_{1,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5%	25	25
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	21	4
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	4	21
Bonferonni p-value for $H1$ across stocks	0.000	0.000

Note: This table presents the relation between the immediate effects of the slopes of the higher and lower levels on returns and results for our first set of hypotheses on the equality of the immediate effects of lower and higher levels ($H1a$ and $H1b$). All numbers with the exception of the last row are out of 30 stocks in our sample. The row "$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which the immediate effect of the lower levels is smaller (greater) than that of the higher levels. The row "$H1 : \boldsymbol{\alpha}_{z,1}(S_{1,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5%" presents the number of stocks for which our first hypothesis on the equality of the immediate effects of the slopes of the higher and lower levels. The row "$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which our hypothesis is rejected at the 5% significance level in favor of the alternative that the immediate effect of the slope of the lower levels is smaller (greater) than that of the higher levels. The row "Bonferonni p-value for $H1$ across stocks" presents the p-value for testing $H1$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H1$ for the i^{th} stock.

Table A.11: The Relation between the Overall Effects of the Slopes of the Higher and Lower Levels on Returns

	Ask	Bid
Mean of F-statistic for $H2$	115.726	129.245
Median of F-statistic for $H2$	121.515	123.057
Std. Dev. of F-statistic for $H2$	65.511	56.976
Nb. of Stocks for which $H2$ is rejected at 5%	30	30
Bonferroni p-value	0.000	0.000

Note: This table presents results on whether the overall effects (based on all lags considered) of the slopes of the lower and higher levels of the same side are equal ($H2a$ and $H2b$). We test $H2$ for ask and bid sides and for each stocks separately. The mean, median and standard deviation are computed over these individual F-statistics for $H2$. The row "Nb. of Stocks for which $H2$ is rejected at 5% " presents the number of stocks (out of 30) for which our second hypothesis that the overall effects of the slopes of the lower and higher levels of the same side are equal. The row "Bonferonni p-value for $H2$ across stocks" presents the p-value for testing $H2$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H2$ for the i^{th} stock.

Table A.12: The Difference between the Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask		Bid	
	Low	High	Low	High
Mean	-0.001	0.182	0.000	-0.211
Median	-0.006	0.188	0.003	-0.193
Std. Dev.	0.016	0.101	0.018	0.108
Positive	9	28	19	1
Sign. Positive	4	26	6	1
Negative	21	2	11	29
Sign. Negative	8	0	4	29
Bonferonni pvalue	0.012	0.000	0.000	0.000

Note: This table presents summary statistics on the difference between the immediate and cumulative effects of a given slope variable on returns. The difference between the immediate and cumulative effects of a given slope variable on returns is computed as the difference between the coefficient estimate on the first lag of a slope variable in the return equation and the sum of the coefficients on all its five lags. It is also given as the sum of the coefficient estimates on the second to fifth lags of this slope variable. We compute these differences for each stock separately and the mean, median and standard deviation are computed across the 30 stocks in our sample. Positive (Negative) and Sign. Positive (Sign. Negative) present the number of stocks for which the difference is positive (negative) and significantly so at the 5% significance level, respectively. These differences correspond to our third hypothesis, related to the over- or under-reaction of returns to information embedded in the limit order book. The results under Ask-Low correspond to $H3a : \boldsymbol{\alpha}_{z,1}(S^A_{1,5,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{1,5,t})$ and Ask-High to $H3a^* : \boldsymbol{\alpha}_{z,1}(S^A_{5,20,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{5,20,t})$. The results under Bid-Low and Bid-High are for the corresponding hypotheses for the bid sides, i.e. $H3b$ and $H3b^*$. The row "Bonferonni p-value for $H3$ across stocks" presents the p-value for testing $H3$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H3$ for the i^{th} stock.

The following four tables (Tables A.13-A.16) present the estimation and hypothesis test results from the estimation of the empirical model with 8 lags. Tables A.13-A.16 can be directly compared to Tables 2-5 of the main text.

Table A.13: Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask Slope (Levels 1-5)		Ask Slope (Levels 5-20)		Bid Slope (Levels 1-5)		Bid Slope (Levels 5-20)	
	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.
Mean	0.352	0.151	0.446	0.033	-0.372	-0.148	-0.473	-0.033
Median	0.332	0.149	0.446	0.026	-0.361	-0.150	-0.483	-0.027
Std. Dev.	0.128	0.039	0.168	0.039	0.143	0.034	0.202	0.036
Positive	30	30	29	27	0	0	1	5
Significantly Positive	30	28	28	19	0	0	1	2
Negative	0	0	1	3	30	30	29	25
Significantly Negative	0	0	0	3	30	28	29	17
Bonferonni p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the immediate and cumulative effects of limit order book slopes on returns from the estimation of the model in Equation 5. We estimate the VAR system for each stock separately using all tick-by-tick data available for that stock in our sample over July 2010, May 2011, and June 2011. The immediate effect is the coefficient estimate on the first lag of the corresponding slope variable in the return equation 5a ($\alpha_{z,1}$). The cumulative effect is the sum of coefficient estimates on all lags of the slope variable ($\sum_{\tau=1}^{5} \alpha_{z,\tau}$). The mean, median and standard deviations are computed across the 30 stocks in our sample. Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is positive (negative). Significantly Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is significantly positive (negative) at the 5% significance level.

