Expert Opinion

„The Impact of Payment for Order Flow on Market Quality"

by

Prof. Dr. Ralf Elsas[1]

Publicly certified expert („Sachverständiger") for „Securities and Derivatives, Exchanges" at the Chamber of Industry and Commerce (IHK München/Oberbayern) and Institute for Finance & Banking, LMU Munich School of Management, 80539 Munich, Germany

Prof. Dr. Lutz Johanning[2]

Chair of Empirical Capital Market Research, WHU – Otto Beisheim School of Management, Burgplatz 2, 56179 Vallendar, Germany

Prof. Dr. Erik Theissen[3]

Chair of Finance, University of Mannheim, Business School, L 9, 1–2, 68161 Mannheim, Germany

Expert opinion on behalf of Trade Republic Bank GmbH, Kastanienallee 32, 10435 Berlin, Germany

No. 2022_03

Germany, May 25, 2022

(Prof. Dr. Ralf Elsas) (Prof. Dr. Lutz Johanning) (Prof. Dr. Erik Theissen)

[1] Email: elsas@lmu.de, phone: +49 89 3303 6955.
[2] Email: Lutz.Johanning@whu.edu, phone: 261 6509 720
[3] Email: theissen@uni-mannheim.de, phone: +49 621 181 1518.

Inhaltsverzeichnis

Abstract

So-called neo-brokers like Trade Republic or Robinhood offer retail customers access to trading in capital markets at very low (or even zero) transaction costs, enabled by payment for order flow (PFOF). Because retail customers tend to be uninformed, their orders are less risky for market makers (or suppliers of liquidity in general) and thus cheaper to fill. Consequently, retail customers should pay lower spreads than institutions. Yet in traditional markets all traders are pooled and pay the same spread. Payment for order flow separates retail customer orders from the overall order flow and thus allows for price differentiation. The market maker (in regard of the low risk of retail orders) pays a fee to the broker. The broker, in turn, passes on some of the fee to the customer in the form of lower commissions. Critics argue, however, that payment for order flow harms liquidity and price discovery on public exchanges because it diverts retail order flow away from these venues.

To test the impact of payment for order flow on market quality, we conduct a unique field experiment in which payment for order flow for retail orders of one large neo-broker is switched off and on again for a randomly selected sample of stocks. This setting creates a truly exogenous shock to the amount of payment for order flow and thus enables *causal* inference on the implications for market quality of payment for order flow.

The investigation is carried out as a differences-in-differences analysis. Randomly selected from a group of stocks where the German neo-broker Trade Republic has a high market share in daily trade volume, we deliberately "switched off" payment for order flow for a treatment group of 10 stocks by routing retail orders directly to the main market Xetra as the treatment. On treatment days, this leads c.p. to an *increase* in trading in Xetra due to additional orders from investors who tend to be uninformed. The number of additional traded shares was in the millions and therefore is economically significant. This treatment was not applied to a control group of twelve similar stocks in the same period, so that we can compare market quality measures before, during and after treatment, for the treatment and control groups.

With this experimental set-up, we are able to test the hypotheses raised by PFOF critics that liquidity and informational efficiency of prices are systematically worse under a PFOF system than they would otherwise (i.e. without PFOF) be. The results of our empirical study do not support the arguments of the PFOF critics. Our main findings can be summarized as follows:

- all the measures of market quality examined (capturing both liquidity and price discovery) did not change in a statistically significant way on the treatment days, when compared to the control group,
- despite the high proportion of additional retail orders for treated shares during the experiment, trading in Xetra is not more liquid and informational efficiency is not higher as without the treatment.

The most important policy implication of our analysis is that a unilateral ban on payment for order flow, as currently discussed in the EU, is not warranted. Our results show that even large volumes of retail orders withdrawn by payment for order flow do not affect market quality. Also, in our theoretical analysis, we show that the existence of other arrangements which allow

for price differentiation between informed and uninformed traders (like internalization and dedicated retail exchanges) are fundamentally equivalent to payment for order flow, such that these arrangements should be treated equally in terms of regulation.

Finally, our analysis implies that the most crucial desirable regulatory instrument to achieve fair stock trading for all market participants in the EU is the creation of pre-trade price transparency about the trading opportunities for participants in all markets at any point in time, similar to the so-called National Best Bid and Offer price (NBBO) in the U.S.

1 Objectives

So-called neo-brokers like Trade Republic or Robinhood offer retail customers access to trading in capital markets at very low (or even zero) transaction costs. This is possible because neo-brokers exploit economies of scale through their web-based customer service and modern IT infrastructure. In addition, neo-brokers route their retail customers' orders to a limited number of specific brokers (wholesalers), who in return pay the neo-broker for the order flow. The wholesaler earns the bid-ask-spread. As retail trades are unlikely to be driven by information unknown to the market,[4] these trades can be exercised at lower bid-ask spreads as the wholesaler (or market makers) does not have to include an adverse selection cost into the bid-ask-spread. Recent studies indicate that this leads to significant reductions in direct and indirect costs of order executions for neo-brokers' retail customers.[5]

The rise of the neo-brokers has, however, renewed the debate about the impact of payment for order flow (PFOF) on market quality, as some parties allege that withdrawing retail orders from the primary exchange harms investors in general due to reduced market quality, traditionally measured by liquidity and price discovery. These allegations have triggered a regulatory debate which led, for example, the EU to make a proposal for an amendment to the EU directive 2014/65/EU ("Mifid") aiming at prohibiting payment for order flow in general. However, allegations related to a harmful impact of PFOF on market quality, are, so far, not based on empirical evidence. Such evidence would require observation and analysis of a counterfactual world to compare market quality with and without the withdrawal of retail customers' orders from the main market, which is difficult to achieve.

In this context, the German neo-broker Trade Republic mandated us to provide an independent expert opinion to assess the impact of payment for order flow on market quality. The expert opinion accordingly presents the results of this analysis. To our knowledge, our analysis is the first and only empirical study that can provide a causal analysis of the impact of PFOF on market quality based on state-of-the-art econometric methods and economic theory. In cooperation with Trade Republic, we have conducted a field experiment where the neo-broker did *not* route its retail customers' orders for a randomly selected group of stocks over a pre-specified time-period to a specific market place that pays for the order flow.[6] Instead, for these selected stocks and trading days, the orders were routed to the main exchange, in this case the electronic trading platform Xetra maintained by Deutsche Börse AG.

[4] See e.g. Barber/Odean (2000, 2008). Others, like Barrot/Karniel/Sraer (2016) or Boehmer et al. (2021), provide some evidence consistent with retail order flow containing private information in the U.S.

[5] Some recent studies show that retail customers benefit from neo-brokers through reduced transaction fees and significant price improvements to their orders (relative to a major trading venue or the National Best Bid and Offer NBBO in the U.S.), thus having lower direct and indirect costs of trading. See for example Meyer et al. (2021) or Kothari/So (2021).

[6] Under normal circumstances (i.e. without the experimental treatment), Trade Republic routes all its orders to the stock exchange LSX against payment for order flow. LSX stands for Lang & Schwarz Exchange, a segment of the Hamburg Stock Exchange. LSX is operated by Lang & Schwarz TradeCenter AG & Co. KG, a private brokerage firm. However, LSX trading is monitored by the governmental Trading Surveillance Office of the Hamburg Stock Exchange, to ensure high execution quality for retail investors. During the treatment days of the experiment, orders for treated stocks have still been transmitted to LSX by Trade Republic, whose market makers then routed these order directly to the Xetra trading system, without payment for order flow. See Section 4 for a test whether Trade Republic orders actually have been routed to Xetra or not.

This experimental design allows inferring the causal impact of PFOF on market quality:

- The treated group of stocks were selected from the group of stocks for which the neo-broker had an overall market share of trading of at least 2% and up to 5% of daily trade volume. About half of these stocks were treated (and therefore constitute the treatment group) while the other half was not (the control group). Thus, the treatment group is based on stocks where, without treatment, the market share of the broker is large enough. This ensures that an economically significant impact on market quality could be observed in the first place.
- Both the control group and the treatment group can be observed before, during and after the treatment. By "switching" payment for order flow on and off under controlled circumstances, comparing measures or market quality allows to infer the potential impact on market quality by estimating the average treatment effect. The comparison of the treatment group with the control group via a differences-in-differences (thereafter DiD) econometric analysis thus allows a causal assessment of the impact of PFOF on market quality.

The analysis is structured as follows. In Section 2 we discuss the economic characteristics of payment for order flow. Economically, it leads to a better price differentiation between informed and uninformed trading. Because a separation of uniformed from informed trading may reduce liquidity of the main exchange, we also theoretically derive the potential impact of this liquidity reduction and postulate corresponding hypotheses. The section also comprises a survey of previous empirical studies on payment for order flow.

In Section 3, we describe the market quality measures used to assess information processing and liquidity in stock markets which serve as our main variables to assess the impact of payment for order flow. Liquidity measures are the time-weighted relative spread, the tick-based effective spread, the tick-based 1-minute realized spread, and the tick-based 1-minute price impact. Price discovery measures are the absolute difference of one with the variance ratio of 5-minute over 1-minute mid-point log-returns, the 1-minute midpoint autocorrelation, and the 1-minute market return cross-correlation.

We then explain the methodology of a differences-in-differences analysis, describe the sample of stocks used in the experiment and provide key descriptive statistics. Our main experiment comprises 22 stocks, out of which 10 stocks were treated on five successive trading days. We can observe market quality measures before, during and after the treatment days, where the before and after period both comprise about two trading weeks each.

Finally, Section 3 shows the results from differences-in-differences analyses using panel regressions with stock and day fixed effects. For each of the market quality measures, one regression is conducted, testing for a causal impact of (the absence of) payment for order flow on treatment days.

Section 4 provides results from robustness tests. In particular, we test whether Trade Republic's customer orders were actually routed to Xetra instead of being handled on LSX against payment for order flow. Using a sample of 1,000 randomly drawn customer orders provided by the neo-broker, we were able to match about 95% of these orders to actual Xetra transactions in a 2-hour-window, confirming that the treatment in fact occurred.

Section 5 concludes and describes regulatory policy implications of our analysis.

2 Background and Related Literature

2.1 The Economics of Payment for Order Flow

Payment for order flow is a practice by which a broker receives customer orders and routes these orders to a market making firm for execution. The brokerage firm receives a reward (the "payment for order flow") for the orders it executes via the market maker. This reward, in turn, allows the brokerage firm to charge low or even zero commissions to its customers.

The market making firm usually executes the orders it receives in-house (i.e. against its own book) rather than routing them to an exchange. Consequently, payment for order flow reduces the on-exchange trading volume and increases the degree of fragmentation of the market. The market making firm executes buy orders [sell orders] it receives at the ask [bid] price and consequently earns the bid-ask spread.

An obvious question is if (and why) such an arrangement can be advantageous for the customer, the brokerage firm, and the market maker.[7] The main source of revenue is the bid-ask-spread earned by the market maker. Some of the spread revenue is used for the payment for order flow, a source of revenue for the broker. Finally, because of the payment for order flow the brokerage firm receives, it can offer low trading commissions to the customer.

Of course, if the market-making firm were free to choose the bid and ask prices it charged to customers, it could overcharge spreads, thereby generating profits at the expense of the customer. In practice, however, the bid and ask prices are constrained by the bid and ask prices of a reference market. In the case of Trade Republic, the reference market is Xetra.[8] The customer is guaranteed execution at prices which are no worse than the prices at which an equally sized order would execute on Xetra at the same time.

With this restriction in place, it is not obvious how the payment for order flow arrangement can generate profits. The bid-ask spreads on Xetra are a reflection of the cost of market making.

[7] Payment for order flow arrangements emerged in the US in a time when the minimum tick size in US equity markets was 12.5 cents (one eighths of a dollar). The bid-ask spread in such an environment can only be a multiple of 12.5 cents, i.e. 12.5 cents, 25 cents, etc. Now consider a stock with an equilibrium spread of 15 cents (for the determinants of the equilibrium spread, see further down in the main text). Because a spread of 12.5 cents would result in losses to the suppliers of liquidity, the spread in this market will be 25 cents, 10 cents above the equilibrium spread. Consequently, executing trades is profitable to suppliers of liquidity and they will be willing to pay for additional order flow. Because of the low minimum tick sizes in today's equity markets (the tick size for stocks trading at prices between € 50 and € 100 is between 0.5 and 0.01 cents, depending on the average daily number of transactions) we do not consider the minimum tick size to be a major reason for the profitability of payment for order flow arrangements.

