Are the Young Traders taking over? Evidence from an Analysis of Information Shares in the German Equity Market

Term paper for the
Seminar 'S510/S520 – Master Seminar on Econometrics'
of Prof. Dr. Martin Biewen and Prof. Dr. Joachim Grammig

January 7, 2022

Leonard Berger
Collegiumsgasse 14
72070 Tuebingen
Student-ID: 5629475
M.Sc. Data Science in Business and Economics

Contents

List of Figures

List of Tables

1 Introduction

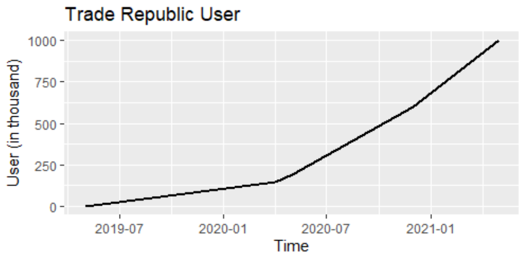
The year is 2021. A pandemic has silenced the world, cinemas are closed and no one knows when they will open again or when public life can return back to normal. Meanwhile, AMC, an almost insolvent operator of movie theaters, surges over 400% in market capitalization within just one month. At its peak, the stock of AMC brought a return of 2700% in the period from January until July. Only a few days earlier, the stock of an American video game retailer called GameStop, whose best days definitely lay in the past, gained momentum and increased its value by a multiple of 20. While reaching prices of over 400 USD in January, the stock plumped back to 40 USD just one month later. Both stocks shared a high amount of short-seller and hedgefunds such as Melvin Capital or Citron Research Group had to close their positions with a huge loss – a serious blow for Wallstreet.

Both stories are somewhat extreme examples but they might be the most prominent ones within a development seen over the last two years in the stock market where prices seem to be set rather through emotions and internet communities than through fundamentals. This development shifted the attention to a new group of trader that entered the market not that long ago. Young retail investors between 18 and 30 started to gain interest in the stock market and they had a useful tool at hand which led to a disruption within the trading community: zero commission broker. These so-called *neobroker* enabled young people to start trading in small quantities while not paying any fees. The most famous one is without a doubt the American startup Robinhood but the concept is not limited to the US. A similar development is taking place in Germany where app-based zero commission broker like Trade Republic or Scalable Capital offer easy access to stock and option trading for everyone. Only within the past two years, Trade Republic gained more than one million customers (see Fig. 1) and now manages over six billion euros in customer assets, while being the most valuable German startup[1]. And the average user is not even 30 years old[2]. But although these new traders came with a bang, literature on how they disrupted the market is scarce, especially for German speaking countries.

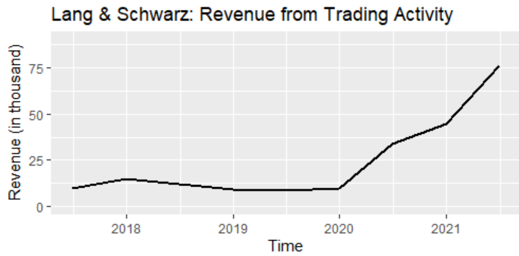
The rest of the paper continues as follows. Section 2 introduces to the economical research questions answered in the analysis. Section 3 gives a short overview over the existing literature on price discovery and the market entry of neobroker. The methodology and the data used are described in section 4. Results are presented in section 5 and section 6 concludes.

[1]https://t3n.de/news/broker-app-trade-republic-sammelt-1380109/

[2]https://www.spiegel.de/wirtschaft/service/trade-republic-robinhood-und-co-app-boom-die-jungen-zocker-kommen-a-5f701590-5a19-43d7-a97f-03f10584c1b0



(a) Development of users of Trade Republic



(b) Lang & Schwarz Trading Activity

Figure 1: Market entry of new traders via neobroker

2 Research Question

In the following I assess to what extend these new mom-and-pop investors[3] participate in the price discovery process in Germany on the basis of DAX-companies. Price discovery examines how markets incorporate new information and enables to quantify the share of information that is brought by each market to the general trend. Most studies that deal with price discovery focus on different venues where a security is traded but do not differentiate between the agents that trade at these venues. In general, these exchanges have homogeneous traders that mostly only differ by nationality, e.g. for cross-listed stocks. However, the setting of how Trade Republic implemented their brokerage allows to examine two different groups of buyers and sellers. As mentioned before, the people trading on the app are not the average German stockholder, but rather belong to the generation of people aged 35 and under, inexperienced and risk-tolerant. And due to the construction of Trade Republic, these traders are restricted to only one exchange in Germany: the Lang & Schwarz exchange. All trades conducted through the app are executed at Lang & Schwarz via the stock exchange in Hamburg. As it can be seen in Fig 1 (b), the revenue from trading activity for the Lang & Schwarz exchange went up significantly after Trade Republic paved the way for a whole new group of traders to this exchange. Comparing price discovery metrics of the Lang & Schwarz exchange now to another exchange can be seen as a proxy for comparing the information brought to the market by the young retail investors to the information brought by the rest of the market participants.

