The Information Content of Retail Investors' Order Flow

Ingmar Nolte[*]   Sandra Nolte[†]

Lancaster University   Lancaster University

This Version: November 7, 2014

The Information Content of Retail Investors' Order Flow

Abstract

In this paper we provide evidence that the trading activity of small retail investors carries significant genuine information that can be exploited for the short term out-of-sample forecasting of foreign exchange rates. Our findings are based on a unique dataset of around 2000 retail investors from the OANDA FXTrade electronic trading platform.

Our results are consistent with the view that in the foreign exchange market private information is highly dispersed, but can be extracted by observing customer order flow. Previous studies, however, focused on the information content of costumer order flow of dealers in the interbank market, whose clients are themselves large institutional and professional investors. Our study is the first that analyses a crowd of small retail investors and shows that even the trading activity of these investors contains, on aggregate, important non-public information, that can be exploited for short term exchange rate forecasting. Our findings lead us to conjecture that retail investors (on aggregate) are not pure noise traders but process dispersed information at least partially in a similar way as large institutional investors and hence place their orders accordingly.

1 Introduction

In market microstructure theory, key ingredients for a large number of models are two types of investors: informed investors that are endowed with some sort of private information and uninformed noise traders. In practice, informed investors are usually associated with professional and big institutional investors, while retail investors and smaller institutional investors are associated with noise traders, who possess no (important) private information and trade at best merely for liquidity reasons or with price contingent strategies (e.g. momentum or contrarian). In this paper, we provide evidence that *on aggregate* retail investors possess additional information that is otherwise not publicly available and can be exploited to significantly improve short term out-of-sample price change forecasts. More specifically, additional information refers to information beyond the information already incorporated in past prices and interbank order flow.

Most of the research on order flow and currencies has concentrated on the interbank market (e.g., Evans & Lyons (2002a,b, 2005, 2006), Payne (2003), Bjønnes & Rime (2005), Berger et al. (2008), Daníelsson, Luo & Payne (2012) among many others) and analysed the information content of inter-dealer order flow. Here two main theories are put forward why order flow contains information about future price changes. One theoretical foundation is given by the portfolio allocation model of Evans & Lyons (2002a,b), in which exchange rate movements are explained by investors optimally re-balancing an underlying portfolio, because of shifts in risk preferences for different markets. Their model considers two different markets: the customer-dealer market and the dealer-dealer (interbank) market. Dealers trading in the interbank market learn about order flow in the customer-dealer market and this customer order flow helps in predicting currency price changes and order flow in the interbank market. Another theoretical foundation can be based on the argument of Sarno & Taylor (2001), who consider order flow as an aggregation device of macroeconomic fundamental information, which is highly dispersed among market participants. Both models require that market participants interpret information, either about portfolio changes or about changes of macroeconomic fundamentals, in the correct way, that they adjust their expectations for future prices in light of this information, and therefore place their orders accordingly.

In practice, the way how information is incorporated into the price processes is, how-

ever, much more complex as traders can resort to different trading strategies, order types and can use their information to correct for perceived mispricing at different time horizons. Another important factor that needs to be kept in mind is that when we use order flow to try to proxy for information, different aggregation frequencies will potentially capture different types of information.

There seems to be a consensus that in the FX market customer order flow (trading between a dealer bank and non-dealer clients) is one of the most fruitful source of private information for a dealer bank. The papers of Osler (2005), Osler & Vandrovych (2009) and Marsh & Miao (2012) analyse such customer-dealer order flow and find evidence for the predictive power of such type of order flow. Osler & Vandrovych (2009) analyse the customer order flow of the Royal Bank of Scotland and break it down into different types of end-user order flows. They find that only trades of leveraged investors, who are assumed to actively acquire and interpret information, carry significant information for forecasting horizons of 30 minutes up to 1 week.

These client order flow streams in the analyses above differ, however, significantly from the retail customer order flow that we analyse in this paper. The studies above consider players in the dealer market, whose clients are themselves large institutional customers such as hedge funds, governments, broker-dealers and commercial banks, who often trade volumes which are several times higher than the standardised order sizes in the interbank market. Our paper is, therefore, not the first that analyses the information content of order flow in the foreign exchange market, but it is the first that analyses *retail* investors' order flow in the FX market on the basis of a unique dataset from an electronic currency trading platform: OANDA FXTrade. Our dataset contains the whole record of all trading actions of around 2000 traders on a second-by-second basis. Most of these traders are small retail investors and companies without access to private information such as observing own customer order flow for example. A central element of our analysis is to capture the information flow process on a high-frequency level for these investors; or put slightly differently to measure their expectations revealed by their trading actions. Those expectations may well be driven by specific news events, risk preferences, hedging demands and investors' particular day-to-day business. The key question asked in this paper is: Does the trading activity of these retail investors (on aggregate) provide new information to the market?

The characteristics of our dataset enable us to fill two existing gaps in the literature. First, in contrast to Marsh & Miao (2012), who for example not even include order volumes below 500,000 EUR in their database, we focus explicitly on the information content of order flow of a large set of small retail investors (non-financial customers) operating with comparatively tiny trading volumes (order flow with median transaction volumes of 2,000 EUR) for the prediction of future price changes. Such retail investors are typically placed among the lowest in the pecking order of liquidity traders and considered as pure noise traders. To our knowledge there has been no analysis of such customer data obtained from an electronic trading platform, as of yet. Second, the structure of the OANDA FXTrade dataset enables us not only to investigate the information flow process and information content of each and every order category on a high frequency level, but also whether orders were submitted to open or close a position. We use this information to look at the information content of transaction category specific order flows on the micro-level. This further enriches the existing literature, which considers so called combined order flow measures (e.g. Lyons (1995)) that can be constructed from datasets with less information, but consequently aggregate potentially valuable information out.

We start our analysis by taking the specific nature of the OANDA FXTrade platform, which is partially organised as a crossing network and the detailed information on individual order type order flows into account, to assess the forecasting performance of the category specific order flows. In an in-sample and out-of-sample forecasting study we then analyse the forecasting power of these order flow measures for the prediction of future currency price changes over horizons ranging from 1 minute up to 1 hour. More precisely, we investigate whether forecasting models for future price changes relying on category specific order flow information outperform models relying on historical price changes and interbank order flow information. We provide evidence that *on aggregate* retail investors trading flow can be exploited to significantly improve *short* term out-of-sample price change forecasts, and that both market order and limit order order flows predict future price changes at high frequencies up to 5 minutes.

In a next step we then construct a combined order flow measure that aims to capture the genuine information of the crowd of OANAD FXTrade investors and compare it to the forecasting performance of the standard order flow measure that has traditionally been used in the foreign exchange (e.g., Lyons (1995)) and equity (e.g.,

Chordia, Roll & Subrahmanyam (2002)) market literature. The results of this analysis confirm that on aggregate retail investors trading conveys new information about future prices and that the use of the standard order flow measure mixes up information processes of OANDA FXTrade customers and interbank market dealers, which in our study leads to an inferior forecasting performance.

We also investigate whether the traders on OANDA FXTrade form consistently correct beliefs about future price movements or whether they are consistently wrong. To carry out this analysis, we abstract from more complex information processing mechanisms and rely on a strong simplifying assumption that a positive (negative) correlation between order flow and future price changes indicates correct (wrong) beliefs. We find that there is no clear cut answer to the question and that it highly depends on the type of order used and the forecasting horizon that we consider. There is, however, a clear tendency that for short forecasting horizons their beliefs seem to be correct, while for longer forecasting horizons they seem to be wrong.

Our analysis is related to the papers of Nolte & Voev (2011), Nolte (2012) and Nolte & Nolte (2012) who analyse the same trading activity dataset from OANDA FXTrade, but focus on how the individual investors make trading decisions and are influenced by behavioral effects such as monitoring, endowment effects, over-confidence and disposition effects.

This paper is organised as follows: Section 2 gives an overview of the literature about small retail investors primarily in equity markets. In Section 3 we explain our dataset and the trading mechanism on the OANDA FXTrade platform. Section 4 presents the in- and out-of-sample analyses of the category specific order flow measures. While Section 5 provides a detailed discussion about alternative order flow measures, Section 7 presents some robustness checks. Finally Section 8 concludes.

2   Retail Investors

Although this paper is, to the best of our knowledge, the first paper to examine the prediction power of a collective of small retail investors in the foreign exchange market, it contributes to the ongoing debate prevailing in other markets on whether retail trading conveys information about future prices. In the finance literature, and more specifically in the market microstructure and behavioral finance literature, small retail investors usually have the image of unsophisticated noise traders suffering from psychological biases and having almost no (relevant) private information to trade on (Odean (1998), Barber & Odean (2000)).

This view is deeply rooted in traditional finance reasoning and present in the literature since decades. Recently, it has again been confirmed by Foucault, Sraer & Thesmar (2011) who show that retail trading activity has only a positive effect on the volatility of stock returns without generating any price impact which suggests that retail investors behave like noise traders, as they trade for non-informational reasons.

Despite this unflattering image, more and more researchers try to address the future return predictability based on retail investors' order flow, by investigating whether stock returns can be predicted using retail investors' net buying statistics. In fact, as pointed out by Kaniel, Liu, Saar & Titman (2012), the aggregation of the individual information contained in all trades executed by a crowd of retail investors might send a quite precise signal. They focus on individual trades around earnings announcements and argue further that it might be much easier for them to trade quickly and aggressively small quantities when they are informed than for institutions, as institutions might face trading restrictions and are cautious to incur litigation risk and adverse publicity by trading too aggressively.

Nevertheless, this debate in the literature is far from over with many existing studies coming to different results with no clear cut conclusions. Dorn, Huberman & Sengmueller (2008) analyse daily transaction records of 37,000 retail clients from a single German retail broker from 1998 to 2000, where they distinguish between speculative and other trades, and between market orders and executed limit orders. They show that both aggressive and passive order imbalances positively predict stock returns at monthly horizons. Moreover they also find that these retail investors behave both like trend followers (positive feedback trading) and contrarians.

Hvidkjaer (2008) investigates the relationship between retail investor trading behavior and future stock returns by using monthly small-trade volume data on listed New York Stock Exchange (NYSE) and American Stock Exchange stocks in the period from January 1983 to December 2005. He shows that small-trade volume contains information about future stock returns, and more specifically, that stocks with high small trade buying pressure are outperformed by those with strong selling pressure.

Kaniel et al. (2012) provide evidence that, between 2000 to 2003, aggregated retail orders executed around earnings announcements on the NYSE are able to predict both large positive and negative daily abnormal firms' stock returns. Their results support the argument that on aggregate retail investors' trading prior to earnings announcements contains information. Moreover, by decomposing the cumulative abnormal returns into a component attributed to liquidity provision and a component attributed to trading on private information, they show that retail investors act as contrarians and provide liquidity to the market and that almost 50% of the trades are based on some kind of private information.