[8] This holds for instruments that are tradable on Xetra.

Market microstructure theory has identified three cost components that are reflected in the bid-ask spread.

1. *Order processing cost*: Traders who supply liquidity face costs, such as staff expenses, IT costs, etc. These costs must be recovered through the spread revenue.

2. *Inventory holding costs*: Traders who supply liquidity to others do not have full control over the composition of the portfolio they hold. They have to sell when customers buy and they have to buy when customers sell. As a result, they will often hold a portfolio that is different (and more importantly, riskier) than their optimal portfolio. Risk averse traders will require a compensation for these risks. This compensation is integrated into the bid-ask spread.

3. *Adverse selection costs*: Some traders may possess better information on the value of a financial instrument than the liquidity suppliers. These traders will only trade when they expect a profit. Because trading is a zero-sum gain, the trader's profit is the loss of the liquidity supplier. Because liquidity suppliers cannot identify better informed traders,[9] they cannot avoid trading with them. What they *can* do, though, is to increase the bid-ask spread (i.e. to increase the ask price and/or to decrease the bid price). Increasing the spread will not avoid losses in trades with better informed traders, but it will generate profits in trades with uninformed traders. These profits cover the losses incurred when trading with informed counterparties.

   As a simple numerical example, consider a case in which the value of a financial instrument will be either 80 or 120, with equal probabilities. The trader population consists of 80% uninformed traders (who do not know whether the value is 80 or 120) and 20% informed traders (who do know whether the value is 80 or 120). Consider an uninformed market maker who wishes to set bid and ask prices for a single trade such that she breaks even (i.e. makes no losses in expectation). She faces no other costs. If the market maker would quote bid and ask prices of 100 each (the unconditional expectation of the asset value) she would suffer losses. There is a 20% chance that the trader she trades with is informed. If the asset value were 80 [120], that trader would sell [buy] at 100 an asset that is worth 80 [120].

   How can the market maker break even? Assume the next trader is a buyer. There is an 80% chance that the trader is uninformed. In this case the best guess of the asset value (even after the trade) is 100. However, there is a 20% chance that the trader is informed. In that case, the best guess of the asset value is 120. In expectation, conditional on the next trade being a customer buy, the best guess of the asset value is $0.8 \cdot 100 + 0.2 \cdot 120 = 104$. By the same logic, the best guess of the asset value conditional on the next trade being a customer sale is 96. Consequently, if the market maker quotes a bid price of 96 and an ask price of 104, she will always trade at a price that is equal to the best guess on the asset value conditional on the direction (buy or sell) of the trade. In

[9] Remember that trading in most equity markets is anonymous.

other words, the market maker will break even in expectation. In the resulting equilibrium, the bid-ask spread of 8 is entirely a compensation for losses to informed traders. The academic literature refers to this component of the spread as the adverse selection component.

For a payment for order flow arrangement to be profitable, it must be the case that the purchasing market maker has lower costs than the marginal liquidity supplier on Xetra. Otherwise, it would not be possible to pay for the order flow and still earn a profit while matching the bid and ask prices of the reference market.

It is possible that the purchasing market maker has lower order processing costs and/or is less risk averse and therefore has lower inventory holding costs. However, in these cases, the market making firm would not need the payment for order flow arrangement to make profits. Rather, it could, based on its lower costs, simply undercut the best bid and ask quotes on the reference market and earn a profit there.

Therefore, lower adverse selection costs are a more plausible reason for a cost advantage of the purchasing market maker. Neo-brokers using a payment for order flow business model usually target retail customers. It is well-documented that retail investors are less likely to possess superior information on the asset value than institutional investors.[10] Consequently, the adverse selection component of the bid-ask spread as defined above will be lower for the retail customer orders that are the subject of payment for order flow arrangements.

To see this, reconsider the example from above. Assume that the trader composition has 50% retail traders and 50% other traders (institutions). Assume further that the fraction of informed traders is 10% among retail traders and 30% among institutions. If there were separate markets for retail traders and institutions the equilibrium spread, by the logic introduced above, would be 4 (ask price 102, bid price 98) for the retail sub-market and 12 (106 and 94) for the institutional sub-market.

In reality, we do not have strictly separated markets for retail and institutional customers. However, as noted above, neo-brokers with a payment for order flow business model such as Trade Republic specialize in serving retail customers. To pursue with the example, assume that neo-brokers attract 10% of the retail customers while the remaining retail customers continue to trade in the reference market. The market maker purchasing the neo-broker order flow breaks even at a spread of 4 (102 and 98, as above). The reference market now has 52.63% (50/95) institutional customers and 47.37% (45/95) retail customers. The overall fraction of informed traders is $0.5263 \cdot 0.3 + 0.4737 \cdot 0.1 = 0.20526$. With this overall fraction of informed traders,

[10] See section 2.3 for more detail. We note that the fact that retail orders are typically both small and uninformed makes it unlikely that market makers front-run these orders. Front running means that someone knows about a pending customer order and expects that order to move prices. The front runner then trades ahead of the order, with the intention to close the position at a profit after the pending customer order has been executed and has moved prices. Leaving aside the fact that front running is usually illegal, the profitability of front running hinges on the price impact of the impending order. Orders that have the potential to move prices are either large or were submitted by traders who hold private information on the value of the asset. Neither is likely for retail customer orders. Of course, a large number of retail orders on the same side of the market could move prices. However, exploiting such a price impact requires the ability to forecast retail order flow. Just waiting for orders to accumulate is not feasible because the customer orders have to be executed immediately.

the equilibrium spread on the reference market is 8.21. It is thus cheaper to execute the purchased order flow than to execute orders on the reference market. It is this cost advantage that makes payment for order flow schemes viable.

The considerations of this sub-section can be summarized as follows.

1. Because of lower adverse selection risk, retail customer orders are cheaper to execute.
2. If retail customer orders are executed at the spread established in the reference market, the purchasing market maker earns a profit equal to the difference between the spread on the main market and the equilibrium spread in the retail sub-market (the profit is 4.21 in the example above). This profit can be distributed among the participants of the payment for order flow scheme. The purchasing market maker pays the broker for the order flow, and the broker passes on some of that payment to its customers in the form of lower trading commissions.
3. The practice of payment for order flow, by bypassing the reference market for a group of uninformed (or less well informed) traders, results in an increase in the fraction of informed traders, an increase in the bid-ask spread (in our example from 8.00 to 8.21), and a decrease in trading volume on the reference market.

The payment for order flow model described thus far corresponds to the business model of Trade Republic and is thus the relevant setting for our empirical analysis. We wish to note, though, that there are two alternative settings that have similar economic implications and therefore deserve, in our opinion, similar scrutiny from the regulator. These settings are (a) internalization and (b) the evolution of trading venues that cater to retail traders. We discuss both settings briefly.

The practice of internalizing customer orders is economically similar, if not equivalent, to payment for order flow. Internalization means that a brokerage firm receiving a (retail) customer order executes this order against its own book rather than routing it to a trading venue. This practice is exemplified by the following quote taken from the "Order Execution Policy" document of Deutsche Bank (available at https://www.db.com/legal-resources/oep-documents/Umbrella-Order-Execution-Policy.pdf, accessed April 27, 2022).

> 4.1 Internalisation of Transactions
>
> Unless instructed otherwise, in some cases Deutsche Bank may choose to “internalise” your Order by executing it in part or wholly from our own principal book. In circumstances in which Deutsche Bank internalises Client Orders, Deutsche Bank will act on its own behalf as a counterparty to the Client. In such circumstances, Deutsche Bank will treat its principal book as an Execution Venue and apply this Policy accordingly. This means that, instead of Deutsche Bank passing on indicative prices or quotes from third party brokers or from affiliates (as to which see section 4.2 below), it will instead communicate its own prices to you directly.

Obviously, the internalization of transactions in the way described here is similar to a payment for order flow business model, except that the neo-broker and the purchasing market maker are the same institution. Internalization can thus be thought of as a situation in which the broker and the executing market maker are a single firm. In this case, no explicit payment for order flow is required. However, the economic rationale for the arrangement and the potential incentive problems associated with it are similar to those in payment for order flow arrangements. Consequently, regulation should treat payment for order flow and internalization similarly. If payment for order flow is subjected to more stringent regulation than internalization (or if payment for order flow is prohibited) this will create incentives for neo-brokers and executing market makers to vertically integrate (i.e. to merge) and engage in internalization.

The creation of specialized trading venues for retail orders, from a fundamental economic perspective, also resembles payment for order flow arrangements. Consider a trading venue which, by appropriately selected institutional details, attracts retail traders but is not accessible or unattractive to institutions. Now, by the very same logic as in the adverse selection example above, the equilibrium spread in the retail venue would be 4 (as compared to 8 in a market were all investors are pooled). Now assume that (in analogy to the example above) such a venue attracts 10% of the retail orders. The main market now has more institutional (52.63%, as above) and fewer retail traders (47.37%). Consequently, the equilibrium bid-ask spread in the main market will increase from 8 to 8.21. Now let us return to the retail market. Most trading venues catering to retail customers operate as market maker markets.[11] Often there is only one market maker. That market maker may issue a price guarantee that ties its bid and ask prices to those of a reference market (typically Xetra). The outcome of such an arrangement is comparable to the outcome of a payment for order flow arrangement. The (retail) customer trades at a price that is not worse than that in the reference market, but is above the actual cost of servicing the customer (because retail customers tend to be uninformed). The difference between the spread that the customer pays and the equilibrium spread in the retail customer segment is the source of profits, to be shared between the venue operator, the market maker and the customer (who may benefit from lower exchange fees). The outcome is evidently similar to the outcome of a payment for order flow arrangement. Retail order flow is diverted away from the main market. Trading volume on the main market decreases, the proportion of informed traders and the equilibrium spread increase, and the degree of market fragmentation increases.

We conclude that a regulatory approach that singles out payment for order flow but leaves other economically similar institutional settings untouched may be inappropriate.

2.2 Arguments Against Payment for Order Flow

There are essentially two "lines of attack" against payment for order flow business models, one which focuses on the execution quality of those orders that are subject to payment for order flow, and one that focuses on market quality at large.

The first counter argument states that, while customers of neo-brokers typically pay low commissions, their total execution costs may nevertheless be higher because they may be trading at worse (bid and ask) prices.[12] The argument is based on the simple insight that there is an inherent conflict of interest between the customer and the purchasing market maker. While the customer obviously wants to trade at a low spread, the market maker earns the spread and thus

[11] See e.g. https://www.finanzfluss.de/geldanlage/boersen-unterschiede/ for details (accessed May 4, 2022).
[12] See e.g. Specht (2022).

earns higher revenue when the spread is high. As already noted above, the market making firm is not free in selecting the bid and ask prices it charges the customers. Rather, these prices are constrained by the prices in the reference market. Consequently, if the reference market is the most liquid market for the financial instrument under consideration, and if trades on the reference market are indeed executed at the quoted prices, the customer will not be hurt by a payment for order flow arrangement. This statement contains several "ifs" that deserve a discussion.

1. The selection of the reference market is crucial. For most domestic German stocks, Xetra is the most liquid market. However, for foreign stocks, Xetra may be less liquid than the home market of the stock.

2. The reference market only offers reference prices while it is open. Therefore, there is no obvious benchmark to judge the execution quality of trades executed outside of the trading hours of the reference market.

3. The reference market may offer execution at prices that are better than those on the trading screen. This may, for example, happen when a market order partially executes against the hidden part of an iceberg order.

4. Customers cannot supply liquidity by submitting limit orders, as they can in an open limit order book such as Xetra. While customers formally can submit limit orders, these limit orders do not compete with the market maker quotes. Rather, they are stored and execute only when the market maker's quotes change such that the stored limit order becomes executable.

The second counter argument states that payment for order flow, particularly when implemented in large scale, negatively affects the quality of the main market. This effect on market quality, in turn, has two dimensions:

5. Bid-ask spreads in the reference market may increase. This effect is illustrated in the "adverse selection costs" example in the previous subsection. If all orders are directed to the reference market the spread is 8. With payment for order flow in place, the spread increases to 8.21. The increased spread in the reference market may have repercussions on the retail orders that are subject to payment for order flow because the execution price guarantee relates to the (now higher) bid and ask prices in the reference market.

6. The informativeness of prices in the reference market is reduced because price discovery is based on fewer orders and fewer trades. Additionally, if spreads in the reference market increase, it may become unprofitable to trade on "small" pieces of information because the transaction costs may exceed the value of the information.