Price discovery metrics are assessed using information shares of the particular markets. The analysis is then conducted via two dimensions. First, information shares are compared before and after the introduction of Trade Republic in the beginning of 2019. If information shares shift towards the Lang & Schwarz exchange over time,

[3]The term mom-and-pop investors is adopted from an article of Osipovich (2020) in The Wall Street Journal. Osipovich talks about young, inexperienced retail investors who trade in small quantities and are willing to take higher risks in return for higher capital gains on their investments.

this might be an indication for new market dynamics brought by the younger retail traders. Different people meet at different markets, so if differences are found between those markets, these differences could very well be attributed to the people involved in these markets. Second, how do information shares vary between certain stocks? Young retail investors tend to invest in *lottery stocks* (see e.g. Frino et al. (2019) or stocks for which the product of the company is familiar (Welch (2021)). These companies often stand in the focus of financial internet communities such as Twitter or Reddit. The hypothesis here is that the information share for stocks that draw more attention of young people is higher than for the rest of the market. It is feasible to assume that mom-and-pop investors spend more time on the internet and consume information in higher frequencies. Therefore, information that might move the stock price should be brought into the market by the young first while other securities that draw less attention stay unaffected. These popular stocks of interest - mostly tech-stocks - often trade at higher multiples. The relation between price discovery and the descripted phenomenon is modelled through the correlation of information shares and the EV/sales-ratio. Both research questions are assessed using *End-of-Day*-data (EOD-data) for DAX-companies throughout the last years. The latter hypothesis is also evaluated on a set of stocks sampled at 1-minute intervals for a shorter period.

3 Literature Review

When it comes to the question of which market determines the prices there is no getting around Joel Hasbrouck's study from 1995. Until then, most researchers focused on "lead-lag" return regressions to examine which market sets the price and which one follows. However, Hasbrouck (1995) criticizes that these models might be generally miss-specified and introduced what would further be known as the *Hasbrouck information share*. Hasbrouck (1995) assumed that asset prices are driven by one efficient price, and that different prices at different exchanges materialize based on the new information that is incorporated within these exchanges. This underlying assumption made it possible to assign a measure of price discovery to certain markets. The study of Hasbrouck (1995) was motivated by the fact that more and more trades seemed to shift from the big exchanges to alternative markets. Hasbrouck (1995) then took a look at the 30 stocks listed in the Dow Jones Industrial Average over a three-month period in 1993 that were traded at different US exchanges. Hasbrouck (1995) found that - even with increasing volumes for other markets - price discovery still appeared to be concentrated at the NYSE with a median information share of around 93 percent.

After this pioneer-study, more recent research started to focus on cross-listed stocks, i.e. stocks that trade on different venues in different countries simultaneously. Bacidore and Sofianos (2002) argue that price discovery should mainly take place in the home market since this is the place where most information about the company is generated. A study by Grammig et al. (2001) a year before found the same effects, with the extension that the shares also depend on the level to which a company is engaged in multi-national markets. The Xetra information share of Deutsche Telekom, a firm mostly active within Germany, was higher than the one of SAP compared to NYSE quotes. Although a larger share of information could still be attributed to the home market in Frankfurt, SAP itself is considered to be more of an international company, with most of its competitors being located in the US, whereas Deutsche Telekom was mostly active within Germany at that time. Ghadhab and Hellara (2016) complement that, for cross-listed firms, the home market might still be more dominant whereas for multi-listed firms, foreign markets contribute more to price discovery. Within the globalized environment where it is easy to trade at multiple exchanges, it makes sense that informed trader choose the market with best conditions. That might not necessarily be the home market but rather the one with lower trading costs (see e.g. Ghadhab and Hellara (2016), Frijns et al. (2015)). Also, Ghadhab and Hellara (2016) show that the US exchanges contribute more to the price discovery process overall than the European ones, regardless of where the firm is located. In one of the few studies without US component, Agarwal et al. (2007) raise the interesting thought that if trading on the big exchanges is liquidity rather than information driven, and that most trades are conducted for portfolio rebalancing purposes, international market prices should fully incorporate home market prices but not vice versa. Therefore, the London market does not seem to have an impact on stocks originally traded in Hong-Kong (Agarwal et al. (2007)).

So far, all studies focused on static information shares. In the latter analysis, the price discovery process will also be examined regarding a time component. Frijns et al. (2015) point out that only very few studies have looked at how price discovery evolves over time. Their study involved Gonzalo and Granger Component Shares for a sample of Canadian cross-listed firms including a time component - pre and post 2002. Comparing the information shares before and after, Frijns et al. (2015) find that once a market gained a high share in the price discovery process it tends to stay at this market. Fuess et al. (2018) look at price discovery before and after the financial crisis of 2008 and find co-movements between the markets during the crisis period. However, their study lacks specific results regarding the development of information shares before and after the cutoff point.

Analyzing stock quotes using the Hasbrouck information share brings two major drawbacks. First, the method needs high frequency data. For example, Hasbrouck (2021) compares information shares using quotes down to 10 microseconds intervals. Second, the computation results in upper- and lower-bounds. Narayan and Smyth (2015) provide a nice overview over studies that used the Hasbrouck method and criticize that, especially for low-level frequency data, these bounds can become quite large and interpretability diminishes. To tackle these challenges, Grammig and Peter (2013) developed a metric to compute unique information shares if only EOD-data is available based on tail dependence. The method is explained in more detail in section 4.2. However, Lien and Wang (2016) find that the Hasbrouck information share provides good results on simulated data and should be preferred over the approach by Grammig and Peter (2013).

For now, most studies shown concentrate on different market venues (i.e. *locations*). The aim of this study is to also attribute the price discovery process to different *groups* of traders. To my knowledge, the study of Bjonnes et al. (2021) is the only one that allocates information shares also to different agents participating within a Forex-setting. According to them, new innovations in the stock market mostly come from a group of financial institutions, such as hedge funds or investment managers. Non-financial corporations, governments or insurance firms do not bring any information (Bjonnes et al. (2021)). However, Bjonnes et al. (2021) totally exclude retail investors in their analysis although this group is becoming more and more important. Hu et al. (2021) argue that through zero commission trading and the so-called *gamification of investing* retail traders started to take their fair share in the price discovery process. Analyzing Reddit data – a forum where especially young trader exchange financial information – Hu et al. (2021) find that more positive content and comments among these retail traders are connected to higher returns, higher order flow and lower shorting flows in the future.