Table A.14: The Relation between the Immediate Effects of the Slopes of the Higher and Lower Levels on Returns

Number of stocks for which	Ask Side	Bid Side
$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	24	6
$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	6	24
$H1 : \boldsymbol{\alpha}_{z,1}(S_{1,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5%	25	26
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	20	5
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	5	21
Bonferonni p-value for $H1$ across stocks	0.000	0.000

Note: This table presents the relation between the immediate effects of the slopes of the higher and lower levels on returns and results for our first set of hypotheses on the equality of the immediate effects of lower and higher levels ($H1a$ and $H1b$). All numbers with the exception of the last row are out of 30 stocks in our sample. The row "$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which the immediate effect of the lower levels is smaller (greater) than that of the higher levels. The row "$H1 : \boldsymbol{\alpha}_{z,1}(S_{1,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5%" presents the number of stocks for which our first hypothesis on the equality of the immediate effects of the slopes of the higher and lower levels. The row "$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which our hypothesis is rejected at the 5% significance level in favor of the alternative that the immediate effect of the slope of the lower levels is smaller (greater) than that of the higher levels. The row "Bonferonni p-value for $H1$ across stocks" presents the p-value for testing $H1$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H1$ for the i^{th} stock.

Table A.15: The Relation between the Overall Effects of the Slopes of the Higher and Lower Levels on Returns

	Ask	Bid
Mean of F-statistic for $H2$	158.302	157.931
Median of F-statistic for $H2$	152.284	154.472
Std. Dev. of F-statistic for $H2$	66.478	64.931
Nb. of Stocks for which $H2$ is rejected at 5%	30	30
Bonferroni p-value	0.000	0.000

Note: This table presents results on whether the overall effects (based on all lags considered) of the slopes of the lower and higher levels of the same side are equal ($H2a$ and $H2b$). We test $H2$ for ask and bid sides and for each stocks separately. The mean, median and standard deviation are computed over these individual F-statistics for $H2$. The row "Nb. of Stocks for which $H2$ is rejected at 5% " presents the number of stocks (out of 30) for which our second hypothesis that the overall effects of the slopes of the lower and higher levels of the same side are equal. The row "Bonferonni p-value for $H2$ across stocks" presents the p-value for testing $H2$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H2$ for the i^{th} stock.

Table A.16: The Difference between the Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask		Bid	
	Low	High	Low	High
Mean	0.201	0.413	-0.224	-0.440
Median	0.192	0.418	-0.208	-0.452
Std. Dev.	0.107	0.147	0.125	0.184
Positive	30	29	0	1
Sign. Positive	29	29	0	1
Negative	0	1	30	29
Sign. Negative	0	0	29	29
Bonferonni pvalue	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the difference between the immediate and cumulative effects of a given slope variable on returns. The difference between the immediate and cumulative effects of a given slope variable on returns is computed as the difference between the coefficient estimate on the first lag of a slope variable in the return equation and the sum of the coefficients on all its five lags. It is also given as the sum of the coefficient estimates on the second to fifth lags of this slope variable. We compute these differences for each stock separately and the mean, median and standard deviation are computed across the 30 stocks in our sample. Positive (Negative) and Sign. Positive (Sign. Negative) present the number of stocks for which the difference is positive (negative) and significantly so at the 5% significance level, respectively. These differences correspond to our third hypothesis, related to the over- or under-reaction of returns to information embedded in the limit order book. The results under Ask-Low correspond to $H3a : \boldsymbol{\alpha}_{z,1}(S^A_{1,5,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{1,5,t})$ and Ask-High to $H3a^* : \boldsymbol{\alpha}_{z,1}(S^A_{5,20,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{5,20,t})$. The results under Bid-Low and Bid-High are for the corresponding hypotheses for the bid sides, i.e. $H3b$ and $H3b^*$. The row "Bonferonni p-value for $H3$ across stocks" presents the p-value for testing $H3$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H3$ for the i^{th} stock.

The following four tables (Tables A.17-A.20) present the estimation and hypothesis test results where we change the cutoff level to two and define lower and higher levels based on the first two and second to twentieth levels, respectively. Tables A.17-A.20 can be directly compared to Tables 2-5 of the main text.