In what follows, we will discuss the six points listed above in more detail. The main issues we focus on are (a) whether payment for order flow is indeed disadvantageous and (b) which regulatory measures are appropriate to avoid execution at unfavorable prices and/or negative implications for market quality of payment for order flow.

Issue 1: Selection of reference market

It is undeniably true that the customer and the purchasing market maker have diverging interests with respect to price determination. A verifiable execution guarantee for the customer is thus required. Tying execution prices to the prices of a reference market can, in principle, achieve that objective. From the point of view of the customer, the reference market should be the most liquid market for the financial instrument under consideration at the time of order submission. Such a solution could be implemented easily if a consolidated tape was introduced that aggregates pre-trade information (i.e. bid and ask prices) from all trading venues to an "EU best bid-ask" (similar to the "national best bid and offer" NBBO in the United States).

If such an EU best bid-ask existed, it would be guaranteed to the customer that her order is executed at a price no worse than the best price available on *any* market for the financial instrument under consideration. That guarantee would be verifiable because the execution price of the customer order can be compared to the EU best bid or ask at the time of order submission. With such a guarantee in place there would be no need for a prohibition of payment for order flow.

Unfortunately, it appears that an EU best bid-ask is unlikely to be established without regulatory pressure. While the European Union is currently pushing for a "consolidated tape", current proposals concentrate on aggregating post-trade information. We are in favor of also aggregating and publishing pre-trade information and of speeding up the regulatory process.

We note that some trading venues are difficult to integrate into a consolidated tape. This is, for example, true for dark pools that match orders at the quote midpoint of a reference market. However, these venues are typically not accessible to retail customers and are therefore of limited relevance in the present context.

Issue 2: Execution outside of the trading hours on the reference market

If an EU best bid-ask as described above were used as reference market, there would always be a verifiable execution guarantee whenever *any* market is open. The customer would be guaranteed execution not worse than that at the best open market.

Of course, if *no* market is open there would be no benchmark price. However, in that case the neo-broker would be offering a service that is not available elsewhere. We would expect that execution quality for such off-hours trades is slightly worse than that for trades executed within normal trading hours. This is a natural outcome, however, because off-hours trades are associated with additional risks for the executing market maker. This is because (a) there is higher valuation uncertainty when no pricing information is available and (b) because the market making firm cannot hedge its position when all markets are closed and is thus exposed to the risk of price changes.

We do not see a point for a prohibition of payment for order flow here. Customers trading off-hours demand a service unavailable elsewhere. That service is (because of the increased risk mentioned above) more costly to the executing market maker than a standard trade during normal opening hours and thus warrants a higher price.

Issue 3: Hidden liquidity

By its very nature, hidden liquidity is not visible on the trading screens and therefore cannot be integrated into the benchmark price. While it may be the case that, because of the existence of hidden liquidity, a customer might have received a better price on the reference market than she receives from the neo-broker, we do not believe that this is an important issue. Only market orders whose size exceeds the depth of the best bid or ask price can be executed against the hidden part of an iceberg order. Retail customer orders placed with neo-brokers are rarely large. Meyer et al. (2021) have documented that the median size of orders executed by Trade Republic is € 459. We further note that a considerable fraction (21.1% as reported in Meyer et al. 2021) of trades executes at a price that is *better* than the price on the reference market.

Issue 4: Treatment of limit orders

It is true that customer limit orders submitted to Trade Republic, or to neo-brokers more generally, do not compete with the quotes of the purchasing market maker. This treatment of limit orders, however, is not specific to payment for order flow arrangements. Rather, it is specific to trading venues organized as market maker markets.[13] We note that several of the extant trading venues catering to retail traders operate as market maker markets.

At that broader level (i.e., with respect to market maker markets in general), one may ask whether traders are unduly disadvantaged by the treatment of limit orders in market maker markets. We do not think that this is the case as long as traders can make an informed choice. Market maker markets offer certain advantages (e.g. the high immediacy guaranteed by the presence of market makers and their obligation to quote firm prices) but also certain disadvantages. The impossibility for traders to supply liquidity via limit order and thus compete with the market maker is one such disadvantage. Traders who wish to supply liquidity may indeed be better served in an open limit order book while traders trading via market orders may prefer a market maker market.

Issue 5: Increased spreads on the reference market

While it is true that payment for order flow, if implemented at large scale, may result in reduced liquidity of the main market, it is much less clear that this is an undesirable result. To illustrate the argument, consider the following example. Assume barber shops charge the same price for

[13] In a market maker market, traders can only trade with market makers at the prices quoted by the market makers. To illustrate the point, consider a market maker market in which current bid and ask quotes are 40 and 41. Assume that a customer submits a limit order to buy with a price limit of 40.50. In an open limit order book, the publicly displayed bid price would now change to 40.50, and all traders could sell at that price. In a market maker market, the quotes would remain unchanged. The limit order would be held until a market maker quotes a bid price of 40.5 or higher.

all haircuts. Assume further that there are 50% long-haired customers and 50% short-haired customers. The cost to the barber of servicing a long-haired [short-haired] customer is 10 [6]. The break-even uniform price charged by barber shops is 8. Now assume one barber (call her the neo-barber) decides to only serve short-haired customers and charges them 6.50. The latter benefit from the price decrease (they pay 6.50 instead of 8) and the neo-barber also benefits from this situation (she earns a profit of 0.50). Assume that the neo-barber attracts 10% of the short-haired customers. Other barbers, if they continue to charge a uniform price to all customers, now have to increase their price to $\left(0.5 \cdot 10 + 0.45 \cdot 6\right) / 0.95 = 8.11$. The price increase from 8 to 8.11 reflects the fact that the other barbers now have more than 50% (52.6%, to be precise) long-haired customers, who are more expensive to serve. What we have achieved here is price discrimination. The short-haired customers who choose the neo-barber are low-cost customers who now benefit from a lower price. The other customers, as a group, are now more costly to serve and are, therefore, charged a higher price. For the long-haired customers, this is entirely justified because they are expensive to serve. The "losers" here are the short-haired customers who continue to go to the traditional barbers. They are now charged a higher price even though they are actually low-cost customers. However, these customers could also choose the neo-barber. If more neo-barbers emerge and all short-haired customers choose a neo-barber, we would obtain perfect price discrimination. Long-haired customers would be charged 10, reflecting the cost of servicing them, and short-haired customers would be charged 6.50. In fact, if many neo-barbers emerge, competition will likely reduce the price for a haircut to 6. Even though prices of traditional barbers have increased, from an economic perspective the new equilibrium is actually better than the original one[14] because prices now better reflect the cost of servicing different groups of customers.

The analogy to the adverse selection example discussed above is evident. Retail customers are low-cost customers because they are less likely to be informed. If they trade in the reference market, they are lumped together with institutional customers and pay a uniform price that exceeds the cost of servicing them. Neo-brokers single out retail customers and charge them a price that is closer to the cost that they cause. The remaining trader population in the reference market is now tilted towards informed traders, implying that the equilibrium spread in the reference market increases. It is inappropriate to conclude from the mere fact that the spread in the reference market has increased that the market outcome with payment for order flow is inferior. The higher spread in the reference market will simply reflect the fact that the average customer trading there is now more expensive to serve due to its potential information privilege. An improved price discrimination thus leads to a fairer pricing, i.e., reduced subsidies from uniformed to informed traders, which, of course, is also consistent with the regulators' desire to increase investor protection, especially for retail investors.

[14] Consider a short-haired customer whose willingness to pay for a hair cut is 7. In the uniform price case this customer would not get a hair cut even though she is willing to pay a price in excess of the cost of servicing her. The result is a welfare loss that is absent in the price discrimination equilibrium.

Issue 6: Less informative prices on the reference market

Prices in financial markets are informative because traders possess information and submit orders that reflect their information. Retail customers are rarely informed and therefore contribute little to price discovery. It is thus not obvious that removing retail order flow from the market will negatively affect the informativeness of prices. In fact, if retail traders behave like noise traders (as some of the empirical papers to be reviewed in the next section have argued), then removing them from the reference market may result in lower volatility and, consequently, less noisy, and more informative prices. Whether payment for order flow negative effects the informativeness of prices, or the degree of informational efficiency of prices, is an empirical question. We address it in section 3.

2.3 Prior Empirical Evidence

In this section we briefly review prior empirical research on several aspects relating to payment for order flow. We first discuss papers that analyze whether retail traders are indeed less well informed than institutional traders. We then discuss studies that analyze whether the existence of payment for order flow arrangements and / or internalization negatively affect overall market quality. Finally, we present the result of studies that ask whether retail customers benefit or lose when their orders are sold.

Are retail customers uninformed?

Easley et al. (1996) use a (then-new) empirical methodology to estimate the probability of informed trading. They compare order flow on the New York Stock Exchange and the Cincinnati Stock Exchange. The latter exchange was known for executing purchased order flow. The authors find that the probability of informed trading is lower on the Cincinnati Stock Exchange, consistent with purchased order flow being less informative. In a similar vein, Grammig/Theissen (2012) compare (retail customer) trades that were internalized via Xetra Best to trades executed in the Xetra order book and conclude that the internalized trades have lower information content, implying that the retail customers whose orders are internalized tend to be uninformed.

Foucault et al. (2011) analyze an institutional change in the French equity market (the abolishment of the "Règlement Mensuel" in 2000) that resulted in a reduction of retail trading for some (but not all) French stocks. When comparing the stocks that were affected by the rule change to those that were not, they find that return volatility of affected stocks as well as price impact of trades decrease. These results are supportive of the hypothesis that retail investors tend to be uninformed and rather behave as "noise traders".

Barber et al. (2009) provide more direct evidence. They analyze a very detailed data set from Taiwan and conclude (p. 609) that "[i]ndividual investor trading results in systematic and economically large losses". Two more recent papers paint a more differentiated picture. Boehmer et al. (2021) use data on marketable orders submitted by US retail investors and conclude (p. 2249) that there is "suggestive, but only suggestive, evidence that retail marketable orders might contain firm-level information that is not yet incorporated into prices". Coval et al. (2021) find evidence that investors in the top decile of their retail investor sample earn risk-adjusted returns of 6% per year. The data used in that study covers 1991-1996 and is thus not necessarily representative of today's market conditions. We wish to stress that none of our arguments was based on the assumption that all retail traders are uninformed. We just assumed that the *average* retail trader is less informed than the average institutional trader (remember the example in section 2.1).

<u>Are retail customers hurt when their orders are sold?</u>

Battalio et al. (2001) analyze the total trading costs of customers whose orders are sold to Knight Securities, a major Nasdaq dealer. They conclude (p. 54) that "trading with brokers taking payment for order flow is not unambiguously harmful to investors relative to trading through brokers not accepting such order-routing inducements."

Battalio et al. (2003) analyze the execution quality for retail customer market orders across different brokerage firms. They find that orders get better prices at the reference market than from a market making firm (Trimark) that is known to execute purchased order flow. However, after factoring in payments for order flow, brokers get better net prices from Trimark. If brokers pass on a sufficiently large fraction of the payment for order flow to their customers,15 these customers would not be worse off – but potentially be better off, when their order flow is sold.

Meyer et al. (2021) use data on executions in Trade Republic and match them to synchronous bid and ask quotes in Xetra. They find that the price offered by Trade Republic equals the corresponding best bid or ask quotes in Xetra in 78.1% of the cases. In 21.1% of the cases the customer got a better price from Trade Republic than she would have received in the Xetra order book. In 0.9% of the cases, Trade Republic customers received a price which was worse than the best price available in Xetra. On average, the Trade Republic customers benefitted from a price advantage of € 6.85 per 100 shares traded, or 5.2 basis points. This comparison does not yet take into account that Trade Republic customers are also likely to pay lower commissions than traders trading through traditional brokers. The study by Meyer et al. (2021) thus provides convincing evidence that Trade Republic customers trade at prices equal to, or better than, those available in Xetra.

In a study based on US data, Kothari et al. (2021) obtain similar results.[16] They conclude (p. 1) that "retail investors, and especially Robinhood customers, have enjoyed substantial price improvements on trades executed off-exchange and that off-exchange retail trades generally experience better execution quality than trades of similar sizes on public exchanges."