However, the literature is divided when it comes to analyzing the impact of young traders on the market. Pasztor (2021) warns that the gamification and gambling aspects of the Robinhood app paired with its trading volume could affect the integrity of the whole financial system. Tokic (2020) adds that during the Covid-19 pandemic, many young Americans receiving unemployment benefits while "sitting at home unemployed, bored without live sports or other entertainment" (Tokic, 2020, p. 12) played a large role in what he considered to be a big tech bubble in recent months. Stein (2020) points out that retail traders focus most on the stocks that are already in an upward trend and popular within their trading app, a mechanism where the information provided by the brokerage influences the trading behavior of investors while the number of trades also directly affects asset prices in the whole market. Friedman and Zeng (2021) paint a less pessimistic picture. Robinhood users do seem to increase volatility while decelerating the incorporation of new information into prices for companies with larger market capitalization (Friedman and Zeng (2021)). However, for smaller firms with less efficient and costlier price discovery mechanisms, retail traders seem to help when prices need to adapt to new market information. When assessing the impact of retail investors on the market, Pagano et al. (2021) observe mixed effects. During the Covid-19 crisis, market quality measures such as the variance ratio or realized spreads seemed to worsen with increased holdings of Robinhood users. However, during less stressful market periods, retail investors were associated with improved price discovery metrics (Pagano et al. (2021)). Welch (2021) shows that Robinhood investors acted as a market-stabilizing force during the Covid-19 market crash. Although most Robinhood portfolios might seem odd compared to the rest of the market, investors did not panic and – on average – did not underperform (Welch (2021)).

4 Methodology and Data

4.1 Hasbrouck Information Shares

The underlying framework of Hasbrouck (1995) considers one security traded at multiple markets simultaneously with different prices $\mathbf{p}_t = (p_{1,t}, ..., p_{n,t})$. In this article I shed light on assets traded at Xetra and Lang & Schwarz, so in the following $n = 2$ markets are considered. Hasbrouck (1995) further assumes that prices for the same asset at different venues are only variations of one efficient market price m_t, and their deviations from the efficient price can be attributed to the different information available at the different markets in time t. Therefore, the price framework of two markets can be described as the following,

$$\begin{aligned} p_{1t} &= m_t + s_{1t} \\ p_{2t} &= m_t + s_{2t} \end{aligned} \tag{1}$$

where the efficient price follows a typical random walk, $m_t = m_{t-1} + v_t$, with v_t as the serially uncorrelated innovations of the efficient price, and s_{it} as the individual market-specific microstructure effects (Grammig and Peter, 2013). As usual for asset prices, each price series is nonstationary while the price changes are assumed to be stationary (integrated of order one). The *law of one price* then suggests a cointegrating relation between the two time series, i.e. although the processes themselves are not stationary, the linear combination of both is, due to

$$p_{1t} - p_{2t} = s_{1t} - s_{2t} \tag{2}$$

where the microstructure effects s_{1t} and s_{2t} are each market-specific innovations and stationary and therefore the linear combination of them is as well. Here, the theoretical

cointegrating vector is given as $\boldsymbol{\beta'} = (1, -1)$. Empirical estimations of the cointegrating vector in the setting of this paper are shown in the appendix section 7.1 and support the hypothesized cointegrated system.
The price series - or more specific: the price changes $\mathbf{\Delta p}_t$ - can further be described by the following vector error correction model (VECM) with lag q

$$\mathbf{\Delta p}_t \quad = \quad \boldsymbol{\alpha\beta'}\mathbf{p}_{t-1} + \mathbf{\Gamma}_1\mathbf{\Delta p}_{t-1} + ... + \mathbf{\Gamma}_{q-1}\mathbf{\Delta p}_{t-q+1} + \mathbf{u}_t, \tag{3}$$

where $\boldsymbol{\alpha}$ is a matrix of adjustment coefficients attributed to the cointegrating relation of the two prices $\boldsymbol{\beta'}\mathbf{p}_{t-1}$. $\mathbf{\Gamma}_1$ through $\mathbf{\Gamma}_{q-1}$ are the parameter matrices of the VECM and $\mathbf{u}_t = (u_{1,t}, u_{2,t})'$ are serially uncorrelated innovations with mean zero. However, the innovations may be contemporaneously correlated with $\mathbb{E}(\mathbf{u}_t\mathbf{u}_t') = \mathbf{\Sigma_u}$ as covariance matrix (see. Grammig and Peter (2013)).
The innovations of the efficient price v_t can then be linked to the VECM innovations by

$$v_t = \boldsymbol{\xi}'\mathbf{u}_t, \tag{4}$$

with $\boldsymbol{\xi'}$ being one of the identical rows of

$$\mathbf{\Xi} \quad = \quad \boldsymbol{\beta}_\perp \left[\boldsymbol{\alpha'}_\perp \Big(\mathbf{I}_n - \sum_{i=1}^{q-1}\mathbf{\Gamma}_i\Big)\boldsymbol{\beta}_\perp\right]^{-1} \boldsymbol{\alpha'}_\perp, \tag{5}$$

where $\perp$ indicates the orthogonal complements of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. The elements of $\boldsymbol{\xi}'$ then denote the permanent impact of an innovation $u_{i,t}$ on the efficient price m_t. A more detailed derivation of $\mathbf{\Xi}$ can be found in Lehmann (2002).
To allocate market information shares, the innovations of each market need to be related to the *true* innovations of the underlying asset. Therefore, the idiosyncratic innovations $\boldsymbol{\varepsilon_t} = (\varepsilon_{1,t}, \varepsilon_{2,t},)$ need to be derived for each market from the VECM and its correlated innovations $\mathbf{u}_t$. The exogenous innovations can be linked to the VECM via

$$\mathbf{u}_t = \mathbf{B}\boldsymbol{\varepsilon_t}. \tag{6}$$

The object of interest, the variance of the underlying efficient price innovations v_t, is given combining formula 4 and formula 6 to

$$\mathrm{Var}(v_t) \quad = \quad \mathrm{Var}(\boldsymbol{\xi}'\mathbf{u}_t) \quad = \quad \mathrm{Var}(\boldsymbol{\xi}'\mathbf{B}\boldsymbol{\varepsilon}_t) \quad = \quad \boldsymbol{\xi}'\mathbf{BB}'\boldsymbol{\xi}. \tag{7}$$