Table A.17: Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask Slope (Levels 1-2)		Ask Slope (Levels 2-20)		Bid Slope (Levels 1-2)		Bid Slope (Levels 2-20)	
	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.
Mean	0.107	0.099	0.397	0.053	-0.114	-0.104	-0.433	-0.045
Median	0.104	0.098	0.417	0.047	-0.110	-0.104	-0.452	-0.043
Std. Dev.	0.036	0.024	0.186	0.043	0.040	0.029	0.225	0.044
Positive	30	30	29	28	0	0	1	4
Significantly Positive	30	28	27	23	0	0	1	3
Negative	0	0	1	2	30	30	29	26
Significantly Negative	0	0	0	2	30	28	27	18
Bonferonni p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the immediate and cumulative effects of limit order book slopes on returns from the estimation of the model in Equation 5. We estimate the VAR system for each stock separately using all tick-by-tick data available for that stock in our sample over July 2010, May 2011, and June 2011. The immediate effect is the coefficient estimate on the first lag of the corresponding slope variable in the return equation 5a ($\alpha_{z,1}$). The cumulative effect is the sum of coefficient estimates on all lags of the slope variable ($\sum_{\tau=1}^{5} \alpha_{z,\tau}$). The mean, median and standard deviations are computed across the 30 stocks in our sample. Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is positive (negative). Significantly Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is significantly positive (negative) at the 5% significance level.

Table A.18: The Relation between the Immediate Effects of the Slopes of the Higher and Lower Levels on Returns

Number of stocks for which	Ask Side	Bid Side
$\boldsymbol{\alpha}_{z,1}(S_{1,2,t}) < \boldsymbol{\alpha}_{z,1}(S_{2,20,t})$	27	3
$\boldsymbol{\alpha}_{z,1}(S_{1,2,t}) > \boldsymbol{\alpha}_{z,1}(S_{2,20,t})$	3	27
$H1 : \boldsymbol{\alpha}_{z,1}(S_{1,2,t}) = \boldsymbol{\alpha}_{z,1}(S_{2,20,t})$ is rejected at 5%	24	25
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,2,t}) < \boldsymbol{\alpha}_{z,1}(S_{2,20,t})$	21	3
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,2,t}) > \boldsymbol{\alpha}_{z,1}(S_{2,20,t})$	3	22
Bonferonni p-value for $H1$ across stocks	0.000	0.000

Note: This table presents the relation between the immediate effects of the slopes of the higher and lower levels on returns and results for our first set of hypotheses on the equality of the immediate effects of lower and higher levels ($H1a$ and $H1b$). All numbers with the exception of the last row are out of 30 stocks in our sample. The row "$\boldsymbol{\alpha}_{z,1}(S_{1,2,t}) < \boldsymbol{\alpha}_{z,1}(S_{2,20,t})$" ("$\boldsymbol{\alpha}_{z,1}(S_{1,2,t}) > \boldsymbol{\alpha}_{z,1}(S_{2,20,t})$") presents the number of stocks for which the immediate effect of the lower levels is smaller (greater) than that of the higher levels. The row "$H1 : \boldsymbol{\alpha}_{z,1}(S_{1,2,t}) = \boldsymbol{\alpha}_{z,1}(S_{2,20,t})$ is rejected at 5%" presents the number of stocks for which our first hypothesis on the equality of the immediate effects of the slopes of the higher and lower levels. The row "$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,2,t}) < \boldsymbol{\alpha}_{z,1}(S_{2,20,t})$" ("$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,2,t}) > \boldsymbol{\alpha}_{z,1}(S_{2,20,t})$") presents the number of stocks for which our hypothesis is rejected at the 5% significance level in favor of the alternative that the immediate effect of the slope of the lower levels is smaller (greater) than that of the higher levels. The row "Bonferonni p-value for $H1$ across stocks" presents the p-value for testing $H1$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H1$ for the i^{th} stock.

Table A.19: The Relation between the Overall Effects of the Slopes of the Higher and Lower Levels on Returns

	Ask	Bid
Mean of F-statistic for $H2$	165.688	195.355
Median of F-statistic for $H2$	145.593	180.067
Std. Dev. of F-statistic for $H2$	97.379	90.378
Nb. of Stocks for which $H2$ is rejected at 5%	30	30
Bonferroni p-value	0.000	0.000

Note: This table presents results on whether the overall effects (based on all lags considered) of the slopes of the lower and higher levels of the same side are equal ($H2a$ and $H2b$). We test $H2$ for ask and bid sides and for each stocks separately. The mean, median and standard deviation are computed over these individual F-statistics for $H2$. The row "Nb. of Stocks for which $H2$ is rejected at 5% " presents the number of stocks (out of 30) for which our second hypothesis that the overall effects of the slopes of the lower and higher levels of the same side are equal. The row "Bonferonni p-value for $H2$ across stocks" presents the p-value for testing $H2$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H2$ for the i^{th} stock.