Barber, Odean & Zhu (2009) use tick-by-tick transaction level data for US stock markets over the period 1983-2001. Given that their dataset does not contain information on who is a retail investor, they infer their presence by the size of the trade. They find that retail investors are able to predict weekly and annually returns in the way that heavily bought (sold) stocks one week earn strong (low) returns the subsequent week. This effect reverses on an annual basis.

More recently, Kelley & Tetlock (2013) investigate the role of retail investors in stock pricing over the period 2003-2007. They show that collectively retail investors are able to predict monthly returns, and hence cannot only be consider as "noise" traders, as usually assumed in the finance literature (e.g., Black (1986) and many others). Menkhoff, Sarno, Schmeling & Schrimpf (2013) use data on daily customer order flows of asset managers, hedge funds, corporate and private clients of the FX market over a sample period from 2001 to 2011. They show that end users order flow is highly informative for future exchange rate changes, but that order flow from corporations and "private" customers are not or negatively related to price movements. However, even in the presence of a negative correlation of order flow and price movements, order flow is informative from a forecasting point of view.

In the light of the above mentioned papers, one can postulate that retail investors can be regarded as a special class of "informed" traders, possessing private information, in the sense that they trade not only based on their beliefs concerning the price change direction rather than the fundamentals of the underlying asset, as pointed out by Harris (2003), but also in the sense that they use the aggregated information flow in the market to learn and become informed (i.e. chartists, trend followers, contrarians). Taking these characteristics into account, Xue & Gencay (2012) introduce a market microstructure model in which they assume that the aggregated supply of noise traders can be split into two components, a predictable aggregated supply which is a function of past prices and an unpredictable, idiosyncratic liquidity shock. They then introduce different layers of information hierarchies in their model with the result that more information hierarchies among informed traders imply a slower information diffusion rate. In their model, they obtain this result because it is impossible for the partially informed traders in the middle to infer whether the price movements come from a change in the fundamental value/information or from the liquidity shock generated by noise traders. This also means that the best informed traders prefer to trade with each other in the presence of information hierarchies and thereby risk ignoring the layers of less informed traders, who in the meantime might have acquired new valuable information. Using simulation studies and data from China's limit order book market from January to September 2007, they show using their model not only that the trend-following behavior of noise traders can generate momentum in returns, but also that the aggregated information of noise traders can generate excess returns in a perfectly rational way.

In contrast to the existing literature focusing on the future return predictability based on retail investors' order flow in the equity market, studies on order flow in the FX market entirely concentrate on the information content of inter-dealer or aggregated customer order flow, as mentioned previously in the introduction. In the light of the model of Xue & Gencay (2012), the results obtained for retail investors in equity markets and our unique dataset containing information on small retail investors in the FX market AND on detailed flow of different categories of individual trading actions, we postulate that the trading activity of the retail traders on OANDA FXTrade contains valuable information for the prediction of future price changes. To be more precise, we formulate the even stronger hypothesis that *on aggregate* retail investors possess additional information for short term intraday price

change forecasting beyond the information already reflected by past prices and order flow of institutional investors. In this context, we especially focus on the information content of the individual transaction categories' order flows.

3 OANDA FXTrade and Data Description

OANDA FXTrade (www.oanda.com) is an electronic trading platform for currencies operating on a 24 hours, 7 days per week basis. The platform is organised as a market making system and continuously provides bid and ask quotes with spreads as low as 2 to 3 pips for all transaction sizes for all major currency pairs. The quotes are based on a proprietary forecasting algorithm relying on OANDAs' own inventory book and an external (crossing network) data feed. Given various boundary conditions, as for example sufficient margin requirements, orders are always executed. Market orders (buy or sell) are executed immediately and limit orders are maintained in the system for up to one month. The bid and ask prices quoted by OANDA rely only partially on their own inventory (limit order) book and a situation might arise where a newly submitted limit order lies inside the bid-ask spread without visibly improving OANDAs' quoted prices. Therefore a new limit order could be matched by OANDA i) against an incoming market order, as usual; or ii) against the bid or the ask price of OANDAs' quoted prices once they "cross" the limit order which lay inside the bid-ask spread. In addition to standard limit orders, so called special limit orders (stop-loss and take-profit) can be submitted to the system. Stop-loss orders imply the closure of a position when a previously set loss limit is reached. Take-profit orders, on the contrary, close an open position when a certain profit is realised.

The dataset used in our analysis is constructed from the trading activity record of OANDA FXTrade which contains the exact timing of all trading actions on a second-by-second basis. The first column of Table 1 gives an overview of these actions. The data spans from 06/10/2003 to 14/05/2004 (160 weekdays, 32 weeks) and contains all the trading activity of around 2000 investors. Depending on the order type, we receive information on transaction prices (market orders, limit orders executed, stop-loss, take profit, margin call), bid and ask quotes (limit orders pending), associated transaction units, and the limits of stop-loss and take-profit orders. A more detailed description of the dataset is included in Nolte & Voev (2011). In our analysis we focus on the most actively traded currency pair *EUR/USD*, which accounts for nearly 39 % of all records with an average interrecord-duration of 8.5 seconds. Table 1

contains the descriptive statistics for the dataset and the transaction volumes for the specific order categories. All figures are daily averages computed over the whole dataset. The average number of *different* traders per day lies around 750 for the *EUR/USD* currency pair.

As already mentioned above the order flow measures usually considered in the literature are constructed from datasets with less information. In fact, the standard order flow measure of Lyons (1995) only requires knowledge about the number of buy and sell market orders in a given time period. At this aggregation level valuable information (for prediction purposes) contained in the individual order categories could potentially be aggregated out. Our dataset, however, allows us to analyse the information content of each and every order category and to distinguish whether these orders were submitted for the purpose of opening or closing a position.

4 Analysis of Category Specific Order Flows

To assess the forecasting performance of the category specific order flows we carry out an in- and out-of sample forecasting analysis. We check whether forecasting models for interbank price changes relying on category specific order flow information in addition to a benchmark forecasting specification outperform the benchmark. We use two benchmark specifications: 1) a simple white noise model containing no further information about the trading process whatsoever and 2) an advanced model that contains lagged price changes and lagged interbank order flow (computed in the standard way, as in Lyons (1995)). We conduct this analysis for 9 intraday forecasting horizons: 1 min, 2 min, 3 min, 4 min, 5 min, 10 min, 20 min, 30 min, and 1 hour. The interbank order flow measures are computed with Reuters Dealing 3000 data and the price change series are computed from mid-quote series provided by Olsen Financial Technologies. The advanced benchmark model is a quite tough competitor as it contains in addition to past price information also information about the order flow in the interbank market. We decide to include lagged price changes into the advanced model to control for particular autoregressive structures in the price change processes that could be implied by bid-ask bounce and feedback trading effects. Such effects also include short term high-frequency predictability implied by noise traders that rely on trend following or chart-analysis as postulated in the hierarchical information market microstructure model by Xue & Gencay (2012).

Transaction Record	%	Obs	Trading Volume in EUR per Day								
			Total	Mean	Min	5% Quan.	25% Quan.	50% Quan.	75% Quan.	95% Quan.	Max
Buy market (open)	13.10	1322	37930860	25854	82	113	515	2065	9240	85854	2220414
Sell market (open)	10.61	1072	30816226	27218	44	89	592	2138	9861	96214	1759412
Buy market (close)	8.27	835	25074760	27468	163	201	672	2326	9553	95940	1630034
Sell market (close)	10.27	1037	31839764	29534	29	66	564	2164	10063	97248	1930846
Limit order: Buy	5.41	546	14041270	28876	24	63	549	2053	9469	95436	1934417
Limit order: Sell	4.76	482	11080825	34283	237	267	515	1662	7509	117914	1511133
Buy limit order executed (open)	3.22	325	5416146	17484	41	79	422	1410	6267	67127	735479
Sell limit order executed (open)	2.92	295	3231307	10554	58	84	242	824	3652	34607	584303
Buy limit order executed (close)	0.46	46	1382690	32718	4800	4824	5313	7020	17994	80426	506182
Sell limit order executed (close)	0.46	46	1470630	32287	407	436	927	3440	16816	93447	452512
Buy take-profit (close)	3.14	317	2918779	9779	144	170	310	704	2724	30314	583296
Sell take-profit (close)	3.49	352	4404025	12857	61	75	256	796	3960	43028	820876
Buy stop-loss (close)	2.18	220	4488496	16433	126	175	667	2535	9837	70968	513989
Sell stop-loss (close)	2.55	258	5309807	16667	23	59	503	2255	9424	66743	650061
Buy margin call (close)	0.12	12	166375	7263	1006	1010	1185	1817	3718	14211	71133
Sell margin call (close)	0.17	17	275282	6381	1369	1372	1440	2351	4409	17266	77231
Change order	3.01	305	13898910	49771	105	203	1295	4888	18181	162927	1622712
Change stop-loss or take-profit	22.36	2260	60965013	26748	10	79	867	3694	14163	95983	1703030
Cancel by hand	2.41	243	10043949	42295	211	272	1031	4186	16003	148571	1662224
Cancel: insufficient funds	0.28	28	2439586	67905	4938	4953	5431	7354	66280	186280	622650
Cancel bound violation	0.20	20	195118	14803	571	571	627	2650	6860	29909	98308
Order expired	0.65	66	1063061	19942	44	54	443	1682	7204	68648	355982

Table 1: Descriptive statistics of the OANDA FXTrade trading activity dataset for the EUR/USD currency pair. All numbers are daily averages and all transaction volume statistics are denominated in EUR.

The number of lagged price changes and lagged order flows are chosen according to the Schwarz Information Criteria (SIC). Our category specific order flow measures are computed as balances of the number of buy minus sell orders for each category. For example, we compute the order flow of the market order (open) category as the difference between the number of buy market orders (open) and sell market orders (open) for each time period. Thus, we obtain seven category specific order flow measures that are summarised in Table 2.

Category	Description
1	Limit orders
2	Limit orders executed (open)
3	Limit orders executed (close)
4	Market orders (open)
5	Market orders (close)
6	Stop-loss orders (close)
7	Take-profit orders (close)

Table 2: Description of the category specific order flow measures. Col. 1 states the number of the category and col. 2 gives the category description.

We consider a period of 32 weeks starting on Monday the 6th of October 2003 and ending on Friday the 14th of May 2004. We divide our data into an in-sample estimation period containing the first 28 weeks and an out-of-sample period containing the last 4 weeks. We use a conservative forecasting approach and estimate the model parameters only once in-sample and then keep the parameters fixed for the whole out-of-sample period. We compute Root-Mean-Squared-Prediction-Errors (RMSPE) and analyse whether the models incorporating information on the category specific order flows provide significantly better forecasts than the benchmark models with the help of the modified Diebold-Mariano (mDM) test of Harvey, Leybourne & Newbold (1997). We exclude the overnight periods from our data as trading activity is very low both on OANDA FXTrade and on Reuters Dealing 3000.