$\boldsymbol{\varepsilon}_t$ vanishes since $\boldsymbol{\varepsilon}_t \sim (\mathbf{0}, \mathbf{I}_n)$ by construction. To obtain $\mathbf{B}$, Cholesky factorization can be used in the manner that $\mathbf{\Sigma_u} = \mathbf{CC}'$ with $\mathbf{B} = \mathbf{C}$ and $\mathbf{\Sigma_u}$ being the covariance matrix of the residuals. The contribution of a market's innovation to the total innovation is then given by the Hasbrouck information share

$$\mathbf{IS} \quad = \quad \frac{[\boldsymbol{\xi}'\mathbf{C}]^{(2)}}{\mathrm{Var}(v_t)} \quad = \quad \frac{[\boldsymbol{\xi}'\mathbf{C}]^{(2)}}{\boldsymbol{\xi}'\mathbf{C}\mathbf{C}'\boldsymbol{\xi}}, \tag{8}$$

the proportion of the variances of the idiosyncratic innovations to the underlying efficient price innovations. Since the Cholesky decomposition depends on the ordering of rows in the origin matrix (here: the ordering of the residuals of the two markets), the lower triangular matrix $\mathbf{C}$ implies a certain kind of hierarchy. But conceptually, we do not know any relationship of this kind between the two markets. Therefore, two information shares are computed, each with permuted ordering in the covariance matrix of the residuals $\boldsymbol{\Sigma}_{\mathbf{u}}$. The two results can then be interpreted as estimates for an upper and a lower bound of the information share.

4.2 Unique Information Shares

As pointed out in section 3, the Hasbrouck information share entails two major drawbacks. First, the upper and lower bound only give a range for the markets' importance, and second, these bounds can diverge to extends which makes interpretations cumbersome, especially using low-frequency data. Since in the current use case only EOD-data (i.e. very low frequency data) is available for the main part of the study, having unique information shares at hand could provide further insides about the relationship of the markets of interest. In in the final results, unique information shares as introduced by Grammig and Peter (2013) are displayed as well. Since those are only used to support and challenge the results from the Hasbrouck information shares, I would refer those interested in the whole derivation to the study of Grammig and Peter (2013) and focus only on intuition and the specific calculation at this point.

The unique information shares rely on the assumption that financial market return data tends to be characterized by tail dependence (see e.g. Frahm et al. (2005)), a concept in which the correlation of two or more variables is higher in the tails of a distribution than in its center. Grammig and Peter (2013) then build the identification of idiosyncratic innovations on non-normal residuals from the VECM and derive the relationship between composite and idiosyncratic shocks as

$$\mathbf{u}_t \quad = \quad \mathbf{B}\boldsymbol{\varepsilon_t} \quad = \quad \mathbf{W}\mathbf{e}_t, \tag{9}$$

where $\mathbf{W}$ represents a weighting matrix and $\mathbf{e}_t$ a vector of contemporaneously and serially uncorrelated random variables. Those now do not have unit variances but rather result from two random variables

$$\mathbf{e}_t \quad = \quad \begin{cases} \mathbf{e}_{1,t} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n) & \text{with probability } \gamma \\ \mathbf{e}_{2,t} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Psi}) & \text{with probability } 1-\gamma \end{cases}, \tag{10}$$

where the elements of the covariance matrix $\mathbf{\Psi}$ are referred to as regime variances. Then, the covariance matrix of $\mathbf{e}_t$ is denoted as

$$\mathbf{\Sigma}_{\mathbf{e}} \quad = \quad \gamma \mathbf{I}_n + (1-\gamma)\mathbf{\Psi}, \tag{11}$$

such that

$$\mathbf{\Sigma}_{\mathbf{u}} \quad = \quad \mathbf{B}\mathbf{B}' \quad = \quad \mathbf{W}\mathbf{\Sigma}_{\mathbf{e}}\mathbf{W}'. \tag{12}$$

The formula to compute the information share is then similar to formula 8, with the extnsion that the covariance matrix of the residuals is now accompanied by the weighting matrix $\mathbf{W}$

$$\mathbf{IS}_T(\theta_m, \theta_v) \quad = \quad \frac{\left[\boldsymbol{\xi}'\mathbf{W}\mathbf{\Sigma}_{\mathbf{e}}^{0.5}\right]^{(2)}}{\boldsymbol{\xi}'\mathbf{W}\mathbf{\Sigma}_{\mathbf{e}}\mathbf{W}'\boldsymbol{\xi}}. \tag{13}$$