Table A.20: The Difference between the Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask		Bid	
	Low	High	Low	High
Mean	0.008	0.344	-0.010	-0.388
Median	0.002	0.367	-0.003	-0.399
Std. Dev.	0.021	0.162	0.022	0.204
Positive	16	29	11	1
Sign. Positive	8	28	1	1
Negative	14	1	19	29
Sign. Negative	1	0	9	29
Bonferonni pvalue	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the difference between the immediate and cumulative effects of a given slope variable on returns. The difference between the immediate and cumulative effects of a given slope variable on returns is computed as the difference between the coefficient estimate on the first lag of a slope variable in the return equation and the sum of the coefficients on all its five lags. It is also given as the sum of the coefficient estimates on the second to fifth lags of this slope variable. We compute these differences for each stock separately and the mean, median and standard deviation are computed across the 30 stocks in our sample. Positive (Negative) and Sign. Positive (Sign. Negative) present the number of stocks for which the difference is positive (negative) and significantly so at the 5% significance level, respectively. These differences correspond to our third hypothesis, related to the over- or under-reaction of returns to information embedded in the limit order book. The results under Ask-Low correspond to $H3a : \boldsymbol{\alpha}_{z,1}(S^A_{1,5,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{1,2,t})$ and Ask-High to $H3a^* : \boldsymbol{\alpha}_{z,1}(S^A_{2,20,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{2,20,t})$. The results under Bid-Low and Bid-High are for the corresponding hypotheses for the bid sides, i.e. $H3b$ and $H3b^*$. The row "Bonferonni p-value for $H3$ across stocks" presents the p-value for testing $H3$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H3$ for the i^{th} stock.

The following four tables (Tables A.21-A.24) present the estimation and hypothesis test results where we exclude the first level from the definition of lower levels and define it to be levels between the second and fifth, and we define higher levels as levels between the fifth and twentieth. Tables A.21-A.24 can be directly compared to Tables 2-5 of the main text.

Table A.21: Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask Slope (Levels 2-5)		Ask Slope (Levels 5-20)		Bid Slope (Levels 2-5)		Bid Slope (Levels 5-20)	
	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.
Mean	0.092	0.096	0.309	0.045	-0.098	-0.093	-0.337	-0.043
Median	0.100	0.092	0.346	0.039	-0.104	-0.091	-0.346	-0.038
Std. Dev.	0.052	0.031	0.165	0.043	0.055	0.030	0.184	0.037
Positive	29	30	28	28	1	0	1	2
Significantly Positive	25	28	27	23	0	0	1	1
Negative	1	0	2	2	29	30	29	28
Significantly Negative	0	0	1	2	27	28	27	19
Bonferonni p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the immediate and cumulative effects of limit order book slopes on returns from the estimation of the model in Equation 5. We estimate the VAR system for each stock separately using all tick-by-tick data available for that stock in our sample over July 2010, May 2011, and June 2011. The immediate effect is the coefficient estimate on the first lag of the corresponding slope variable in the return equation 5a ($\alpha_{z,1}$). The cumulative effect is the sum of coefficient estimates on all lags of the slope variable ($\sum_{\tau=1}^{5} \alpha_{z,\tau}$). The mean, median and standard deviations are computed across the 30 stocks in our sample. Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is positive (negative). Significantly Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is significantly positive (negative) at the 5% significance level.

Table A.22: The Relation between the Immediate Effects of the Slopes of the Higher and Lower Levels on Returns

Number of stocks for which	Ask Side	Bid Side
$\boldsymbol{\alpha}_{z,1}(S_{2,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	28	2
$\boldsymbol{\alpha}_{z,1}(S_{2,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	2	28
$H1 : \boldsymbol{\alpha}_{z,1}(S_{2,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5%	25	25
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{2,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	23	2
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{2,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	2	23
Bonferonni p-value for $H1$ across stocks	0.000	0.000

Note: This table presents the relation between the immediate effects of the slopes of the higher and lower levels on returns and results for our first set of hypotheses on the equality of the immediate effects of lower and higher levels ($H1a$ and $H1b$). All numbers with the exception of the last row are out of 30 stocks in our sample. The row "$\boldsymbol{\alpha}_{z,1}(S_{2,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$\boldsymbol{\alpha}_{z,1}(S_{2,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which the immediate effect of the lower levels is smaller (greater) than that of the higher levels. The row "$H1 : \boldsymbol{\alpha}_{z,1}(S_{2,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5%" presents the number of stocks for which our first hypothesis on the equality of the immediate effects of the slopes of the higher and lower levels. The row "$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{2,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{2,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which our hypothesis is rejected at the 5% significance level in favor of the alternative that the immediate effect of the slope of the lower levels is smaller (greater) than that of the higher levels. The row "Bonferonni p-value for $H1$ across stocks" presents the p-value for testing $H1$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H1$ for the i^{th} stock.

