Who Knows? Information Differences Between Trader Types

Albert J. Menkveld*   Ion Lucas Saru†

January 15, 2022

Abstract

Classifications of informed traders depend on market conditions and are thus endogenous. We argue that exogenous information on the trader type that is independent of the market conditions should be used: whether the trader is an agent and principal trader. At an hourly frequency, variation in agent flow accounts for 25% of the total variance of innovations in efficient prices while the contribution of principal flow is virtually zero. Informativeness of different order types depends on market conditions as measured by the VIX.

classification codes: 360, 420, 350, 570

1 Introduction

Research on informed trading usually relies on an endogenous classification of which traders are informed. However, which trader is labelled informed may depend on the market circumstances, the trader's order execution strategy, and the interaction with other traders. To avoid such problems, it is desirable to have an exogenous classification of traders that it independent of the market circumstances and link it the informativeness of their orders.

We analyze the trading behavior of different types of traders – agent and principal traders – in terms of their contribution to price discovery and price impact. Even though recent evidence focusing on a subset of arguably informed traders shows that these traders use limit orders (see, for example Collin-Dufresne and Fos (2015) and Kacperczyk and Pagnotta (2019)), it remains unanswered how to link the exogenous classification of a trader with the informativeness of the traders' their order. We focus on differences in information content of trades between traders using the same order type. Are agents' market orders as informed as principals' market orders?

*Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, the Netherlands, albertjmenkveld@gmail.com, +31 20 598 6130, and Tinbergen Institute.

†Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, the Netherlands, saru.ionlucas@gmail.com, +31 20 598 6060, and Tinbergen Institute.

We propose a state space model to estimate the contribution of orders of different traders to price discovery. Our model controls for differences in trading volume. Thus, even though trading volume between agents and principals differs, we estimate the information content in order flow scaled by volume. The state space framework allows us to disentangle transitory and permanent movements in security prices. Based on this, we identify whether price changes are due to information or price pressure. Even if agents and principals use the same order type, the order's contribution to price discovery differs.

We exhibit our main results in Figure 1. At an hourly frequency, aggressive orders used by agents account for 25% of the variance of innovations in efficient prices, while the contribution of aggressive principal orders is virtually zero. Our findings indicate that besides the order type, also information on the trader using the order is relevant for its informativeness.

We show that order informativeness depends on market conditions as measured by the VIX. Efficient price changes load stronger on agents' market orders in high-VIX periods and weaker in low-VIX periods. In low-VIX periods, efficient price innovations load stronger on agents' limit orders. Also, the variance of innovations in efficient prices is lower in low-VIX periods while the variance of pricing errors is virtually the same across market conditions. Our findings are consistent with at least some agents with information using limit orders in times of low market volatility. This complements the findings of Menkveld, Yueshen, and Zhu (2017) who show that volume migrates between trading venues dependent on the market conditions. We show that information scaled by trading volume varies with market volatility.

In light of a recent literature analyzing closing auctions in equity markets (Bogousslavsky and Muravyev, 2021; Comerton-Forde and Rindi, 2021), we analyze opening and closing auctions in futures markets. Our results indicate no significant differences in terms of price discovery and liquidity provision of the traders between opening auctions and continuous trading. For closing auctions, we do not find differences in terms of price discovery but evidence that agents and principals trade stronger against pricing errors. Rather than creating auction price deviations, this reduces deviations from efficient prices.

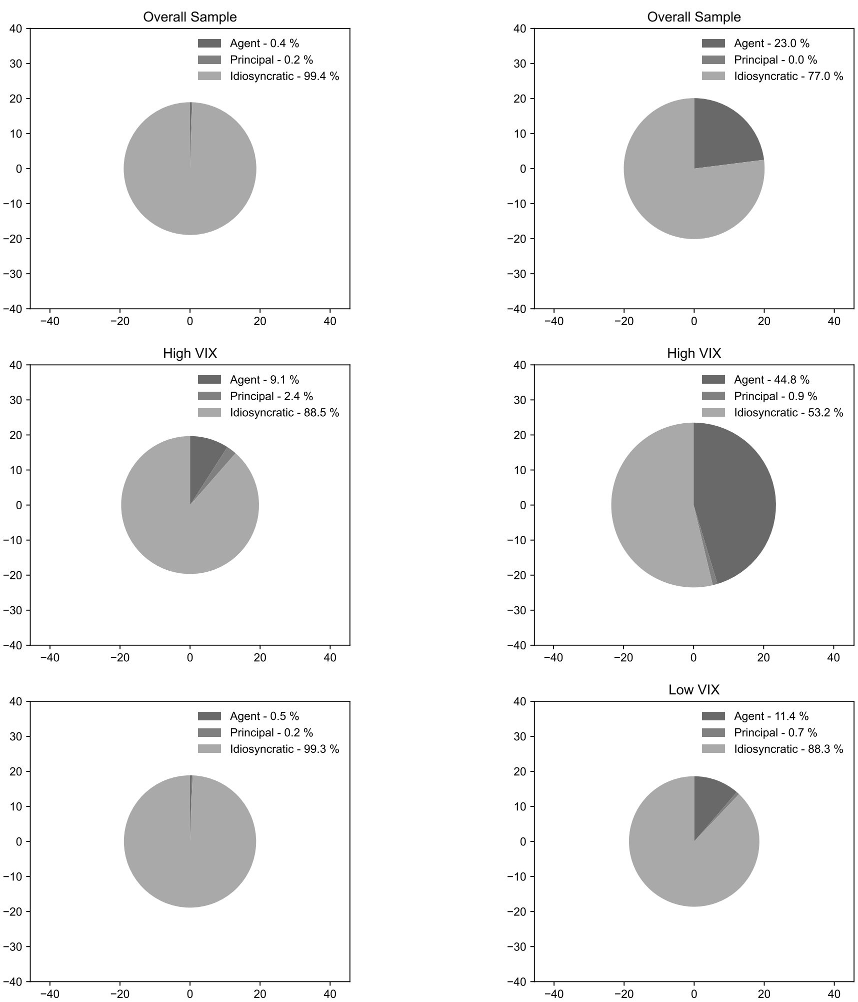
Our analysis focuses on trading in Euro STOXX 50 index futures, capturing the 50 largest companies in the Eurozone from eight countries. We use data on futures trading as futures are leading other instruments in terms of price discovery (Kawaller, Koch, and Koch, 1987; Stoll and Whaley, 1990; Hasbrouck, 2003; Tse, Bandyopadhyay, and Shen, 2006). Since we are asking the question how different traders contribute to price discovery, we want to analyze the instrument where price discovery first occurs rather than analyzing instruments that are lagging in price discovery.[1]

In contrast to existing studies on the use of order types by informed traders that focus on traders with stock-specific information (Collin-Dufresne and Fos, 2015; Kacperczyk and Pagnotta, 2019), we focus on information regarding a

[1]In that case, traders might incorporate information into prices that has been revealed in other instruments. For this, however, the speed of the trader matters and the analysis might confound different effects.

Figure 1: Hourly variances of pricing errors and efficient price innovations

The figure plots hourly variances of pricing errors and efficient price innovations as well as the share of agents' and principals' aggressive orders in these variances. The radius corresponds to the standard deviation of pricing errors and efficient price innovations, respectively, in *bp*.



(a) Pricing Errors (b) Efficient Price Innovations basket of stocks. It is not clear that existing results extend to instruments that reflect a basket of multiple stocks. Prices of futures, such as the Euro STOXX 50 futures, reflect, next to information on the indexes constituents, information on the state of the "European" economy.[2] We address the question how traders with this more general, macro-type of information trade upon their information and provide liquidity using different orders. We find that both information on the trader type as well as the order used matters for the expected informational content of the order relative to analyzing either on a stand-alone basis. Thus, with our work we contribute to the question of price efficiency and information incorporation into prices. Our results suggest that rather than relying on an endogenous classification of informed versus uninformed orders, research should focus on exogenous exchange versus non-exchange member trader types.

Our work relates to a vast literature studying information content of different order types as well as order submission by informed traders. In recent work, Li, Ye, and Zheng (2021) study different order types at the NYSE and evaluate their performance, also in terms of price discovery. However, they do not observe the traders submitting the orders. It is possible that the same trader uses different order types, depending on the market circumstances and the informational horizon. While we do not observe the exact order type, we observe whether a trade record stems from an aggressive or a passive order. However, we have information on the type of trader utilizing the order. Thus, observing how orders perform depending on which type of trader uses them allows us to infer valuable information on the trader's information and motivation. We find a pecking order of order types depending on market volatility. Our findings reveal that aggressive orders are relatively more informative in high-volatility periods and passive orders are less subject to adverse selection in low-volatility periods.

We address the question which order type is used by informed traders. The early literature assumes that informed traders use market order (Harris, 1998). Recent evidence shows that informed investors use limit orders. For example, Collin-Dufresne and Fos (2015) find that 13D activist traders use limit orders. Parlour (1998) and Foucault (1999) study limit and market order submission by uninformed traders. Hollifield, Miller, and Sandås (2004) theoretically and and empirically analyze optimal limit order submissions. In early empirical work, Biais, Hillion, and Spatt (1995) study limit order book dynamics and interactions between market and limit orders.[3] Bloomfield, O'Hara, and Saar (2005), Baruch, Panayides, and Venkataraman (2017) and Kacperczyk and Pagnotta (2019) find that informed traders use limit orders.[4] Other work focuses on order

[2]Traders with stock-specific information can achieve a greater exposure to a stock by trading its shares or derivatives on its shares rather that trading futures on the index that contains the share.

[3]Goettler, Parlour, and Rajan (2009) develop a model of the choice of acquiring information and choosing the order type and Hoffmann (2014) studies order choice in a market with HFTs.

[4]Brogaard, Hendershott, and Riordan (2019) document that HFTs' limit orders contribute to price discovery, and Fleming, Mizrach, and Nguyen (2018) provide evidence consistent with price discovery through limit orders.

choice by informed and uninformed traders, respectively. Bloomfield, O'Hara, and Saar (2015) find that informed trader's order choice depends on whether they can hide liquidity and that they use more limit orders if liquidity not visible. Kaniel and Liu (2006) present a theoretical model in which informed trades can submit limit orders. In equilibrium, limit orders may contain more information than market orders. Also, if private information is more long-lived, the probability of using limit orders increases. Our results are consistent with the latter interpretation. However, the classification of traders into informed and uninformed traders is endogenous. We show that rather than relying on an endogenous classification, exogenous information on trader types contains information on the informativeness of the traders' orders as well price discovery under different market conditions. Overall, we find more information content in market orders, but our evidence suggests that agents' limit order are more informed at longer horizons and in times of low market volatility.

Several studies that analyze the trading behavior of groups of traders focus only on subgroups. Kelley and Tetlock (2013) study the usage of limit and market orders by retail investors. In contrast to us, they analyze trading and return patterns at a lower frequency and relate daily imbalances to monthly returns. They find that retail investors using market orders trade on new information. Also limit orders by retail investors provide liquidity and some limit orders may be informed. Hendershott, Livdan, and Schürhoff (2015) study price discovery of institutional investors and find that they contribute to price discovery regarding news events. However, they do not distinguish between order types. In a recent study, Beason and Wahal (2020) study institutional trading algorithms and find that they mainly use limit orders. Hence, to the extent that institutional investors are informed, they are incorporating their information using limit orders. Anand, Chakravarty, and Martell (2005) analyze the intraday pattern of orders usage by liquidity traders and institutional traders. They find that institutional traders use market orders early during the day and limit orders later during the day. Our analysis incorporates information on different trader types (agents and principals) as well as information on the orders used. This allows us to analyze differences in the information content of orders scaled by volume.

Barber et al. (2009) study the Taiwanese market during the period 1995 – 1999. They distinguish between what they call aggressive and non-aggressive orders as the market only allowed for limit orders. They find that individual traders make losses through "aggressive orders" while aggressive and non-aggressive orders of institutions are profitable. We, in contrast, analyze a modern limit order book market that allows for submitting both market and limit orders.

The remainder of the paper is structured as follows. In Section 2 we provide institutional details on Euro STOXX 50 futures trading before we describe the data in Section 3. Section 4 discusses our state space methodology. We present our main results on relative differences in order flow informativeness by trading type in Section 5 and results on opening and closing auction in Section 6. Finally, Section 7 concludes.

2 Institutional Background

Our analysis focuses on Euro STOXX 50 futures, one of the most actively traded futures contracts in the world. The Euro STOXX 50 is the index for the largest companies of the Eurozone with the constituents registered in Belgium, Finland, France, Germany, Ireland, Italy, the Netherlands, and Spain.[5] The constituents are not all listed at the same exchange, however the trading hours are aligned. Trading takes place in an electronic limit order book.

In our sample period, trading in Euro STOXX 50 futures takes place from 8 in the morning to 10 in the evening, Central European (Summer) Time. Trading starts with an opening auction in the morning. After continuous trading terminates in the evening, it is followed by a call phase of at least three minutes, before a closing auction is held.

2.1 Fee Structure

Eurex (2021) sets a fee structure that differentiates between principals (exchange members) and agents (non-exchange members).[6] Exchange members include internationally operating banks as well a nationally operating European banks, proprietary trading firms, high frequency trading firms, global asset management companies, and other institutional investors. Non-exchnage members are by definition the remaining traders and include other institutional investors, pension funds, et cetera. According to Eurex, there is little migration of agents to principal accounts and retail order flow in agency order flow is negligible.

The standard fee per contract via the limit order book is 0.35 EUR for agents and 0.30 EUR for principals. A reduced fee beyond a certain volume threshold does not apply. Fees for cash settlement are 0.35 EUR and 0.30 EUR for agents and principals, respectively, and position closing adjustment fees are 0.70 EUR and 0.60 EUR. Rebates for liquidity provision are based on several components and amount up to 80%.[7] Volume rebates are calculated on a monthly level and all volume across all index futures is taken into account.

3 Data

Our focus is on studying the information in order flow and liquidity provision by different trader types. We use proprietary trading data on Euro STOXX 50 futures from Eurex. The sample period spans from January 4, 2010 to December 7, 2018.[8] We focus on futures trading as futures are leading in terms of price

[5] https://www.stoxx.com/index-details?symbol=SX5E

[6] A list of exchange members can be found on the Eurex website. We provide a list of all unique exchange members listed on Eurex' website in Appendix A.

[7] These include basic requirements, requirements for a combination of products, requirements for updating quotes in response to quote requests, size requirements, and maximum spread requirements.

[8] We restrict our sample to end in December 7, 2018 in order to avoid any confounding effects with a trading hour extension in the morning on December 8, 2018.

discovery (see Hasbrouck (2003), among others).[9] Central to our analysis is the information on the account role of a trader as well as whether the trader utilizes an aggressive order or not.[10] The data contains the following information:

- Expiration of the futures contract.
- Indicator whether the trade is a buy or sell.
- Trade size.
- Execution price.
- Aggressor flag, whether the trade stems to a market or limit order. In the following, we label trades that pertain to a market or marketable limit order as *aggressive* trades and trades that stem to a limit order as *passive* trades.
- The account role of the trader, whether the trader is a principal or agent.
- Indicator whether the order was fully or partially executed in the trade.