In the United Kingdom, payment for order flow has essentially been prohibited since 2012. In a study published by the CFR Institute, Rosov (2016) analyzes the execution quality of retail customer orders before and after that change. His criterion is the fraction of retail-sized orders

[15] The authors of the study did not have information on which fraction of the payment for order flow is passed on to customers via lower commissions or via other channels.
[16] We note that the authors of the study have a financial relationship with Robinhood, as disclosed in the first footnote of their paper.

executed at the best available prices on the London Stock Exchange (LSE) before and after the ban on payment for order flow. He observes (p. 1) "an increase in the proportion of retail-sized trades executing at best quoted prices between 2010 and 2014 from 65% to more than 90%". This study is sometimes cited as evidence of harmful effects of payment for order flow (e.g. as in Specht 2022). However, we think that this is a misinterpretation of the results in Rosov (2016) for several reasons.

- In terms of methodology, the study by Rosov (2016) is similar to Meyer et al. (2021). In particular, both studies compare execution prices to matched bid and ask prices available in a single reference market (the LSE in the case of Rosov (2016) and Xetra in the case of Meyer et al. (2021)). As laid out above, Meyer et al. (2021) find that 99.1% of the trades executed by Trade Republic execute at prices equal to, or better than, the corresponding Xetra quotes. This a much better result than the 90% that Rosov (2016) reports for the post-prohibition period. Thus, the execution quality offered by Trade Republic (measured relative to the reference market) is better than what retail customers in the UK obtained both before and after the prohibition of payment for order flow. Therefore, we do not think that the findings for the UK support a prohibition of payment for order flow in the European Union.
- Rosov (2016) does not analyze retail orders, but retail-sized orders. There is an important difference between these two concepts. Smart order routing systems often split large institutional orders into chunks (or "child orders") that are retail-sized. Thus, a sample of retail-sized orders will often contain sizeable fractions of institutional order flow.
- The UK operates a retail service provider model (RSP, described in some detail in Rosov 2016) where essentially several market makers compete for the execution of retail trades. There is no direct counterpart to the RSP in Europe. Therefore, prohibiting payment for order flow would not necessarily result in the same execution quality that UK retail customers receive - simply because the institutional setting in the EU is different from that in the UK.

Another frequently cited study is that of the Dutch regulator, the AFM (AFM 2022). This study uses a price-based method and a quote-based method to assess the execution quality of retail orders placed via payment for order flow arrangements.[17] The results of both methods are similar. The price-based method compares the execution price to the prices of trades executed on other (non-payment for order flow) trading venues in the same second. Consequently, the execution quality of a trade can only be assessed when trades on other markets have been executed in the same second, which often is not the case. For example, the study by the Spanish national

[17] A follow-up study by the Spanish national competent authority (CNMV 2022) adopts the price-based methodology developed in the AFM study and arrives at similar conclusions. The CNMV study uses data from a single broker and a single trading venue. The analysis is based on only 4,676 transactions executed during a six-months period in 50 Spanish stocks, amounting to less than one transaction per stock-day. In contrast, Meyer et al. (2021) analyze more than 2 million trades. Another follow-up study by the Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin 2022) finds more differentiated results. Based on a sample of more than 20 million trades the study finds that orders executed via payment for order flow receive similar or better execution than in the reference market in 63% (for non-DAX stocks) to 81% (for small trades in DAX stocks) of the cases. Correspondingly, the share of worse executions ranges between 19% and 37%, markedly lower than the shares reported in the AFM and CNMV studies.

competent authority (CNMV 2022) identifies 41,444 transactions executed via payment for order flow but finds trades on the reference market in the same second only for less than 5,000.

We focus on the quote-based analysis because it is conceptually similar to the methodology used by Meyer et al. (2021). The analysis identifies buy and sell orders in Dutch stocks executed via payment for order flow arrangements and compares the execution prices to the best bid and ask prices on Euronext Amsterdam, the most liquid market for Dutch stocks. They find (in Annex III) that orders executed via payment for order flow arrangements receive worse execution in more than 70% of the cases. This finding is in sharp contrast to the results of Meyer et al. (2021), who find worse execution in only 0.9% of the cases. We believe that there are two potential explanations for these conflicting findings. First, the AFM study compares execution prices to best bid and ask quotes without considering trade size.[18] Only trades the size of which does not exceed the volume available at the best quote can be executed at that price. To judge the execution quality of larger trades it is necessary to obtain order book information for the reference market and calculate the weighted average price at which the trade could be executed there (as Meyer et al. 2021 do). Second, payment for order flow is prohibited in the Netherlands. Consequently, the sample analyzed in the AFM study consists of trades in Dutch stocks executed via non-Dutch institutions. A non-Dutch neo-broker may be using a reference market other than Euronext Xetra. For example, a German retail trader may submit an order to buy Dutch stocks with a German neo-broker, and that broker may be using Xetra as its reference market. Prices for Dutch stocks on Xetra are likely to be worse than those on Euronext Amsterdam. Consequently, when the AFM study relates the execution price to quotes from Euronext Amsterdam, it will conclude that the customer received a bad price. To be very clear here: The conclusion is correct; the customer did receive a bad price. A neo-broker anywhere in the EU should guarantee execution at a price at least equal to that available in the most liquid market for the financial instrument under consideration. A neo-broker using Xetra as a reference market for trades in Dutch stocks does not meet that requirement. We nevertheless strongly believe that the results of the AFM study do not support a call for a ban on payment for order flow for two reasons. First, the study analyzes a sample that is not representative of the typical payment for order flow trade (namely, as noted above, trades in Dutch stocks executed via non-Dutch institutions). Second, the introduction of a EU best bid and ask as called for in section 2.2 would entirely eliminate the problem because it would require brokers anywhere in the EU to use the most liquid market (most likely Euronext Amsterdam in the case of Dutch stocks) as a reference market.

A final remark with respect to the AFM study and the follow-up studies by CNMV and BaFin is in place. These studies do not consider the lower brokerage commissions that neo-brokers often charge their customers. It is thus perfectly possible that these studies categorize an execution as "worse" than in other venues while the customer in fact pays lower total (i.e. cum-commission) fees.

[18] This argument extends to the price-based methodology also employed in the AFM study. That methodology compares prices of orders executed via payment for order flow to prices observed in the same second at other venues but does not consider the size of the trades. The BaFin study (BaFin 2022) differentiates between trade size categories and finds markedly different results in the different trade size brackets.

Does payment for order flow harm market quality?

Battalio (1997) analyzes how the start of large-scale payment for order flow activity by Bernard L. Madoff Investment Securities affected bid-ask spreads of stocks listed on the New York Stock Exchange. He does not find evidence of widened spreads and concludes (p. 341) that "the adverse selection problem associated with allowing agents to selectively execute orders in exchange-listed securities may be economically insignificant." Similarly, Battalio et al. (1997) conclude that the internalization of orders on the Boston and Cincinnati Stock Exchanges had little effect on quoted and effective spreads on the New York Stock Exchange.

Hasbrouck (1995) find that price discovery in the US equity markets takes place almost exclusively at the New York Stock Exchange (NYSE). As noted above, purchased order flow used to be executed at regional exchanges such as the Boston or Cincinnati Stock Exchange. The fact that these trading venues do not contribute to price discovery suggests that payment for order flow does not harm price discovery on the reference market. If the order flow directed to the Boston or Cincinnati Stock Exchange collectively does not contribute to price discovery - why would we expect that price discovery on the NYSE would improve if the order flow was redirected to the NYSE?

Overall, we do not find any previous empirical evidence supporting a negative impact of PFOF on market quality. However, none of these studies can control for potential problems of endogeneity and selectivity in their analysis, as they cannot rely on an experimental setting which allows to directly compare stocks with and without payment for order flow under the exact same economic setting. Our field experiment thus complements the available evidence, as the experiment comprises 22 stocks, out of which 10 randomly selected stocks were treated by “switching off” payment for order flow on five successive trading days, while a control group of 12 stocks was not treated at the same observation period. We therefore can observe market quality measures before, during and after the treatment days, for the treatment and the control group with and without payment for order flow, which allows causal inference about the impact of PFOF on market quality.

3 Empirical Analysis

3.1 Market Quality Measures

Our empirical analysis aims at evaluating the impact of payment for order flow arrangements on the liquidity and informational efficiency of the reference market. We rely on well-established measures of financial market quality, described in detail in the following paragraphs.

Liquidity Measures: The most widely used measure of financial market liquidity is the bid-ask spread. We use two measures of the bid-ask spread, the quoted and the effective spread. We further decompose the effective spread into the price impact and the realized spread.[19]

[19] For a recent application of these measures in a related context, see for example Comerton-Forde et al. (2019). For recent application to Xetra see Johann et al. (2018).

The quoted bid-ask spread is the difference between the prices at which investors can buy (ask price) and sell (bid price) a financial instrument. It is the cost if a "roundtrip trade", i.e. the amount an investor loses if she buys the asset at the ask price and then immediately sells it at the bid price. We express the *quoted bid-ask spread* as a percentage of the quote midpoint and calculate it from intraday data on best bid and ask quotes in Xetra. We then calculate a daily time-weighted average of the quoted spread for each stock in our sample.

The quoted spread is based on prices at which investors could have traded. It does not take into account whether trades actually took place. In contrast, the *effective spread* is a conditional measure that captures the execution costs of trades that actually occurred. It is defined as twice the distance between the price of the transaction observed at time τ, p_τ, and the quote midpoint in effect immediately prior to that transaction, m_τ, i.e.

$$\text{(1) Effective Spread: } \begin{cases} 2(p_\tau - m_\tau) \text{ for buyer-initiated trades} \\ 2(m_\tau - p_\tau) \text{ for seller-initiated trades} \end{cases}.$$

We express the effective spread as a percentage of the quote midpoint and use intraday data which is then aggregated to daily (equally-weighted and trade size-weighted) averages. To classify trades into buyer- or seller-initiated trades, we use the standard Lee/Ready (1991) algorithm. A trade is classified as buyer initiated [seller-initiated] if the transaction price is higher [lower] than the prevailing quote midpoint. If the trade occurs at a price equal to the quote midpoint the assignment follows the so-called tick test, according to which a trade is classified as buyer-initiated [seller-initiated] if the trade price is higher [lower] than the previous transaction price.

Because some traders trade on private information, trades might be informative. An informed trader will buy when she has information indicating that the current ask price is too low, and sell when the current bid price is too high. While suppliers of liquidity cannot generally identify informed trades (if they could, they would avoid them), they do know that there is a non-zero probability that trades were initiated by informed traders. Consequently, they will adjust bid and ask prices after a trade by an amount, which reflects an estimate of the new information conveyed by the trade. By that logic, we can use the change in the quoted midpoint after a trade to obtain an estimate of the information content of the trade.[20] The resulting measure is known as the *market or price impact* of a trade and is calculated as follows:

$$\text{(2) Price impact: } \begin{cases} 2(m_{\tau+1min} - m_\tau) \text{ for buyer-initiated trades} \\ 2(m_\tau - m_{\tau+1min}) \text{ for seller-initiated trades} \end{cases}.$$

The difference between the effective spread and the price impact is known as the *realized spread* and is a measure of the revenue earned by the suppliers of liquidity. It is calculated as follows:

$$\text{(3) Realized Spread: } \begin{cases} 2(p_\tau - m_{\tau+1min}) \text{ for buyer-initiated trades} \\ 2(m_{\tau+1min} - p_\tau) \text{ for seller-initiated trades} \end{cases}.$$

[20] We compare the quote midpoint immediately prior to the trade to the midpoint 1 minute after the trade. This choice is motivated by the empirical results of Conrad/Wahal (2020) who report that the majority of price impacts of trades in small-cap stocks is reflected in the quotes after 60 seconds.

To illustrate the effective spread and its decomposition into the price impact and the realized spread, consider the following example. A market maker believes that the value of a stock is 40.50 and quotes bid and ask prices of 40 and 41, respectively. Now a trader buys at 41. The effective spread on the trade is 1 [2*(41.00-40.50)]. Because that buyer may have been trading on private information, the market maker adjusts upwards her belief about the asset value, say to 40.65. Correspondingly, she changes her quotes to 40.15 and 41.15. If we scale up the quote change to a roundtrip trade, we get the price impact of 0.30. Effectively, the market maker has been selling an asset worth 40.65 at a price of 41. Thus, her actual revenue is 0.35 or, if we scale that up to a roundtrip trade, 0.70. The latter figure is the realized spread on the trade [2*(41-40.65]. The price impact of 0.30 [2*(40.65-40.50] is an estimate of the amount the market maker lost to informed traders, or put differently, an estimate of the adverse selection component of the spread.[21] It is also referred to as the permanent component of the bid-ask spread because it relates to a revaluation of the asset.