Table A.23: The Relation between the Overall Effects of the Slopes of the Higher and Lower Levels on Returns

	Ask	Bid
Mean of F-statistic for $H2$	144.971	156.673
Median of F-statistic for $H2$	119.859	143.654
Std. Dev. of F-statistic for $H2$	93.519	86.462
Nb. of Stocks for which $H2$ is rejected at 5%	30	29
Bonferroni p-value	0.000	0.000

Note: This table presents results on whether the overall effects (based on all lags considered) of the slopes of the lower and higher levels of the same side are equal ($H2a$ and $H2b$). We test $H2$ for ask and bid sides and for each stocks separately. The mean, median and standard deviation are computed over these individual F-statistics for $H2$. The row "Nb. of Stocks for which $H2$ is rejected at 5% " presents the number of stocks (out of 30) for which our second hypothesis that the overall effects of the slopes of the lower and higher levels of the same side are equal. The row "Bonferonni p-value for $H2$ across stocks" presents the p-value for testing $H2$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H2$ for the i^{th} stock.

Table A.24: The Difference between the Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask		Bid	
	Low	High	Low	High
Mean	-0.004	0.264	-0.005	-0.293
Median	-0.008	0.306	0.007	-0.300
Std. Dev.	0.042	0.139	0.046	0.163
Positive	12	28	17	1
Sign. Positive	4	26	9	1
Negative	18	2	13	29
Sign. Negative	8	1	7	28
Bonferonni pvalue	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the difference between the immediate and cumulative effects of a given slope variable on returns. The difference between the immediate and cumulative effects of a given slope variable on returns is computed as the difference between the coefficient estimate on the first lag of a slope variable in the return equation and the sum of the coefficients on all its five lags. It is also given as the sum of the coefficient estimates on the second to fifth lags of this slope variable. We compute these differences for each stock separately and the mean, median and standard deviation are computed across the 30 stocks in our sample. Positive (Negative) and Sign. Positive (Sign. Negative) present the number of stocks for which the difference is positive (negative) and significantly so at the 5% significance level, respectively. These differences correspond to our third hypothesis, related to the over- or under-reaction of returns to information embedded in the limit order book. The results under Ask-Low correspond to $H3a : \boldsymbol{\alpha}_{z,1}(S^A_{2,5,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{2,5,t})$ and Ask-High to $H3a^* : \boldsymbol{\alpha}_{z,1}(S^A_{5,20,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{5,20,t})$. The results under Bid-Low and Bid-High are for the corresponding hypotheses for the bid sides, i.e. $H3b$ and $H3b^*$. The row "Bonferonni p-value for $H3$ across stocks" presents the p-value for testing $H3$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H3$ for the i^{th} stock.

The following four tables (Tables A.25-A.28) present the estimation and hypothesis test results from the estimation of the empirical model using data only from July 2010. Tables A.25-A.28 can be directly compared to Tables 2-5 of the main text.

Table A.25: Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask Slope (Levels 1-5)		Ask Slope (Levels 5-20)		Bid Slope (Levels 1-5)		Bid Slope (Levels 5-20)	
	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.
Mean	0.306	0.150	0.427	0.070	-0.334	-0.149	-0.484	-0.059
Median	0.300	0.151	0.379	0.044	-0.336	-0.147	-0.435	-0.039
Std. Dev.	0.132	0.047	0.278	0.073	0.157	0.049	0.325	0.071
Positive	29	29	29	26	0	0	4	5
Significantly Positive	29	27	24	17	0	0	0	3
Negative	0	0	0	3	29	29	25	24
Significantly Negative	0	0	0	1	29	27	25	13
Bonferonni p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the immediate and cumulative effects of limit order book slopes on returns from the estimation of the model in Equation 5. We estimate the VAR system for each stock separately using all tick-by-tick data available for that stock in our sample over July 2010. The immediate effect is the coefficient estimate on the first lag of the corresponding slope variable in the return equation 5a ($\alpha_{z,1}$). The cumulative effect is the sum of coefficient estimates on all lags of the slope variable ($\sum_{\tau=1}^{5} \alpha_{z,\tau}$). The mean, median and standard deviations are computed across the 30 stocks in our sample. Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is positive (negative). Significantly Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is significantly positive (negative) at the 5% significance level.

Table A.26: The Relation between the Immediate Effects of the Slopes of the Higher and Lower Levels on Returns

Number of stocks for which	Ask Side	Bid Side
$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	21	7
$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	8	22
$H1 : \boldsymbol{\alpha}_{z,1}(S_{1,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5%	17	20
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	11	6
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	6	14
Bonferonni p-value for $H1$ across stocks	0.000	0.000

Note: This table presents the relation between the immediate effects of the slopes of the higher and lower levels on returns and results for our first set of hypotheses on the equality of the immediate effects of lower and higher levels ($H1a$ and $H1b$). All numbers with the exception of the last row are out of 30 stocks in our sample. The row "$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which the immediate effect of the lower levels is smaller (greater) than that of the higher levels. The row "$H1 : \boldsymbol{\alpha}_{z,1}(S_{1,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5%" presents the number of stocks for which our first hypothesis on the equality of the immediate effects of the slopes of the higher and lower levels. The row "$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which our hypothesis is rejected at the 5% significance level in favor of the alternative that the immediate effect of the slope of the lower levels is smaller (greater) than that of the higher levels. The row "Bonferonni p-value for $H1$ across stocks" presents the p-value for testing $H1$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H1$ for the i^{th} stock.