As in the previous section, the $\boldsymbol{\xi}$-coefficients are obtained as in formula 5 using the parameters of the VECM. The unknown parameters $\{\gamma, \mathbf{\Psi}, \mathbf{W}\}$ can be obtained using maximum likelihood estimation. The conditional log-likelihood function is set in the following way

$$\begin{aligned} \mathcal{L}(\theta_m, \theta_v) \quad = \quad & \sum_{t=1}^{T} \ln\Bigg(\gamma \times (2\pi)^{-\frac{n}{2}} \det(\mathbf{W})^{-1} \exp\Bigg\{-\frac{\mathbf{u}_t'(\mathbf{W}\mathbf{W}')^{-1}\mathbf{u}_t}{2}\Bigg\} \\ & + (1-\gamma) \times (2\pi)^{-\frac{n}{2}} \det(\mathbf{\Psi})^(-0.5) \det(\mathbf{W})^{-1} \\ & \times \exp\Bigg\{-\frac{\mathbf{u}_t'(\mathbf{W}\mathbf{\Psi}\mathbf{W}')^{-1}\mathbf{u}_t}{2}\Bigg\}\Bigg). \end{aligned} \tag{14}$$

In the estimation process, $\mathbf{u}_t$ is replaced by the residuals from the VECM. Maximization with respect to the parameters $\{\gamma, \mathbf{\Psi}, \mathbf{W}\}$ yields all necessary ingredients to compute unique information shares as shown in formula 13.

4.3 Data

Stock quotes for EOD-data were obtained from ariva.de for the period from 01.01.2018 until 30.12.2021. Closing prices of various exchanges are provided and the quotes of Lang & Schwarz and Xetra were used. Using 31.12.2019 at a cutoff point, each information share was calculated using approximately 500 observations for each *side* of the analysis. Data was collected for 38 out of the 40 Dax companies listed effectively after September 2021. The companies Siemens Energy and Siemens Healthineers had

to be excluded due to their rather recent IPOs and the lack of quotes before these days. Information on the EV/sales-ratio for 2021 was obtained from boersengefluester.de.
Data on 1-minute intervals for a subset of the Dax companies was obtained using webscraping. Webscraping can be a very sensitive topic and I *strongly* recommend taking a look at the appendix section 7.4 where I outline that no privacy and property rights were violated in the process.
The Lang & Schwarz quotes were scraped directly from its website ls-tc.de. Within the individual company sides (e.g. for the *Daimler stock*), Lang & Schwarz provides a panel *Quotes* which displays bid and ask prices on high frequencies. When obtaining the underlying HTML code for the website, these quotes are aggregated to the full minute. The average between the bid and the ask price was used for the final calculation.
Data on stock quotes for the Xetra exchange was scraped from finanzen.net. In the section *Times and Sales* (e.g. for the *Daimler stock*) all trades are listed that were executed at the Xetra exchange down to the second for a certain interval, accompanied by the volume and the relevant price. Using all trades that were executed within a minute and averaging the prices of those give reliable estimates for the Xetra prices.
Around 4,000 observations could be obtained for each stock in the analysis. However, webscraping can be error-prone and not every stock might cover each minute of every day within the sample period. But the dataset can definitely be seen as a reliable source for the specific task. The scraping process was conducted over a two week period between December 8th and December 23rd. Queries were sent every minute to both data providers during trading hours. The process was conducted on a virtual machine running in the BW-Cloud.
All code and data can be accessed on GitHub under *github.com/Lnrdbrgr*.

5 Results

5.1 Low-Frequency Data

Table 1 shows information shares for 38 Dax companies two years before and table 2 2 years after the rollout of the Trade Republic app in Germany. The cutoff point was set to December 31st, 2019. The lag chosen to estimate the VECM and its residuals was set to $q = 2$. An analysis of the best fitting number of lags can be found in the appendix section 7.2. To assess the hypothesis that Trade Republic user primarily buy and sell stocks that trade at higher multiples, the companies were ranked according to their EV/sales-ratio in 2021. The Hasbrouck information shares are displayed as upper and lower bounds as described in section 4.1. Following the literature, the median value between these bounds is shown as well.
Looking at the information shares prior to 2020 it can be seen that the intervals dis-

play a huge range. This shows the uncertainty of the method about the information advantage of each of the markets. Looking at the two extreme examples Siemens and Covestro, the upper and lower bounds reach from 0 to 1, indicating that the information share of the Lang & Schwarz exchange in this period could be anything between 0% to 100%. This shows the possibility that all information is generated at only one market. Companies with a smaller range where the lower bound starts around 4%-6% and the upper bound ends around 94%-96% include Vonovia, Puma, Volkswagen, and Daimler. This leads to the conclusion that at least *some* information is generated at the Lang & Schwarz exchange while some information can be attributed to Xetra. The majority of the lower Hasbrouck information shares start at 1%-3% while the upper bound mostly ranges from 97%-99%. Accordingly, the median information shares do not draw a more precise picture. This leaves only limited interpretability for the analysis. Very high or very low bounds of the information shares might not be a problem, as long as the tendency towards one side is given. However, the bounds here seem very balanced such that the median information share exclusively lies at 50% for all stocks, with very little deviations. This shows again that, according to the Hasbrouck information shares, neither Xetra nor Lang & Schwarz could claim an information advantage for any stock in the period from 2018-2019.

In the two-year period after 2019 the overall picture remains the same (see table 2). Extreme examples (0-1 range) are missing, but EON, Deutsche Bank and Beiersdorf are close. Surprisingly, the range for Vonovia was smallest before 2020 while afterwards it does not seem feasible to attribute at least some information to any of the two markets. On the other hand, while none of the exchanges claimed any information for Covestro before 2019, afterwards the results suggest that at least 10% of the price discovery process can be linked to either exchange, leading to the conclusion that not one market place grabs all the attention of informed investors. The same applies to a somewhat lesser extend to Sartorius, Zalando, Brenntag, Fresenius, and Daimler. However, a clear pattern across industries or company size is missing. The majority of information shares - again - lies within the interval from 1%-99% and leaves therefore only limited room for interpretation.