The analysis is based on trading in the contract with the highest trading volume. This is usually the nearest-to-maturity contract.[11] Also, we focus on continuous trading as only during continuous trading there is a limit order for every market order.[12] Our focus is on the differences between different traders as well as how they trade on their information. We thus distinguish between aggressive and passive orders. This distinction is only meaningful during continuous trading. Furthermore, this assures that there are no confound effects with dynamics during the opening, closing, and intraday auctions (Bogousslavsky and Muravyev, 2021; Comerton-Forde and Rindi, 2021).

In the next section we describe details on the data cleaning procedure before we provide summary statistics for the data.

3.1 Data Cleaning

We identify continuous trading by requiring that for every timestamp, the volume of market orders equals the volume of limit orders. This classification is

[9] An alternative is analyzing patterns in price discovery based on ETF trading data. Hasbrouck (2003) finds that ETFs contribute significantly to price discovery as well, with the ETFs' contribution varying between instruments. Furthermore, Menkveld and Yueshen (2019) document a cross-market non-arbitrage relationship between S&P 500 futures and ETFs (E-mini and SPY). Moreover, the futures market is less fragmented than the ETF market and very liquid.

[10] We do not observe the exact order used. Orders at Eurex include next to market and limit orders also stop orders, orders for the closing auction, as well as one-cancels-other and book-or-cancel restrictions for limit orders (Eurex, 2021a).

[11] Similarly, Huang (2018) uses the contract with the highest volume. Our approach yields similar results to using the front contract and rolling over to the next contract a pre-specified number of days before expiration (Andersen et al., 2007).

[12] This is in line with, for example, Brogaard, Hendershott, and Riordan (2014), who study the contribution of high-frequency traders to price discovery during continuous trading.

performed based on the aggressor flag: the volume of aggressive orders has to equal the volume of passive orders. This is feasible for the period January 1, 2010 – May 7, 2013. On May 8, 2013, Eurex migrated its products to the T7 trading system. This causes some imprecision in the timestamps in the data of the following form. Orders that were executed against each other are not necessarily recorded at the same timestamp but at consecutive timestamps where the difference between the records is usually within the range of a few tens of milliseconds.[13] Thus, requiring the volume of market orders to equal the volume of limit orders at every timestamp to identify continuous trading is not feasible.[14]

We address this problem using an event time approach.[15] Trades that are recorded at the same price within a defined time-interval are grouped together. Then, the total volume of limit orders and the total volume of market orders over the grouped trades are computed. If the total volume of limit orders equals the total volume of market orders, the trades are labeled as continuous trading and included in the main analysis.

The algorithm starts at the beginning of each trading day. For each price-timestamp combination, a time window starting with that trade record is initialized.[16] All following trades that are executed within the time window at the same trading price are grouped together and assigned the timestamp of the first recorded trade in that group. Once a trade is executed at the same trading price but does not fall within the time window, or at a different trading price, a new time window starting from that trade record is initialized. Again, all trades that occur within the time window at the same execution price are grouped together. This procedure continues until the end of the trading day.

The only parameter that has to be chosen is the length of the time window. Choosing the window length trades off two factors. On the one hand, choosing a longer window length assures that all corresponding trades are grouped together even if there is substantial imprecision in the timestamps and high trading activity. On the other hand, if the window length is chosen too long and the volume of market orders and limit orders is *not* equal, substantial volume is excluded form the main analysis. We consider the possibilities of 100ms, 500ms, 2s, and 4s. In the main analysis, we focus on data that has been cleaned using a window length of 100ms. Our findings are robust to using a different window length.

[13]Also, it appears that this "noise" in the timestamps increases in times of high trading activity.

[14]Also, market clearing does not hold for every timestamp as a results of the imprecision in the timestamps.

[15]Here we outline the main considerations and move a detailed discussion in Appendix C. We discuss the advantages of an event-time approach over a wall-clock time approach in which all trades within a time unit are grouped together.

[16]This approach is comparable to the methodology developed in Aquilina, Budish, and O'Neill (2021), Ernst (2020), and Ernst, Sokobin, and Spatt (2021).

3.2 Variable Construction

We run our analysis at an hourly frequency. The last observed trading price within an hour is assigned to that interval. For order flow, we sign volume using the trade direction indicator from the data set. Then, we sum signed volume for every hour. Thus, order flow has the same frequency as the prices that we observe.

We are focusing on differences in price discovery and liquidity provision of orders with the same characteristics between trader types. In terms of order characteristics, we distinguish between aggressive (market and marketable limit) orders and passive (limit) orders based on the aggressor flag. In terms of trader types, we distinguish between principal and agent traders based on the account role information in the data set. Orders and trades must be identified as principal or agent trades with the distinction not being arbitrary (Eurex, 2021a).[17]

We use the information on the aggressor flag and account role to create order flow variables for every aggressor flag-account role combination. This yields several account role-aggressor flag combinations:

1. agent flow,
2. Principal flow,
3. Agent aggressive flow,
4. Agent passive flow,
5. Principal aggressive flow,
6. Principal passive flow,
7. Aggressive flow,
8. Passive flow.

1 and 2, 3 – 6, as well as 7 and 8 clear the market, thus, in the empirical analysis at least one of the respective account role-aggressor flag combinations has to be omitted from the model for estimation.

We give an overview of the order flow variables and exhibit summary statistics on returns as well as the order flow variables in Table 1. Due to market clearing, summary statistics on agent flow and principal flow as well as on aggressive flow and passive flow mirror each other. Our sample contains many extreme observations for returns and order flow. Differences between the order flow variables give a first indication that orders with the same characteristics are used differently by different trader types. The kurtosis of most order flow variables is considerably higher than that of the returns: there are more tail realizations for order flow than for returns in our sample period.

[17] For accounts classified by Eurex as agent accounts (account code "A"), we retain the label. We label accounts classified by Eurex as proprietary (account code "P") and market maker (account code "M") as principal accounts.

Table 1: Summary statistics returns and order flow

This table reports summary statistics for returns (based on the last trading price) as well as order flow variables at an hourly frequency. Order flow variables are computed based on signed order flow for the account role-aggressor flag combination and accumulated for the respective hour. Order flow variables in EUR.

	Mean	Median	Std. Dev	Skewness	Kurtosis	N Obs
Returns	−0.0008	0.0000	35.2872	−1.5336	70.9444	32479
Agent flow	−174779.9703	0.0000	18559671.5705	0.4328	51.2954	32480
Principal flo	174747.6365	0.0000	18559633.9941	−0.4328	51.2959	32480
Aggressive agent flow	114941.2252	164868.5000	18456698.0294	−1.5293	171.2922	32480
Passive agent flow	−291431.7354	−48771.5000	11830447.2718	−0.2090	183.6812	32480
Aggressive principal flow	191467.2131	25812.0000	21322857.1656	7.3871	547.8034	32480
Passive principal flow	−15009.0367	−6671.5000	24204727.9310	−4.8319	328.4736	32480
Aggressive flow	306408.4383	138030.0000	29200497.5200	2.3092	289.1163	32480
Passive flow	−306440.7722	−138872.5000	29200436.9199	−2.3092	289.1186	32480

4 Methodology

To study information content in order flow and liquidity provision by different traders, we propose a state space model. The framework builds on Hasbrouck's (1993) approach as well as on the state space framework developed in Menkveld, Koopman, and Lucas (2007). The framework decomposes security prices into an efficient price component as well as a deviation from the efficient price, the pricing error. Our approach allows us to decompose observed prices into both components and to get estimates of both the efficient price series as well as the size of the pricing error.

4.1 State Space Model

Following Hasbrouck (1993) and Campbell et al. (1998), efficient prices are modeled to follow a martingale and observed prices are the sum of the efficient price and the pricing error

$$p_t = m_t + s_t \tag{1}$$

$$m_t = m_{t-1} + w_t \tag{2}$$

with p_t denoting log prices, $s_t \sim \mathcal{N}(0, \sigma_s^2)$ being the pricing error, m_t the efficient price, and $w_t \sim \mathcal{N}(0, \sigma_w^2)$ innovations in the efficient price. Identification of pricing errors in this standard model relies on either the assumption of independent pricing errors and innovation in the efficient price series or fixing the correlation between pricing errors and innovations in efficient prices to a specific value (George and Hwang, 2001; Menkveld, Koopman, and Lucas, 2007). Imbalances in order flow help explaining pricing errors and innovations in order flow contain information (Brandt and Kavajecz, 2004; Pasquariello and Vega, 2007; Evans and Lyons, 2008; Hendershott and Menkveld, 2014). We thus use information on signed order flow to identify pricing errors and innovations in efficient prices. It follows the full state space model

$$p_t = m_t + s_t \tag{3}$$

$$m_t = m_{t-1} + \sum_{s \in \mathcal{S}} \gamma_s \tilde{x}_{s,t} + w_t \tag{4}$$

$$s_t = \phi s_{t-1} + \sum_{s \in \mathcal{S}} \delta_s x_{s,t} + \varepsilon_t \tag{5}$$

with $w_t \sim \mathcal{N}(0, \sigma_w^2)$ and $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. $x_{s,t}$ denotes order flow by account role-aggressor flag combination s and $\tilde{x}_{s,t}$ are surprises in order flow. $\mathcal{S}$ denotes the account role-aggressor flag combinations included in the estimation. The identifying assumption is that w_t and ε_t are uncorrelated (Durbin and Koopman, 2012; Hendershott and Menkveld, 2014; Brogaard, Hendershott, and Riordan, 2014). Similar models have been applied in Menkveld, Koopman, and Lucas (2007), Menkveld (2013), Hendershott and Menkveld (2014), Brogaard, Hendershott, and Riordan (2014), Chordia, Green, and Kottimukkalur (2018), and

Yueshen and Zhang (2020), among others, see also Hasbrouck (2007). Economically, the correlation between pricing errors and innovations in order flow is due to trading activity in the market that is captured by the order imbalance. Thus, once we control for order flow and innovations in order flow, the orthogonal part is arguably independent.

4.2 Order Flow Series

Surprises in order flow are obtained as the residual from an VAR model for all order flow series included in the specification, with the number of lags determined using BIC. We perform the same analysis estimating a AR model, again determining the optimal lag length using BIC.[18]

Incorporating information on account role and aggressor flag yields several account role-aggressor flag combinations, as discussed in in Section 3.2.[19] As total order flow clears the market, at least one of the respective account role-aggressor flag combinations has to be omitted from the model for estimation. We thus estimate differences between the included order flow series. Given that our focus is on differences between traders, we use aggressive and passive order flow, respectively, and distinguish between principals and agents.[20]

As we control for trading volume in euros, our estimates on order flow and surprises in order flow account for differences in order flow and measure *relative* differences in order flow. We this estimate information scaled by volume. Additionally, we quantify the contribution of an account role-aggressor flag combination to price discovery by expressing the variation in efficient prices that can be explained by innovations in the respective order flow series relative to the total variation in efficient prices. This yields

$$\frac{\gamma_s^2 var(\tilde{x}_s)}{\gamma' \mathbf{\Sigma} \gamma + \sigma_w^2}$$

where s denotes the account role-aggressor flag combination, γ is the vector of estimated coefficient on innovations in order flow and $\mathbf{\Sigma}$ is the covariance matrix of innovations in order flow.

4.3 Model Estimation

We estimate our state space model over the sample period, similar to Menkveld (2013) and Hendershott and Menkveld (2014), rather than estimating the model day-by-day as in Brogaard, Hendershott, and Riordan (2014). As Easley et al. (2008) document evidence consistent with predictable trade patterns across

[18] When estimating AR and VAR models with different lag lengths, we adjust the number of observations included such that models with the same number of observations are compared.

[19] These are agent flow, principal flow, agent aggressive flow, agent passive flow, principal aggressive flow, principal passive flow, aggressive flow, passive flow.

[20] Thus, for the specification with aggressive order flow omitted, we include agent passive and principal passive order flow. For the specification with passive order flow omitted, we include agent aggressive and agent passive order flow.

trading days and our question addresses differences in price discovery at a lower frequency, this is more appropriate than a day-by-day estimation. In particular, this enables us to identify longer-lasting pricing errors. Also, Easley et al. (2008) find persistence in uninformed order arrival. This motivates our specification order flow.

Our state space model can be mapped into the standard state space representation (Durbin and Koopman, 2012) and standard estimation techniques apply. We describe the mapping in appendix D.

The model is estimated by maximum-likelihood estimation and the Kalman filter is used to evaluate the likelihood function. The Kalman filter requires initial conditions for the state variables, given by a prior mean and a prior variance. Since the efficient price series is assumed to follow a martingale walk, the state for the efficient price series is initialized as diffuse. Therefore, the prior variance is set to κ with $\kappa \to \infty$. The prior for the pricing errors are initialized as stationary using the unconditional variance.

Based on the estimation results, the Kalman smoother is used to obtain estimates of the unobserved states, conditional on all observations. This allows obtaining estimates of the efficient price series as well as of pricing errors at every point in time. Starting values for the maximum likelihood estimation are obtained in three steps. First, we obtain starting values for a reduced form model excluding order flow and innovations in order flow

$$
\begin{aligned}
p_t &= m_t + s_t \\
m_t &= m_{t-1} + w_t \\
s_t &= \phi s_{t-1} + \varepsilon_t
\end{aligned}
$$

with $w_t \sim \mathcal{N}(0, \sigma_w^2)$ and $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. We obtain the starting values based on return variances and autocovariances, details are provided in Appendix E. Second, the estimation results from this reduced form model are used as starting values for the full state space model, but with the coefficients on the order flow variables estimated as states rather than parameters in MLE.[21] Finally, estimation results from the second model are used to estimate the full model by maximum likelihood.[22] Inference is based on robust, quasi-maximum likelihood standard errors (Harvey, 1990) such that inference is still valid under misspecification. Estimation is implemented using the state space models package within statsmodels in python (Seabold and Perktold, 2010; Fulton, 2015).

The methodology has several advantages over alternative approaches, as discussed in Menkveld, Koopman, and Lucas (2007), Menkveld (2013), Hendershott and Menkveld (2014), and Brogaard, Hendershott, and Riordan (2014). Estimation using maximum likelihood is efficient and unbiased under the assumption of correct model specification. We only observe trading prices during the trading hours. Thus missing observations have to be dealt with. The

[21]Therefore, the coefficients are introduced as other latent state variables, for example $\delta_t = \delta_{t-1}$, and initialized as diffuse states.

[22]We perform simulation exercises revealing that this approach yields reliable convergence of the maximum likelihood estimation to known parameters.

Kalman filter deals in a tractable manner with missing observations by extrapolating the state vector from the last observation, while the Kalman smoother interpolates between observations (Durbin and Koopman, 2012). This allows obtaining estimates of the states even for periods without observations. Also, the model can incorporate level shifts and structural breaks in the time series. This is important given that we study almost a decade of trading data.

5 Results

In this section we first present reduced form results to illustrate our empirical approach. Then we turn to the full model on information content and liquidity provision of different trader's order flow. In Appendix F we present additional results showing the robustness of our findings when we allow pricing errors to over- and under-react to past innovations in order flow.