We express both price impacts and realized spreads as percentages of the quote midpoint, calculate both from intraday data and aggregate them to (equally-weighted and trade size-weighted) daily averages.

Price Discovery: The degree to which prices reflect the value of financial assets, and the speed at which changes in the value of an asset are incorporated into its price, is another important characteristic of a financial market. While the true value of an asset is unpredictable, we do know that price changes are unpredictable if prices correctly reflect all available information (Samuelson 1973, Fama 1970). A market in which that is the case is referred to as an informationally efficient market. The random walk model is a statistical model that describes such a market. If transaction prices follow a random walk it holds that:

1. the variance of returns over a long interval (say $k{*}t$ minutes) is k times the variance of returns over the short t-minute interval. Correspondingly, the variance ratio $\frac{\sigma_{kt}^2}{k\sigma_t^2}$ should be equal to one.
2. Successive price changes are not serially correlated.

From these characteristics we derive our first two measures of informational efficiency. The first is the absolute deviation of the variance ratio from one. To compute it we set the long [short] interval to 5 minutes [1 minute] and then calculate:

$$(1)\ \ Variance\ Ratio\colon \left|1-\frac{\sigma_{5min:i,\tau}^2}{5\sigma_{1min:i,\tau}^2}\right|,$$

where the variances are estimated from quoted midpoint (log-)returns.

[21] To see this, assume that the market maker sells at a price equal to the quote midpoint, i.e. 40.50. After the trade the market maker, as described earlier, adjusts upward her belief about the asset value to 40.65. She thus sold a stock with a post-trade estimated value of 40.65 for 40.5, resulting in a 0.15 loss. Scaling up the loss to a roundtrip trade gives the adverse selection component of 0.30.

The second measure is simply the absolute value of the autocorrelation of quote midpoint returns measured at 1-minute time-intervals. Deviations from zero indicate deviations of the price process from a random walk.

Our third measure of informational efficiency is the cross-correlation between the return of a 1-minute-lagged market portfolio proxy and the (quoted-midpoint-based) stock return. In an informationally efficient market, stock prices should adjust instantaneously to observable market-level activity. A non-zero correlation between current stock-level returns and lagged market-level returns is evidence of delayed adjustment of prices to new information.

3.2 Hypotheses on the Impact of the Experiment

The field experiment we conduct is intended to assess the impact of payment for order flow and the resulting partial market segmentation on financial market quality. The theoretical discussion in the previous section has shown that payment for order flow may result in increased (quoted and effective) bid-ask spreads on the reference market. In our experiment, the treatment is the disappearance of payment for order flow during the treatment period. If payment for order flow indeed reduces the liquidity on the reference market, one should expect to find lower Xetra spreads for treated stocks during the treatment period due to the increased liquidity by the experiment.

> H1: *Quoted spreads on Xetra decrease for treated stocks during the treatment period.*
>
> H2: *Effective spreads on Xetra decrease for treated stocks during the treatment period.*

We have argued that the retail orders subject to payment for order flow arrangements are predominantly uninformed. Thus, during the treatment there are more uninformed orders executed on Xetra. Consequently, price impacts should decrease.

> H3: *Market impacts on Xetra decrease for treated stocks during the treatment period.*

We do not expect an effect on realized spreads of the treatment (but nevertheless test whether such an effect exists).

Critics of payment for order flow arrangements argue that payment for order flow results in reduced informational efficiency. Although given our discussion in Section 2.1 this claim may not be true, we will empirically test the corresponding hypothesis H4 (using the three measures of informational efficiency introduced in section 3.1 above).

> H4: *Informational efficiency improves for the treatment stocks during the treatment period.*

3.3 Sampling Design and Selection Criteria

The selection of our treatment and control groups was implemented in two stages. We first conducted a pilot study (test sample), and learning from the resulting experience, designed the main experiment (main sample).

The general idea was to sample stocks for which the neo-broker has an economically significant market share in daily trading volume, so that a significant trade volume will be routed to Xetra on treatment days and potential impacts on market quality will be observable.[22] Therefore, only stocks representing at least 2% of the daily trade volume by the neo-broker have been selected. This share was determined based on the daily average trade volume participation in the calendar year until January 31, 2022 and identified 59 stocks traded on Xetra.

In the main experiment of our study, stocks were randomly selected by the authors, according to the filter rule that the market share of the neo-broker has to be in the range of 2% and 5% of daily trading volume,[23] and quote updates occur in at least 80% of updates of the DAX index proxy.[24] This main sample comprises 22 stocks, out of which 10 were randomly selected for the treatment of order re-routing to Xetra. Treatment occurred from April 8 - April 14, 2022 (five trading days) for all 10 stocks of the treatment group.

Table 1 provides general descriptive statistics for the stocks included in the pilot study and the main experiment.

Table 1: Descriptive statistics on the sample firms

Note: The table shows pre-experiment characteristics of the firms that were used to select the treatment and control group. All measures are taken up until January 31, 2022. The experiment comprises overall a sampling period (before, during and after treatments) from February 01, - April 28, 2022. The main sample is analyzed in the period from March 21 – April 28, 2022.

	Market Capitalization as of 01/31/2022 [€ Mill.]	**Average daily trading volume (6 month until 31.01.2022)**	**Share of neo-broker in daily trading volume (average, 1 year until 01/31/2022)**	**Market Capitalization as of 01/31/2022 [€ Mill.]**	**Average daily trading volume (6 month until 31.01.2022)**	**Share of neo-broker in daily trading volume (average, 1 year until 01/31/2022)**
	Untreated			**Treated**		
Main Sample of the Fiueld Experiment						
No of Stocks	12			10		
Mean	5,005	430,099	2.73%	11,042	306,824	2.76%
Median	1,145	129,809	2.57%	2,230	193,619	2.43%
Std.	13,108	731,162	0.63%	27,276	374,817	0.81%
Min	428	8,235	2.07%	754	5,780	2.02%
Max	46,576	2,241,957	3.83%	88,592	1,285,213	4.32%

[22] Xetra trading hours start at 9:00 o'clock UTC+1 and end at 17:30 UTC+1. For the main experiment, summer time was started on March 27, such that trading hours then are at UTC+2. Except for the opening auction, a potential noon auction in cases of high voilatility, and the closing auction at about 17:36 each day, trading in the Xetra system is continuous.

[23] The 5% cap on Trade Republic's participation rate was chosen because the pilot study showed that stocks with participation rates of the neo-broker above 5% are very illiquid, such that only few trades and only few quote updates occurred on Xetra. Results on the pilot study for a subsample of 12 eligible stocks (7 treated) are documented in Appendix A.

[24] This amounts on average to about 1.4 quote updates per second. However, many quote updates just occur due to changes in offered quantities of stocks in one of 5 observed levels of the order book, i.e. mid-quote and spread remain constant in most cases. Spread-updates occur much less frequently, thereby preventing us from using shorter time resolutions than the 5-min/1-min framework described in Section 3.2.

3.4 Differences-in-Differences Analysis

3.4.1 Methodology

A differences-in-differences (DiD) analysis serves to test the causal impact of an exogenous shock on a treatment group versus a control group, assumed untreated. Optimally, and this is the case in a controlled experiment i.e. in a setting where the researcher "designs" the experiment, the exogenous shock is ensured to be truly exogenous. Also, achieving a true random selection of the individuals entering the two groups is easier in an experimental setting – as opposed to a DiD analysis implemented with observational data (so-called "natural experiments"). Random selection of the treatment and control groups assures that the (unobserved) characteristics of subjects do not interfere with the outcome of interest. Such interference, in turn, could invalidate the results of the analysis.

The DiD approach to make causal inference has been long established and is in particular ubiquitous in medical research.[25] For example, to check if a certain drug has a positive impact on patients' health, researchers will give a drug to a random sample of patients (treated patients), and placebo pills to the others (untreated patients, or control group), and check how the medical status of patients evolve over time after the treatment. The difference in the average change of some health measures between the two groups of patients due to the drug is referred to as the "average treatment effect" (ATE). Provided that the assignment of the actual drug versus placebo among patients is random, the ATE is an estimate of the causal effect of the drug on average.

In its most simple form, a regression measuring the causal impact of a treatment has the form as shown in equation (2):

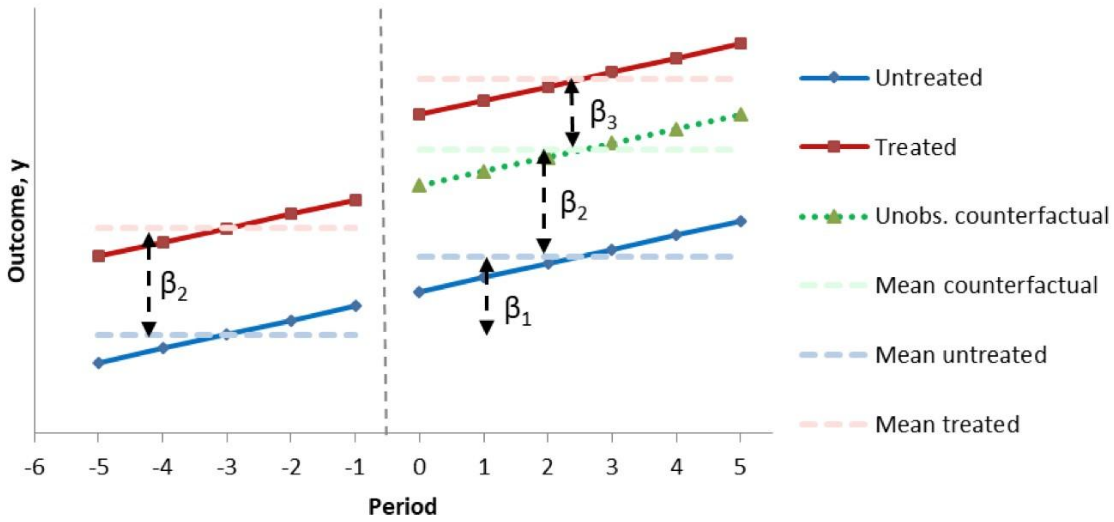
$$\text{(2)} \qquad \text{outcome}_{\text{i,t}} = \beta_0 + \beta_1 \text{shock}_{\text{t}} + \beta_2 \text{treated}_{\text{i}} + \beta_3 (\text{shock}_{\text{t}} \text{ x treated}_{\text{i}}) + \text{u}_{\text{i,t}}$$

Where $outcome_{i,t}$ is the measurement of interest for individual *i* at time *t*, $shock_t$ is an indicator variable, taking the value of one if a shock is prevalent at time *t* (and zero otherwise), and $treated_i$ is an indicator variable indicating whether the individual i have received the treatment at some point in time (and zero otherwise). $Shock_t \ x \ treated_i$ is equal to one on days where treatment firms are being treated, i.e., in our context, when their orders are routed to Xetra instead of LSX.

The coefficients β_0, β_1, β_2 and β_3 need to be estimated over individuals and time. In terms of interpretation, β_0 measures the average outcome for the control group (the intercept of the curve), β_1 captures the change resulting from the shock for the control group. β_2 is the difference in the outcome levels observed between the control group and the treatment group, before the the shock and in the counterfactual world where the treatment has no impact. β_3 is the average treatment effect, i.e. the difference between control group and treatment group due to the shock.

[25] The first application of a DiD analysis is usually attributed to Snow (1855), who showed that the transmission channel of cholera works via water instead of air, as at those times was commonly believed. A well-known early example of an application of DiD analysis in economic research is Card/Krueger (1994), analyzed the impact of minimum wages on employment by comparing two neighboring U.S. states before and after a minimum-wage increase.

Figure 1: Visual interpretation of the coefficients of the DiD model with trends[26]



In the context of this expert opinion, the individuals are the stocks, comprising stocks that are never treated (the control group), and the ones that were treated (the treatment group). The treatment group is treated in the sense that retail customers' orders are not sold against payment for order flow to a broker but instead are routed to the main exchange Xetra (*XetraRouting*). Note that we observe treated stocks and the control group before, during and after the shock (i.e. the Xetra routing), which is the setting needed to implement the DiD approach.

As the data are panel data (i.e. consisting of individuals repeatedly observed over time), we can estimate a two-way fixed effects panel model, with fixed effects for individuals (stocks) and time (trading days). The benefit of this model is that the individual fixed effects automatically control for all observable *and* unobservable time-invariant characteristics of the stocks. Given the short time period of observation of our experiment of only about three months, the individual fixed-effects can control for the average market capitalization of the firms, the stock price level, the industry of firms, etc. Similarly, the time fixed effects control for all observable and unobservable effects that are the same across individuals but vary over time, like the market development, the business cycle, interest rates, and all other macroeconomic factors.