Let y_t denote the FX price process. The benchmark white noise specification is then given by:

$$\Delta y_t = \varepsilon_t, \tag{WN}$$

and the advanced benchmark model specification, with x_t^{OF} denoting the interbank

order flow measure, takes the following form:

$$\left(1 - B_p^y(L)\right) \Delta y_t = c + B_q^x(L) x_t^{\text{OF}} + \varepsilon_t, \qquad\qquad \text{(AR-OF)}$$

where $B_p^z(L)$ denotes the associated lag-polynomials specified as $B_p^z(L) = \beta_{z1}L + \ldots + \beta_{zp}L^p$, and ε_t is a white noise process. In the advanced benchmark model we try to control for autoregressive price change patterns such as bid-ask bounce effects and the effect of lagged interbank order flow. In the second columns of Tables 3 and 4 we report the RMSPEs of both benchmark specifications for all aggregation levels. As the advanced benchmark model nests the white noise specification it is clear that it provides a better in-sample fit than the white noise specification. The relative improvement is higher for higher frequencies on which market micro structure issues matter more. The same pattern is also visible for the out-of-sample RMSPEs, although for frequencies of 4, 20 and 30 minutes the advanced specification performs slightly worse than the white noise specification.

In order to test whether each category k specific order flow measure x_t^k with $k = 1, \ldots, 7$ contains new information for forecasting, we extend the (AR-OF) specification with lagged category k order flow measures. This yields in total k models of the form:

$$\left(1 - B_p^y(L)\right) \Delta y_t = c + B_q^x(L) x_t^{\text{OF}} + B_r^{xk}(L) x_t^k + \varepsilon_t. \qquad\qquad \text{(CAT-}k\text{)}$$

The results of the in-sample estimation are presented in Table 3. The RMSEs for the models containing the category k order flow measures are smaller than those of both benchmark specifications across all frequencies. The p-values of the mDM-test, however, show that especially the specifications relying on order flow of 6-Stop-loss orders and 7-Take-profit orders outperform the benchmark model up to the 20 min sampling frequency. Also, the order flows of 2-Limit orders executed (open) and 4-Market orders (open) show a significant forecasting improvement up to the 4 min frequency, while the order flow in all the other categories seems to be helpful only up to the 2 min frequency. The strong forecasting power of stop-loss and take-profit orders for up to 20 min can be explained by the fact that these are usually submitted as safeguards and been triggered by more pronounced price movements.

A weaker but similar pattern emerges from the analysis of the out-of-sample RMSPE presented in Table 4. Here one notable difference is that also the order flow of category 1-Limit orders submitted but not yet executed is clearly helpful for prediction

and helps to significantly outperform both benchmark specifications up to the 5 min frequency.

The bottom line from this analysis is that i) the order flow on OANDA FXTrade, even in disaggregated form (as individual order flow categories), contains significant information beyond the information already contained in past price changes and lagged interbank order flow; and ii) the order flows of both 3-Limit orders executed (close) and 5-Market orders (close) which are transactions with the purpose to close an open existing position are least informative. These results support our hypothesis that *collectively* small retail investors' trading actions have prediction power, and therefore confirm the findings that Kelley & Tetlock (2013), Kaniel et al. (2012) and Barber et al. (2009) obtained for equity markets. This directly implies that also in market microstructure models for FX markets, retail investors mustn't be considered as pure "noise" traders.

In addition, our results show that *on aggregate* retail investors trading flow can be exploited to significantly improve *short* term out-of-sample price change forecasts which allows us to fill a second gap in the literature and to go even a step further than previous studies including Dorn et al. (2008) that consider longer horizons and non high frequency forecasting setups. We find that both market order and limit order order flows predict future price changes at high frequencies up to 5 minutes.

Freq	BM	CAT1	CAT2	CAT3	CAT4	CAT5	CAT6	CAT7
1 min	.	0.2587	0.2585	0.2586	0.2586	0.2585	0.2574	0.2576
(WN)	0.2623	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**
(AR-OF)	0.2593	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**
2 min	.	0.3857	0.3854	0.3856	0.3855	0.3856	0.3853	0.3855
(WN)	0.3865	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**
(AR-OF)	0.3859	**(0.0005)**	**(0.0000)**	**(0.0001)**	**(0.0000)**	**(0.0001)**	**(0.0006)**	**(0.0001)**
3 min	.	0.4692	0.4692	0.4692	0.4691	0.4692	0.4689	0.4692
(WN)	0.4696	**(0.0012)**	**(0.0012)**	**(0.0006)**	**(0.0002)**	**(0.0009)**	**(0.0000)**	**(0.0003)**
(AR-OF)	0.4692	**(0.4709)**	**(0.3728)**	**(0.1561)**	**(0.0204)**	**(0.3063)**	**(0.0122)**	**(0.2691)**
4 min	.	0.5409	0.5408	0.5409	0.5407	0.5409	0.5407	0.5409
(WN)	0.5416	**(0.0028)**	**(0.0018)**	**(0.0018)**	**(0.0007)**	**(0.0023)**	**(0.0003)**	**(0.0022)**
(AR-OF)	0.5409	**(0.3713)**	**(0.1121)**	**(0.1551)**	**(0.0351)**	**(0.3207)**	**(0.1015)**	**(0.4118)**
5 min	.	0.5957	0.5956	0.5956	0.5956	0.5956	0.5953	0.5955
(WN)	0.5959	**(0.0322)**	**(0.0273)**	**(0.0125)**	**(0.0310)**	**(0.0235)**	**(0.0018)**	**(0.0066)**
(AR-OF)	0.5957	**(0.3060)**	**(0.2610)**	**(0.1232)**	**(0.2412)**	**(0.2460)**	**(0.0416)**	**(0.1207)**
10 min	.	0.8301	0.8300	0.8299	0.8301	0.8300	0.8293	0.8293
(WN)	0.8303	**(0.1534)**	**(0.1260)**	**(0.0857)**	**(0.1826)**	**(0.1244)**	**(0.0029)**	**(0.0165)**
(AR-OF)	0.8301	**(0.2677)**	**(0.2026)**	**(0.1354)**	**(0.3234)**	**(0.2482)**	**(0.0174)**	**(0.0691)**
20 min	.	1.1568	1.1568	1.1570	1.1569	1.1567	1.1549	1.1543
(WN)	1.1570	**(0.2058)**	**(0.2298)**	**(0.3598)**	**(0.2838)**	**(0.2085)**	**(0.0248)**	**(0.0229)**
(AR-OF)	1.1570	**(0.2118)**	**(0.2487)**	**(0.4083)**	**(0.3134)**	**(0.2110)**	**(0.0266)**	**(0.0223)**
30 min	.	1.4095	1.4093	1.4095	1.4077	1.4095	1.4089	1.4093
(WN)	1.4112	**(0.0935)**	**(0.0775)**	**(0.0925)**	**(0.0232)**	**(0.0921)**	**(0.0551)**	**(0.0870)**
(AR-OF)	1.4095	**(0.4977)**	**(0.3132)**	**(0.4278)**	**(0.0992)**	**(0.4330)**	**(0.1647)**	**(0.2762)**
1 hr	.	2.0210	2.0208	2.0207	2.0192	2.0211	2.0201	2.0206
(WN)	2.0225	**(0.2945)**	**(0.2806)**	**(0.2785)**	**(0.1649)**	**(0.3043)**	**(0.2402)**	**(0.2702)**
(AR-OF)	2.0211	**(0.4147)**	**(0.3607)**	**(0.3392)**	**(0.1875)**	**(0.4508)**	**(0.2435)**	**(0.2940)**

Table 3: Results for the **in-sample** price change prediction on different sampling frequencies (Freq) for the 7 different category specific order flow measures. The first cell entry is the Root-Mean-Squared-Prediction Error (RMSPE) of the associated forecasting model.The second and third cell entries in parenthesis are the p-value from the modified Diebold-Mariano (mDM) test with the null hypothesis that the RMSPE of the associated forecasting model is not smaller than the RMSPE of the corresponding White Noise or AR-OF benchmark model (WN, AR-OF). P-values in bold correspond to those cases where the RMSPE of the associated forecasting model is smaller than the RMSPE of the corresponding benchmark model (WN, AR-OF).

Freq	BM	CAT1	CAT2	CAT3	CAT4	CAT5	CAT6	CAT7
1 min	.	0.2413	0.2411	0.2414	0.2413	0.2412	0.2391	0.2391
(WN)	0.2448	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0098)**	**(0.0029)**
(AR-OF)	0.2420	**(0.0348)**	**(0.0000)**	**(0.0055)**	**(0.0323)**	**(0.0002)**	**(0.0788)**	**(0.0404)**
2 min	.	0.3681	0.3672	0.3681	0.3679	0.3680	0.3687	0.3683
(WN)	0.3686	**(0.0221)**	**(0.0480)**	**(0.0189)**	**(0.0057)**	**(0.0353)**	(0.5546)	**(0.3714)**
(AR-OF)	0.3684	**(0.0230)**	**(0.1011)**	**(0.0205)**	**(0.0367)**	**(0.0433)**	(0.6620)	**(0.4908)**
3 min	.	0.4519	0.4517	0.4521	0.4516	0.4519	0.4521	0.4522
(WN)	0.4521	**(0.1433)**	**(0.1080)**	**(0.4330)**	**(0.1467)**	**(0.1557)**	**(0.4855)**	**(0.5414)**
(AR-OF)	0.4519	**(0.0381)**	**(0.1466)**	(0.8848)	**(0.2063)**	(0.5959)	(0.7431)	(0.8246)
4 min	.	0.5005	0.5002	0.5012	0.5006	0.5006	0.5004	0.5006
(WN)	0.5000	(0.7660)	(0.6744)	(0.9142)	(0.8164)	(0.8051)	(0.7874)	(0.8241)
(AR-OF)	0.5006	**(0.0219)**	**(0.1662)**	(0.9921)	(0.5522)	(0.7583)	**(0.3336)**	(0.9247)
5 min	.	0.5926	0.5921	0.5926	0.5925	0.5927	0.5939	0.5942
(WN)	0.5938	**(0.0511)**	**(0.1216)**	**(0.0704)**	**(0.0822)**	**(0.0471)**	(0.5946)	(0.8539)
(AR-OF)	0.5928	**(0.0429)**	**(0.1997)**	(0.2153)	(0.1508)	(0.4652)	(0.8663)	(0.9428)
10 min	.	0.8260	0.8243	0.8267	0.8260	0.8259	0.8261	0.8278
(WN)	0.8273	**(0.0966)**	**(0.1206)**	**(0.3455)**	**(0.1483)**	**(0.0856)**	**(0.0663)**	(0.6511)
(AR-OF)	0.8260	(0.5927)	**(0.1492)**	(0.8475)	(0.5641)	**(0.3752)**	(0.5493)	(0.7955)
20 min	.	1.0724	1.0701	1.0713	1.0710	1.0708	1.0656	1.0678
(WN)	1.0711	(0.9153)	**(0.1143)**	(0.6457)	**(0.3941)**	**(0.3749)**	**(0.0056)**	**(0.2304)**
(AR-OF)	1.0713	(0.8873)	**(0.0715)**	(0.4570)	(0.2732)	(0.2919)	**(0.0038)**	**(0.2172)**
30 min	.	1.4275	1.4237	1.4269	1.4240	1.4274	1.4270	1.4278
(WN)	1.4202	(0.9539)	(0.8709)	(0.9451)	(0.8780)	(0.9497)	(0.9011)	(0.9219)
(AR-OF)	1.4275	(0.6644)	**(0.0689)**	**(0.2343)**	**(0.1346)**	**(0.4277)**	**(0.3711)**	(0.5815)
1 hr	.	1.7603	1.7605	1.7616	1.7659	1.7600	1.7578	1.7592
(WN)	1.7648	**(0.1161)**	**(0.1660)**	**(0.3110)**	(0.5803)	**(0.0951)**	**(0.0638)**	**(0.0897)**
(AR-OF)	1.7602	(0.5605)	(0.5981)	(0.6091)	(0.9010)	**(0.3117)**	**(0.1904)**	**(0.3333)**