5.1 Reduced Form Results

We first present a reduced form version of our state space model omitting order flow and innovations in order flow. We obtain estimates of the efficient price series and pricing errors as smoothed states from the model and relate the order flow variables to changes in efficient prices and pricing errors. Therefore, we estimate the model

$$\begin{aligned} p_t &= m_t + s_t \\ m_t &= m_{t-1} + w_t \\ s_t &= \phi s_{t-1} + \varepsilon_t \end{aligned}$$

with $w_t \sim \mathcal{N}(0, \sigma_w^2)$ and $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. As described in Section 4, the identifying assumption for estimation is that w_t and ε_t are uncorrelated. Estimation results are presented in Table 2.

Then, we relate changes in the efficient price series as well as pricing errors to order flow as well as innovations in order flow. Both order flow variables are expressed in EUR. We compute correlations between the order flow and price variables over time and plot them with the corresponding confidence intervals in Figures 2 and 3. For each month, quarter, and year in the sample, we compute correlations based on hourly data.[23]

The correlations presented in Figures 2 and 3 are quarterly. The patterns suggest that principals, even if using market orders, trade less in the direction of price pressures than agents. At the same time, agents' limit orders are overall positively correlated with pricing errors in the second half of the sample period.

[23] To compute the confidence intervals, we first apply a Fisher transformation to the correlation coefficients

$$\begin{aligned} z &= \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) \\ &= \tanh^{-1}(r) \end{aligned}$$

Table 2: Estimation results of the reduced form state space model

We present estimation results for the reduced form state space model with autocorrelation in transitory pricing errors and omitting the order flow series given by

$$p_t = m_t + s_t$$
$$m_t = m_{t-1} + w_t$$
$$s_t = \phi s_{t-1} + \varepsilon_t$$

at an hourly frequency. Standard deviations are in *bp*. Robust standard errors are computed. * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level.

Variable	Estimate
σ_w	22.1363*** (0.4222)
σ_ε	21.3933*** (0.4555)
ϕ	0.8644*** (0.0088)
#Observations	32,480

Similar patterns are true for innovations in order flow and changes in the efficient price series. Innovations in agent aggressive flow are positively correlated with changes in efficient prices. The correlations between principal aggressive flow and changes in efficient prices decreases over the sample period to zero. Passive order flow is negatively correlated with changes in the efficient price series over all account roles as well as for agents and principals.

where r is the correlation. Then, two sided confidence limits are computed as

$$z_U = z + z_{1-\alpha/2}\sqrt{\frac{1}{N-3}}$$
$$z_L = z - z_{1-\alpha/2}\sqrt{\frac{1}{N-3}}$$

where N denotes the number of observations used to compute the correlation and $z_{1-\alpha/2}$ is the critical value of the normal distribution at an α significance level. Finally, critical values for the correlation are obtained by transforming the confidence limits

$$r_U = \frac{\exp(2z_U - 1)}{\exp(2z_U + 1)} = \tanh(z_U)$$
$$r_L = \frac{\exp(2z_L - 1)}{\exp(2z_L + 1)} = \tanh(z_L).$$

Graphical inspection of the correlations suggests that aggressive agent flow contains information. The positive correlation between pricing errors and aggressive agent flow is both consistent with agents trading in the direction of price pressures – thus demanding liquidity – as well as with prices overreacting. This evidence gives further motivation for the state space model presented in Section 4. As pricing errors and innovations in efficient prices tend to be correlated, assuming the innovations in a simple state space model without order flow variables to be uncorrelated is not sufficient. Instead, order flow has to be accounted for to get meaningful results. The full state space model results presented in the following Section address these points.

5.2 Full Results

In this section we present results on the full state space model incorporating order flow and surprises in order flow. We first only include information on the account role and the aggressor flag, before incorporating information on both account role and aggressor flag.

Estimation by aggressor flag

Estimation results including aggressive order flow – we thus estimate relative differences in information scaled by volume between aggressive and passive flow – are presented in column (1) of Table 3. The results suggest that, in comparison to passive flow, aggressive order flow contains more information. This is consistent with informed traders using market orders as in Harris (1998).

Kaniel and Liu (2006) argue that if information is more short-lived, traders rather use market orders. Our finding that overall market orders contain more information scaled by volume suggests that the information traded on in the futures market is relatively short lived.

Pricing errors loading positively on aggressive order flow is consistent with overreactions to information as well as market orders demanding liquidity.

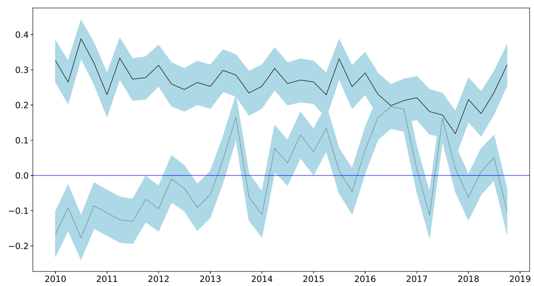
Estimation by account role

Next, we estimate the state space model including agent flow (column (2) of Table 3). We thus estimate relative differences between agents and principals. In comparison to principal flow, agent flow contains more information in the sense that changes in the efficient price series load significantly stronger on innovations in agent flow. At the same time, there are no differences in terms of trading behavior in the direction of (or against) price pressures. The finding that agent flow is relatively more informative is in line with the findings of Menkveld, Sarkar, and Wel (2012) for the US treasury market.

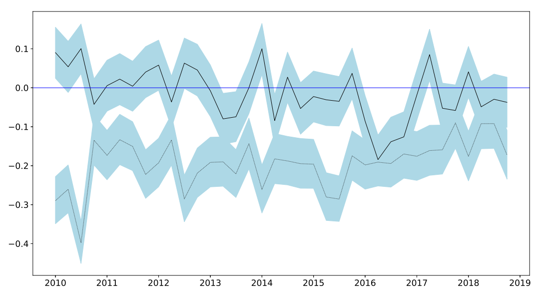
In terms of order informativeness, agent flow resembles aggressive order flow. Our findings suggest that agents are both informed and more informed than principal traders. It is noteworthy that the relative difference in order informativeness between agent flow and principal flow is larger than the relative

Figure 2: Quarterly correlations between order flow and pricing errors

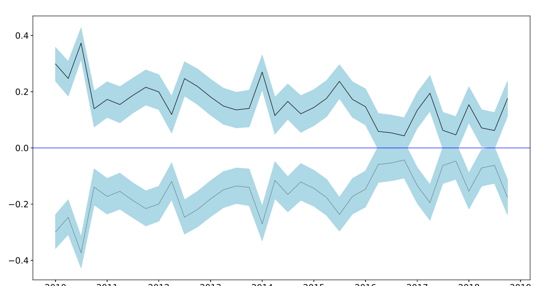
The figure plots quarterly correlations between order flow and pricing errors by account role. The solid line depicts aggressive order flow and the dotted line passive order flow. Blue areas are 95% confidence intervals. Note the different scales of the y-axes.



(a) Agent flow



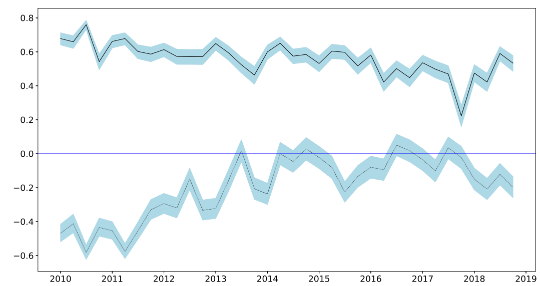
(b) Principal flow



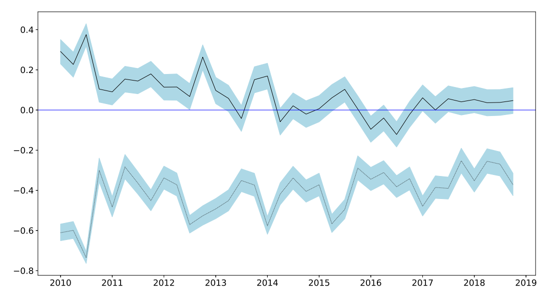
(c) All account roles

Figure 3: Quarterly correlations between innovations in order flow and efficient price changes

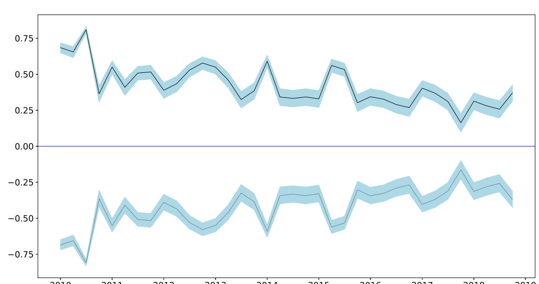
The figure plots quarterly correlations between innovations in order flow and changes in the efficient price series by account role. The solid line depicts aggressive order flow and the dotted line passive order flow. Blue areas are 95% confidence intervals. Note the different scales of the y-axes.



(a) Agent flow



(b) Principal flow



(c) All account roles

Table 3: State space model with order flow

This table presents estimation results for the full state space model

$$p_t = m_t + s_t$$
$$m_t = m_{t-1} + \sum_{s\in\mathcal{S}} \gamma_s \tilde{x}_{s,t} + w_t$$
$$s_t = \phi s_{t-1} + \sum_{s\in\mathcal{S}} \delta_s x_{s,t} + \varepsilon_t$$

at an hourly frequency. x_t is order flow and $\tilde{x}_t$ are surprises in order flow obtained as the residual from a VAR model. The subscripts on γ and δ denote the account role with c (p) standing for agent (principal) and the aggressor flag with a (n) standing for aggressive (passive) order flow. Variances are in *bp* and δ as well as γ in $bp/1,000,000$ EUR. Robust standard errors in parentheses. * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level.

Variable	(1)	(2)	(3)	(4)
σ_w	19.7390***	18.8255***	17.6391***	19.8213***
	(0.2453)	(0.1484)	(0.2315)	(0.2183)
σ_ϵ	20.2853***	21.0704***	18.8810***	20.1506***
	(0.3829)	(0.3096)	(0.4438)	(0.3856)
ϕ	0.8888***	0.8271***	0.8806***	0.8768***
	(0.0128)	(0.0140)	(0.0153)	(0.0137)
efficient price				
γ_a	0.3710***			
	(0.0304)			
γ_c		0.7214***		
		(0.0332)		
$\gamma_{c,a}$			0.8213***	
			(0.0460)	
$\gamma_{p,a}$			0.0057	
			(0.0306)	
$\gamma_{c,n}$				−0.1823***
				(0.0694)
$\gamma_{p,n}$				−0.4502***
				(0.0402)
pricing error				
δ_a	0.0576**			
	(0.0254)			
δ_c		−0.0380		
		(0.0311)		
$\delta_{c,a}$			0.0996***	
			(0.0365)	
$\delta_{p,a}$			0.0582**	
			(0.0265)	
$\delta_{c,n}$				−0.1579***
				(0.0550)
$\delta_{p,n}$				−0.0086
				(0.0303)
#Observations	32,225	32,219	32,223	32,221

difference between aggressive and passive order flow. Thus, the relative informational advantage of agents over principals is larger than that of traders using market orders over trader using limit orders. Rather than comparing the performance of different order types, we thus advocate to compare different trader types. Dependent on the circumstances, agents might feel inclined to use different order types to maximize their profits. If we analyze the performance of order types, we do not know *who* is using these orders. Given these findings, we combine the information on account role and order type in the next section.

Also, the difference in informativeness of aggressive order flow on the one hand and agent order flow on the other hand suggests that there are at least some informed traders that are using limit orders, consistent with Collin-Dufresne and Fos (2015) and Kacperczyk and Pagnotta (2019). However, rather than looking at stock-specific information, we are concerned with economy-wide information.

Estimation by account role and aggressor flag

Given the results from estimating the model by aggressor flag and account role separately, we next incorporate information on both. First, we estimate the state space model for aggressive agent flow and principal flow (column (3) of Table 3). Then, we estimate the model for passive agent and principal flow (column (4) of Table 3). This specification allows us to compare relative differences between trader types – agents and principals – that are using the same order types. Hence, rather than asking the question how different orders perform, we are focusing on differences between trader types.

Recall that aggressive orders are either market orders or marketable limit orders. In line with the previous results, changes in efficient prices load positive on innovations in aggressive agent flow. The coefficient on aggressive agent flow is statistically highly significant and economically relevant. The same does not hold true for aggressive principal flow, where we cannot reject the null hypothesis that innovations in efficient prices do not load on innovations in aggressive principal flow.

We express the contribution to price discovery as the fraction of variation in efficient prices that can be explained by the variance in aggressive agent flow, that is

$$\frac{\gamma_{c,a}^2 var(\tilde{x}_{c,i)}}{\gamma' \mathbf{\Sigma} \gamma + \sigma_w^2}$$

where γ is the vector of estimated coefficient on innovations in order flow and $\mathbf{\Sigma}$ is the covariance matrix of innovations in order flow. At an hourly frequency, this yields a value of approximately 0.25. For principal aggressive flow it is virtually 0.[24]

These differences in contribution to price discovery are economically meaningful. In general, our results are consistent with informed agents using market orders (Harris, 1998). Agents are relatively more informed. Principals may infer

[24] We exhibit these contribution graphically in Figure 1.

information from order flow (Evans and Lyons, 2002). For example, principal dealers observe the order flow of their customers and may also deduce information from the state of the order book. Hortaçsu and Kastl (2012) argue that dealers may extract information from the orders of their customers to either compete with their customers or deduce fundamental information from order flow. Our results suggest that at an hourly frequency, principals do not learn enough information to place informative market orders. This may be because optimal order placement of informed traders (for example as in Collin-Dufresne and Fos (2016)) does not allow principals to deduce information from the order book and trade flows.

The estimation results for agents and principals using passive – i.e. limit – orders, reveal that efficient price changes load negatively on innovations in passive agent and principal flow. This is consistent with traders using limit orders being adversely selected (see, for example, Gârleanu and Pedersen (2004) and Linnainmaa (2010)). Principals are stronger subject to adverse selection that clients. Together with the results on aggressive agent and principal flow, these results suggest that both informed agents and principals use market orders, while uninformed traders rather use limit orders and are adversely selected.

A potential mechanism causing informed traders to use limit orders instead of market orders are rebates for supplying liquidity. If rebates are sufficiently high compared to the execution risk of limit orders relative to market orders, informed traders prefer to submit limit orders. Our results indicate that this is, in general, not the case as limit orders are on average subject to adverse selection while efficient price innovations load positively on market orders.

The results for principal traders are consistent with principals making the market and offering quotes to other traders. Agents using passive orders are less informed than traders using aggressive orders, but more informed that principal's passive orders. Together with the results on aggressive order flow, a possible explanation is that agents with a longer information horizon not only trade using market orders, but also use limit orders, consistent with Kaniel and Liu (2006). Another interpretation that is consistent with our results is that the group of agent traders is diverse. On the one hand, the group consists of uninformed traders whose limit orders are adversely selected and who do not contribute to price discovery. On the other hand, the group consists of informed traders with long-lived information who are using limit orders that contribute to price discovery, as documented by Bloomfield, O'Hara, and Saar (2005), Collin-Dufresne and Fos, 2015, Baruch, Panayides, and Venkataraman (2017), and Kacperczyk and Pagnotta (2019). Furthermore, our results suggest that principals do not learn enough from the orders of their customers to prevent their limit orders from being adversely selected (Hortaçsu and Kastl, 2012).