Equation (3) shows the regression model using the two-way fixed effect specification, where μ_i are individual fixed effects, δ_t the time fixed effects and $u_{i,t}$, is a general iid error term. The outcome variable varies between the regressions. It is one of our measures of market liquidity or price discovery, as introduced in Section 3.1. The coefficient β_3 on the interaction term *XetraRouting x Treated* is the estimated average treatment effect ATE.[27] We do not need to estimate coefficients β_1 or β_2 as presented in the previous equation because these effects are captured by the fixed effects for individuals and time periods.

$$(3) \qquad outcome_{i,t} = \beta_0 + \beta_3(XetraRouting_t \, x \, Treated_i) + \mu_i + \delta_t + u_{i,t}$$

[26] The figure is taken from Schiozer et al. (2020), their Figure 1B.
[27] We include individual (stocks) and time (trading days) fixed effects and cluster standard errors for individuals and time as well.

3.4.2 Descriptive Statistics and Parallel Trends Assumption

Our main sample consists of 10 randomly selected stocks in the treatment group with high liquidity and a market share of the neo-broker between 2%-5%. The control group comprises 12 stocks that met the same selection criteria as the treatment group. Table 2 shows averages of trade statistics in the sample period of the experiment. In particular, it allows for a comparison of average trade participation of the neo-broker during the experiment's sample period with its average trade participation in the calendar year before the selection of stocks for the experiment. Also, for both time periods, averages are shown for the treatment and control group of stocks.

The table illustrates that the neo-broker's trade participations rates are comparable before and during the experiment, and for the control and treatment group. About 3.18 million shares have been traded in the 27-day sample period due to Trade Republic's customer orders, and about 2.75 million shares in the stocks of the control group. The experiment thus analyzes groups of stocks with high trade participation of Trade Republic and a significant amount of absolute trade volume due to the neo-broker. Hence, if payment for order flow has a market quality impact, the impact should be measurable, in particular for the analyzed group of stocks in the experiment.

Table 2: Trade Statistics in the Sample Period

Note: The table shows averages of the trade participation of the neo-broker in the sample period of the main experiment form March 21-Apiril 28, 2022, differentiated between treatment and control groups of stocks. The trade participation on non-treatment days is neo-broker's shares divided by (Xetra volume + neo-broker volume), and on treatment days neo-broker's shares divided by Xetra volume. Median daily trade value and median daily shares traded per stock are provided as well.

.	Treatment Group (10 stocks)	Control Group (12 stocks)
Trade Participation Neo-Broker (at selection)	2.59%	2.54%
Trade Participation Neo-Broker (sample period experiment)	2.58%	2.82%
Median Daily Trade Value per stock [€]	334,194	130,061
Median Daily Shares Traded per stock	5,340	5,013
Traded Shares from neo-broker's customer orders (in million]	3.18	2.75

Table 3 shows descriptive statistics for the main sample with regard to the liquidity and price discovery measures used to analyze market quality. The table contains two panels, which show averages of liquidity and price discovery measures differentiated between the period where no treatment occurred (Panel A) and the period with the actual treatment (Panel B), i.e. where orders for the treatment group were routed to Xetra instead of being internalized by LSX against payment for order flow. Panel C shows p-values for significance tests of differences in the median (using a Wilcoxon rank sum test) between different groups of comparison.

Significant differences between the treatment and control group medians occur for some quality measures, both in the periods before and during the treatment. Note that this cannot be interpreted as the result of the experiment. This does not inform us on the impact of payment for order flow because all observations get simply pooled over time and then compared to each other. In addition, the univariate significance tests do not control for other factors, like the individual stock's sensitivity to macro variables (e.g. interest rates, market volatility etc.) or stock characteristics (market capitalization, industry, transparency etc.). Only the differences-in-differences regressions in the next section control for all of these effects simultaneously

Hence, the univariate level comparisons just indicate that treatment and control group are on different median levels for some variables (*Autocorrelation*, *Cross Correlation*, *Realized Spread*, *Price Impact* and *Effective Spread*). Significantly different measures outside the treatment days also remain significantly different during the treatment (except for the cross-correlation measure for the index), thereby not indicating any change. The significant differences also do not violate required assumptions of the DiD analysis, because what matters here is whether treatment and control group have parallel trends (and not levels), and whether they change differently in the treatment.

Table 3: Descriptive Statistics Main Experiment

Note: The table shows averages of measures for stock market price discovery and liquidity for our main sample of analysis. The main sample consists of 10 randomly selected stocks with high liquidity and a market share of the no-broker between 2%-5%. The control group comprises 12 stocks that met the same selection criteria as the treatment group. The sample period is from March 21 – April 28, 2022. For treated stocks, in the treatment period from April 08 – April 14, 2022, the neo-brokers' customer orders were routed to the Xetra trading system, instead of being sold to the neo-brokers' trading venue (LSX) against payment for order flow. Price discovery measures are the absolute difference of one with the variance ratio of 5-minute over 1-minute mid-point log-returns, the 1-minute midpoint autocorrelation, and the 1-minute market return cross-correlation. Liquidity measures are the time-weighted relative spread, the tick-based effective spread, the tick-based 1-minute realized spread, and the tick-based 1-minute price impact.
Panel A of the table shows statistics for the market quality measures in non-treatment times (from March 21 – April 07, 2022, before the treatment, and from April 19 – April 28, 2022, after the treatment). Panel B show statistics during the treatment period from April 8 – April 14, 2022, for treated and untreated stocks. Note that there was no trading from April 15 – April 18, 2022 due to Easter bank holidays in Germany. Panel C shows results of Wilcoxon rank sum significance tests for differences in medians between treated and untreated stocks.

Column	Group of Stocks	Variance Ratio	Auto-Correlation	Cross-Correlation Index (lagged)	Relative Spread	Price Impact	Realized Spread	Effective Spread	Number of Observations
		Price Discovery			Liquidity				
		Panel A: Non-Treatment Days (Payment for Order Flow)							
1	Benchmark group: Not treated	18.882%	11.958%	9.355%	0.244%	0.355%	0.008%	0.361%	264
2	Treated	17.393%	10.579%	8.111%	0.217%	0.317%	0.013%	0.326%	220
		Panel B: Treatment Days (Order Routing to Xetra, no PFOF for treated stocks)							
3	Benchmark group: Not treated	19.607%	15.495%	9.228%	0.231%	0.398%	-0.025%	0.371%	60
4	Treated while routed to Xetra	17.359%	12.417%	7.706%	0.200%	0.292%	0.011%	0.301%	50
		Panel C: Wilcoxon rank sum test for median equality							
1 vs 2	treated vs non-treated, non-treatment period	34.611%	3.195%	2.342%	0.166%	0.027%	17.720%	0.159%	---
3 vs 4	treated vs non-treated, treatment period	53.835%	22.413%	24.056%	4.593%	2.696%	11.233%	4.528%	---

The parallel trends assumption is one fundamental assumption of DiD analysis, however. Therefore, Figure 1 and Figure 3 show two graphs plotting smoothed means of two key measures over time (*Variance Ratio* for price discovery and *Effective Spread* for liquidity).[28] The vertical dashed lines in the graphs show the beginning and the end of the treatment period, the lines with circle-indicators show the control group, lines with "x"-indicators show the treatment group.

Figure 2: Parallel Trends Graph: Variance Ratio

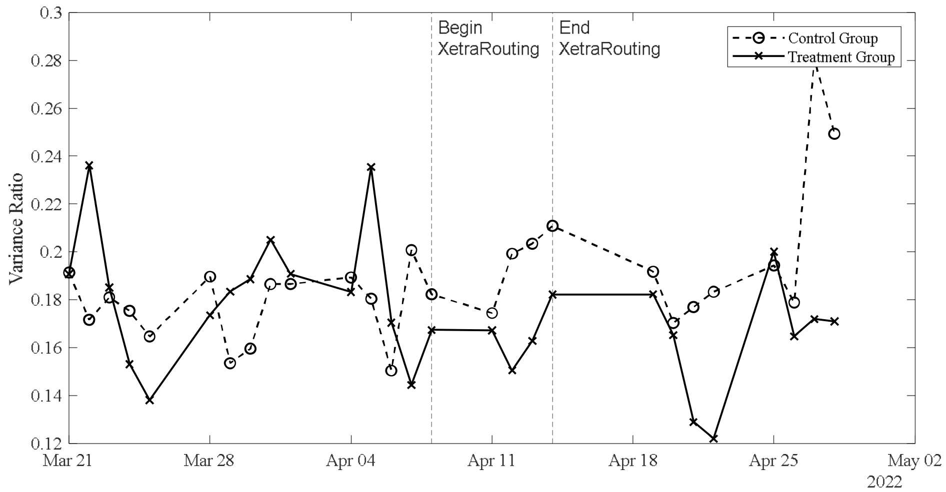
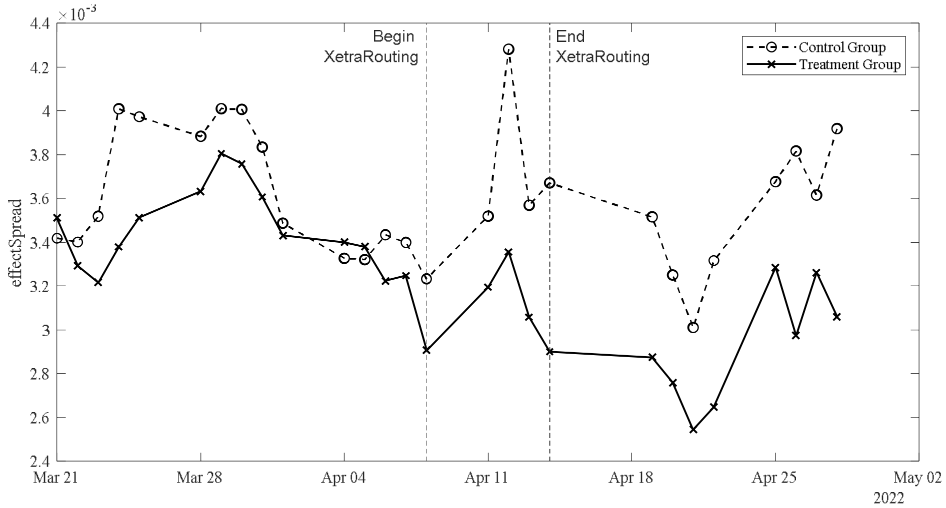


Figure 3: Parallel Trends Graph: Effective Spread



As can be seen from the figures, there is some pronounced volatility in the time-series of averages of the treatment and control group with irregular spikes, but no apparent pattern of a different behavior relative to each other before, during and after the treatment. Thus, a differences-in-differences analysis appears reasonable.

[28] As both groups only consist of a relatively small group of stocks (10 and 12 for treatment and control, respectively), the time-series of the quality measures exhibit a pronounced volatility. We therefore smooth all time-series in the figures with an exponentially weighted moving average, using a three-trading-day window.

3.4.3 DiD Regression Results

Table 4 shows regression results obtained with the DiD approach relying on a model with stocks and days fixed effects (see equation (3)), clustering standard errors at the day and stock levels. As additional control variables, we use the inverse stock price as well as the natural log of trading volume, each one-day lagged.[29] For interpretation purposes, note again that the stock fixed effects control for all time-invariant observable and unobservable characteristics like firm size, industry etc., and the time fixed effects control for all observable and unobservable characteristics that are constant across stocks, like macro-variables, the market volatility etc.

Accordingly, the key variable of interest to assess the impact of payment for order flow on market quality for each of the analyzed measures is the interaction term between the indicator for days with Xetra routing and the treatment group. For brevity, only the estimated coefficients for this interaction (previously referred to as β_3) are shown in Table 4.

The table shows that none of the estimated coefficients for *XetraRouting x Treatment* are statistically different from zero.[30] Thus, the average treatment effects for all liquidity and for all price discovery measures in the treatment period do not differ between the treatment and the control group.

Alleged harmful impacts of payment for order flow on market quality when routing retail customers' orders to the main market should have resulted in higher liquidity, smaller spreads and more efficient information processing in our field experiment. However, none of these hypotheses are consistent with the evidence from the experiment. There is no improvement on information processing, as Hypothesis 4 predicted. Also, the increased trading due to a significant increase in retail orders in the Xetra market does not affect liquidity measures in a statistically significant way, opposed to Hypothesis 1-3 on liquidity effects.