Table 4: Results for the **out-of-sample** price change prediction on different sampling frequencies (Freq) for the 7 different category specific order flow measures. The first cell entry is the Root-Mean-Squared-Prediction Error (RMSPE) of the associated forecasting model.The second and third cell entries in parenthesis are the p-value from the modified Diebold-Mariano (mDM) test with the null hypothesis that the RMSPE of the associated forecasting model is not smaller than the RMSPE of the corresponding White Noise or AR-OF benchmark model (WN, AR-OF). P-values in bold correspond to those cases where the RMSPE of the associated forecasting model is smaller than the RMSPE of the corresponding benchmark model (WN, AR-OF).

5 Discussion about Order Flow Measures

The results of the forecasting study in the last section shows that the category specific order flows on OANDA FXTrade contain significant information for the prediction of high frequency price changes, beyond the information that is already contained in past price changes and interbank order flow. Moreover, different order categories are informative for different forecasting horizons. Given these striking observations, two questions arise from a forecasting point of view: i) can we find a

combined (or aggregated across categories) order flow measure that is even better suited as a forecasting variable than any of the individual categories and ii) what is the performance of such a combined measure in comparison to standard order flow measures available in the literature (e.g. Lyons (1995)). From an information point of view, we mustn't forget that OANDA FXTrade is organised partially as a crossing network which means that the bid and ask quotes on OANDA FXTrade depend in addition to its own inventory book on bid and ask prices from the dealer market through an external datafeed. The precise algorithm of how bid and ask prices on OANDA FXTrade are quoted, however, is a company secret and unknown to us. It is our understanding that the dealer market prices are not simply put through as it would happen in a pure crossing network. Given the setup of OANDA FXTrade and many other electronic trading platforms that rely to some degree on external datafeeds from the interbank market, it should, however, be immediately clear that all transactions (regardless of their type) on OANDA FXTrade combine information from both markets, the dealer market and OANDA FXTrade. Hence – from our forecasting point of view – the underlying key question is, how do we extract valuable information for forecasting from both markets with the order flows that we observe on OANDA FXTrade in the best possible way. But, we are also interested in constructing an order flow measure that mostly aggregates trading intentions from the crowd of OANDA FXTrade traders alone. Such a measure would shed some light on whether any improved forecasting performance stems from a 'learning channel' in the sense that order flow on OANDA FXTrade simply extracts counterparty information from the dealer market, of course, differently than the above employed interbank order flow measure; or an 'information endowment channel' in the sense that the crowd of traders on OANDA FXTrade has new information that has not reached the dealer market, yet. The latter is consistent with the common view that new information for players in the dealer market is generated through their own trading with their own customers, which naturally happens before they trade themselves in the dealer market. The difference here is that the investors on OANDA FXTrade are mainly retail investors.

To have a benchmark measure we decide to follow the literature and construct the standard order flow measure in the spirit of Lyons (1995). His standard definition of an aggregated net order flow measure is the difference between buyer initiated and seller initiated trades within a given period; or stated differently, the cumulative sum of signed orders where buyer initiated and seller initiated orders receive positive and

negative signs, respectively. The key idea behind this measure is to capture very recent changes in the information of the market participants, by focusing on the initiating party of a trade (submission of market orders), deciding to trade because of new (private) information. Hence, for example, an executed buy limit order is treated as a seller initiated trade, since it has to be merged with a sell market order. Here the very recent information of the seller about the future direction of the price process is treated as more important than the information of the buyer, who might not have the latest information (or interpreted it differently) and previously decided to wait by placing a buy limit order. This standard order flow measure is very well suited for explaining future prices as far as the interbank market is considered, as demonstrated by Daníelsson et al. (2012).

Let us now consider trades on OANDA FXTrade where bid and ask prices depend partially on an external data-feed from the interbank market and its own inventory book. Hence the bid and ask prices on OANDA actually depend on the information in two markets: OANDA itself and the interbank market. For the sake of the argument let us assume for the moment that OANDAs' inventory book is empty and bid and ask prices are simply put through one-to-one from the interbank market. A newly submitted buy limit order (bid) on OANDA is therefore sitting in the "book" and will get matched in the absence of further market orders against the ask price of OANDA FXTrade (the external data-feed) once the ask price just reaches (crosses) the bid of this buy limit order. As a result, the execution of this buy limit order is triggered by selling pressure in the primary market shortly before, e.g. a large sell market order in the interbank market, consuming the previous best bids there and also causing an adjustment of the best ask quotes to lower prices sightly thereafter. Measuring order flow on OANDA FXTrade with the standard net order flow measure described above yields a mixture of price expectations from traders on OANDA FXTrade (mainly through market orders) and price expectations from the interbank market (mainly through executed limit orders). Stated differently, the standard order flow measure doesn't allows us to make a distinction between the before mentioned "learning" and "information endowment" channels.

To disentangle these two types of information processes, we decide to construct an additional order flow measure that only tries to aggregate information about the direction of a future price change on OANDA FXTrade. This measure helps to test the statement of Harris (2003), that retail investors trade based on their beliefs concern-

17

ing a price change direction rather than fundamental information of the underlying asset. As explained below for the construction of the OANDA order flow measure we, therefore, give up the priority of the trades of the initiating party in determining the sign of the order flow measure. The key assumption that we make is, that as in the example above, limit orders on OANDA will not get matched against OANDA market orders, but rather the external data-feed. In Table 5, we summarise definitions of the standard and the additional order flow measure, which we denote as "OANDA order flow". In column one we list the different types of actions, in column two the signs for the standard order flow measure and in column three the signs for the OANDA order flow measure.

Buy market orders, irrespective of whether they are submitted to open or close a position, get positive signs in both order flow measures since the traders on OANDA FXTrade initiate these trades and believe that the price will go up. Hence, sell market orders receive negative signs.

In the standard order flow measure, submitted (pending) limit orders are not considered, since they are not yet executed, which means that there is not yet an initiating party. They are, however, taken into account in the OANDA order flow measure since the trader, who submits a limit order, expresses his belief that the price will go up (buy, positive sign) or down (sell, negative sign).

Executed buy limit orders are treated as seller initiated in the standard order flow measure (see the discussion above) and are thus assigned negative signs. In the OANDA order flow measure they receive positive signs, since the submitter still believes that the price will increase. Otherwise he could have cancelled the order before execution. Executed sell limit orders are treated analogously.

Buy take-profit orders (close) are buy limit orders that receive negative signs in the standard order flow measure. In the OANDA order flow measure, they get positive signs, because the trader believes that the price will fall further. A buy take-profit order (close) can only be executed if the trader has a short position in a currency pair (short position in the base currency). Sell take-profit orders receive the symmetric signs. Buy stop-loss orders (close) get negative signs in both measures. For the standard order flow measure the explanation is that it is a special buy limit order. For the OANDA order flow measure the explanation is that the trader believes that

the price will fall further. Again, sell stop-loss orders are treated analogously.

Buy margin call orders (close) are not used in both order flow measures. On the one hand, one can argue that they should get positive signs since they are buy market orders. On the other hand, one can argue that they should receive negative signs as they are not motivated by new information and that the traders were just proven to have wrong expectations about the price. Due to their scarce occurrence (0.12% and 0.17%, see Table 1) we don't expect them to play any role.

Transaction Record	Standard Order Flow Signs	OANDA Order Flow Signs
Buy market (open)	+	+
Sell market (open)	-	-
Buy market (close)	+	+
Sell market (close)	-	-
Limit order: Buy	not used	+
Limit order: Sell	not used	-
Buy limit order executed (open)	-	+
Sell limit order executed (open)	+	-
Buy limit order executed (close)	-	+
Sell limit order executed (close)	+	-
Buy take-profit (close)	-	+
Sell take-profit (close)	+	-
Buy stop-loss (close)	-	-
Sell stop-loss (close)	+	+
Buy margin call (close)	not used	not used
Sell margin call (close)	not used	not used
Change order	not used	not used
Change stop-loss or take-profit	not used	not used
Cancel order by hand	not used	not used
Cancel order: insufficient funds	not used	not used
Cancel order: bound violation	not used	not used
Order expired	not used	not used

Table 5: Definitions of the order flow measures. Col. 1 states the record entries, col. 2 contains the signs for the construction of the standard net order flow measure and col. 3 contains the signs for the construction of the OANDA order flow measure.

In order to provide a comparative basis for the in- and out-of-sample analyses performed in Section 4, we also investigate the performance of a purely data driven order flow measure which is not based on any theoretical motivation of how information is aggregated into a single order flow measure. Here we simply include all category k specific order flow measures as explanatory variables into the regression setup.

Similar to the previous section, we again extend the (AR-OF) benchmark specification with our lagged combined order flow measures x_t^k with $k = \{\text{SOF}, \text{OOF}\}$ where the superscripts SOF and OOF denote the standard and OANDA order flow measures, respectively. This yields:

$$\left(1 - B_p^y(L)\right) \Delta y_t = c + B_q^x(L)x_t^{\text{OF}} + B_r^{xk}(L)x_t^k + \varepsilon_t. \qquad \text{(CAT-}k\text{)}$$

The pure data driven combined order flow model is given by:

$$\left(1 - B_p^y(L)\right) \Delta y_t = c + B_q^x(L)x_t^{\text{OF}} + \sum_{k=1}^{7} B_r^{xk}(L)x_t^k + \varepsilon_t. \qquad \text{(ALL-CAT)}$$

where x_t^k with $k = 1, \dots, 7$ denote the category k specific order flow measure as before.