Note that the signs of the coefficients on the impact of order flow on efficient price innovations and pricing errors are not fully aligned with the results of Brogaard, Hendershott, and Riordan (2014) for HFTs' and non-HFTs' liquidity demanding and supplying orders. Similarly to our results on aggressive order flow, they find liquidity demanding trades to be positively associated with efficient price innovations. At the same time, in their results, pricing errors load negatively on liquidity demanding trades. Thus, such trades push observed prices in the direction of efficient prices. We find that pricing errors load positively on aggressive order flow. Similarly for passive order flow, we find negative coefficients for the impact of order flow on efficient prices and negative coefficients for the impact on pricing errors. Brogaard, Hendershott, and Riordan (2014) find negative and positive coefficients, respectively. Our result that pricing errors load positively on aggressive order flow is consistent with prices overshooting to new information. Moreover, as noted by Brogaard, Hendershott, and Riordan (2014), in the state space formulation, traders managing risk causes order flow to relate positively to pricing errors. In our setting, this has the intuitive interpretation of traders using market orders for risk management.

Overall, our findings suggest that it does not only matter which order type is used, but also *who* uses which order. We document relative differences in information scaled by volume of the same order type between different groups of traders. Combining information on both the order type used and the account role of the trader using the order gives a substantially more differentiated picture of order informativeness than analyzing either separately.

5.3 Estimation by market volatility

We are interested in how trading patterns change depending on the market conditions. Therefore, we include information of the CBOE's volatility index (VIX). The VIX serves as a background variable that is plausibly exogenous to trading in Euro STOXX 50 futures as it is calculated based on options on the S&P 500. At the same time, it captures general market conditions that influence trading in Euro STOXX 50 futures. Thus, we prefer this specification over alternative specifications using measures such as returns in certain time intervals or realized volatilities as these are endogenous to the trading process.

We augment our state space model as follows. We include an indicator variable that equals one if the VIX on the respective trading day falls in the lowest and highest decile, respectively, of the distribution over our sample period and interact it with the order flow variables.[25] As trading in Euro STOXX 50 futures might not instantaneously react to changes in the VIX and our focus is on price discovery over longer time horizons, we assign the indicator based on daily VIX levels. Then it follows for the state space model:

$$p_t = m_t + s_t \tag{6}$$

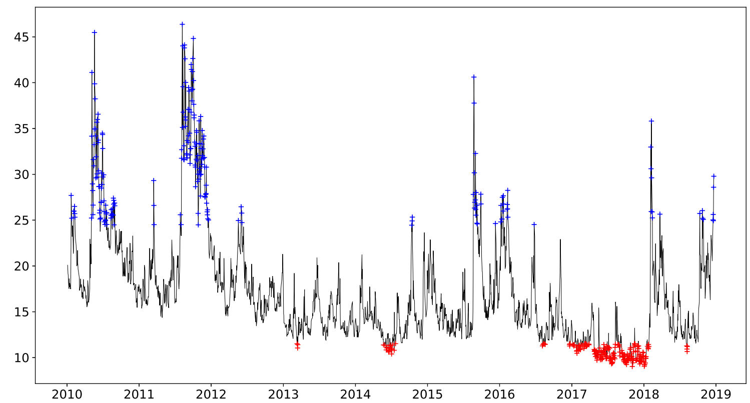
$$m_t = m_{t-1} + \sum_{s\in\mathcal{S}} \gamma_s \tilde{x}_{s,t} + \sum_{s\in\mathcal{S}} \gamma_{s,VIX} \mathbb{1}(VIX_t \in \mathcal{D}) \tilde{x}_{s,t} + w_t \tag{7}$$

$$s_t = \phi s_{t-1} + \sum_{s\in\mathcal{S}} \delta_s x_{s,t} + \sum_{s\in\mathcal{S}} \delta_{s,VIX} \mathbb{1}(VIX_t \in \mathcal{D}) x_{s,t} + \varepsilon_t. \tag{8}$$

[25] We compute the distribution of the VIX over our sample period based on daily closing prices. Then, using each daily closing price, we assign the indicator variable to all observations on the respective trading day, depending on in which percentile of the sample distribution the respective closing price falls.

Figure 4: Daily VIX closing prices over the sample period

The figure plots daily VIX closing prices over the sample period. Blue crosses indicate observations in the top decile of the sample distribution and red crosses indicate observations in the bottom decile of the sample distribution.



In the specification, $\mathcal{D}$ denotes the respective decile. This specification captures differences in price discovery and liquidity provision in normal periods versus high-VIX and low-VIX periods. Again, we ask the question which traders are informed and how they are trading on their information. Thus, we again focus on differences between trader types conditional on order type.

Observations in the bottom and top decile of the VIX distribution are distributed unevenly over our sample period (Figure 4). Most of the the observations in the bottom decile of the sample distribution cluster in 2017. Most observations that fall in the top decile of of the sample distribution cluster early in our sample period in the years 2010 and 2011.

As in the previous analysis, we run the state space model for aggressive and passive order flow separately and include order flow from agents and principals. Overall, our results indicate differences in trading patterns in times of high and low market volatility. Also, the differences are larger for agents than for principals.

The results for aggressive flow are presented in Table 4. In comparison to the results presented in the previous section, our results are remarkably stable, indicating the robustness of our results. In low-volatility regimes, the relative contribution of aggressive agent flow to price discovery is lower than over the whole sample period. Still, agents' contribution to price discovery using market and marketable limit orders dominates the contribution of principals using the same order type.

The coefficient on client flow in the pricing error equation, δ, is lower in low-volatility periods. The combination of a positive coefficient on innovations in order flow and a negative coefficient on the impact on pricing errors is more in line with the results of Brogaard, Hendershott, and Riordan (2014) for liquidity demanding HFTs.

While we find differences in agents' usage of aggressive orders between low-volatility periods and the overall sample, our main conclusions remain unchanged. Agents contribute relatively more to price discovery using their aggressive orders than principals do. For principals using aggressive orders, most of the coefficients for the low-volatility dummy are insignificant. Overall, there is no evidence for a higher contribution of principals' aggressive orders to price discovery in low-volatility periods.

For high-volatility periods, the patterns are reversed in comparison to low-volatility periods – except that we find small changes for principals. Efficient price innovations load stronger on innovations in aggressive agent flow in high-volatility periods relative to the overall sample. That is, aggressive agent flow contains more information scaled by volume and contributes more to price discovery in high-VIX periods than in the overall sample. A plausible explanation for this pattern is that informed agents rather use market or marketable limit orders in high-volatility periods while they rely on limit orders in low volatility periods. This is consistent with a pecking order of order types dependent on the market conditions, in a spirit of Menkveld, Yueshen, and Zhu (2017).

The impact of agent flow on pricing errors increases – if anything – in high volatility periods, reinforcing our results for the overall sample. We interpret this result as overreactions to information that are especially pronounced in highly volatile periods.

For principal trades, the results are mostly unchanged. Efficient price innovations do not load significantly stronger on innovations in aggressive principal flow in high-VIX periods than in the overall sample. This is consistent with the previous finding that most information is incorporated into prices through agent flow. In high volatility periods, the variance in agent flow accounts for roughly 45% of the variance in efficient price innovations, while the variance in principal flow accounts for less than 1%. We do not find evidence that in highly volatile times, principals are better able to extract information from the state of the order book and trade on this information (Parlour, 1998).

Next, we turn to the results for passive order flow (Table 5). In low-volatility periods, innovations in efficient prices load stronger on passive agent flow than over the entire sample period. Recall that over the whole sample, changes in efficient prices load negatively on passive agent flow, indicating that agents' limit orders are adversely selected. In low volatility periods, agents are less exposed to adverse selection. We do not find evidence that principals' limit order are less exposed to adverse selection in low volatility periods than in the overall sample period.

These findings can be explained by agents being better able to manage their orders in times of low volatility. Also, some informed agents may use passive orders rather than aggressive orders in less volatile times. As a result, on aver-

Table 4: State space model for aggressive order flow including VIX

This table presents estimation results for the full state space model including VIX

$$
\begin{aligned}
p_t &= m_t + s_t \\
m_t &= m_{t-1} + \sum_{s \in \mathcal{S}} \gamma_s \tilde{x}_{s,t} + \sum_{s \in \mathcal{S}} \gamma_{s,VIX} \mathbb{1}(VIX_t \in \mathcal{D}) \tilde{x}_{s,t} + w_t \\
s_t &= \phi s_{t-1} + \sum_{s \in \mathcal{S}} \delta_s x_{s,t} + \sum_{s \in \mathcal{S}} \delta_{s,VIX} \mathbb{1}(VIX_t \in \mathcal{D}) x_{s,t} + \varepsilon_t
\end{aligned}
$$

at an hourly frequency. x_t is order flow and $\tilde{x}_t$ are surprises in order flow obtained as the residual from a VAR model. $\mathbb{1}(VIX_t \in \mathcal{D})$ is an indicator that equals one if the closing VIX on the respective trading day is in the lowest decile or highest decile, respectively, of the distribution over the sample period. Passive order flow is omitted from the specification. The subscripts on γ and δ denote the account role with c (p) standing for agent (principal) and the aggressor flag with a standing for aggressive order flow. Standard deviations are in *bp* and δ as well as γ in $bp/1,000,000$ EUR. Robust standard errors in parentheses. * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level.

	low	high
σ_w	17.5000***	17.1452***
	(0.2370)	(0.2382)
σ_ϵ	18.8085***	18.5078***
	(0.4708)	(0.3984)
ϕ	0.8859***	0.8901***
	(0.0150)	(0.0133)
efficient price		
$\gamma_{c,a}$	0.8285***	0.7126***
	(0.0508)	(0.0453)
$\gamma_{p,a}$	0.0376	−0.0076
	(0.0345)	(0.0294)
$\gamma_{c,a,VIX}$	−0.2932***	0.6286***
	(0.0865)	(0.1391)
$\gamma_{p,a,VIX}$	−0.1533***	0.1736
	(0.0574)	(0.1516)
pricing error		
$\delta_{c,a}$	0.1395***	0.1028***
	(0.0408)	(0.0330)
$\delta_{p,a}$	0.0360	0.0540**
	(0.0292)	(0.0260)
$\delta_{c,a,VIX}$	−0.2400***	0.2673**
	(0.0705)	(0.1236)
$\delta_{p,a,VIX}$	0.0813	0.1298
	(0.0522)	(0.1412)
#Obs	32,223	32,223

Table 5: State space model for passive order flow including VIX

This table presents estimation results for the full state space model including VIX

$$
\begin{aligned}
p_t &= m_t + s_t \\
m_t &= m_{t-1} + \sum_{s \in \mathcal{S}} \gamma_s \tilde{x}_{s,t} + \sum_{s \in \mathcal{S}} \gamma_{s,VIX} \mathbb{1}(VIX_t \in \mathcal{D}) \tilde{x}_{s,t} + w_t \\
s_t &= \phi s_{t-1} + \sum_{s \in \mathcal{S}} \delta_s x_{s,t} + \sum_{s \in \mathcal{S}} \delta_{s,VIX} \mathbb{1}(VIX_t \in \mathcal{D}) x_{s,t} + \varepsilon_t
\end{aligned}
$$

at an hourly frequency. x_t is order flow and $\tilde{x}_t$ are surprises in order flow obtained as the residual from a VAR model. $\mathbb{1}(VIX_t \in \mathcal{D})$ is an indicator that equals one if the closing VIX on the respective trading day is in the lowest decile or highest decile, respectively, of the distribution over the sample period. Aggressive order flow is omitted from the specification. The subscripts on γ and δ denote the account role with c (p) standing for agent (principal) and the aggressor flag with n standing for passive order flow. Standard deviations are in bp and δ as well as γ in $bp/1,000,000$ EUR. Robust standard errors in parentheses. * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level.

	low	high
σ_w	19.7916***	19.3578***
	(0.2371)	(0.2363)
σ_ϵ	20.0336***	19.7299***
	(0.3998)	(0.3703)
ϕ	0.8846***	0.8875***
	(0.0136)	(0.0127)
efficient price		
$\gamma_{c,n}$	−0.3269***	−0.0912
	(0.0810)	(0.0632)
$\gamma_{p,n}$	−0.4340***	−0.3949***
	(0.0442)	(0.0385)
$\gamma_{c,n,VIX}$	0.2691**	−0.7087***
	(0.1104)	(0.2409)
$\gamma_{p,n,VIX}$	0.0508	−0.4234***
	(0.0745)	(0.1434)
pricing error		
$\delta_{c,n}$	−0.0849	−0.1415***
	(0.0616)	(0.0482)
$\delta_{p,n}$	−0.0372	−0.0159
	(0.0336)	(0.0275)
$\delta_{c,n,VIX}$	0.1962**	−0.2140
	(0.0982)	(0.2324)
$\delta_{p,n,VIX}$	0.1249*	−0.1927
	(0.0648)	(0.1332)
#Obs	32,221	32,221

age, agents are less exposed to adverse selection. At the same time, principal traders may mainly provide liquidity and are subject to adverse selection.

The results for the high-VIX periods are in line with this intuition. In comparison to the overall sample period, efficient price innovations load more negative on both agents' and principals' passive orders. That indicates that limit orders of both trader types are stronger subject to adverse selection in volatile times than in the overall sample period. This result mirrors the results for aggressive orders in high-VIX periods: efficient price innovations load stronger on aggressive agent flow. This suggests that informed traders use predominantly aggressive orders in periods of high volatility. Traders supplying liquidity, in contrast, trade for non-informational reasons.

In high-volatility periods, the change in exposure to adverse selection is higher for agents than for principals. In such periods, principals might have to post limit orders to provide liquidity within the exchange's requirements. At the same time, they might have superior information from the order book and are thus less subject to adverse selection (in line with Hortaçsu and Kastl (2012)). Our results provide suggestive evidence in line with the latter channel. They also suggest that the finding that agent's passive orders are adversely selected is mainly due to high-volatility periods.

These results suggest a pecking order of order types dependent on the market conditions, akin to Menkveld, Yueshen, and Zhu (2017). Also, these results are consistent with the intuition of Kaniel and Liu (2006). Higher market volatility can be interpreted as decreasing the horizon on which investors can trade on their information. As markets are volatile, movements in the disadvantage of a trader may occur more frequent and are less predictable. A reduction in their horizon causes informed traders to use market orders rather than limit orders. This is what we observe in our data.

The results of Collin-Dufresne and Fos (2015) and Kacperczyk and Pagnotta (2019) show that informed traders are using limit orders. Their results speak to insiders that possess firm-specific information. We analyze trading in futures on a Pan-European index, the Euro STOXX 50. Thus, even though traders active in these futures contracts may be motivated by firm-specific information on the constituents, prices of Euro STOXX 50 futures also reflect information on the state of the "European" economy. Our results suggest that also traders possessing information of this nature use limit orders, dependent on the market conditions. Thus, our evidence is consistent with the results of Collin-Dufresne and Fos (2015) and Kacperczyk and Pagnotta (2019) extending to a wider set of information and asset classes.