Overall, these results show that even the addition of large volumes of uninformed trading, as in this experiment, are still too small to affect market quality.

[29] Omitting the inverse price and log-trade volume as explanatory variables does not qualitatively change any results regarding the market quality impact of payment for order flow.

[30] The p-values given in Table 4 are used to assess statistical significance. They indicate the probability of error when one rejects the assumption (null-hypothesis) of a coefficient of zero on the basis of the observed data, if the null is true. In economics, only cases in which the probability of error is less than 5% to 10% are considered statistically significant.

Table 4: Differences-in-Differences Analysis Main Experiment

Note: The table shows estimation results from a fixed effects differences-in-differences analysis for our main sample of 22 stocks, using day and stock fixed effects. The dependent variables are measures for price discovery as well as market liquidity. Price discovery measures are the absolute difference of one with the variance ratio of 5-minute over 1-minute mid-point log-returns, the 1-minute midpoint autocorrelation, and the 1-minute market return cross-correlation. Liquidity measures are the time-weighted relative spread, the tick-based effective spread, the tick-based 1-minute realized spread, and the tick-based 1-minute price impact. See Section 3.1 for definitions. In all regressions, additional control variable used are the (one day lagged) inverse stock price and the natural log of the daily trading volume.
Coefficients are only shown for the estimate of the average treatment effect (*XetraRouting x Treated*). The treatment group are ten stocks treated in the sample period from March 21 – April 28, 2022. For treated stocks in the treatment period from April 08 – April 14, 2022, the neo-brokers' customer orders were routed to the Xetra trading system instead of being sold to the neo-brokers' trading venue (LSX) against payment for order flow. Standard errors are clustered for stocks and trading days.

Dependant	Variable	Coefficient	t-Statistic	p-Value	nObs	R^2
Price Discovery						
Variance Ratio	XetraRouting x Treated	-0.0065	-0.2007	0.8410	594	0.032
Autocorrelation	XetraRouting x Treated	-0.0151	-0.7784	0.4367	594	0.076
Cross-correlation Index	XetraRouting x Treated	-0.0065	-0.5614	0.5748	594	0.086
Liquidity (equal-weighted)						
Relative Spread	XetraRouting x Treated	-0.0001	-1.0887	0.2767	594	0.278
Price Impact	XetraRouting x Treated	-0.0007	-1.2671	0.2057	594	0.094
Effective Spread	XetraRouting x Treated	-0.0004	-1.0841	0.2788	594	0.096
Realized Spread	XetraRouting x Treated	0.0003	1.3014	0.1937	594	0.064
Liquidity (value-weighted)						
Price Impact	XetraRouting x Treated	-0.0007	-1.0948	0.2741	594	0.089
Effective Spread	XetraRouting x Treated	-0.0004	-1.0422	0.2977	594	0.092
Realized Spread	XetraRouting x Treated	0.0003	0.8932	0.3721	594	0.037

4 Robustness: Testing for Actual Treatment

Our finding that payment for order flow has no impact on liquidity and price discovery on Xetra during the treatment days can result for two reasons: (i) there is no causal effect of routing the retail orders to the main market Xetra, (ii) the orders were actually not routed to Xetra. At least in principle, there exists a potential conflict of interest with respect to routing the orders to Xetra instead of taking them on the own book for the market makers of the LSX exchange. If the field experiment had shown that, on treatment days, market quality is better due to the routing of the neo-brokers' order to Xetra, that could have significantly affected LSX exchange and its profitability for the operator Lang & Schwarz.

To verify that such a distortion of the experiment did not occur, we have collected a sample of 1,000 customer orders, randomly sampled across treated stocks on treatment days by the neo-broker. For these transactions, we can check whether the orders were actually routed to Xetra by matching them with the Xetra tick data. Table 5 shows results of the matching attempt.

Matching is based on stock ISIN, order volume, and order price, in a two-hours window after order initialization at the neo-broker. The table shows that, differentiating for market versus limit orders, one can match about 95% of market orders using Trade Republic customer order information to the corresponding Xetra tick trades in the two-hour trading windows. Some orders will be not be captured due to

- trading took place after Xetra close,
- trading occurred on the next trading day,
- orders were split into smaller blocks (with the same execution price) due to market conditions,
- or a failure by Lang & Schwarz (the broker) to route orders to Xetra.

The broker itself reported to Trade Republic that it was able to route 97% of all customer orders to Xetra, and the test sample of orders confirms this information.

The median execution time, measured as the difference between the Xetra tick timestamp and the order initialization timestamp, is 11 seconds. For limit orders, about 79% of orders are matched within a two-hour window, where the gap to 100% is of course predominantly attributable to the limit not having been reached in the matching window.

Table 5: Matching Trade Republic Retail Orders to Xetra Tick Data

Note: The table shows results from matching a random sample of Trade Republic customers' orders for the treatment group of stocks during the treatment days from April 08 – April 14, 2022, during which customer orders were routed to trading system instead of being sold to the neo-brokers' trading venue (LSX) against payment for order flow. Matching is based on stock ISIN, order volume and order price, in a two-hour window after order initialization at the neo-broker.

	All	**Market**	**Limit**
Number of Transactions	1,000	817	183
Matched in Xetra [%, within 2 hours after initiation]	---	92.78	86.34
Number of Firms (all, matched)	10	10	9
Median Execution Time (seconds)	---	11.18	511.51
Median Trade Value [€]	---	415.13 €	798.20 €

The matching exercise demonstrates that the treatment of stocks on treatment days actually took place. As a second, related robustness test, Table @ shows results from a differences-in-differences analysis of the average treatment effect for (the natural logarithm of) trade volume and the number of transactions. Due to the experiment, we should expect that on treatment days trading in the treated stocks increases as additional retail orders are added to Xetra. Since retail orders are on average small and millions of additional shares were traded on treatment days, one should in particular find that the number of trades in Xetra increases significantly on treatment days.

From Table 6 can be seen that this is indeed the case. The estimated coefficients on the interaction term between the indicator for days with Xetra routing and the treatment group are positive for both *log(Volume)* as well as *log(Number of Trades)*. With a p-value of 0,95% the ATE for the *log(Number of Trades)* is statistically significant, showing that the treatment indeed had a significant impact on trading in Xetra on treatment days.

Table 6: DiD-Analysis of Volume and the Number of Trades

Note: The table shows estimation results from a fixed effects differences-in-differences analysis for our main sample of 22 stocks, using day and stock fixed effects. The dependent variables are the natural logarithm of trade volume, *log(Volume)*, and the natural logarithm of the number of trades, *log(Number of Trades)*, for a given stock at a given trading day in the sample period form March 21 – April 28, 2022.
Coefficients are only shown for the estimate of the average treatment effect (*XetraRouting x Treated*). The treatment group are ten stocks out of 22 overall. For treated stocks in the treatment period from April 08 – April 14, 2022, the neo-brokers' customer orders were routed to the Xetra trading system instead of being sold to the neo-brokers' trading venue (LSX) against payment for order flow. Standard errors are clustered for stocks and trading days.

Model	Dependant	Variable	Coefficient	t-Statistic	p-Value	nObs	R^2
FE for stocks and days, controls	log(Volume)	XetraRouting x Treated	0.0934	0.8686	0.3854	594	0.158
FE for stocks and days, controls	log(Number of Trades)	XetraRouting x Treated	0.2159	2.6015	0.0095	594	0.102

Overall, the two robustness tests of matching Trade Republic orders to Xetra tick data and analyzing the impact of the treatment on trade volume in Xetra show that the treatment of stocks on treatment days took place as intended by the experiment.

5 Conclusions

So-called neo-brokers like Trade Republic or Robinhood offer retail customers access to trading in capital markets at very low (or even zero) transaction costs, enabled by payment for order flow. The rise of the neo-brokers has, however, re-sparked the debate about the impact of payment for order flow (PFOF) on market quality. Opponents argue that withdrawing retail orders from the primary exchange harms investors in general through reduced market quality, where market quality is traditionally measured by the liquidity of the market and the degree to which it fulfills its price discovery function. In this context, the German neo-broker Trade Republic mandated us to provide an independent expert opinion to assess the impact of payment for order flow on market quality.

Our analysis of the impact of payment for order flow on the quality of stock markets in the context of so-called neo-brokers first provides a thorough review of theoretical arguments. It shows that a major consequence of payment for order flow arrangements is price differentiation. Without payment for order flow all investors are pooled and pay the same bid-ask spread when trading via market orders. An important component (known as the adverse selection component) of the spread is the compensation that suppliers of liquidity require for the risk of trading with better-informed counterparties. Retail investors are less likely to possess superior information than institutional investors. Consequently, the appropriate adverse selection component is lower for retail than for institutional traders. Yet, in a pooled market all investors pay the same adverse selection component, implying that retail customers implicitly subsidize institutions. Payment for order flow arrangement single out retail order flow and execute it at better conditions. This is achieved by charging a lower bid-ask spread than the main market and/or lower or even zero brokerage commissions. By implication, when retail orders are diverted away from the main market, the adverse selection risk in the main market increases, possibly resulting in wider spreads in the main market. While this widening of spreads is often used as an argument against payment for order flow, it may simply be a reflection of the increased price discrimination discussed above.

In spite of the theoretical prediction that payment for order flow results in increased spreads in the main market, prior empirical research has not produced evidence of such an increase in spreads. Our own analysis adds to this evidence. We conduct a unique field experiment, allowing us to test the causal relationship between payment for order flow and the market quality measures for the main market.

We identify stocks for which the neo-broker Trade Republic has a significant market share (2%-5% of the daily volume) and then randomly select a sample of 10 stocks for which payment for order flow was deliberately switched "off" and then "on" again. This was achieved by routing all orders that Trade Republic received in these stocks to the main market (Xetra) for a treatment period of five trading days.

Hence, for these stocks on treatment days, trading for Trade Republic retail customers occurred without payment for order flow and without the otherwise usual execution on the exchange segment of Lang & Schwarz (LSX). Rather, the orders were executed on the main market. This re-routing leads c.p. to an *increase* in Xetra trading on treatment days due to additional orders from retail investors who tend to be uninformed. The number of additionally traded shares was in the millions. The control group of stocks (those 12 out of the stocks for which Trade Republic has a significant market share that were *not* randomly selected for the treatment) were not treated, meaning that orders in these stocks were executed via the usual payment for order flow arrangement. This setting allows us to directly infer the causal impact of payment for order flow on market quality by comparing quality measures before, during and after treatment for the treatment and control groups. The field experiment thus allows us to observe a world with and without payment for order flow, which is not usually possible.

We find that all our measures of market quality for the main market Xetra (related to market liquidity and to the informational efficiency of the market) did not change in a statistically significant way on the treatment days. Despite the large proportion of additional retail orders, trading in Xetra did not get more liquid, and the informational efficiency of the market did not improve.

These results imply that payment for order flow, even if implemented at large scale (remember that the market share of Trade Republic in the stocks under consideration ranged between 2% and 5%), does *not* lead to lower market quality in the main market. Even the increase in bid-ask spreads predicted by theoretical considerations cannot be observed. The latter finding, by the way, is consistent with earlier empirical evidence, surveyed in section 2.3 above, from the US.

In our field experiment we switched on and off payment for order flow by a single (yet large) neo-broker. We can only speculate what would happen if *all* payment for order flow were switched off. A reasonable estimate of the proportion of retail trading in the total trading volume is 10%.[31] Thus, 10% is a reasonable upper limit to the fraction of volume executed via payment for order flow arrangements. In our experiment we switched on and off between 2% and 5% of the volume, with no effect on market quality. We believe (but cannot prove) that market quality in the main market would not deteriorate even if all retail flow were executed off the main market via payment for order flow arrangements.

Our analysis also has important implications for the regulation of exchange markets in general and payment for order flow arrangements in particular:

- Different institutional settings that aim at separating retail order flow from the total order flow have similar economic implications. Thus, from an economic perspective, neo-brokers and payment for order flow are fundamentally similar to the internalization of retail orders at financial intermediaries such as banks, or the establishment of specific exchanges for private investors only. Any regulation should therefore address *all* of these settings equally, rather than focusing on just one of them (as is the case in the current debate on payment for order flow).
- In our view, the most sensible regulatory measure to achieve fair stock trading for all market participants is the creation of pre-trade transparency through the introduction of a European best bid-offer (similar to the NBBO in the US). To create such a European best bid-offer, pre-trade information would need to be collected from all publicly accessible trading venue, aggregated on a per-instrument level, and published in real time. It could then be made mandatory (and easily verifiable!) that off-exchange executions in the context of payment for order flow and internalization arrangements must occur at prices at least equal to the European best bid-offer.
- Once the European best bid-offer has been established, brokers who use payment for order flow or internalize could be required to regularly publish execution statistics. These would report the execution quality relative to the European best bid-offer. Such a measure would promote competition between neo-brokers and would arguably result in lower trading cost for retail investors.
- We strongly believe that a prohibition of payment for order flow is not warranted. It would ultimately cut off retail investors from low-cost execution of their trades without improving the quality of the main market. In addition, payment for order flow arrangements, if prohibited, might simply be replaced with internalization schemes, which, as pointed out above, have similar economic implications.