Table 6 presents the RMSEs of our three combined order flow measure specifications for all 9 prediction horizons. The p-values of the mDM-tests show that the standard order flow model only outperforms the (AR-OF) benchmark model up to a 4 min frequency, while the OANDA order flow measure outperforms the benchmark up to a frequency of 30 minutes and the purely data driven order flow model outperforms it even up to a frequency of 1 hour. This observation shows that even in-sample the standard order flow measure doesn't seem to do such a good job in aggregating important forecasting information and the comparison with the OANDA order flow measure or the data driven model shows that it lacks the ability to capture the slightly longer term forecasting information. It is not surprising that the pure data driven model performs best in the in-sample analysis as it is the most flexible specification, with no a priori assumptions imposed on how the individual categories are aggregated up. The OANDA order flow measure, constructed to aggregate the information of the crowd of OANDA FXTrade retail traders, performs almost as good as the pure data driven measure, which shows again that there is clearly collectively valuable information among OANDA FXTrade retail traders. Given these results, we can further conclude that information is not alone coming from a learning, but most likely also from an information endowment channel.

The out-of-sample results in Table 7 confirm the conclusions drawn from the in-sample study and although they are in general weaker they support the finding that the OANDA order flow measure contains valuable forecasting information especially for the longer forecasting horizons of 20 and 30 minutes. The purely data driven

20

order flow model is also not the uniformly best model anymore.

To sum up, these results are again in line with the previous findings in this paper and the literature stating that *on aggregate* retail trading conveys information about future prices (see Dorn et al. (2008), Kaniel et al. (2012), Menkhoff et al. (2013), among others).

Freq	BM	CAT-SOF	CAT-OOF	ALL-CAT
1 min	.	0.2587	0.2574	0.2567
(WN)	0.2623	**(0.0000)**	**(0.0000)**	**(0.0000)**
(AR-OF)	0.2593	**(0.0000)**	**(0.0000)**	**(0.0000)**
2 min	.	0.3856	0.3852	0.3849
(WN)	0.3865	**(0.0000)**	**(0.0000)**	**(0.0000)**
(AR-OF)	0.3859	**(0.0004)**	**(0.0000)**	**(0.0000)**
3 min	.	0.4692	0.4688	0.4688
(WN)	0.4696	**(0.0005)**	**(0.0000)**	**(0.0000)**
(AR-OF)	0.4692	**(0.0856)**	**(0.0021)**	**(0.0128)**
4 min	.	0.5408	0.5407	0.5404
(WN)	0.5416	**(0.0010)**	**(0.0002)**	**(0.0001)**
(AR-OF)	0.5409	**(0.0857)**	**(0.0658)**	**(0.0141)**
5 min	.	0.5956	0.5954	0.5951
(WN)	0.5959	**(0.0304)**	**(0.0026)**	**(0.0007)**
(AR-OF)	0.5957	**(0.2446)**	**(0.0462)**	**(0.0142)**
10 min	.	0.8301	0.8293	0.8285
(WN)	0.8303	**(0.2052)**	**(0.0052)**	**(0.0007)**
(AR-OF)	0.8301	**(0.4394)**	**(0.0265)**	**(0.0065)**
20 min	.	1.1569	1.1550	1.1524
(WN)	1.1570	**(0.3112)**	**(0.0214)**	**(0.0021)**
(AR-OF)	1.1570	**(0.3362)**	**(0.0220)**	**(0.0021)**
30 min	.	1.4083	1.4079	1.4064
(WN)	1.4112	**(0.0317)**	**(0.0254)**	**(0.0134)**
(AR-OF)	1.4095	**(0.1399)**	**(0.0793)**	**(0.0435)**
1 hr	.	2.0205	2.0192	2.0161
(WN)	2.0225	**(0.2525)**	**(0.1838)**	**(0.0735)**
(AR-OF)	2.0211	**(0.2999)**	**(0.1665)**	**(0.0715)**

Table 6: Results for the **in-sample** price change prediction on different sampling frequencies (Freq) for the standard (CAT-SOF), OANDA (CAT-OOF) and aggregated categories (ALL-CAT) order flow measures. The first cell entry is the Root-Mean-Squared-Prediction Error (RMSPE) of the associated forecasting model.The second and third cell entries in parenthesis are the p-value from the modified Diebold-Mariano (mDM) test with the null hypothesis that the RMSPE of the associated forecasting model is not smaller than the RMSPE of the corresponding White Noise or AR-OF benchmark model (WN, AR-OF). P-values in bold correspond to those cases where the RMSPE of the associated forecasting model is smaller than the RMSPE of the corresponding benchmark model (WN, AR-OF).

Freq	BM	SOF	OOF	ALL-CAT
1 min	.	0.2413	0.2393	0.2380
(WN)	0.2448	**(0.0000)**	**(0.0000)**	**(0.0021)**
(AR-OF)	0.2420	**(0.0472)**	**(0.0015)**	**(0.0223)**
2 min	.	0.3682	0.3681	0.3678
(WN)	0.3686	**(0.0710)**	**(0.1866)**	**(0.0013)**
(AR-OF)	0.3684	**(0.1476)**	**(0.2776)**	**(0.0001)**
3 min	.	0.4518	0.4518	0.4522
(WN)	0.4521	**(0.1242)**	**(0.1830)**	(0.5896)
(AR-OF)	0.4519	**(0.2506)**	**(0.3612)**	(0.7991)
4 min	.	0.5005	0.5010	0.5004
(WN)	0.5000	(0.7685)	(0.9060)	(0.7633)
(AR-OF)	0.5006	**(0.3908)**	(0.9474)	**(0.4228)**
5 min	.	0.5925	0.5931	0.5932
(WN)	0.5938	**(0.0425)**	**(0.1696)**	**(0.1532)**
(AR-OF)	0.5928	**(0.0318)**	(0.8770)	(0.7521)
10 min	.	0.8260	0.8258	0.8276
(WN)	0.8273	**(0.0985)**	**(0.1560)**	(0.6050)
(AR-OF)	0.8260	(0.6114)	**(0.4465)**	(0.8090)
20 min	.	1.0712	1.0658	1.0638
(WN)	1.0711	(0.5392)	**(0.0235)**	**(0.0633)**
(AR-OF)	1.0713	**(0.3974)**	**(0.0202)**	**(0.0581)**
30 min	.	1.4269	1.4231	1.4236
(WN)	1.4202	(0.9402)	(0.7345)	(0.7259)
(AR-OF)	1.4275	**(0.3997)**	**(0.0936)**	**(0.1465)**
1 hr	.	1.7638	1.7635	1.7670
(WN)	1.7648	**(0.4074)**	**(0.4017)**	(0.6003)
(AR-OF)	1.7602	(0.9467)	(0.8237)	(0.7828)

Table 7: Results for the **out-of-sample** price change prediction on different sampling frequencies (Freq) for the standard (CAT-SOF), OANDA (CAT-OOF) and aggregated categories (ALL-CAT) order flow measures. The first cell entry is the Root-Mean-Squared-Prediction Error (RMSPE) of the associated forecasting model. The second and third cell entries in parenthesis are the p-value from the modified Diebold-Mariano (mDM) test with the null hypothesis that the RMSPE of the associated forecasting model is not smaller than the RMSPE of the corresponding White Noise or AR-OF benchmark model (WN, AR-OF). P-values in bold correspond to those cases where the RMSPE of the associated forecasting model is smaller than the RMSPE of the corresponding benchmark model (WN, AR-OF).

6 Are Retail Investors Right or Wrong?

In the previous section we showed with the help of predictive regressions and out-of-sample tests that the order flow of OANDA investors carries significant genuine

information for short term exchange rate forecasting. We showed (even after controlling for interbank order flow) that the information of the at least one period *lagged* category specific and aggregated order flow measures lead to significant forecasting improvements. This result on its own shows that the investors on OANDA generate through their trading behavior valuable exploitable information for forecasting that is not pure noise. It is, however, not clear how exactly this information is incorporated into the price process. Is it because investors place consistently correct orders in the sense that buying pressure predicts upward price movements, or is it because investors are consistently wrong and forecasting power is generated because their wrong orders are gratefully been picked up by the interbank traders and information is incorporated through this channel?

Given the fact that OANDA FXtrade possess certain features of a crossing network as discussed in Section 5 and especially limit order flows carry information from both OANDA FXTrade and the interbank market, it seems natural to address this question by distinguishing between order flow categories and forecasting horizons. In Table 8 we present cross-autocorrelations between lagged order flow (up to lag 5) and price changes for the category specific order flow and the combined order flow measures for all aggregation frequencies. For signing these order flow categories, we follow the same logic as in Section 5 where we discussed the construction of the combined OANDA order flow measure, that aims to measure the trading intention of the OANDA traders.[1] Hence, a positive cross-autocorrelation coefficient indicates correct information, while a negative one indicates wrong information for forecasting the future price change direction.

There are a number of interesting observations that we can make: Firstly, when traders open positions with market orders (category 4) they seem to be only right for a very short term horizon (up to 1 min) about the direction of the future price movement, but otherwise consistently wrong. When they, however, close an existing position with market orders (category 5) it seems that they possess the right information about price movements ranging between 80 minutes and 5 hours and leave the market early. A possible explanation for this observation is that when traders submit market orders to close a position, they already had an exposure to this currency pair and their incentives to monitor their investment could be higher

---

[1]Please, note that for the definition of the combined measure we made no assumptions about the intended forecasting horizon of the embedded information.

in comparison to when they simply open a new position. Secondly, when traders try to open a position by submitting limit orders (category 1) to the system they seem to be right for horizons up to around 10 minutes, for longer horizons the picture is mixed. Together with the first observation, it might well be that traders reveal their correct short term information not by placing market orders but limit orders.

The order flows discussed so far are clean in the sense that they are an immediate consequence of the actions of the OANDA traders themselves and are not directly triggered by events in the interbank market. It is already clear, however, that our traders are not consistently wrong, but also not consistently right and different order flow categories seem to reveal correct information for different forecasting horizons.

We now turn our attention to the order flow categories of executed limit orders. Executed limit orders to open a position (category 2) show that they are informative for the correct price movement only for short horizons around 3 minutes, for longer horizons they are consistently wrong. A similar pattern emerges for take profit orders (category 7) and the opposite pattern for stop-loss orders (category 6). For executed limited orders to close a position the pattern is quite mixed. Please note again that all executed limit order flows contain information about trading directions and intentions of both the interbank market and the OANDA traders and that they were constructed in such a way that they reveal the correct intentions of the OANDA traders. For all categories except the stop-loss category (category 6), where expectations of the interbank market and the OANDA traders are aligned, they measure by construction opposite trading intentions. Nonetheless, we observe that for short term forecasting horizons correct beliefs are revealed.