6 Opening and Closing Auctions

In this Section we analyze trading during opening and closing auctions. With this, we add to the contemporaneous literature focusing on closing auctions in American (Bogousslavsky and Muravyev, 2021) and European (Comerton-Forde and Rindi, 2021) equity markets. We analyze auctions in a futures market that tracks the leading index of the Eurozone.

We identify auctions from the trading data as follows. Trades are grouped using the event time approach described in Section C, using a window length of 100ms. For the opening auction, the trade record with the highest volume during the first 20 seconds of trading, for which the volume of aggressive and passive orders does not match is identified.[26] All other trades executed at the same price within the first 20 seconds of trading with non-matching aggressive and passive volume are labeled as opening auctions. For the closing auction, we identify auctions based on trade records for the last three minutes of trading. This is motivated by the market structure at Eurex. After continuous trading, the call phase for the auction lasts at least three minutes with a random end time. Thus, by focusing on the trade records within the last 3 minutes of trading, we are confident that we do not capture trade records during continuous trading.[27]

In comparison to to continuous trading, the share of agent trading in opening and closing auctions is substantially higher (70% and 48%, respectively, versus 32%).[28] That is, the share of agent trading during opening auctions is higher than during closing auctions. From this follows the question whether participation of different trader types in the opening and closing auction relative to continuous trading have a positive or negative impact on price discovery. In comparison to equity markets, the share or auction volume in daily trading volume is low and usually in the magnitude of a few basis points.

The finding of a low volume share of opening and closing auctions in continuous trading is robust over the sample period and to the identification of auctions.[29] Also, this is in contrast to high closing-auction shares described in the equity markets literature (Bogousslavsky and Muravyev, 2021; Comerton-Forde and Rindi, 2021).

To address the question of price discovery during auctions, we analyze auction prices relative to the prices 15 minutes around the auctions, similar to Bogousslavsky and Muravyev (2021). Also, to account for potential longer lasting pricing errors, price reversals, and long-term patterns in price discovery, we incorporate auction prices into our state space model. Similar the the analysis for periods with high and low market volatility presented in Section 5.3, we allow pricing errors and innovations in efficient prices to depends differently on order flow during auctions in comparison to continuous trading. The focus on longer-term patterns in prices are motivated by the price patterns described in

[26] All records for the auctions should have an aggressor flag that indicates a passive order. In fact, this is not true for all trade records. Ideally, we would like to exclude all trade records with an aggressor flag that indicates an aggressive order as well as all trade records that pertain to limit orders that these market orders executed against. Since various trade records are recorded at the same timestamp and the timestamps are imprecise from May 8, 2013, onward, this is however not possible.

[27] We perform robustness checks with a window of five minutes for the closing auctions and windows up to 3 minutes for the opening auction which leaves the results unchanged.

[28] We provide more summary statistics on the auctions in Appendix G.

[29] We consider a different window length in our event-time approach and allow auctions to be recorded further into continuous trading.

Bogousslavsky (2021). Thus, we also account for noisy prices in the first minutes of the trading day (Bogousslavsky, 2021).

As a first step, we focus on deviations of auction prices similar to Bogousslavsky and Muravyev (2021) and we compute the auction prices deviation as

$$|\ln(p_{cnt}/p_{auc})|$$

for the opening auction and

$$|\ln(p_{auc}/p_{cnt})|$$

for the closing auction where p_{auc} denotes the auction price and p_{cnt} denotes the first (last) price record during continuous trading for the opening (closing) auction. Alternatively, we compute the auction price deviation based on the price 15 minutes after (before) the auction.

We regress the auction price deviation on the share of agent trading during the auction, the VIX on the respective trading day, the auction share in continuous trading, and the share of agent volume in trading volume in 15 minutes around the auction (results are presented in Table 13 in Appendix G). Higher auction price deviations are consistently correlated with higher VIX levels, as also found in Bogousslavsky and Muravyev (2021). Furthermore, when comparing opening auction prices to the price 15 minutes after the opening auction and closing auction prices to the last price recorded during continuous trading, we find that auction price deviations are positively related to the share of agent trading in the auction.

Are these deviations transitory noise? Or do they reflect fundamental information? We address this within our state space model. This allows us to explicitly distinguish between the information and noise components. As our model controls for trading volume, we can compare the the results for the opening and closing auctions even though they differ in volume. As auction volume is by definition passive, we estimate our state space model for passive order flow, distinguishing between principals and agents.

Our results are presented in Table 6. They indicate that trading during opening and closing auctions does not differ substantially from continuous trading in terms of price discovery. Note that we found that passive order flow of both agents and principals is adversely selected and thus efficient price innovations load negatively on innovations in the order flow series. There is some evidence that trading during the closing auction of both agents are principals is more informative than during continuous trading. Taking this into account, order flow during closing auctions is even informative in the sense that efficient price innovations load positively on innovations in the order flow series.

Furthermore, we find that during closing auctions, agents and principals trade more against pricing errors than during continuous trading. This is not in line with the idea that closing auctions are increasing pricing errors. For opening auctions, the evidence is at best weak.

Thus, our results suggest that opening auctions do not differ significantly from continuous trading, irrespective of whether we consider agent or principal trading. Closing auctions do in the sense that both agents and principals trade more against pricing errors. A possible explanation for this is that traders in futures markets align their inventory in the end of the trading day and thus reduce the pricing error (Hendershott and Menkveld, 2014). Thus, our results differ from the equity market literature. A potential explanation are differences in the market we are analyzing relative to equity markets. Trading in the end of the trading day may be substantially impacted by hedging demands that create pricing errors. Also, trading volume in the end of the trading day is low and the market is likely less liquid. In this setting, closing auctions can be seen as a more liquid trading opportunity in a spirit of Budish, Cramton, and Shim (2015).

7 Conclusions

How does information in futures markets get incorporated into prices? How does this depend on market conditions? Do auctions deviate from continuous trading? We address these questions based on trading in Euro STOXX 50 futures and a state space framework. Our results indicate that the classification of who is informed is endogenous to market conditions. We link exogenous information on the trader type to their order's informativeness. Our results indicate that aggressive orders contain more information scaled by volume. Relative order informativeness differs between agent's and principal's orders, with agent's orders being more informative.

Changes in efficient prices load stronger on agent's aggressive orders than on principal's aggressive orders. At an hourly frequency, 25% of the variation in efficient price innovations can be explained by the variance in aggressive agent flow. At the same time, agent's passive orders are less exposed to adverse selection than principal's passive orders. Thus, in futures markets, principals do not deduce sufficient information from the orders of their customers to better protect their limit orders from adverse selection.

Comparing different market conditions as measured by the VIX indicates a pecking order of (informed) traders' order choice. This complements the pecking order of trading venues documented in Menkveld, Yueshen, and Zhu (2017). In low-volatility regimes, agents' passive orders are less subject to adverse selection than in normal times and in high-volatility regimes, aggressive orders are relatively more informative. The results suggest that some informed traders use limit orders not only to trade on stock-specific information, but also trade on economy-wide information in futures markets.

Auctions in futures markets appear different than auction in equity markets (Bogousslavsky and Muravyev, 2021; Comerton-Forde and Rindi, 2021). For opening auctions, we do not find differences in terms of price discovery and liquidity provision relative to continuous trading. For closing auctions, while we do not find differences in terms of price discovery, but we find that both principal's and agent's orders trade stronger against pricing errors than during continuous trading. Thus, rather than creating price deviations, orders in the closing auction push observed prices closer to efficient prices.

Table 6: State space model for non-aggressive order flow including auctions

This table presents estimation results for the full state space model for non-aggressive order flow including opening and closing auctions at an hourly frequency. Aggressive order flow is omitted. The subscripts on γ and δ denote the account role with c (p) standing for agent (principal) and the aggressor flag with n standing for non-aggressive order flow. Standard deviations are in bp and δ as well as γ in $bp/1,000,000$ EUR. Robust standard errors in parentheses. * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level.

	open	close
σ_w	19.8510***	19.9972***
	(0.2046)	(0.1899)
σ_ϵ	20.0147***	18.6370***
	(0.3857)	(0.4070)
ϕ	0.8700***	0.8541***
	(0.0136)	(0.0174)
efficient price		
$\gamma_{c,n}$	−0.2145***	−0.1760**
	(0.0694)	(0.0691)
$\gamma_{p,n}$	−0.4511***	−0.4629***
	(0.0393)	(0.0410)
$\gamma_{c,n,auction}$	1.4936	2.4531**
	(4.7538)	(0.9882)
$\gamma_{p,n,auction}$	0.2744	0.4029
	(4.7195)	(0.3333)
pricing errors		
$\delta_{c,n}$	−0.1545***	−0.1647***
	(0.0547)	(0.0523)
$\delta_{p,n}$	0.0084	0.0048
	(0.0298)	(0.0285)
$\delta_{c,n,auction}$	−7.3689	−7.2249***
	(5.8456)	(0.8272)
$\delta_{p,n,auction}$	−5.6077	−5.5375***
	(5.8678)	(0.8272)
#Obs	34,510	34,510

References

Anand, Amber, Sugato Chakravarty, and Terrence Martell (2005). "Empirical Evidence on the Evolution of Liquidity: Choice of Market versus Limit Orders by Informed and Uninformed Traders". In: *Journal of Financial Markets* 8.3, pp. 288–308.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Clara Vega (2007). "Real-time Price Discovery in Global Stock, Bond and Foreign Exchange Markets". In: *Journal of International Economics* 73.2, pp. 251–277.

Aquilina, Matteo, Eric Budish, and Peter O'Neill (July 2021). *Quantifying the High-Frequency Trading "Arms Race"*. NBER Working Papers 29011. National Bureau of Economic Research, Inc.

Barber, Brad M., Yi-Tsung Lee, Yu-Jane Liu, and Terrance Odean (2009). "Just How Much Do Individual Investors Lose by Trading?" In: *Review of Financial Studies* 22.2, pp. 609–632.

Baruch, Shmuel, Marios Panayides, and Kumar Venkataraman (2017). "Informed Trading and Price Discovery Before Corporate Events". In: *Journal of Financial Economics* 125.3, pp. 561–588.

Beason, Tyler and Sunil Wahal (2020). *The Anatomy of Trading Algorithms.* Working Paper.

Biais, Bruno, Pierre Hillion, and Chester Spatt (1995). "An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse". In: *Journal of Finance* 50.5, pp. 1655–1689.

Bloomfield, Robert, Maureen O'Hara, and Gideon Saar (2005). "The "Make or Take" Decision in an Electronic Market: Evidence on the Evolution of Liquidity". In: *Journal of Financial Economics* 75.1, pp. 165–199.

Bloomfield, Robert, Maureen O'Hara, and Gideon Saar (2015). "Hidden Liquidity: Some New Light on Dark Trading". In: *Journal of Finance* 70.5, pp. 2227–2274.

Bogousslavsky, Vincent (2021). "The Cross-Section of Intraday and Overnight Returns". In: *Journal of Financial Economics* 141.1, pp. 172–194.

Bogousslavsky, Vincent and Dmitriy Muravyev (2021). *Who Trades at the Close? Implications for Price Discovery and Liquidity.* Working Paper.

Brandt, Michael W. and Kenneth A. Kavajecz (2004). "Price Discovery in the U.S. Treasury Market: The Impact of Orderflow and Liquidity on the Yield Curve". In: *The Journal of Finance* 59.6, pp. 2623–2654.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan (2014). "High-Frequency Trading and Price Discovery". In: *Review of Financial Studies* 27.8, pp. 2267–2306.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan (2019). "Price Discovery without Trading: Evidence from Limit Orders". In: *Journal of Finance* 74.4, pp. 1621–1658.

Budish, Eric, Peter Cramton, and John Shim (2015). "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response". In: *The Quarterly Journal of Economics* 130.4, pp. 1547–1621.

Campbell, John Y., Andrew W. Lo, A. Craig MacKinlay, and Robert F. Whitelaw (1998). "The Econometrics of Financial Markets". In: *Macroeconomic Dynamics* 2.4, pp. 559–562.

Chordia, Tarun, T Clifton Green, and Badrinath Kottimukkalur (2018). "Rent Seeking by Low-Latency Traders: Evidence from Trading on Macroeconomic Announcements". In: *Review of Financial Studies* 31.12, pp. 4650–4687.

Collin-Dufresne, Pierre and Vyacheslav Fos (2015). "Do Prices Reveal the Presence of Informed Trading?" In: *Journal of Finance* 70.4, pp. 1555–1582.

Collin-Dufresne, Pierre and Vyacheslav Fos (2016). "Insider Trading, Stochastic Liquidity, and Equilibrium Prices". In: *Econometrica* 84, pp. 1441–1475.

Comerton-Forde, Carole and Barbara Rindi (2021). *Trading @ the Close*. Working Paper.

Durbin, James and Siem Jan Koopman (2012). *Time Series Analysis by State Space Methods*. OUP Catalogue 9780199641178. Oxford University Press.

Easley, David, Robert F. Engle, Maureen O'Hara, and Liuren Wu (2008). "Time-Varying Arrival Rates of Informed and Uninformed Trades". In: *Journal of Financial Econometrics* 6.2, pp. 171–207.

Ernst, Thomas (2020). *Stock-Specific Price Discovery from ETFs*. Working Paper.

Ernst, Thomas, Jonathan Sokobin, and Chester Spatt (2021). *The Value of Off-Exchange Data*. Working Paper.

Eurex (Mar. 2021a). *Conditions for Trading at Eurex Deutschland*. Tech. rep.

Eurex (May 2021b). *Price List of Eurex Clearing AG*. Tech. rep.

Evans, Martin D. D. and Richard K. Lyons (2002). "Order Flow and Exchange Rate Dynamics". In: *Journal of Political Economy* 110.1, pp. 170–180.

Evans, Martin D.D. and Richard K. Lyons (2008). "How is Macro News Transmitted to Exchange Rates?" In: *Journal of Financial Economics* 88.1, pp. 26–50.

Fleming, Michael J., Bruce Mizrach, and Giang Nguyen (2018). "The Microstructure of a U.S. Treasury ECN: The BrokerTec Platform". In: *Journal of Financial Markets* 40.C, pp. 2–22.

Foucault, Thierry (1999). "Order Flow Composition and Trading Costs in a Dynamic Limit Order Market". In: *Journal of Financial Markets* 2.2, pp. 99–134.

Fulton, Chad (2015). "Estimating Time Series Models by State Space Methods in Python: Statsmodels". In:

George, Thomas J and Chuan-Yang Hwang (2001). "Information Flow and Pricing Errors: A Unified Approach to Estimation and Testing". In: *Review of Financial Studies* 14.4, pp. 979–1020.

Goettler, Ronald L., Christine A. Parlour, and Uday Rajan (2009). "Informed Traders and Limit Order Markets". In: *Journal of Financial Economics* 93.1, pp. 67–87.

Gârleanu, Nicolae and Lasse Heje Pedersen (2004). "Adverse Selection and the Required Return". In: *Review of Financial Studies* 17.3, pp. 643–665.