Table A.27: The Relation between the Overall Effects of the Slopes of the Higher and Lower Levels on Returns

	Ask	Bid
Mean of F-statistic for $H2$	53.392	59.344
Median of F-statistic for $H2$	45.519	50.695
Std. Dev. of F-statistic for $H2$	34.876	38.790
Nb. of Stocks for which $H2$ is rejected at 5%	27	28
Bonferroni p-value	0.000	0.000

Note: This table presents results on whether the overall effects (based on all lags considered) of the slopes of the lower and higher levels of the same side are equal ($H2a$ and $H2b$). We test $H2$ for ask and bid sides and for each stocks separately. The mean, median and standard deviation are computed over these individual F-statistics for $H2$. The row "Nb. of Stocks for which $H2$ is rejected at 5% " presents the number of stocks (out of 30) for which our second hypothesis that the overall effects of the slopes of the lower and higher levels of the same side are equal. The row "Bonferonni p-value for $H2$ across stocks" presents the p-value for testing $H2$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H2$ for the i^{th} stock.

Table A.28: The Difference between the Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask		Bid	
	Low	High	Low	High
Mean	0.155	0.357	-0.185	-0.424
Median	0.144	0.343	-0.157	-0.416
Std. Dev.	0.109	0.230	0.139	0.281
Positive	29	29	0	4
Sign. Positive	28	25	0	0
Negative	0	0	29	25
Sign. Negative	0	0	27	25
Bonferonni pvalue	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the difference between the immediate and cumulative effects of a given slope variable on returns. The difference between the immediate and cumulative effects of a given slope variable on returns is computed as the difference between the coefficient estimate on the first lag of a slope variable in the return equation and the sum of the coefficients on all its five lags. It is also given as the sum of the coefficient estimates on the second to fifth lags of this slope variable. We compute these differences for each stock separately and the mean, median and standard deviation are computed across the 30 stocks in our sample. Positive (Negative) and Sign. Positive (Sign. Negative) present the number of stocks for which the difference is positive (negative) and significantly so at the 5% significance level, respectively. These differences correspond to our third hypothesis, related to the over- or under-reaction of returns to information embedded in the limit order book. The results under Ask-Low correspond to $H3a : \boldsymbol{\alpha}_{z,1}(S^A_{1,5,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{1,5,t})$ and Ask-High to $H3a^* : \boldsymbol{\alpha}_{z,1}(S^A_{5,20,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{5,20,t})$. The results under Bid-Low and Bid-High are for the corresponding hypotheses for the bid sides, i.e. $H3b$ and $H3b^*$. The row "Bonferonni p-value for $H3$ across stocks" presents the p-value for testing $H3$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H3$ for the i^{th} stock.

The following four tables (Tables A.29-A.32) present the estimation and hypothesis test results from the estimation of the empirical model using data only from May 2011. Tables A.29-A.32 can be directly compared to Tables 2-5 of the main text.

Table A.29: Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask Slope (Levels 1-5)		Ask Slope (Levels 5-20)		Bid Slope (Levels 1-5)		Bid Slope (Levels 5-20)	
	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.
Mean	0.363	0.185	0.358	0.028	-0.383	-0.187	-0.391	-0.032
Median	0.313	0.186	0.350	0.036	-0.332	-0.182	-0.417	-0.023
Std. Dev.	0.141	0.057	0.236	0.044	0.162	0.053	0.228	0.043
Positive	30	30	28	24	0	0	1	7
Significantly Positive	30	28	24	14	0	0	0	4
Negative	0	0	2	6	30	30	29	23
Significantly Negative	0	0	1	3	30	28	26	11
Bonferonni p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the immediate and cumulative effects of limit order book slopes on returns from the estimation of the model in Equation 5. We estimate the VAR system for each stock separately using all tick-by-tick data available for that stock in our sample over May 2011. The immediate effect is the coefficient estimate on the first lag of the corresponding slope variable in the return equation 5a ($\alpha_{z,1}$). The cumulative effect is the sum of coefficient estimates on all lags of the slope variable ($\sum_{\tau=1}^{5} \alpha_{z,\tau}$). The mean, median and standard deviations are computed across the 30 stocks in our sample. Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is positive (negative). Significantly Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is significantly positive (negative) at the 5% significance level.