To address the research questions whether traders trading via Trade Republic at the Lang & Schwarz exchange are more informed about companies with high multiples, one might consider the correlation between the (median) post information share with the EV/Sales-ratio. The Pearson correlation lies at -0.0086, indicating that there is no such relationship between the valuation of a company and informed trading at either

[4] As part of the Volkswagen Konzern and due to a large holding of Volkswagen stocks, the Porsche stock trades at significantly higher EV/Sales-ratio than usually expected.

[5] The optimization algorithm failed to find meaningful parameters for the Linde stock in the pre-period for the unique information share measure.

Reference Entity	Hasbrouck IS Lower	Upper	Median	Unique IS	EV/Sales-Ratio
Porsche	0.0367	0.9684	0.5025	0.6976	161.43[4]
Sartorius	0.0022	0.9976	0.4999	0.7340	17.54
Vonovia	0.0404	0.9603	0.5004	0.3931	11.97
Delivery Hero	0.0014	0.9988	0.5001	0.7361	9.95
Deutsche Boerse	0.0003	0.9997	0.5000	0.7382	7.94
Qiagen	0.0243	0.9859	0.5051	0.6589	7.57
Linde	0.0334	0.9700	0.5017	(-)[5]	7.45
Merck	0.0001	0.9999	0.5000	0.6949	5.63
SAP	0.0035	0.9968	0.5002	0.6294	5.61
Symrise	0.0098	0.9917	0.5007	0.4796	5.17
Infineon	0.0002	0.9999	0.5000	0.7808	4.81
Beiersdorf	0.0035	0.9968	0.5002	0.7651	3.24
HelloFresh	0.0032	0.9971	0.5002	0.6375	3.13
Puma	0.0470	0.9529	0.5000	0.4952	3.10
Adidas	0.0028	0.9972	0.5000	0.6227	2.45
MTU	0.0043	0.9964	0.5003	0.8784	2.41
Zalando	0.0193	0.9837	0.5015	0.8330	2.34
Siemens	0.0000	1.0000	0.5000	0.7546	2.08
RWE	0.0308	0.9694	0.5001	0.6485	1.83
Airbus	0.0034	0.9971	0.5003	0.8918	1.76
Henkel	0.0067	0.9939	0.5003	0.5947	1.59
Bayer	0.0114	0.9914	0.5014	0.8668	1.12
Deutsche Post	0.0005	0.9996	0.5000	0.7407	1.05
Brenntag	0.0383	0.9662	0.5023	0.6293	1.04
Covestro	0.0000	1.0000	0.5000	0.8045	0.98
BASF	0.0239	0.9769	0.5004	0.4166	0.96
Deutsche Bank	0.0055	0.9951	0.5003	0.3949	0.95
Fresenius Medical Care	0.0009	0.9992	0.5000	0.7145	0.94
Deutsche Telekom	0.0165	0.9836	0.5000	0.2961	0.80
Heidelberg Cement	0.0096	0.9917	0.5007	0.6369	0.67
MunichRe	0.0025	0.9972	0.4998	0.9293	0.66
Allianz	0.0022	0.9977	0.4999	0.9482	0.60
BMW	0.0295	0.9728	0.5011	0.3916	0.58
Fresenius	0.0085	0.9913	0.4999	0.8647	0.55
EON	0.0089	0.9911	0.5000	0.5925	0.53
Volkswagen	0.0434	0.9600	0.5017	0.4259	0.51
Continental	0.0369	0.9635	0.5002	0.3430	0.49
Daimler	0.0615	0.9385	0.5000	0.7468	0.47

Table 1: Information shares on low frequency data 2018-2019

exchange.

To look at how trading activity changed after the market entry of Trade Republic, table 3 also displays the change of the medium Hasbrouck information share in the pre- and post-era. Positive values indicate a shift of information towards Lang & Schwarz, negative values suggest that informed traders might have moved to the Xetra exchange. At first notice, the changes are small to an extend that they are almost negligible. All differences stay below one percentage point. Therefore, it cannot be said that more informed trading happened at Lang & Schwarz after a new generation of traders joined the market through the neobroker. This possibly supports the thesis of Frijns et al. (2015) who find that once a market obtained a share in the price discovery process it tends to stay at this level. When relating the change in information shares to the valuation of the companies, the Pearson correlation lays at -0.3843. With meaningful information shares, this result would contradict the research question and suggest that traders involved in the price discovery process of highly priced companies moved to the Xetra market after 2020. However, as said before, the change in information shares is of only limited interpretability and therefore the correlation seems more like a random outcome than a meaningful metric.