[31] Puckett/Yan (2011) estimate the proportion of retail orders to be at about 20%. More recent data rather predicts a lower proportion. The Xetra tick data provide an indicator for algorithmic trading, i.e. trades initiated on behalf of institutional traders, which has a proportion of about 90% for our overall sample of 59 stocks and in our observation period. Boehmer et al. (2021) find that about 7% of daily average trade volume in all U.S. stocks arises from retail market orders.

Appendix A: Results on the Test Experiment

In the first stage of the experiment, we conducted a pilot study. Nine stocks out of 37 were randomly treated, while the remainder 28 served as the control group. The treatment of routing to Xetra instead of using the trading venue with payment for order flow (Lang & Schwarz Exchange, LSX) occurred for 4 stocks from Feb. 22, 2022 until Feb 24, 2022 (three trading days). For another 5 stocks, treatment occurred in the following week, i.e. from Feb 28 – March 04, 2022 (five trading days).

This first experiment served as a testing ground to gather experience with the experiment, in particular for Trade Republic and its broker who usually pays for the order flow, as customer orders from the neo-broker had to be systematically redirected to Xetra by LSX. As it turned out, three of the treated stocks and 22 stocks of the control group were too illiquid, such that only few trades and only few quote updates occurred on Xetra. For twelve stocks (7 in treatment, 5 in control), however, liquidity on Xetra was high enough to conduct an analysis on the intraday level. Henceforth we label this group of stocks as the "test sample". Table 7 shows basic characteristics of the firms in the test sample, similar to Table 2.

Table 7: Characteristics of Firms in the Pilot Study and Test Sample

Note: The table shows pre-experiment characteristics of firms, used to select the treatment and control group in the pilot study of our experiment (test sample). All measures are taken up until January 31, 2022. The experiment comprises overall a sampling period (before, during and after treatments) from February 01, - April 28, 2022, treatment of 7 firms in the test sample occurred on Feb. 22 - Feb 24, 2022 or Feb 28 - Mar 04, 2022.

	Market Capitalization as of 01/31/2022 [€ Mill.]	**Average daily trading volume (6 month until 31.01.2022)**	**Share of neo-broker in daily trading volume (average, 1 year until 01/31/2022)**	**Market Capitalization as of 01/31/2022 [€ Mill.]**	**Average daily trading volume (6 month until 31.01.2022)**	**Share of neo-broker in daily trading volume (average, 1 year until 01/31/2022)**
	Untreated			**Treated**		
	Pilot Study: All stocks					
No of Stocks	28			9		
Mean	222	22,496	10.19%	4,476	2,258,265	6.56%
Median	104	12,590	7.69%	334	76,192	5.87%
Std.	228	25,535	12.44%	11,203	6,454,548	2.36%
Min	9	1,742	5.10%	91	2,766	5.03%
Max	660	105,538	72.46%	34,231	19,466,938	12.58%
	Pilot Study: Test Sample					
No of Stocks	8			7		
Mean	424	33,345	7.38%	5,728	2,901,181	6.91%
Median	445	25,896	6.47%	468	114,177	5.92%
Std.	207	32,716	2.19%	12,613	7,306,039	2.60%
Min	110	1,742	5.16%	287	6,214	5.05%
Max	640	105,538	10.86%	34,231	19,466,938	12.58%

The main characteristic of the stocks included in the test sample is that the neo-broker has a high market share in daily trading volume in excess of 5% trade volume participation, but still some of these stocks are characterized by rather low liquidity. Therefore, for the test sample, we analyze liquidity in 10-minute and 2-minute intervals for the Variance Ratio, and a 2-minute interval for all other price discovery measures. Liquidity measures are based on contemporaneous and 1-minute forward mid-quotes, as for the main sample (see Section 3.1).

Table 8 shows estimation results according to equation (3) for all seven market quality indicators described in Section 3.1. Compared to the analysis of the main sample of liquid stocks analyzed in Section 3.4, qualitative results do not change even for this sample of much more illiquid stocks.

Price discovery measures do not change on treatment days, implying that information processing is not affected by adding liquidity to the Xetra by increasing the number of uniformed retail orders. The only difference with the results obtained with the main sample is that for equal weighted liquidity measures, the price impact is statistically significant negative and the realized spread is significantly positive. As discussed before, a finding that the price impact is reduced is consistent with the fact that more retail orders are on treatment days in the Xetra market. If market makers identify the increase in uniformed orders, this could lead to a decrease in the average price impact (see Hypothesis 3). The higher realized spread would imply a higher temporary impact of orders on average. Overall, these effects cancel out, as can be seen from the insignificant effect on the effective spread. However, since the effects on the spread components are not observable with value weighting of liquidity measures, the results are not robust and most likely due to artifacts attributable to the low liquidity and the small sample size of the test sample.

Table 8: Differences-in-Differences Analysis: Test Sample

Note: The observation period is from Feb 01 until Mar 18, 2022. Treatments in terms of order routing to Xetra without (PFOF) instead of LSX (with PFOF) occurred for six stocks on Feb 22- Feb 28, 2022 or Feb 28 - Mar 04, 2022, respectively. The control group consists of six stocks as well. Price discovery measures are the absolute difference of one with the variance ratio of 10-minute over 2-minute mid-point log-returns, the 2-minute midpoint autocorrelation, and the 2-minute market return cross-correlation. Liquidity measures are the time-weighted relative spread, the tick-based effective spread, the tick-based 1-minute realized spread, and the tick-based 1-minute price impact. The two-way fixed effects panel regressions include fixed effects for stocks and trading days (time). Standard errors are clustered for stocks and trading days.

Dependant	Variable	Coefficient	t-Statistic	p-Value	nObs	R^2
Price Discovery						
Variance Ratio	XetraRouting x Treated	-0.0104	-0.1995	0.8420	408	0.066
Autocorrelation	XetraRouting x Treated	-0.0022	-0.0817	0.9349	408	0.066
Cross-correlation Index	XetraRouting x Treated	0.0226	0.9192	0.3586	408	0.093
Liquidity (equal-weighted)						
Relative Spread (time-weighted)	XetraRouting x Treated	0.0000	-0.1293	0.8972	408	0.320
Price Impact	XetraRouting x Treated	-0.0010	-2.2797	0.0232	408	0.377
Effective Spread	XetraRouting x Treated	-0.0005	-1.1671	0.2439	408	0.444
Realized Spread	XetraRouting x Treated	0.0005	2.2343	0.0261	408	0.032
Liquidity (value-weighted)						
Price Impact	XetraRouting x Treated	-0.0008	-1.4275	0.1543	408	0.315
Effective Spread	XetraRouting x Treated	-0.0005	-0.8202	0.4127	408	0.396
Realized Spread	XetraRouting x Treated	0.00029	0.7041	0.4818	408	0.026

Literature

AFM (2022): Assessing the quality of executions on trading venues. Version 2, March.

BaFin Bundesanstalt für Finanzdienstleistungsaufsicht (2022): Studie zur Ausführungsqualität an ausgewählten deutschen Handelsplattformen, April.

Barber, B. / Odean, T. (2000): Trading is hazardous to your wealth: The common stock investment performance of individual investors, Journal of Finance 55, 773-806.

Barber, B. / Odean, T. (2008): All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, Review of Financial Studies 21, 785-818.

Barber, B. / Lee, Y.-T. / Liu, Y.-J. Liu / Odean, T. (2009): Just How Much Do Individual Investors Lose by Trading?, Review of Financial Studies 22, 609-632.

Barrot, J. / Kaniel, R. / Sraer, D. (2016): Are retail traders are compensated for providing liquidity?, Journal of Financial Economics 120, 146-168.

Battalio, R. (1997): Third market broker-dealers: Cost competitors or cream skimmers?, Journal of Finance 52, 341-352.

Battalio, R. / Greene, J. / Jennings, R. (1997): Do competing specialists and preferencing dealers affect market quality?, Review of Financial Studies 1997 (10), 969-993.

Battalio, R. / Hatch, B. / Jennings, R. (2003): All else equal? A multidimensional analysis of retail, market order execution quality, Journal of Financial Markets 6, 143-162.

Battalio, R. / Jennings, R. / Selway, J. (2001): The relationship among market-making revenue, payment for order flow, and trading costs for market orders, Journal of Financial Services Research 19 (1), 39-56.

Bessembinder, H. / Venkataraman, K. (2010): Bid-ask spreads: measuring trade execution costs in financial markets, in: Cont, R. (Ed.), Encyclopedia of Quantitative Finance. Wiley, Hoboken, New Jersey, pp. 184–190.

Boehmer, E. / Jones, C. / Zhang, X. / Zhang, X. (2021): Tracking Retail Investor Activity, Journal of Finance 76, 2249-2305.

Card, D. / Krueger, A. (1994): Minimum wages and Employment: A Case study of the Fast Food Industry in New Jersey and Pennsylvania, American Economic Review 84, 772-784.

CNMV Comisión Nacional del Mercado de Valores (2022): Payment for order flow: An analysis of the quality of execution of a zero-commission broker on Spanish stocks, March.

Comerton-Forde, C. / Grégoire, V. / Zhong, Zhou (2019): Inverted fee structures, tick size and market quality, Journal of Financial Economics 134, 141–164.

Coval, J. / Hirshleifer, D. / Shumway, T. (2021): Can individual investors beat the market?, Review of Asset Pricing Studies 11, 552-579.

Easley, D. / Kiefer, N. / O'Hara, M. (1996): Cream-skimming or profit sharing? The curious role of purchased order flow, Journal of Finance 51, 811-833.

Fama, E. (1970): Efficient capital markets: A review of theory and empirical work, Journal of Finance 25 (2), 383-417.

Foucault, T. / Sraer, D. / Thesmar, D. (2011): Individual investors and volatility, Journal of Finance 66, 1369-1406.

Grammig, J. / Theissen, E. (2012): Is BEST really better? Internalization of orders in an open limit order book, Schmalenbach Business Review 64(2), 82-100.

Hasbrouck, J. (1995): One security, many markets. Journal of Finance 50(4), 1175-1199.

Johann, T. / Scharnowski, S. / Theissen, E. / Westheide, C. / Zimmermann, L. (2018): Liquidity in the German Stock Market, Working Paper.

Kothari, S.P. / So, E. (2021): Commission Savings and Execution Quality for Retail Trades, Working Paper.

Lee, Ch. / Ready, M. (1991): Inferring Trade Direction from Intraday data, Journal of Finance 46, 733-746.

Meyer, S. / Uhr, C. / Johanning, L. (2021): Private investors and the emergence of neo-brokers: Does payment for order flow harm private investors?, Working Paper.

Mittal, H. / Berkow, K. (2021): The Bad, The Good and The Ugly of Payment for Order Flow, White Paper, Bestex Research, White Paper.

O'Hara, M. / Ye, M. (2011): Is market fragmentation harming market quality?, Journal of Financial Economics 100 (3), 459–474.

Puckett, A. / Yan, X. (2011): The interim trading skills of institutional investors, Journal of Finance 66, 601–633.Rosov, S. (2022): Payment for order flow in the United Kingdom, CFA Institute.

Samuelson, P. (1965): Proof that properly anticipated prices fluctuate randomly, Industrial Management Review 6 (2), 41-49.

Schiozer, R. / Mourad, F / Martins, T. (2020): "A tutorial on the use of differences-in-differences in management, finance, and accounting", Revista de Administração Contemporânea 25.

Snow, J. (1855): *On the Mode of Communication of Cholera*. 2nd ed., John Churchill.

Specht, M. (2022): Die dunkle Seite des Aktienkaufs via App. https://www.finanzwende-recherche.de/2022/02/16/die-dunkle-seite-des-aktienkaufs-via-app/, accessed 4.4.2022

Theissen, E. (2002): Internalisierung und Marktqualität: Was bringt Xetra Best?, Kredit und Kapital 35 (4), 550-571.