Hamilton, James D. (1986). "State-Space Models". In: *Handbook of Econometrics*. Ed. by R. F. Engle and D. McFadden. Vol. 4. Handbook of Econometrics. Elsevier. Chap. 50, pp. 3039–3080.

Harris, Lawrence (1998). "Optimal Dynamic Order Submission Strategies in Some Stylized Trading Problems". In: *Financial Markets, Institutions & Instruments* 7.2, pp. 1–76.

Harvey, Andrew C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Hasbrouck, Joel (1993). "Assessing the Quality of a Security Market: A New Approach to Transaction-Cost Measurement". In: *Review of Financial Studies* 6.1, pp. 191–212.

Hasbrouck, Joel (2003). "Intraday Price Formation in U.S. Equity Index Markets". In: *Journal of Finance* 58.6, pp. 2375–2400.

Hasbrouck, Joel (2007). *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*. OUP Catalogue 9780195301649. Oxford University Press.

Hendershott, Terrence, Dmitry Livdan, and Norman Schürhoff (2015). "Are Institutions Informed About News?" In: *Journal of Financial Economics* 117.2, pp. 249–287.

Hendershott, Terrence and Albert J. Menkveld (2014). "Price Pressures". In: *Journal of Financial Economics* 114.3, pp. 405–423.

Hoffmann, Peter (2014). "A Dynamic Limit Order Market With Fast and Slow Traders". In: *Journal of Financial Economics* 113.1, pp. 156–169.

Hollifield, Burton, Robert A. Miller, and Patrik Sandås (2004). "Empirical Analysis of Limit Order Markets". In: *Review of Economic Studies* 71.4, pp. 1027–1063.

Hortaçsu, Ali and Jakub Kastl (2012). "Valuing Dealers' Informational Advantage: A Study of Canadian Treasury Auctions". In: *Econometrica* 80.6, pp. 2511–2542.

Huang, Xin (2018). "Macroeconomic News Announcements, Systemic Risk, Financial Market Volatility, and Jumps". In: *Journal of Futures Markets* 38.5, pp. 513–534.

Kacperczyk, Marcin and Emiliano S Pagnotta (2019). "Chasing Private Information". In: *Review of Financial Studies* 32.12, pp. 4997–5047.

Kaniel, Ron and Hong Liu (2006). "So What Orders Do Informed Traders Use?" In: *Journal of Business* 79.4, pp. 1867–1914.

Kawaller, Ira G, Paul D Koch, and Timothy W Koch (1987). "The Temporal Price Relationship between S&P 500 Futures and the S&P 500 Index". In: *Journal of Finance* 42.5, pp. 1309–1329.

Kelley, Eric K. and Paul C. Tetlock (2013). "How Wise Are Crowds? Insights from Retail Orders and Stock Returns". In: *Journal of Finance* 68.3, pp. 1229–1265.

Li, Sida, Mao Ye, and Miles Zheng (Feb. 2021). *Financial Regulation, Clientele Segmentation, and Stock Exchange Order Types*. NBER Working Papers 28515. National Bureau of Economic Research, Inc.

Linnainmaa, Juhani T. (2010). "Do Limit Orders Alter Inferences about Investor Performance and Behavior?" In: *Journal of Finance* 65.4, pp. 1473–1506.

Menkveld, Albert J. (2013). "High Frequency Trading and the New-Market Makers". In: *Journal of Financial Markets* 16.4, pp. 712–740.

Menkveld, Albert J., Siem Jan Koopman, and Andre Lucas (2007). "Modeling Around-the-Clock Price Discovery for Cross-Listed Stocks Using State Space Methods". In: *Journal of Business & Economic Statistics* 25, pp. 213–225.

Menkveld, Albert J., Asani Sarkar, and Michel van der Wel (2012). "Customer Order Flow, Intermediaries, and Discovery of the Equilibrium Risk-Free Rate". In: *Journal of Financial and Quantitative Analysis* 47.4, 821–849.

Menkveld, Albert J. and Bart Zhou Yueshen (2019). "The Flash Crash: A Cautionary Tale About Highly Fragmented Markets". In: *Management Science* 65.10, pp. 4470–4488.

Menkveld, Albert J., Bart Zhou Yueshen, and Haoxiang Zhu (2017). "Shades of Darkness: A Pecking Order of Trading Venues". In: *Journal of Financial Economics* 124.3, pp. 503–534.

Parlour, Christine A (1998). "Price Dynamics in Limit Order Markets". In: *Review of Financial Studies* 11.4, pp. 789–816.

Pasquariello, Paolo and Clara Vega (2007). "Informed and Strategic Order Flow in the Bond Markets". In: *Review of Financial Studies* 20.6, pp. 1975–2019.

Seabold, Skipper and Josef Perktold (2010). "Statsmodels: Econometric and Statistical Modeling with Python". In: *Proceedings of the 9th Python in Science Conference*. Vol. 57. Austin, TX, p. 61.

Stoll, Hans R. and Robert E. Whaley (1990). "The Dynamics of Stock Index and Stock Index Futures Returns". In: *Journal of Financial and Quantitative Analysis* 25.4, pp. 441–468.

Tse, Yiuman, Paramita Bandyopadhyay, and Yang-Pin Shen (2006). "Intraday Price Discovery in the DJIA Index Markets". In: *Journal of Business Finance & Accounting* 33.9-10, pp. 1572–1585.

Yueshen, Bart Z and Jinyuan Zhang (2020). *Dynamic Trade Informativeness.* Working Paper.

A Exchange Members

The following list contains all unique exchange members listed on the Eurex website.

3Red Partners LLC
Aardvark Trading, L.L.C.
ABC arbitrage Asset Management
ABN AMRO Bank N.V.
ABN AMRO Clearing Bank N.V.
ADG Europe Ltd
ADG Market Making LLP
ADG Markets Ltd.
ADM Investor Services Inc.
ADM Investor Services International Ltd.
Advantage Futures LLC
AFS Equity & Derivatives B.V.
All Options International B.V.
Allston Capital LLC
Allston Trading LLC
Allston Trading UK Limited
AlphaGrep Pte Ltd
Altura Markets Sociedad de Valores SA
AMP Global Clearing LLC
AP Capital Investment Limited
ARB Trading Group North, LP
Atlantic Trading London Limited
Aurel BGC
Auriga Capital Limited
B. Metzler seel. Sohn & Co. Aktiengesellschaft
Baader Bank Aktiengesellschaft
Banca Akros Spa
Banca Profilo SPA
Banca Sella Holding S.p.A.
Banca Simetica S.p.A.
Banco Bilbao Vizcaya Argentaria S.A.
Banco Comercial Portugues S.A.
Banco Santander SA
Bank J. Safra Sarasin AG
Bank Julius Bär & Co. AG
Bank Vontobel AG
Bankhaus Lampe KG
Bankinter
Banque de Luxembourg
Banque Lombard Odier & Cie SA
Banque Pictet & Cie SA

Barak Capital G.T. LTD.
Barclays Bank Ireland Plc
Barclays Bank PLC
Barclays Capital Securities Ltd.
Basler Kantonalbank
Bayerische Landesbank
BCS Prime Brokerage Limited
Belfius Banque SA
Berner Kantonalbank AG
Bethmann Bank AG
BGC Brokers L.P.
Blue Fire Capital LLC
Bluefin Capital Management, LLC
BNP Paribas
BNP Paribas (Suisse) SA
BNP PARIBAS Arbitrage SNC
BNP Paribas Fortis SA/NV
BNP Paribas S.A. Niederlassung Deutschland
BNP Paribas Securities Services S.C.A. Zweigniederlassung Frankfurt
Boerboel Trading L.P.
BofA Securities Europe SA
BRED Banque Populaire
BSMA Limited
CACEIS Bank SA
Caixabank S.A.
Cantor Fitzgerald Europe
Capital Fund Management
Capital Futures Corp.
Capital MarketsTrading UK LLP
Capital Ventures International
Capitalead Pte. Ltd.
Cast Trading L.P.
Centercross B.V.
China Construction Bank Corporation Niederlassung Frankfurt
China Xin Yongan Futures Company Limited
Citadel Securities (Europe) Ltd.
Citadel Securities GCS (Ireland) Limited
Citigroup Global Markets Europe AG
Citigroup Global Markets Limited
CM Capital Markets Bolsa S.A. A.V.
CN FIRST INTERNATIONAL FUTURES LIMITED
Commerzbank AG
Concord Futures Corp.
Contech LP
Coöperatieve Rabobank U.A.
Corner Banca SA

Corretaje e Información Monetaria y de Divisas, Sociedad de Valores SA
Credit Agricole Corporate and Investment Bank
Crédit Industriel et Commercial
Credit Suisse (Schweiz) AG
Credit Suisse AG
Credit Suisse Bank (Europe) SA.
Credit Suisse International
Criterion Arbitrage & Trading BV
CSC Futures (HK) Limited
CTC London Limited
Cunningham Commodities LLC
D. E. Shaw Asymptote Portfolios LLC
Da Vinci Derivatives B.V.
Daiwa Capital Markets Europe Limited
Danske Bank A/S
De Riva Asia Limited
DekaBank Deutsche Girozentrale
Deutsche Bank AG
Directa SIM
Dom Group AG
Donner & Reuschel Aktiengesellschaft
Dorman Trading L.L.C.
DRW Europe B.V.
DRW Europe Derivatives B.V.
DRW Global Markets Ltd
DRW Investments (UK) Limited
DRW Investments LLC
DRW Singapore Pte Ltd
DV Trading LLC
DZ BANK AG Deutsche Zentral-Genossenschaftsbank
DZ Privatbank S.A.
E D & F Man Capital Markets MENA Limited
Eagle Labs (HK) Limited
Eagle Seven LLC
ED & F Man Capital Markets Ltd
EFG Bank AG
Epoch Capital Pty Ltd
Equita Societa Di Intermediazione Mobiliare SPA
Erste Group Bank AG
Exane Derivatives
Exane S.A.
FCT Europe Limited
Fenics GO Holdings Limited
Fermion Investments Limited
Financial Market Engineering Limited
FinecoBank Banca Fineco S.p.A

Finovesta GmbH
flatexDEGIRO Bank AG
Flow Traders Asia Pte Ltd
Flow Traders B.V.
Flow Traders U.S. LLC
Freeman Commodities Limited
Fubon Futures Co., Ltd
G. H. Financials Ltd.
Gallardo Securities Limited
Gelber Coöperatief U.A.
Gelber Group LLC
Geneva Ireland Financial Trading Ltd.
GFI Securities Ltd.
Global Execution Limited
Goldman Sachs Bank Europe SE
Goldman Sachs International
Grammont Finance SA
GTS Securities Europe Ltd
Hamburg Commercial Bank AG
Hamburger Sparkasse AG
Hard Eight Futures LLC
Hardcastle Trading AG
Hauck & Aufhäuser Privatbankiers AG
HC Technologies LLC
Headlands Technologies Europe B.V.
Headlands Technologies LLC
HGNH INTERNATIONAL FUTURES CO. LIMITED
HNK ALPHA PTE. LTD.
HPC S.A.
HRTEU Limited
HSBC Bank plc
HSBC Continental Europe
HSBC Trinkaus & Burkhardt AG
Hudson River Trading Europe Ltd.
IBKR Financial Services AG
IBROKER GLOBAL MARKETS, S.V., S.A.
IBVV Trading DMCC
ICAP CORPORATES LLC
IMC Trading B.V.
ING Bank N.V.
Ingensoma Arbitrage PTE LTD
Interkapital vrijednosni papiri d.o.o.
Intermonte SIM S.p.A.
Intesa Sanpaolo S.p.A.
Invest Banca SPA
J.P. Morgan AG

Jane Street Capital, LLC
JB DRAX HONORE (UK) LIMITED
Jefferies GmbH
Jefferies International Ltd.
Joh. Berenberg Gossler & Co. KG
Jump Trading Europe B.V.
Jump Trading Futures LLC
Jump Trading Pacific Pte Ltd
KBC Bank N.V.
Kemp Trading B.V. ta Nino Options
Kepler Chevreux (Suisse) SA
Kerdos Investment-AG TGV
KGI Futures Co. Ltd.
Korea Investment & Securities Co. Ltd.
Kreissparkasse Köln
Kutxabank S.A.
Kyte Broking Limited
Landesbank Baden-Württemberg
Landesbank Berlin AG
Landesbank Hessen-Thüringen Girozentrale
Lang & Schwarz AG
Lang & Schwarz TradeCenter AG & Co. KG
Leonteq Securities AG
Liquid Capital Australia Pty. Ltd.
Liquid Capital Markets Ltd.
LR Financial LLC
M.M. Warburg & CO (AG & Co.) Kommanditgesellschaft auf Aktien
Macquarie Bank Europe Designated Activity Company
Mako Derivatives Amsterdam B.V.
Mako Financial Markets Partnership LLP
Mako Global Derivatives Partnership LLP
Marex Financial
Marex North America LLC
Marex Spectron Europe Limited
Mariana UFP LLP
Market Securities (FRANCE) SA
Market Wizards BV
Maven Derivatives Amsterdam B.V
Maven Europe Limited
Mediobanca Banca di Credito Finanziario S.p.A
Melanion Volatility Fund
Mercury Derivatives Trading Limited
Merrill Lynch International
Method Investments & Advisory LTD
Mint Tower Capital Management B.V.
Mizuho Securities USA LLC

MMX Trading B.V.
Morgan Stanley & Co. International PLC
Morgan Stanley Europe SE
Mosaic Finance SAS
MUFG Securities (Europe) N.V.
MUFG Securities EMEA plc
National Bank of Greece SA
Natixis
Natwest Markets NV
Natwest Markets Plc
NH FUTURES CO. LTD.
Nomura Financial Products Europe GmbH
Nomura International plc.
Norddeutsche Landesbank - Girozentrale
Nordea Bank Abp
NRW.BANK
Nyenburgh Holding B.V.
ODDO BHF Aktiengesellschaft
ODDO BHF SCA
Old Mission Capital, LLC
Optiver Australia Pty Limited
Optiver V.O.F.
Panthera Investment GmbH
Phillip Capital Inc.
PNT Financial LLC
Prime Trading, LLC
Q1E LP
Quant.Capital Verwaltungs GmbH
QuantRes Fund SPC
Qube Research & Technologies Limited
Quintet Private Bank (Europe) S.A.
R.J. O'Brien Limited
R.J.O Brien France S.A.S.
Radix Trading Europe B.V.
Radix Trading LLC
Raiffeisen Bank International AG
Raiffeisen Centrobank AG
Raiffeisenlandesbank Oberösterreich Aktiengesellschaft
RBC Capital Markets (Europe) GmbH
RBC Europe Limited
RCUBE ASSET MANAGEMENT
RSJ Securities a.s.
Saccade Capital Limited
Scotiabank Europe Plc
Sea Cliff Investments Limited
Sequoia Capital LLP