Reference Entity	Hasbrouck IS Lower	Upper	Median	Unique IS	EV/Sales-Ratio
Porsche	0.0275	0.9734	0.5004	0.4724	161.43
Sartorius	0.0935	0.9056	0.4996	0.8965	17.54
Vonovia	0.0006	0.9995	0.5000	0.7579	11.97
Delivery Hero	0.0321	0.9689	0.5005	0.5093	9.95
Deutsche Boerse	0.0063	0.9949	0.5006	0.6735	7.94
Qiagen	0.0180	0.9858	0.5019	0.4967	7.57
Linde	0.0327	0.9739	0.5033	0.6367	7.45
Merck	0.0129	0.9878	0.5003	0.6590	5.63
SAP	0.0151	0.9862	0.5007	0.6392	5.61
Symrise	0.0170	0.9857	0.5014	0.4877	5.17
Infineon	0.0134	0.9875	0.5004	0.6196	4.81
Beiersdorf	0.0008	0.9992	0.5000	0.7327	3.24
HelloFresh	0.0035	0.9969	0.5002	0.8314	3.13
Puma	0.0047	0.9962	0.5004	0.6623	3.10
Adidas	0.0027	0.9974	0.5000	0.6464	2.45
MTU	0.0041	0.9959	0.5000	0.7020	2.41
Zalando	0.0508	0.9520	0.5014	0.4292	2.34
Siemens	0.0202	0.9791	0.4997	0.3990	2.08
RWE	0.0444	0.9574	0.5009	0.5322	1.83
Airbus	0.0217	0.9779	0.4998	0.4883	1.76
Henkel	0.0034	0.9970	0.5002	0.6867	1.59
Bayer	0.0314	0.9712	0.5013	0.3866	1.12
Deutsche Post	0.0106	0.9894	0.5000	0.4644	1.05
Brenntag	0.0640	0.9384	0.5012	0.4175	1.04
Covestro	0.1007	0.9030	0.5019	0.5680	0.98
BASF	0.0160	0.9841	0.5000	0.4822	0.96
Deutsche Bank	0.0003	0.9997	0.5000	0.7257	0.95
Fresenius Medical Care	0.0016	0.9987	0.5002	0.7466	0.94
Deutsche Telekom	0.0082	0.9927	0.5004	0.3303	0.80
Heidelberg Cement	0.0238	0.9769	0.5003	0.4486	0.67
MunichRe	0.0127	0.9879	0.5003	0.5794	0.66
Allianz	0.0175	0.9828	0.5002	0.6583	0.60
BMW	0.0238	0.9766	0.5002	0.3702	0.58
Fresenius	0.0779	0.9228	0.5004	0.6027	0.55
EON	0.0007	0.9994	0.5001	0.4209	0.53
Volkswagen	0.0072	0.9931	0.5002	0.6202	0.51
Continental	0.0409	0.9598	0.5003	0.3525	0.49
Daimler	0.0561	0.9440	0.5001	0.4100	0.47

Table 2: Information shares on low frequency data 2020-2021

Why were there not any meaningful results in the analysis of (medium) Hasbrouck information shares? As others pointed out before, applying the Hasbrouck method on low-level frequency data usually results in large bounds and less meaningful results (see e.g. Narayan and Smyth (2015)). I can confirm this observation, as for now only EOD-data was used. In addition, logically it should probably not be possible to detect an information advantage at one of two markets within the same country using daily closing prices. For example, consider the meaning of an information advantage at the Lang & Schwarz exchange that would still systematically be visible after a whole day of trading compared to the prices at Xetra. The arbitrage opportunities would be limitless and the concept of stocks trading at different venues questionable. Also, traders using Trade Republic might in the end not be as informed as assumed, preventing a shift in the price discovery process after 2019. Young people might consume more information on the internet in higher frequency while investing in highly priced stocks (see Frino et al. (2019)), but that does not necessarily mean they use the information in a right manner at the right time. No meaningful shifts in the information shares support this hypothesis.

To overcome some of the obstacles of the Hasbrouck information share I also calculated

Reference Entity	Hasbrouck IS Median Change	Unique IS Change	EV/Sales-Ratio
Porsche	-0.0021	-0.2252	161.43
Sartorius	-0.0003	0.1625	17.54
Vonovia	-0.0003	0.3648	11.97
Delivery Hero	0.0004	-0.2268	9.95
Deutsche Boerse	0.0006	-0.0647	7.94
Qiagen	-0.0032	-0.1621	7.57
Linde	0.0016	(-)	7.45
Merck	0.0003	-0.0359	5.63
SAP	0.0005	0.0098	5.61
Symrise	0.0006	0.0081	5.17
Infineon	0.0004	-0.1611	4.81
Beiersdorf	-0.0001	-0.0324	3.24
Hello Fresh	0.0000	0.1938	3.13
Puma	0.0005	0.1671	3.10
Adidas	0.0001	0.0238	2.45
MTU	-0.0003	-0.1764	2.41
Zalando	-0.0001	-0.4038	2.34
Siemens	-0.0003	-0.3556	2.08
RWE	0.0008	-0.1163	1.83
Airbus	-0.0004	-0.4035	1.76
Henkel	-0.0001	0.0920	1.59
Bayer	-0.0001	-0.4802	1.12
Deutsche Post	-0.0001	-0.2763	1.05
Brenntag	-0.0011	-0.2118	1.04
Covestro	0.0019	-0.2365	0.98
BASF	-0.0004	0.0656	0.96
Deutsche Bank	-0.0003	0.3308	0.95
Fresenius Medical Care	0.0001	0.0322	0.94
Deutsche Telekom	0.0004	0.0342	0.80
Heidelberg Cement	-0.0003	-0.1883	0.67
MunichRe	0.0005	-0.3499	0.66
Allianz	0.0003	-0.2899	0.60
BMW	-0.0009	-0.0215	0.58
Fresenius	0.0005	-0.2620	0.55
EON	0.0001	-0.1716	0.53
Volkswagen	-0.0015	0.1943	0.51
Continental	0.0001	0.0096	0.49
Daimler	0.0001	-0.3369	0.47