To sum up, we found that there is no straightforward answer to the general question whether traders on OANDA FXTrade are consistently right or wrong, but instead that the answer depends highly on the order flow category used and the forecasting horizon considered. There is a tendency in the above results that for short term forecasting horizons up to around 5 minutes OANDA traders seem to have the correct information and that for longer term forecasting horizons they seem to be wrong. This is also quite visible by looking at the aggregated OANDA order flow measure. These results are consistent with the results of the forecasting study presented in Sections 4 and 5 and show again that there is exploitable information among the set of OANDA FXTrade investors.

Freq	#lag	SOF	OOF	CAT1	CAT2	CAT3	CAT4	CAT5	CAT6	CAT7
1 min	1	0.0077	-0.1095	0.0050	-0.0596	-0.0246	0.0083	-0.0249	0.1323	-0.1138
.	2	-0.0179	0.0110	0.0040	0.0022	0.0001	-0.0223	0.0028	-0.0182	0.0276
.	3	-0.0086	0.0108	-0.0040	0.0049	0.0023	-0.0112	0.0035	-0.0161	0.0200
.	4	-0.0062	0.0087	0.0050	0.0048	0.0022	-0.0087	0.0024	-0.0117	0.0141
.	5	-0.0047	-0.0002	0.0069	-0.0003	-0.0014	-0.0063	-0.0013	-0.0019	0.0031
.	.	.	.	.	.	.	.	.	.	.
2 min	1	-0.0186	-0.0275	0.0032	-0.0334	-0.0113	-0.0290	-0.0053	0.0217	-0.0042
.	2	-0.0125	0.0132	0.0038	0.0108	0.0041	-0.0150	0.0055	-0.0167	0.0220
.	3	-0.0063	-0.0023	0.0104	-0.0001	-0.0051	-0.0053	-0.0033	0.0021	0.0012
.	4	-0.0008	-0.0042	-0.0014	-0.0069	-0.0013	-0.0014	0.0003	0.0060	-0.0020
.	5	-0.0014	-0.0011	-0.0001	-0.0007	-0.0001	-0.0041	0.0033	0.0032	-0.0025
.	.	.	.	.	.	.	.	.	.	.
3 min	1	-0.0228	-0.0086	0.0047	-0.0061	-0.0057	-0.0335	0.0002	-0.0012	0.0119
.	2	-0.0054	0.0018	0.0117	-0.0001	-0.0009	-0.0050	-0.0006	-0.0010	0.0041
.	3	-0.0045	-0.0063	0.0007	-0.0068	-0.0027	-0.0051	0.0003	0.0097	-0.0037
.	4	0.0035	0.0003	0.0009	0.0027	0.0001	-0.0023	0.0029	-0.0025	-0.0055
.	5	-0.0096	0.0003	0.0045	0.0115	0.0119	-0.0106	-0.0014	-0.0044	-0.0007
.	.	.	.	.	.	.	.	.	.	.
4 min	1	-0.0236	-0.0020	0.0095	-0.0088	-0.0028	-0.0323	0.0054	-0.0012	0.0192
.	2	0.0003	-0.0058	0.0075	-0.0052	-0.0084	-0.0007	-0.0000	0.0109	-0.0082
.	3	-0.0079	-0.0012	0.0023	0.0106	0.0042	-0.0087	-0.0040	-0.0061	-0.0022
.	4	-0.0046	-0.0092	0.0043	-0.0050	0.0096	-0.0083	-0.0020	0.0085	-0.0099
.	5	-0.0066	-0.0183	0.0033	-0.0100	0.0030	-0.0034	-0.0155	0.0110	-0.0179
.	.	.	.	.	.	.	.	.	.	.
5 min	1	-0.0173	-0.0072	0.0114	-0.0113	-0.0152	-0.0216	-0.0030	0.0012	0.0102
.	2	0.0018	-0.0032	0.0026	-0.0072	-0.0004	-0.0057	0.0027	0.0012	-0.0054
.	3	-0.0075	-0.0103	0.0039	0.0077	0.0086	-0.0092	-0.0053	0.0067	-0.0151
.	4	-0.0160	-0.0207	0.0016	-0.0151	0.0053	-0.0103	-0.0173	0.0127	-0.0113
.	5	0.0042	-0.0127	-0.0009	-0.0087	0.0046	0.0048	-0.0087	0.0098	-0.0169
.	.	.	.	.	.	.	.	.	.	.
10 min	1	-0.0023	-0.0153	0.0114	-0.0183	-0.0112	-0.0140	-0.0015	0.0116	-0.0114
.	2	-0.0191	-0.0271	-0.0000	-0.0131	0.0176	-0.0180	-0.0152	0.0188	-0.0241
.	3	-0.0123	-0.0323	-0.0030	-0.0205	-0.0035	-0.0023	-0.0242	0.0273	-0.0216
.	4	-0.0028	-0.0146	-0.0112	-0.0169	0.0032	-0.0220	0.0022	-0.0005	-0.0083
.	5	0.0033	-0.0104	0.0076	-0.0056	-0.0012	-0.0006	-0.0030	0.0119	-0.0147
.	.	.	.	.	.	.	.	.	.	.
20 min	1	-0.0150	-0.0475	0.0129	-0.0233	0.0063	-0.0172	-0.0216	0.0467	-0.0543
.	2	-0.0144	-0.0246	-0.0080	-0.0160	0.0166	-0.0235	-0.0075	0.0107	-0.0148
.	3	0.0034	-0.0160	0.0131	-0.0205	-0.0138	-0.0069	0.0015	0.0239	-0.0165
.	4	0.0041	-0.0029	-0.0259	0.0019	-0.0083	-0.0215	0.0221	0.0010	-0.0024
.	5	0.0000	0.0027	0.0053	-0.0014	-0.0049	-0.0197	0.0160	-0.0046	0.0012
.	.	.	.	.	.	.	.	.	.	.
30 min	1	-0.0421	-0.0550	-0.0039	-0.0424	0.0079	-0.0479	-0.0187	0.0386	-0.0275
.	2	-0.0033	-0.0163	-0.0002	-0.0087	-0.0041	-0.0139	0.0026	0.0185	-0.0144
.	3	0.0001	-0.0070	-0.0129	-0.0164	-0.0241	-0.0340	0.0271	0.0062	0.0016
.	4	0.0156	-0.0134	-0.0114	-0.0141	-0.0153	-0.0109	0.0115	0.0111	-0.0179
.	5	-0.0076	-0.0011	-0.0150	-0.0108	0.0082	-0.0248	0.0184	0.0003	0.0086
.	.	.	.	.	.	.	.	.	.	.
1 hr	1	-0.0157	-0.0339	0.0074	-0.0183	-0.0167	-0.0325	-0.0009	0.0311	-0.0257
.	2	0.0070	-0.0223	-0.0235	-0.0296	-0.0114	-0.0391	0.0345	0.0337	-0.0145
.	3	-0.0189	-0.0195	-0.0434	-0.0105	-0.0157	-0.0369	0.0162	0.0147	-0.0001
.	4	-0.0184	-0.0103	-0.0024	-0.0108	-0.0005	-0.0265	0.0086	0.0111	0.0012
.	5	-0.0272	-0.0145	-0.0243	-0.0138	-0.0036	-0.0417	0.0181	0.0146	0.0113

Table 8: Cross-correlations between the price change and the lagged order flow measures (standard (SOF), OANDA (OOF) and the 7 different category specific order flow) on different sampling frequencies (Freq) up to 5 lags. The cells highlighted in green indicate that positive significant cross-autocorrelation coefficients are significant, while cells highlighted in dark red indicate significant negative cross-autocorrelation coefficients.

7   Robustness Check

In order to check how robust our results are we reproduce in Appendix A the tables above for an alternative forecasting setup. Here our sample period of 32 weeks is not divided into 28 weeks of an in-sample estimation and 4 weeks of an out-of-sample forecasting period as before, but rather into 8 periods of 4 weeks each, where the first 3 weeks are always treated as the in-sample estimation period and the 4th week used as the out-of-sample forecasting period. The forecasting study described above is then carried out independently 8 times for each of the 4 weeks sub-samples. Table 9 summarises this setup of the in- and out-of-sample periods. We choose this forecasting setup with alternating in-sample and out-of-sample periods in order to account for potential structural breaks.

Period	In-Sample	Out-of-Sample
1	Mo. 6. Oct. 2003 – Fr. 24. Oct. 2003	Mo. 27. Oct. 2003 – Fr. 31. Oct. 2003
2	Mo. 3. Nov. 2003 – Fr. 21. Nov. 2003	Mo. 24. Nov. 2003 – Fr. 28. Nov. 2003
3	Mo. 1. Dec. 2003 – Fr. 19. Dec. 2003	Mo. 22. Dec. 2003 – Fr. 26. Dec. 2003
4	Mo. 29. Dec. 2003 – Fr. 26. Jan. 2004	Mo. 19. Jan. 2004 – Fr. 23. Jan. 2004
5	Mo. 26. Jan. 2004 – Fr. 13. Feb. 2004	Mo. 16. Feb. 2004 – Fr. 20. Feb. 2004
6	Mo. 23. Feb. 2004 – Fr. 12. Mar. 2004	Mo. 15. Mar. 2004 – Fr. 19. Mar. 2004
7	Mo. 22. Mar. 2004 – Fr. 9. Apr. 2004	Mo. 12. Apr. 2004 – Fr. 19. Apr. 2004
8	Mo. 19. Apr. 2004 – Fr. 7. May 2004	Mo. 10. May 2004 – Fr. 14. May 2004

Table 9: In-sample and out-of-sample periods of the forecasting study.

The results of the in-sample estimations are presented in Tables 10 and 12. As before, the RMSEs for the models containing the category k order flow measures are smaller than those of both benchmark specifications across all frequencies. Considering the p-values of the mDM-test, however, shows that as before especially the specifications relying on order flow of 6-Stop-loss orders and 7-Take-profit orders outperform the benchmark model up to the 1 hour sampling frequency. Nevertheless, it turns out that also the order flows of 1-Limit orders, 2-Limit orders executed (open), 4-Market orders (open) and 5-Market orders (close) show a significant forecasting improvement up to an hour frequency, while the order flow in all the other categories seems to be helpful only up to the 3 min frequency. When considering the RMSEs of our three combined order flow measure specifications (Table 12) we see that all three combinations outperform the (AR-OF) benchmark model up to a 1 hour frequency.