Table A.30: The Relation between the Immediate Effects of the Slopes of the Higher and Lower Levels on Returns

Number of stocks for which	Ask Side	Bid Side
$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	21	12
$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	9	18
$H1 : \boldsymbol{\alpha}_{z,1}(S_{1,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5		
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	14	7
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	8	13
Bonferonni p-value for $H1$ across stocks	0.000	0.000

Note: This table presents the relation between the immediate effects of the slopes of the higher and lower levels on returns and results for our first set of hypotheses on the equality of the immediate effects of lower and higher levels ($H1a$ and $H1b$). All numbers with the exception of the last row are out of 30 stocks in our sample. The row "$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which the immediate effect of the lower levels is smaller (greater) than that of the higher levels. The row "$H1 : \boldsymbol{\alpha}_{z,1}(S_{1,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5%" presents the number of stocks for which our first hypothesis on the equality of the immediate effects of the slopes of the higher and lower levels. The row "$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which our hypothesis is rejected at the 5% significance level in favor of the alternative that the immediate effect of the slope of the lower levels is smaller (greater) than that of the higher levels. The row "Bonferonni p-value for $H1$ across stocks" presents the p-value for testing $H1$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H1$ for the i^{th} stock.

Table A.31: The Relation between the Overall Effects of the Slopes of the Higher and Lower Levels on Returns

	Ask	Bid
Mean of F-statistic for $H2$	81.180	83.762
Median of F-statistic for $H2$	79.590	73.841
Std. Dev. of F-statistic for $H2$	41.234	40.250
Nb. of Stocks for which $H2$ is rejected at 5%	30	30
Bonferroni p-value	0.000	0.000

Note: This table presents results on whether the overall effects (based on all lags considered) of the slopes of the lower and higher levels of the same side are equal ($H2a$ and $H2b$). We test $H2$ for ask and bid sides and for each stocks separately. The mean, median and standard deviation are computed over these individual F-statistics for $H2$. The row "Nb. of Stocks for which $H2$ is rejected at 5% " presents the number of stocks (out of 30) for which our second hypothesis that the overall effects of the slopes of the lower and higher levels of the same side are equal. The row "Bonferonni p-value for $H2$ across stocks" presents the p-value for testing $H2$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H2$ for the i^{th} stock.

Table A.32: The Difference between the Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask		Bid	
	Low	High	Low	High
Mean	0.178	0.330	-0.197	-0.359
Median	0.143	0.320	-0.160	-0.365
Std. Dev.	0.115	0.209	0.140	0.209
Positive	30	28	0	1
Sign. Positive	29	26	0	1
Negative	0	2	30	29
Sign. Negative	0	1	29	27
Bonferonni pvalue	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the difference between the immediate and cumulative effects of a given slope variable on returns. The difference between the immediate and cumulative effects of a given slope variable on returns is computed as the difference between the coefficient estimate on the first lag of a slope variable in the return equation and the sum of the coefficients on all its five lags. It is also given as the sum of the coefficient estimates on the second to fifth lags of this slope variable. We compute these differences for each stock separately and the mean, median and standard deviation are computed across the 30 stocks in our sample. Positive (Negative) and Sign. Positive (Sign. Negative) present the number of stocks for which the difference is positive (negative) and significantly so at the 5% significance level, respectively. These differences correspond to our third hypothesis, related to the over- or under-reaction of returns to information embedded in the limit order book. The results under Ask-Low correspond to $H3a : \boldsymbol{\alpha}_{z,1}(S^A_{1,5,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{1,5,t})$ and Ask-High to $H3a^* : \boldsymbol{\alpha}_{z,1}(S^A_{5,20,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{5,20,t})$. The results under Bid-Low and Bid-High are for the corresponding hypotheses for the bid sides, i.e. $H3b$ and $H3b^*$. The row "Bonferonni p-value for $H3$ across stocks" presents the p-value for testing $H3$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H3$ for the i^{th} stock.

The following four tables (Tables A.33-A.36) present the estimation and hypothesis test results from the estimation of the empirical model using data only from June 2011. Tables A.33-A.36 can be directly compared to Tables 2-5 of the main text.

Table A.33: Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask Slope (Levels 1-5)		Ask Slope (Levels 5-20)		Bid Slope (Levels 1-5)		Bid Slope (Levels 5-20)	
	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.	Imm. Eff.	Cum. Eff.
Mean	0.448	0.200	0.508	0.032	-0.449	-0.201	-0.509	-0.036
Median	0.381	0.194	0.522	0.031	-0.419	-0.205	-0.510	-0.030
Std. Dev.	0.218	0.071	0.226	0.059	0.198	0.072	0.260	0.062
Positive	30	30	29	21	0	0	1	6
Significantly Positive	30	28	27	14	0	0	1	4
Negative	0	0	1	9	30	30	29	24
Significantly Negative	0	0	0	5	30	28	27	17
Bonferonni p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the immediate and cumulative effects of limit order book slopes on returns from the estimation of the model in Equation 5. We estimate the VAR system for each stock separately using all tick-by-tick data available for that stock in our sample over June 2011. The immediate effect is the coefficient estimate on the first lag of the corresponding slope variable in the return equation 5a ($\alpha_{z,1}$). The cumulative effect is the sum of coefficient estimates on all lags of the slope variable ($\sum_{\tau=1}^{5} \alpha_{z,\tau}$). The mean, median and standard deviations are computed across the 30 stocks in our sample. Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is positive (negative). Significantly Positive (Negative) presents the number of stocks for which the corresponding effect of the slope variable in the column heading is significantly positive (negative) at the 5% significance level.