SIB (Cyprus) Limited
Sigma Broking Limited
Skandinaviska Enskilda Banken AB
Société Générale
Sparkasse Pforzheim Calw
Squarepoint Master Fund Limited
SSW-Trading GmbH
St. Galler Kantonalbank AG
Star Beta Pty Ltd
StoneX Financial Europe S.A.
StoneX Financial Inc.
StoneX Financial Ltd
Sucden Financial Limited
Sunrise Futures LLC
Susquehanna International Securities Ltd.
Swedbank AB
Swissquote Bank S.A.
Tanius Technology LLC
Tensor Technologies AG
Teza Capital Management LLC
TFS Derivatives HK Ltd
TFS Derivatives Ltd.
Tibra Trading Europe Limited
TMG Trading FZE
Tower Research Capital Europe B.V.
Tower Research Capital Europe Limited
TP ICAP (Europe) SA
TP ICAP Markets Limited
Tradegate AG Wertpapierhandelsbank
TradeLink LLC
TradeLink Worldwide Limited
TradeWeb Europe Ltd
Tradition Securities and Derivatives Inc
Tradition Securities and Futures S.A.
Transtrend B.V.
TTG Capital Limited
Tullett Prebon (Securities) Limited
Tullett Prebon Financial Services LLC
Tyler Capital Ltd.
UBS AG
UBS Europe SE
UniCredit Bank AG
UniCredit S.p.A.
Vallum Trading LLC
Vantage Capital Markets HK Limited
Vantage Capital Markets LLP

Vatic Fund I LLC
Vectalis
Vector Trading LLC
Vegasoul Opus Fund SPC High Street Segregated Portfolio
Virtu Financial Ireland Limited
Virtu Financial Singapore Pte. Ltd.
Volatility Performance Fund SA
VOLKSBANK WIEN AG
Vortex Street Fund Limited
VTB Capital plc
WEBB Traders B.V.
Wedbush Securities Inc.
Wells Fargo Securities International Limited
Wells Fargo Securities, LLC
WH Trading LLC
Whitney Capital Series Fund LLC
Wolfgang Steubing AG Wertpapierdienstleister
Xconnect Market Maker LLP
XConnect Trading Limited
XR Trading EU B.V.
XR Trading LLC
XTX Markets Limited
XTX Markets SAS
Yuanta Futures Co. Ltd.
Zürcher Kantonalbank

B Additional Descriptive Statistics

Figure 5: Evolution of prices in the sample period

This figure depicts the evolution of the price series over our sample period from January 4, 2010 to December 7, 2018.

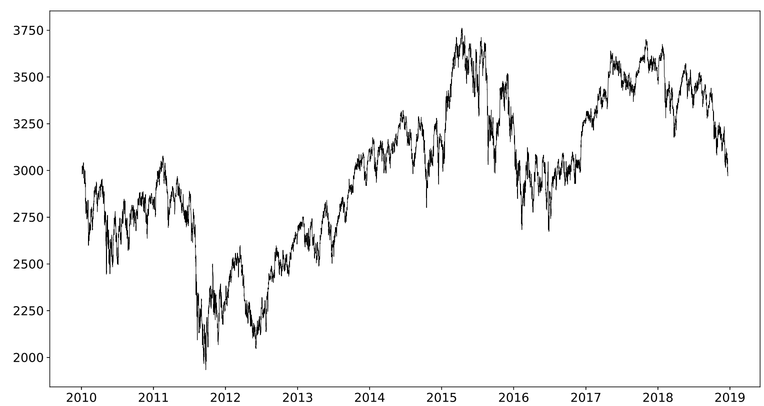


Figure 6: Intraday volume pattern

The figure presents the intraday pattern of relative trading volume. Volume in each minute is divided by the total volume of the respective trading day. The solid line depicts overall volume, the dashed line agent volume and the dotted line principal volume. For agent and principal volume, relative volume is computed based on total daily volume for the respective account role. The time indicated for the US open assumes synchronization of daylight saving time. In fact, there are every year two to three weeks in which the US open is shifted by one hour.

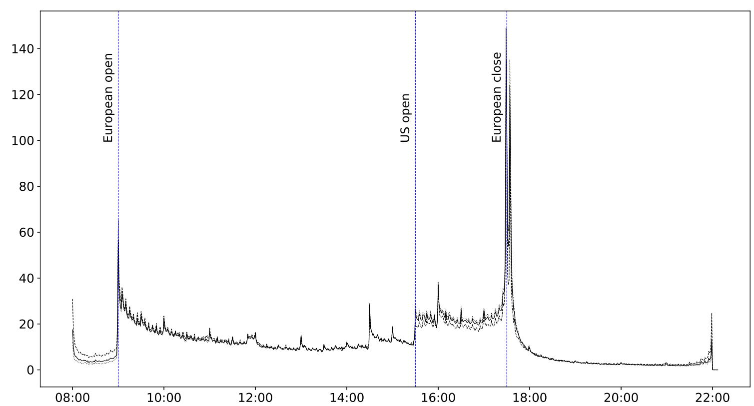


Table 8: Autocorrelations hourly returns and order flow

This table presents autocorrelations for hourly returns (based on the last trading price in that hour) and hourly order flow variables. Order flow variables are computed based on signed order flow for the account role-aggressor flag combination and accumulated for each hour. Order flow variables in EUR. P-values are reported in parentheses.

Lag	1	2	3	4	5	6	7	8	9	10
Returns	−0.0298	0.0052	−0.0182	0.0069	−0.0063	−0.0098	0.0029	−0.0019	0.0216	0.0143
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Agent flow	0.2184	0.1136	0.0650	0.0592	0.0408	0.0433	0.0300	0.0029	0.0099	0.0156
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Principal flow	0.2184	0.1136	0.0650	0.0592	0.0408	0.0433	0.0300	0.0029	0.0099	0.0156
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Aggressive agent flow	0.1364	0.0594	0.0282	0.0297	0.0296	0.0276	0.0216	0.0177	0.0164	0.0199
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Passive agent flow	0.1163	0.0440	0.0218	0.0158	0.0089	0.0085	−0.0025	−0.0015	0.0107	0.0045
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Aggressive principal flow	0.0184	0.0045	−0.0079	−0.0111	0.0007	0.0091	−0.0107	−0.0111	0.0075	0.0014
	(0.0009)	(0.0029)	(0.0033)	(0.0014)	(0.0033)	(0.0024)	(0.0011)	(0.0005)	(0.0004)	(0.0008)
Passive principal flow	0.0785	0.0366	0.0189	0.0043	0.0218	0.0276	0.0185	0.0062	0.0130	0.0098
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Aggressive flow	0.0371	0.0103	−0.0002	−0.0073	0.0111	0.0133	0.0024	0.0058	0.0124	0.0085
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Passive flow	0.0371	0.0103	−0.0002	−0.0073	0.0111	0.0133	0.0024	0.0058	0.0124	0.0085
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Table 9: Autocorrelations in returns

The table presents autocorrelations in returns at different frequencies up to 10 lags. Returns are computed based on the last price in the respective time interval. P-values are reported in parentheses.

Lag	1 min	5 min	15 min	30 min	1 hour	2 hours	1 day	2 days
1	−0.1674	−0.0748	−0.0466	−0.0140	−0.0084	0.0032	−0.0535	−0.0454
	(0.0000)	(0.0000)	(0.0000)	(0.0005)	(0.1442)	(0.7081)	(0.0226)	(0.1248)
2	−0.0075	−0.0173	0.0020	−0.0076	0.0176	0.0005	−0.0143	−0.0368
	(0.0000)	(0.0000)	(0.0000)	(0.0004)	(0.0034)	(0.9307)	(0.0617)	(0.1419)
3	−0.0013	0.0014	0.0029	0.0131	−0.0076	0.0293	0.0066	−0.0009
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0044)	(0.0071)	(0.1299)	(0.2718)
4	−0.0039	0.0002	−0.0029	0.0087	0.0049	0.0082	−0.0592	−0.0446
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0078)	(0.0112)	(0.0171)	(0.1857)
5	0.0012	0.0016	−0.0075	−0.0035	−0.0031	0.0047	0.0081	−0.0178
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0148)	(0.0205)	(0.0328)	(0.2565)
6	−0.0038	0.0025	0.0121	−0.0078	0.0091	−0.0149	0.0367	0.0150
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0108)	(0.0118)	(0.0236)	(0.3392)
7	−0.0044	−0.0025	0.0043	−0.0007	0.0113	0.0006	−0.0432	0.0070
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0047)	(0.0217)	(0.0120)	(0.4434)
8	−0.0017	0.0055	0.0022	0.0067	0.0059	−0.0123	−0.0044	−0.0063
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0060)	(0.0178)	(0.0210)	(0.5466)
9	−0.0021	−0.0038	0.0028	0.0002	0.0145	−0.0049	0.0230	0.0362
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0011)	(0.0266)	(0.0252)	(0.4932)
10	0.0017	−0.0026	−0.0026	0.0017	−0.0102	0.0149	−0.0427	−0.0060
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0006)	(0.0155)	(0.0135)	(0.5846)

C Data Cleaning

In this section we describe details on the data cleaning procedure and discuss our approach in comparison to other approaches.

Between May 6, 2013 and May 13, 2013, Eurex migrated its products to its T7 trading architecture. Euro STOXX 50 futures were migrated on May 8, 2013. From this day onward, there is imprecision (or "noise") in the timestamps. In particular, orders that appear to be executed against each other are not necessarily recorded at the same timestamp. Rather, orders are recorded at consecutive timestamps, with the difference usually being within a few tens of milliseconds. As a result, for each timestamp, buying and selling volume as well as aggressive and non-aggressive volume do not necessarily satisfy market clearing. As our analysis focuses on continuous trading only (Section 3) and we are inferring continuous trading periods from the data, we have to deal with the imprecision in the timestamps in an efficient manner.

We clean the data using an event-time approach that is akin to the methodology of Aquilina, Budish, and O'Neill (2021), Ernst (2020), and Ernst, Sokobin, and Spatt (2021). A natural way to solve the problems arising from the imprecision in the timestamps is grouping trade records with the same execution price that are recorded closely together. This can be addressed in both clock time and event time. A clock time approach, however, comes with the caveat that trades that are executing against each other and are recorded in different intervals, for example seconds, are not grouped together and thus neither of the trades enter the main analysis.[30] The problem can be addressed by allowing volumes of market and limit order to deviate up to a threshold within each time interval. This threshold, however, is arbitrary and hard to infer from the data and the classification remains noisy.

An event time approach offers a tractable and more precise solution. Therefore, trades that are recorded at the same price within a short period of time a grouped together. Then, the total volume of limit orders and the total volume of market orders over these trades are computed. If the total volume of limit orders equals the total volume of market orders, the trades are labeled "continuous trading" and included in the main analysis. The algorithm for classifying trades runs from the start of each trading day. For each trade s_0 that is executed at price p_i, a time window starting with that trade record is initialized. All following trades $s_1, s_2, \ldots$ that are executed within the time window at the same trading price p_i are grouped together and assigned the same time. Once a trade s_{break} is executed at the same trading price p_i but does not fall within the time window, a new time window starting from that trade record is defined. Again, all trades that occur within the time window at the same execution price are grouped together with this trade. This procedure continues until the end of the trading day.

[30] Suppose, for example trades are grouped by seconds. The first trade is recorded at $t.900000$ with t denoting seconds, and the next is recorded at $t + 1.100000$. In an clock time approach, market clearing would not be satisfied for either of the seconds.

The only parameter that has to be chosen is the length of the time window. In general, choosing the window length trades off two factors. On the one hand, choosing a longer window length assures that all corresponding trades are grouped together even if there is substantial noise in the timestamps and high trading activity, that might further delay recording of some of the trades. On the other hand, by choosing a shorter window only trades that were actually executing against each other are captured. If the window length is chosen too long and the volume of market orders and limit orders does not equal, a substantial volume does not enter the main analysis. We consider the possibilities of 100ms, 500ms, 2s, and 4s. In our particular dataset, inspection reveals that the imprecision is usually within the magnitude of a few tens of microseconds. Thus, in comparison to the clock-time approach discussed before, our approach allows to determine the appropriate choice to the parameter based on the data.

Another alternative to clean the data is by removing opening and closing auctions. This does however not account for potential intraday auctions that are held and thus have to be identified differently. Also, the identification of auctions is not clear cut due to the "noise" in the timestamps.[31]

D State Space Representation

In this Section we show how the state space model presented in Section 4 can be mapped into the standard linear state space model form. Our state space model is given by

$$p_t = m_t + s_t \tag{9}$$

$$m_t = m_{t-1} + \sum_{s \in \mathcal{S}} \gamma_s \tilde{x}_{s,t} + w_t \tag{10}$$

$$s_t = \phi s_{t-1} + \sum_{s \in \mathcal{S}} \delta_s x_{s,t} + \varepsilon_t \tag{11}$$

[31]There are several ways to do so. First, opening auctions can be defined as the first trade record on each trading day and closing auctions as the last trade record on each trading day. This is, however, only feasible for the period before May 8, 2013, as after the first recorded trade is not necessarily the opening auction and opening auctions were occasionally recorded at several timestamps. In principle, the aggressor flag for auction trades indicates that these trades pertain to a passive order. This suggests identifying auctions as timestamps for with only passive trades are recorded. In practice, there are timestamps around the (potential) opening or closing auctions for which trades with an aggressor flag indicating an aggressive order as well as trades with an aggressor flag indicating a passive order are recorded. Even if market clearing holds taking all orders into account, the volume of aggressive orders does not equal the volume of passive orders for these timestamps (this is, for example, the case on March 11, 2011). From the trade data, it is not possible to identify against which limit orders this market order has executed. Since the account flag is of first-order importance for the main analysis, such an identification would however be necessary if the market order were to be included.

with $w_t \sim \mathcal{N}(0, \sigma_w^2)$ and $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. The standard linear state space model is given by

$$\mathbf{y}_s = \mathbf{Z}_s \boldsymbol{\alpha}_s + \boldsymbol{\epsilon}_s, \tag{12}$$

$$\boldsymbol{\alpha}_{s+1} = \mathbf{T}_s \boldsymbol{\alpha}_s + \mathbf{R}_s \boldsymbol{\eta}_s, \tag{13}$$

with index $s = 1, \ldots, S$, and the disturbances $\boldsymbol{\epsilon}_s \sim \mathcal{N}(0, \mathbf{H}_s)$ and $\boldsymbol{\eta}_s \sim \mathcal{N}(0, \mathbf{Q}_s)$, following the notation of Durbin and Koopman (2012).