The same weaker pattern emerges here again for the the analysis of the out-of-sample RMSPEs presented in Tables 11 and 13, where the category specific and the combined order flow measures models outperform both benchmark specifications up to the 2 min frequency.

Nevertheless, the results obtained from this alternative forecasting setup corroborate our previous findings, that there is clearly *collectively* valuable information among FXTrade retail traders, that can be exploited to significantly improve short term out-of-sample price change forecasts. In addition, these results underpin the view that retail traders *on aggregate* might be more than only "noise" traders.

8 Conclusion

In this paper we provide evidence that the trading activity of small retail investors carries significant information beyond the information contained in interbank order flow and past price changes. This information is not only able to explain future currency prices changes, but forecast them for short forecasting horizons between 1 and 20 minutes. We don't claim that each and every retail investor possess such non-publicly available information, but that on aggregate and properly extracted even a crowd of small retail investors, generates enough additional information that can be exploited successfully for price change forecasts.

Our paper is the first which analyses the trading activity of retail foreign exchange market investors in such a detail and entirety. Although used in the academic literature mainly as a modelling device, our findings question the status of retail investors as pure noise traders in market microstructure models for FX markets. While this view certainly prevails for single or a crowd of only a few retail investors, a large crowd of them seems to possess enough information processing ability to generate and extract important additional information.

The results presented in this paper are in line with the view that in the foreign exchange market information is highly dispersed among market participants and contained in disaggregated form in customer order flow. While previous studies showed that interbank market dealers should be able to extract important information from their institutional clients order flow, our study shows that even retail customer order flow observed on an electronic trading platform is informative.

In general, our findings are consistent with a hierarchical market setup (Xue & Gencay (2012)) for the foreign exchange market in which dispersed information is processed from the bottom to the top through a system of fewer and fewer players (which are more and more informed) until it reaches the interbank market.

Moreover, our analysis also complies with the literature on investors sentiment (e.g., Baker & Wurgler (2006, 2007)) which provides evidence that widely dispersed news information can be successfully aggregated into sentiment indicators which have significant forecasting power for future price changes of individual companies.

References

BAKER, M. & J. WURGLER (2006): "Investor Sentiment and the Cross-Section of Stock Returns," *The Journal of Finance*, 61 (4), 1645–1680.

—— (2007): "Investor Sentiment in the Stock Market," *Journal of Economic Perspectives*, 21 (2), 129–152.

BARBER, B. M. & T. ODEAN (2000): "Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *The Journal of Finance*, 55, 773–806.

BARBER, B. M., T. ODEAN, & N. ZHU (2009): "Do Retail Trades Move Markets?" *Review of Financial Studies*, 22, 151–186.

BERGER, D., A. CHABOUD, S. CHERNENKO, E. HOWORKA, & J. WRIGHT (2008): "Order Flow and Exchange Rate Dynamics in Electronic Brokerage System Data," *Journal of International Economics*, 75, 93–109.

BJØNNES, G. H. & D. RIME (2005): "Dealer Behavior and Trading Systems in Foreign Exchange Markets," *Journal of Financial Economics*, 75, 571–605.

BLACK, F. (1986): "Noise," *The Journal of Finance*, 41, 529–543.

CHORDIA, T., R. ROLL, & A. SUBRAHMANYAM (2002): "Order Imbalance, Liquidity, and Market Returns," *Journal of Financial Economics*, 65, 111–130.

DANÍELSSON, J., J. LUO, & R. PAYNE (2012): "Exchange Rate Determination and Inter-Market Order Flow Effects," *The European Journal of Finance*, 18 (9-10), 823–840.

DORN, D., G. HUBERMAN, & P. SENGMUELLER (2008): "Correlated Trading and Returns," *The Journal of Finance*, 63, 885–920.

EVANS, M. D. & R. K. LYONS (2002a): "Informational Integration and FX Trading," *Journal of International Money and Finance*, 21, 807–831.

—— (2002b): "Order Flow and Exchange Rate Dynamics," *Journal of Political Economy*, 110, 170–180.

—— (2005): "Meese-Rogoff Redux: Micro-Based Exchange-Rate Forecasting," *American Economic Review*, 95, 405–414.

—— (2006): "Understanding Order Flow," *International Journal of Finance and Economics*, 11, 2–23.

FOUCAULT, T., D. SRAER, & D. THESMAR (2011): "Individual Investors and Volatility," *The Journal of Finance*, 66 (4), 1369–1406.

HARRIS, L. E. (2003): *Trading and Exchange*, Oxford University Press.

HARVEY, D., S. LEYBOURNE, & P. NEWBOLD (1997): "Testing the Equality of Prediction Mean Squared Errors," *International Journal of Forecasting*, 13, 281–291.

HVIDKJAER, S. (2008): "Small Trades and the Cross-Section of Stock Returns," *Review of Financial Studies*, 21 (3), 1123–1151.

KANIEL, R., S. LIU, G. SAAR, & S. TITMAN (2012): "Individual Investor Trading and Return Patterns around Earnings Announcements," *The Journal of Finance*, 67, 639–680.

KELLEY, E. K. & P. C. TETLOCK (2013): "How Wise are Crowds? Insights from Retail Orders and Stock Returns," *The Journal of Finance*, 68, 1229–1265.

LYONS, R. (1995): "Tests of Microstructural Hypotheses in the Foreign Exchange Market," *Journal of Financial Economics*, 39, 321–351.

MARSH, I. W. & T. MIAO (2012): "High Frequency Information Content in End-User Foreign Exchange Order Flows," *The European Journal of Finance*, 18 (9-10), 865–884.

MENKHOFF, L., L. SARNO, M. SCHMELING, & A. SCHRIMPF (2013): "Information Flows in Dark Markets: Dissecting Customer Currency Trades," The Journal of Finance, forthcoming.

NOLTE, I. (2012): "A Detailed Investigation of the Disposition Effect and Individual Trading Behavior: A Panel Survival Approach," *The European Journal of Finance*, 18(9-10), 885–919.

NOLTE, I. & S. NOLTE (2012): "How do Individual Investors Trade?" *The European Journal of Finance*, 18 (9-10), 921–947.

NOLTE, I. & V. VOEV (2011): "Trading Dynamics in the Foreign Exchange Market: A Latent Factor Panel Intensity Approach," *Journal of Financial Econometrics*, 9 (4), 685–716.

ODEAN, T. (1998): "Are investors reluctant to realize their losses?" *The Journal of Finance*, 53, 1775–1798.

OSLER, C. & V. VANDROVYCH (2009): "Hedge Funds and the Origins of Private Information in Currency Markets," Working Paper, Brandeis International Business School.

OSLER, C. L. (2005): "Stop-Loss Orders and Price Cascades in Currency Markets," *Journal of International Money and Finance*, 24, 219–241.

PAYNE, R. (2003): "Informed Trade in Spot Foreign Exchange Markets: An Empirical Investigation," *Journal of International Economics*, 61, 307–329.

SARNO, L. & M. P. TAYLOR (2001): "The Microstructure of the Foreign-Exchange Market: A Selective Survey of the Literature," Princeton Studies in International Economics, No. 89.

XUE, Y. & R. GENCAY (2012): "Hierarchical Information and the Rate of Information Diffusion," *Journal of Economic Dynamics & Control*, 36, 1372–1401.

A   Appendix

Freq	BM	CAT1	CAT2	CAT3	CAT4	CAT5	CAT6	CAT7
1 min	.	0.2605	0.2602	0.2604	0.2603	0.2603	0.2592	0.2592
(WN)	0.2646	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**
(AR-OF)	0.2611	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**
2 min	.	0.3913	0.3909	0.3915	0.3913	0.3913	0.3910	0.3913
(WN)	0.3924	**(0.0000)**	**(0.0011)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**
(AR-OF)	0.3917	**(0.0001)**	**(0.0301)**	**(0.0001)**	**(0.0000)**	**(0.0000)**	**(0.0077)**	**(0.0029)**
3 min	.	0.4763	0.4759	0.4763	0.4761	0.4764	0.4761	0.4762
(WN)	0.4771	**(0.0000)**	**(0.0472)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0001)**	**(0.0009)**
(AR-OF)	0.4765	**(0.0047)**	(0.1823)	**(0.0122)**	**(0.0088)**	(0.0682)	**(0.0316)**	(0.0723)
4 min	.	0.5481	0.5479	0.5481	0.5480	0.5481	0.5479	0.5478
(WN)	0.5491	**(0.0032)**	**(0.0015)**	**(0.0025)**	**(0.0018)**	**(0.0017)**	**(0.0054)**	**(0.0036)**
(AR-OF)	0.5483	**(0.0464)**	(0.0638)	(0.1158)	(0.0617)	(0.0829)	(0.1647)	(0.0984)
5 min	.	0.6098	0.6092	0.6098	0.6097	0.6097	0.6093	0.6093
(WN)	0.6113	**(0.0022)**	**(0.0703)**	**(0.0028)**	**(0.0091)**	**(0.0028)**	**(0.0080)**	**(0.0116)**
(AR-OF)	0.6099	(0.0572)	(0.2352)	(0.1467)	(0.1470)	(0.1490)	(0.0806)	(0.0777)
10 min	.	0.8481	0.8465	0.8476	0.8481	0.8476	0.8473	0.8469
(WN)	0.8500	**(0.0219)**	**(0.0798)**	**(0.0101)**	**(0.0264)**	**(0.0091)**	**(0.0048)**	**(0.0077)**
(AR-OF)	0.8483	(0.1432)	(0.1546)	**(0.0391)**	(0.2558)	**(0.0411)**	**(0.0317)**	**(0.0391)**
20 min	.	1.1669	1.1654	1.1671	1.1671	1.1659	1.1630	1.1634
(WN)	1.1712	**(0.0039)**	**(0.0041)**	**(0.0033)**	**(0.0101)**	**(0.0025)**	**(0.0013)**	**(0.0025)**
(AR-OF)	1.1676	(0.0717)	**(0.0474)**	(0.1043)	(0.2479)	**(0.0414)**	**(0.0061)**	**(0.0147)**
30 min	.	1.4428	1.4381	1.4438	1.4395	1.4424	1.4406	1.4438
(WN)	1.4491	**(0.0128)**	**(0.0277)**	**(0.0140)**	**(0.0015)**	**(0.0096)**	**(0.0040)**	**(0.0134)**
(AR-OF)	1.4446	(0.1077)	(0.0767)	(0.1634)	**(0.0066)**	**(0.0444)**	**(0.0117)**	(0.1440)
1 hr	.	2.0142	2.0129	2.0144	2.0125	2.0192	2.0152	2.0157
(WN)	2.0363	**(0.0052)**	**(0.0054)**	**(0.0050)**	**(0.0034)**	**(0.0124)**	**(0.0070)**	**(0.0072)**
(AR-OF)	2.0203	(0.0687)	**(0.0218)**	(0.0795)	**(0.0148)**	(0.1857)	(0.0542)	(0.0523)

Table 10: Results for the **in-sample** price change prediction on different sampling frequencies (Freq) for the 7 different category specific order flow measures for the alternative forecasting setup. The first cell entry is the Root-Mean-Squared-Prediction Error (RMSPE) of the associated forecasting model. The second and third cell entries in parenthesis are the p-value from the modified Diebold-Mariano (mDM) test with the null hypothesis that the RMSPE of the associated forecasting model is not smaller than the RMSPE of the corresponding White Noise or AR-OF benchmark model (WN, AR-OF). P-values in bold correspond to those cases where the RMSPE of the associated forecasting model is smaller than the RMSPE of the corresponding benchmark model (WN, AR-OF).