Table A.34: The Relation between the Immediate Effects of the Slopes of the Higher and Lower Levels on Returns

Number of stocks for which	Ask Side	Bid Side
$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	23	7
$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	7	23
$H1 : \boldsymbol{\alpha}_{z,1}(S_{1,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5%	19	22
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	15	5
$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$	4	17
Bonferonni p-value for $H1$ across stocks	0.000	0.000

Note: This table presents the relation between the immediate effects of the slopes of the higher and lower levels on returns and results for our first set of hypotheses on the equality of the immediate effects of lower and higher levels ($H1a$ and $H1b$). All numbers with the exception of the last row are out of 30 stocks in our sample. The row "$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which the immediate effect of the lower levels is smaller (greater) than that of the higher levels. The row "$H1 : \boldsymbol{\alpha}_{z,1}(S_{1,5,t}) = \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$ is rejected at 5%" presents the number of stocks for which our first hypothesis on the equality of the immediate effects of the slopes of the higher and lower levels. The row "$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) < \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$" ("$H1$ is rejected in favor of $\boldsymbol{\alpha}_{z,1}(S_{1,5,t}) > \boldsymbol{\alpha}_{z,1}(S_{5,20,t})$") presents the number of stocks for which our hypothesis is rejected at the 5% significance level in favor of the alternative that the immediate effect of the slope of the lower levels is smaller (greater) than that of the higher levels. The row "Bonferonni p-value for $H1$ across stocks" presents the p-value for testing $H1$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H1$ for the i^{th} stock.

Table A.35: The Relation between the Overall Effects of the Slopes of the Higher and Lower Levels on Returns

	Ask	Bid
Mean of F-statistic for $H2$	89.512	88.648
Median of F-statistic for $H2$	78.129	90.620
Std. Dev. of F-statistic for $H2$	44.269	38.703
Nb. of Stocks for which $H2$ is rejected at 5		
Bonferroni p-value	0.000	0.000

Note: This table presents results on whether the overall effects (based on all lags considered) of the slopes of the lower and higher levels of the same side are equal ($H2a$ and $H2b$). We test $H2$ for ask and bid sides and for each stocks separately. The mean, median and standard deviation are computed over these individual F-statistics for $H2$. The row "Nb. of Stocks for which $H2$ is rejected at 5% " presents the number of stocks (out of 30) for which our second hypothesis that the overall effects of the slopes of the lower and higher levels of the same side are equal. The row "Bonferonni p-value for $H2$ across stocks" presents the p-value for testing $H2$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H2$ for the i^{th} stock.

Table A.36: The Difference between the Immediate and Cumulative Effects of Limit Order Book Slopes on Returns

	Ask		Bid	
	Low	High	Low	High
Mean	0.248	0.476	-0.247	-0.473
Median	0.198	0.474	-0.204	-0.464
Std. Dev.	0.181	0.202	0.159	0.236
Positive	30	30	0	2
Sign. Positive	29	27	0	1
Negative	0	0	30	28
Sign. Negative	0	0	29	28
Bonferonni pvalue	0.000	0.000	0.000	0.000

Note: This table presents summary statistics on the difference between the immediate and cumulative effects of a given slope variable on returns. The difference between the immediate and cumulative effects of a given slope variable on returns is computed as the difference between the coefficient estimate on the first lag of a slope variable in the return equation and the sum of the coefficients on all its five lags. It is also given as the sum of the coefficient estimates on the second to fifth lags of this slope variable. We compute these differences for each stock separately and the mean, median and standard deviation are computed across the 30 stocks in our sample. Positive (Negative) and Sign. Positive (Sign. Negative) present the number of stocks for which the difference is positive (negative) and significantly so at the 5% significance level, respectively. These differences correspond to our third hypothesis, related to the over- or under-reaction of returns to information embedded in the limit order book. The results under Ask-Low correspond to $H3a : \boldsymbol{\alpha}_{z,1}(S^A_{1,5,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{1,5,t})$ and Ask-High to $H3a^* : \boldsymbol{\alpha}_{z,1}(S^A_{5,20,t}) = \sum_{\tau=1}^{5} \boldsymbol{\alpha}_{z,\tau}(S^A_{5,20,t})$. The results under Bid-Low and Bid-High are for the corresponding hypotheses for the bid sides, i.e. $H3b$ and $H3b^*$. The row "Bonferonni p-value for $H3$ across stocks" presents the p-value for testing $H3$ jointly across all 30 stocks in our sample adjusted for the issue of multiple testing based on the Bonferroni correction. It is calculated as $\min(1, \min(p_1, \cdots, p_{30}) \times 30)$ where p_i is the p-value for $H3$ for the i^{th} stock.