We follow Hamilton (1986) and include exogenous variables in the state vector. We collect the variables γ_s for $s \in \mathcal{S}$ in the $S \times 1$ vector $\boldsymbol{\gamma}$ and the variables δ_s for $s \in \mathcal{S}$ in the $S \times 1$ vector $\boldsymbol{\delta}$. Similarly, we collect order flow in the $S \times 1$ vector $\mathbf{x}_t$ and innovations in order flow in the $S \times 1$ vector $\tilde{\mathbf{x}}_t$. Note that the dimension S of the vectors depends on which order flow variables are included, as discussed in Secton 4.2. Then we obtain

$$\mathbf{y}_s = p_{t-1}, \tag{14}$$

$$\boldsymbol{\alpha}_s = (m_{t-1}, s_{t-1}, \boldsymbol{\gamma}'_{t-1}, \boldsymbol{\delta}'_{t-1})', \tag{15}$$

$$\boldsymbol{\eta}_s = (w_t, \varepsilon_t)', \tag{16}$$

It is set $\mathbf{H}_s \to \mathbf{0}$ and thus $\boldsymbol{\epsilon}_s = 0$. Then the design matrix[32] is given by

$$\mathbf{Z}_s = \begin{array}{c} \begin{array}{cccc} 1 & 1 & S & S \end{array} \\ \left[\begin{array}{cccc} 1 & 1 & 0 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \end{array} \tag{17}$$

Furthermore, the transition matrix

$$\mathbf{T}_s = \begin{array}{c} \begin{array}{cccc} 1 & 1 & S & S \end{array} \\ \left[\begin{array}{cccc} 1 & 0 & \tilde{\mathbf{x}}'_t & 0 \\ 0 & \phi & 0 & \mathbf{x}'_t \\ 0 & 0 & \mathbf{I}_S & 0 \\ 0 & 0 & 0 & \mathbf{I}_S \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 1 \\ S \\ S \end{array} \tag{18}$$

is time-varying, with $\mathbf{I}_S$ being and $S \times S$ identity matrix. The selection matrix is given by

$$\mathbf{R}_s = \begin{array}{c} \begin{array}{cc} 1 & 1 \end{array} \\ \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 1 \\ S \\ S \end{array}, \tag{19}$$

and the state covariance matrix by

$$\mathbf{Q}_s = \begin{array}{c} \begin{array}{cc} 1 & 1 \end{array} \\ \left[\begin{array}{cc} \sigma_w^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{array}\right] \end{array} \begin{array}{c} \\ 1 \\ 1 \end{array}. \tag{20}$$

[32]For all matrices we denote the dimensions in the first row and last column

In the state space model given by (14) – (20), the parameters on order flow and innovations in order flow are estimated by state estimation. This model is used as second step when determining the starting values, as described in Section 4. For implementation, state estimation is replaced by parameter estimation such that the coefficients on the order flow variables are estimated by maximum likelihood estimation. Therefore, a constant 1 is assigned to exogenous variables. Both the exogenous variables as well as the parameters are included in the system matrices. This yields

$$\mathbf{y}_s = p_{t-1}, \tag{21}$$

$$\boldsymbol{\alpha}_s = (m_{t-1}, s_{t-1}, \boldsymbol{\iota}_S')', \tag{22}$$

$$\boldsymbol{\eta}_s = (w_t, \varepsilon_t)', \tag{23}$$

with $\boldsymbol{\iota}_S$ being an $S \times 1$ vector of ones. Again, it is set $\mathbf{H}_s \to \mathbf{0}$ and thus $\boldsymbol{\epsilon}_s = 0$. The design matrix is unchanged and given by (17). Furthermore, the transition matrix changes and is now given by

$$\mathbf{T}_s = \begin{array}{c} \begin{array}{ccc} 1 & 1 & S \end{array} \\ \left[\begin{array}{ccc} 1 & 0 & \boldsymbol{\gamma}' diag(\tilde{\mathbf{x}}_t) \\ 0 & \phi & \boldsymbol{\delta}' diag(\mathbf{x}_t) \\ 0 & 0 & \mathbf{I}_S \end{array}\right] \begin{array}{c} 1 \\ 1 \\ S \end{array} \end{array} . \tag{24}$$

is time-varying. The selection matrix is given by

$$\mathbf{R}_s = \begin{array}{c} \begin{array}{cc} 1 & 1 \end{array} \\ \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{array}\right] \begin{array}{c} 1 \\ 1 \\ S \end{array} \end{array} , \tag{25}$$

and the state covariance matrix is unchanged and given by (20).

E Implementation of Estimation

This section describes details on the implementation of the model estimation described in Section 4. The model is estimated by maximum likelihood estimation and the Kalman filter recursion is used to valuate the likelihood function. For the maximum likelihood estimation, starting values are required. Here, we describe how these starting values are obtained. Also, we discuss restrictions on the parameters in estimation.

Starting values for the maximum likelihood estimation are obtained in three steps. First, a simple state space model excluding order flow and innovations in order flow is estimated. Thus, the model is given by

$$p_t = m_t + s_t \tag{26}$$

$$m_t = m_{t-1} + w_t \tag{27}$$

$$s_t = \phi s_{t-1} + \varepsilon_t \tag{28}$$

with $w_t \sim \mathcal{N}(0, \sigma_w^2)$ and $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. Since the model is estimated with log prices, return variances and autocovariances can be expressed as a function of the model's parameters. It can be shown that the autocovariances of the returns are

$$\gamma(0) = \sigma_w^2 + \frac{1 - 2\phi}{1 - \phi^2}\sigma_\varepsilon^2 \tag{29}$$

$$\gamma(1) = \frac{\phi - 1}{1 + \phi}\sigma_\varepsilon^2 \tag{30}$$

$$\gamma(2) = \frac{\phi(1 - \phi)}{1 + \phi}\sigma_\varepsilon^2. \tag{31}$$

Using this, starting values for the maximum likelihood estimation are given by

$$\phi = \frac{\gamma(2)}{\gamma(1)} \tag{32}$$

and using the starting value for ϕ it follows for σ_ε^2

$$\sigma_\varepsilon^2 = \gamma(1)\frac{1 + \phi}{\phi - 1} \tag{33}$$

and finally for σ_w^2

$$\sigma_w^2 = \gamma(0) - \frac{1 - 2\phi}{1 - \phi^2}\sigma_\varepsilon^2. \tag{34}$$

Using these starting values the reduced form state space model given by (26) – (28) is estimated. We estimate parameters for σ_ε and σ_w rather than for σ_ε^2 and σ_w^2. Thus, our starting values are given by the square root of (33) and (34), respectively. The estimates are stored and used as starting values for estimating the full state space model, with parameter estimation replaced by state estimation. The model is discussed in Appendix D, (14) – (20). This model introduces the parameters on order flow and innovations in order flow as latent state variables given by

$$\delta_t = \delta_{t-1} \tag{35}$$

and

$$\gamma_t = \gamma_{t-1}. \tag{36}$$

For estimation, the states for the parameters are initialized as diffuse by setting a prior variance of κ with $\kappa \to \infty$.

After estimation, the estimated parameters for σ_ε, σ_w, and ϕ as well as the state estimates for δ and γ are stored and used as starting values for estimating the full state space model (9) – (11) by maximum likelihood.

The model contains parameters that are required to be positive (the standard deviations of the error terms in the state and observation equation) or to be in the interval $[-1, 1]$ (the autocorrelation in pricing errors). To ensure that these restrictions are satisfied in estimation, we transform the restricted parameters before optimization by applying the function $f(x)$ to parameter x. After optimization, we untransform the parameters by applying the function $g(y)$ to the transformed parameter y. For the variance parameters, we use

$$f(x) = x^2 \tag{37}$$

and

$$g(y) = y^{\frac{1}{2}}. \tag{38}$$

For the autocorrelations in pricing errors that are required to be in the interval $[-1, 1]$, the functions are given by

$$f(x) = tanh(x) \tag{39}$$

and

$$g(y) = arctanh(y). \tag{40}$$

F Additional Robustness Checks

We perform additional robustness checks for our results obtained based on the full state space model presented in Section 4. With this, we address the potential concern that pricing errors in a period are not only a function of the order flow in the respective period, but also include a component that is due to under-reactions or over-reactions to past news. To capture this effect, we allow the pricing error s_t in period t to depend on innovations in order flow $\tilde{x}_{t-1}$ in period $t-1$:

$$p_t = m_t + s_t \tag{41}$$

$$m_t = m_{t-1} + \sum_{s \in \mathcal{S}} \gamma_s \tilde{x}_{s,t} + w_t \tag{42}$$

$$s_t = \phi s_{t-1} + \sum_{s \in \mathcal{S}} \delta_s x_{s,t} + \varepsilon_t \tag{43}$$

As before, $w_t \sim \mathcal{N}(0, \sigma_w^2)$, $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, $x_{s,t}$ denotes order flow by account role-aggressor flag combination s, and $\tilde{x}_{s,t}$ are surprises in order flow. $\mathcal{S}$ denotes the account role-aggressor flag combinations included in the estimation. The coefficient $\delta_{s,lag}$ in equation (43) captures the degree to which pricing errors under-react or over-react to past innovations in the respective order flow series.

Results are presented in Table 10. Overall, the evidence for both under-reactions and over-reactions in weak and at most suggestive for innovations in agent flow.

Table 10: State space model estimation results

This table presents estimation results for the full state space model

$$\begin{aligned} p_t &= m_t + s_t \\ m_t &= m_{t-1} + \sum_{s\in\mathcal{S}} \gamma_s \tilde{x}_{s,t} + w_t \\ s_t &= \phi s_{t-1} + \sum_{s\in\mathcal{S}} \delta_s x_{s,t} + \sum_{s\in\mathcal{S}} \delta_{s,lag} \tilde{x}_{s,t-1} + \varepsilon_t \end{aligned}$$

at an hourly frequency. x_t is order flow and $\tilde{x}_t$ are surprises in order flow obtained as the residual from a VAR model. Passive order flow is omitted from the specification. The subscripts on γ and δ denote the account role with c (p) standing for agent (principal) and the aggressor flag with a (p) standing for aggressive (passive) order flow. Variances are in *bp* and δ as well as γ in $bp/1,000,000$ EUR. Robust standard errors in parentheses. * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level.

	(1)	(2)
σ_w	17.6117***	19.6547***
	(0.2642)	(0.2492)
σ_ϵ	19.0850***	20.4210***
	(0.4433)	(0.3832)
ϕ	0.8926***	0.8886***
	(0.0146)	(0.0128)
efficient price		
$\gamma_{c,a}$	0.7928***	
	(0.0460)	
$\gamma_{p,a}$	0.0204	
	(0.0338)	
$\gamma_{c,n}$		−0.1383*
		(0.0773)
$\gamma_{p,n}$		−0.4864***
		(0.0436)
pricing error		
$\delta_{c,a}$	0.1550***	
	(0.0463)	
$\delta_{p,a}$	0.0356	
	(0.0287)	
$\delta_{c,a,lag}$	−0.0131	
	(0.0127)	
$\delta_{p,a,lag}$	0.0013	
	(0.0095)	
$\delta_{c,n}$		−0.2320***
		(0.0642)
$\delta_{p,n}$		0.0261
		(0.0349)
$\delta_{c,n,lag}$		0.0607***
		(0.0195)
$\delta_{p,n,lag}$		−0.0169*
		(0.0099)
#Obs	32,276	32,272

Table 11: Summary statistics on trading during continuous trading and auctions

This table presents summary statistics for trading activity during continuous trading and trading during the opening and closing auctions. Trading volume of agents during continuous trading is computed as the average of hourly shares of agent volume in total volume. Agent share ($q = 0.9$) and agent share ($q = 0.99$) denote the share of agent volume in total auction volume for auctions with volume in the 90% and 99% quantile of the distribution of auction volumes. Agent share around auction denotes the share of agent volume in the 15 minutes of continuous trading around the opening and closing auctions. Auction share denotes the share of auctions in continuous trading. Price deviation denotes the absolute deviation of the auction price from the first (last) recorded price in continuous trading for opening (closing) auctions and price deviation 15 min denotes the deviation of the auction price from the price 15 minutes later (earlier) for opening (closing) auctions.

	Cont' trading	Open	Close
Agent share	0.3159	0.7023	0.4775
Agent share ($q = 0.9$)		0.5506	0.4965
Agent share ($q = 0.99$)		0.5426	0.4654
Agent share around auction		0.4926	0.4926
Auction volume		2832.7121	419.7658
Auction share (bp)		15.07	2.3925
Price deviation (bp)		2.8540	4.5542
Price deviation 15 min (bp)		14.4548	9.7480

G Additional Results on Auction Trading

Table 12: Correlations between auction variables

This table presents correlations between the different variables of interest for the opening and closing auction. Agent share denotes the share of agent volume in total auction volume. Agent share around auction denotes the share of agent volume in the 15 minutes of continuous trading around the opening and closing auctions. Price deviation denotes the absolute deviation of the auction price from the first (last) recorded price in continuous trading for opening (closing) auctions.

	price deviation	agent share around auction	agent share auction	vix
		Opening auction		
price deviation	1.000000	0.052215	0.018860	0.223308
agent share around auction	0.052215	1.000000	0.317057	0.268108
agent share auction	0.018860	0.317057	1.000000	0.090456
vix	0.223308	0.268108	0.090456	1.000000
		Closing auction		
price deviation	1.000000	0.085660	0.154351	0.349090
agent share around auction	0.085660	1.000000	0.345702	0.229460
agent share auction	0.154351	0.345702	1.000000	0.293631
vix	0.349090	0.229460	0.293631	1.000000

Table 13: Auction deviations: regression results

This table presents regression results for auction price deviations relative to the first price after (last price before) opening (closing) auctions as well as the price 15 minutes after (15 minutes before) opening (closing) auctions. agent share denotes the share of agent volume in total auction volume. Agent share around auction denotes the share of agent volume in the 15 minutes of continuous trading around the opening and closing auctions. Auction share denotes the share of auction volume in continuous trading volume on the respective trading day. Clustered standard errors using 20 lags in parentheses. * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level.

	open	open (15 min)	close	close (15 min)
constant	-0.1200	-7.8140***	-0.9508*	-6.0120***
	(0.579)	(1.893)	(0.493)	(1.272)
Agent share auction	0.317843	7.1313***	1.3599***	0.5643
	(0.765)	(1.947)	(0.519)	(0.975)
Auction share	0.010983***	0.0210*	0.1251	0.0819*
	(0.002)	(0.011)	(0.093)	(0.045)
Agent share around auction	-0.785586	3.1205	-0.4377	2.4498
	(1.617)	(3.581)	(0.951)	(1.896)
VIX	0.174428***	0.9041***	0.2793***	0.8307***
	(0.039)	(0.090)	(0.026)	(0.089)
Obs	2289	2289	2289	2289
Adj R^2	0.058	0.154	0.136	0.224

Table 14: State space model for agent flow including auctions

This table presents estimation results for the full state space model for agent flow including opening and closing auctions at an hourly frequency. Principal order flow is omitted. The subscripts on γ and δ denote the account role with c standing for agent. Standard deviations are in bp and δ as well as γ in $bp/1,000,000$ EUR. Robust standard errors in parentheses. * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level.

	open	close
σ_w	18.5076***	19.1561***
	(0.1617)	(0.1282)
σ_ϵ	21.3162***	19.4433***
	(0.3099)	(0.3198)
ϕ	0.8396***	0.7923***
	(0.0127)	(0.0181)
efficient price		
γ_c	0.7039***	0.7235***
	(0.0343)	(0.0318)
$\gamma_{c,auction}$	−0.6341	0.1338
	(0.8398)	(0.6404)
pricing error		
δ_c	−0.0325	−0.0415
	(0.0315)	(0.0294)
$\delta_{c,auction}$	−0.4653	−1.3338
	(0.6246)	(1.3147)
#Obs	34,512	34,512