Freq	BM	CAT1	CAT2	CAT3	CAT4	CAT5	CAT6	CAT7
1 min	.	0.2436	0.2433	0.2436	0.2435	0.2433	0.2422	0.2422
(WN)	0.2464	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**
(AR-OF)	0.2441	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**
2 min	.	0.3588	0.35889	0.3586	0.3586	0.3587	0.3586	0.3586
(WN)	0.3590	**(0.2543)**	**(0.3205)**	**(0.0712)**	**(0.0959)**	**(0.1254)**	**(0.1254)**	**(0.0735)**
(AR-OF)	0.3588	(0.6108)	(0.5600)	**(0.0225)**	(0.1551)	(0.1972)	(0.2301)	(0.0804)
3 min	.	0.4374	0.43774	0.4373	0.4374	0.4373	0.4374	0.4374
(WN)	0.4375	**(0.4008)**	(0.7110)	**(0.3527)**	**(0.4495)**	**(0.3194)**	**(0.3677)**	**(0.4363)**
(AR-OF)	0.4372	(0.8866)	(0.9723)	(0.7971)	(0.7914)	(0.7987)	(0.7394)	(0.8226)
4 min	.	0.4965	0.49642	0.4965	0.4961	0.4964	0.4963	0.4966
(WN)	0.4971	**(0.0322)**	**(0.0469)**	**(0.1104)**	**(0.0050)**	**(0.0276)**	**(0.0101)**	**(0.1065)**
(AR-OF)	0.4963	(0.9148)	(0.6742)	(0.7538)	**(0.2241)**	(0.8127)	**(0.5030)**	(0.8524)
5 min	.	0.5472	0.54742	0.5469	0.5471	0.5471	0.5473	0.5467
(WN)	0.5461	(0.9490)	(0.9394)	(0.8942)	(0.8600)	(0.9162)	(0.9472)	(0.7670)
(AR-OF)	0.5470	(0.9522)	(0.8934)	**(0.4808)**	(0.6645)	(0.7854)	(0.8689)	**(0.2728)**
10 min	.	0.7684	0.76894	0.7697	0.7687	0.7684	0.7675	0.7691
(WN)	0.7664	(0.9773)	(0.9696)	(0.9972)	(0.9864)	(0.9767)	(0.8045)	(0.9712)
(AR-OF)	0.7682	(0.7047)	(0.7809)	(0.9791)	(0.9231)	(0.6703)	(0.1538)	(0.7985)
20 min	.	1.0746	1.07790	1.0757	1.0748	1.0775	1.0738	1.0752
(WN)	1.0697	(0.9837)	(0.9969)	(0.9868)	(0.9864)	(0.9992)	(0.8921)	(0.9543)
(AR-OF)	1.0742	(0.7477)	(0.9735)	(0.8592)	(0.7709)	(0.9993)	**(0.4475)**	(0.6676)
30 min	.	1.3006	1.30007	1.2967	1.2919	1.2969	1.2992	1.2993
(WN)	1.2958	(0.8778)	(0.7932)	(0.6064)	**(0.1978)**	(0.6112)	(0.8236)	(0.8054)
(AR-OF)	1.2940	(0.9936)	(0.9217)	(0.9641)	**(0.2377)**	(0.9326)	(0.9823)	(0.9759)
1 hr	.	1.8713	1.87467	1.8862	1.8703	1.8716	1.8737	1.8835
(WN)	1.8533	(0.9910)	(0.9598)	(0.9918)	(0.9343)	(0.9922)	(0.9926)	(0.9978)
(AR-OF)	1.8685	(0.6834)	(0.7228)	(0.9428)	(0.5797)	(0.9449)	(0.8918)	(0.9765)

Table 11: Results for the **out-of-sample** price change prediction on different sampling frequencies (Freq) for the 7 different category specific order flow measures for the alternative forecasting setup. The first cell entry is the Root-Mean-Squared-Prediction Error (RMSPE) of the associated forecasting model. The second and third cell entries in parenthesis are the p-value from the modified Diebold-Mariano (mDM) test with the null hypothesis that the RMSPE of the associated forecasting model is not smaller than the RMSPE of the corresponding White Noise or AR-OF benchmark model (WN, AR-OF). P-values in bold correspond to those cases where the RMSPE of the associated forecasting model is smaller than the RMSPE of the corresponding benchmark model (WN, AR-OF).

Freq	BM	SOF	OOF	ALL-CAT
1 min	.	0.2604	0.2592	0.2581
(WN)	0.2646	**(0.0000)**	**(0.0000)**	**(0.0000)**
(AR-OF)	0.2611	**(0.0000)**	**(0.0000)**	**(0.0000)**
2 min	.	0.3913	0.3910	0.3902
(WN)	0.3924	**(0.0000)**	**(0.0000)**	**(0.0020)**
(AR-OF)	0.3917	**(0.0028)**	**(0.0000)**	**(0.0188)**
3 min	.	0.4762	0.4762	0.4753
(WN)	0.4771	**(0.0000)**	**(0.0001)**	**(0.0307)**
(AR-OF)	0.4765	**(0.0043)**	**(0.0424)**	**(0.0928)**
4 min	.	0.5479	0.5479	0.5468
(WN)	0.5491	**(0.0009)**	**(0.0024)**	**(0.0004)**
(AR-OF)	0.5483	**(0.0374)**	**(0.1066)**	**(0.0077)**
5 min	.	0.6097	0.6095	0.6075
(WN)	0.6113	**(0.0015)**	**(0.0020)**	**(0.0365)**
(AR-OF)	0.6099	**(0.0915)**	**(0.0893)**	**(0.0745)**
10 min	.	0.8478	0.8472	0.8428
(WN)	0.8500	**(0.0123)**	**(0.0065)**	**(0.0073)**
(AR-OF)	0.8483	**(0.0888)**	**(0.0305)**	**(0.0071)**
20 min	.	1.1665	1.1631	1.1559
(WN)	1.1712	**(0.0158)**	**(0.0008)**	**(0.0001)**
(AR-OF)	1.1676	**(0.1641)**	**(0.0046)**	**(0.0002)**
30 min	.	1.4403	1.4400	1.4259
(WN)	1.4491	**(0.0045)**	**(0.0038)**	**(0.0006)**
(AR-OF)	1.4446	**(0.0218)**	**(0.0128)**	**(0.0007)**
1 hr	.	2.0145	2.0136	1.9851
(WN)	2.0363	**(0.0067)**	**(0.0053)**	**(0.0001)**
(AR-OF)	2.0203	**(0.0640)**	**(0.0451)**	**(0.0001)**

Table 12: Results for the **in-sample** price change prediction on different sampling frequencies (Freq) for the standard (CAT-SOF), OANDA (CAT-OOF) and aggregated categories (ALL-CAT) order flow measures for the alternative forecasting setup. The first cell entry is the Root-Mean-Squared-Prediction Error (RMSPE) of the associated forecasting model. The second and third cell entries in parenthesis are the p-value from the modified Diebold-Mariano (mDM) test with the null hypothesis that the RMSPE of the associated forecasting model is not smaller than the RMSPE of the corresponding White Noise or AR-OF benchmark model (WN, AR-OF). P-values in bold correspond to those cases where the RMSPE of the associated forecasting model is smaller than the RMSPE of the corresponding benchmark model (WN, AR-OF).

Freq	BM	SOF	OOF	ALL-CAT
1 min	.	0.2435	0.2418	0.2417
(WN)	0.2464	**(0.0000)**	**(0.0000)**	**(0.0000)**
(AR-OF)	0.2441	**(0.0000)**	**(0.0000)**	**(0.0000)**
2 min	.	0.3588	0.3584	0.3593
(WN)	0.3590	**(0.2092)**	**(0.0371)**	(0.7062)
(AR-OF)	0.3588	**(0.4548)**	**(0.0474)**	(0.8711)
3 min	.	0.4374	0.4376	0.4379
(WN)	0.4375	**(0.3929)**	(0.5998)	(0.7967)
(AR-OF)	0.4372	(0.7480)	(0.9128)	(0.9405)
4 min	.	0.4966	0.4964	0.4973
(WN)	0.4971	**(0.1173)**	**(0.0404)**	(0.5979)
(AR-OF)	0.4963	(0.8931)	(0.7214)	(0.9336)
5 min	.	0.5470	0.5468	0.5479
(WN)	0.5461	(0.8989)	(0.7999)	(0.9226)
(AR-OF)	0.5470	(0.6081)	**(0.2966)**	(0.8376)
10 min	.	0.7692	0.7681	0.7722
(WN)	0.7664	(0.9939)	(0.9105)	(0.9963)
(AR-OF)	0.7682	(0.9683)	**(0.4480)**	(0.9789)
20 min	.	1.0795	1.0763	1.0866
(WN)	1.0697	(0.9991)	(0.9825)	(0.9989)
(AR-OF)	1.0742	(0.9919)	(0.8450)	(0.9945)
30 min	.	1.3036	1.2976	1.3254
(WN)	1.2958	(0.9602)	(0.6517)	(0.9998)
(AR-OF)	1.2940	(0.9988)	(0.8684)	(1.0000)
1 hr	.	1.8807	1.8725	1.9173
(WN)	1.8533	(0.9925)	(0.9847)	(0.9979)
(AR-OF)	1.8685	(0.9295)	(0.7073)	(0.9902)

Table 13: Results for the **out-of-sample** price change prediction on different sampling frequencies (Freq) for the standard (CAT-SOF), OANDA (CAT-OOF) and aggregated categories (ALL-CAT) order flow measures for the alternative forecasting setup. The first cell entry is the Root-Mean-Squared-Prediction Error (RMSPE) of the associated forecasting model. The second and third cell entries in parenthesis are the p-value from the modified Diebold-Mariano (mDM) test with the null hypothesis that the RMSPE of the associated forecasting model is not smaller than the RMSPE of the corresponding White Noise or AR-OF benchmark model (WN, AR-OF). P-values in bold correspond to those cases where the RMSPE of the associated forecasting model is smaller than the RMSPE of the corresponding benchmark model (WN, AR-OF).