Table 3: Changes in information shares pre and post 2020

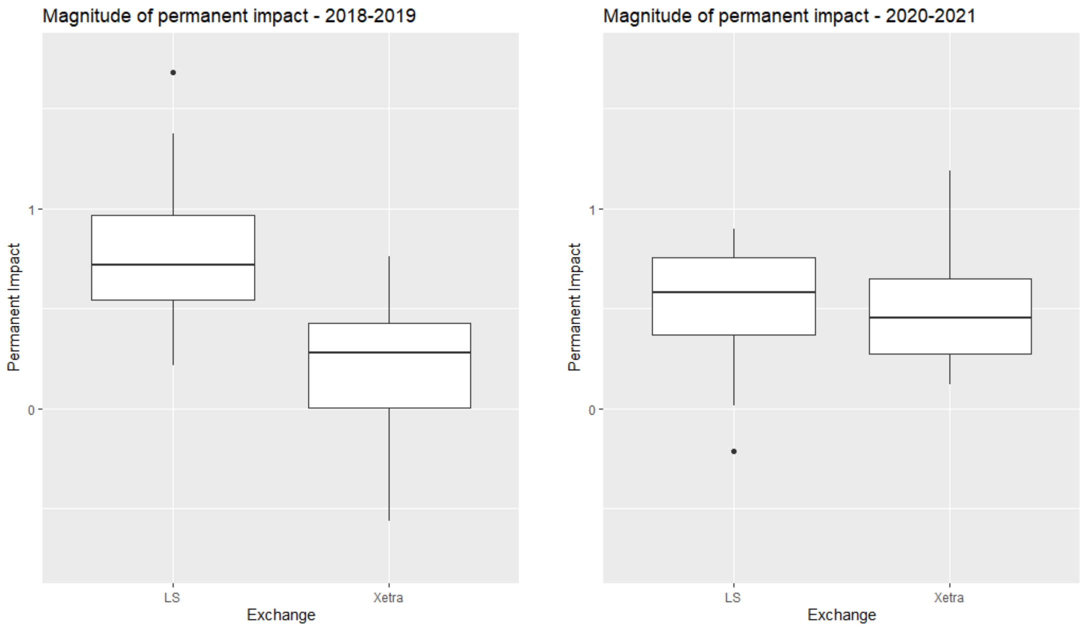
unique information shares as described by Grammig and Peter (2013) and as shown in section 4.2. The unique information shares draw a more biased picture between the two exchanges. The highest information shares pre 2020 are attributed to the two Dax-listed insurance companies Allianz (94.82%) and MunichRe (92.93%) (see table 1). Information shares of over 90 percent suggest that almost all agents in the market that are well informed about the German insurance environment choose to trade at Lang & Schwarz rather than Xetra. Considering the pure size and importance of the Xetra exchange, these large information shares are at least questionable. The striking results for insurance companies cannot sufficiently be explained at this moment. Lang & Schwarz seems to play the smallest part in the price discovering process for Deutsche Telekom (29.61%), Continental (34.30%), and Vonovia (39.31%). The mean (65.96%) and the median (69.49%) information share across all companies from 2018 to 2019 suggest that the Lang & Schwarz exchange plays a more important role in the price discovery process for most Dax companies than the Xetra exchange. 28 out of 42 companies exhibit an information share above 50%. I did not expect that to be the case. Again, considering the national and international importance of Xetra, one would assume that market-moving news would rather be incorporated through the biggest German exchange first. There does not seem to be any relation between the information shares obtained by the Hasbrouck method and the unique information shares.

To address whether the price discovery process changed after 2019, unique information shares were computed for this time as well. The biggest information shares are now attributed to Sartorius (89.65%) and Vonovia (75.79%). The smallest information shares of the sample stayed at Deutsche Telekom (33.03%) and Continental (35.25%), in line with the results for both companies before the cutoff point. Xetra seems to play the bigger role for both stocks nonetheless the time. All in all, the unique information shares fluctuate less after 2020 than they did before suggested by a smaller standard deviation (0.17 before vs. 0.14 after). Also, the mean (56.69%) and median (57.37%) information share as well as 22 companies with information shares about 50% suggest that Xetra gained a more important role in the price discovery process, contradicting the research question.

To assess whether the valuation of stocks plays a role in where informed traders trade their stocks, the correlation between the (post) unique information shares and the EV/sales-ratio is computed as before. A large positive correlation would suggest that traders informed about highly priced companies rather trade at the Lang & Schwarz exchange. However, the correlation of -0.0398 between the two variables does not show a relation as suggested by the research question. The negative sign rather suggests the opposite but - as before - values that close to zero do not carry enough meaning to come to a final conclusion.

Looking at how price discovery changed over time with respect to the unique information shares, it is noted that the lowest information shares again stay with Deutsche Telekom and Continental, following the hypothesis of Frijns et al. (2015) that high information shares tend to stay at certain markets. Both Allianz and MunichRe lose parts of their high share but results suggest that the majority of information is still generated at Lang & Schwarz. Large changes in information shares for Bayer (-48.02pp), Airbus (-40.35pp), and Zalando (-40.38pp) are questionable in their interpretation and should be viewed carefully. It seems somewhat contra-intuitive that the role in the price discovery process changes at these scales over such a short period. The correlation between the changes in the information shares and the EV/sales-ratio is -0.0583, suggesting that the new retail traders did not bring more information about *Hot-Stocks* compared to *Blue-Chips*.

The results of the unique information shares should be viewed with caution for various reasons. First, some of the magnitudes of the effects simply seem not plausible. Huge changes in the pre- and post-era as for Bayer or very large information shares such as the one for Allianz and MunichRe do not seem feasible given general knowledge



(a) Period 2018-2019 (b) Period 2020-2021

Figure 2: Distribution of permanent impacts before and after the cutoff

about the price discovery process and the importance of Lang & Schwarz compared to Xetra. Second, as mentioned before, finding significant information advantages using EOD-data at a certain market could theoretically result in arbitrage possibilities which intuitively cannot exist to this extent in two German markets. Third, Grammig and Peter (2013) introduce two specification tests to assess whether tail dependence is actually present in the data. Tail dependence is the crucial assumption in the unique information share setting. However, these tests were not conducted at this point and the possibility remains that the essential assumption might be violated for some stocks. Fourth, the calculation of unique information shares requires maximization of a conditional log-likelihood function with respect to seven parameters. During the estimations, the maximization algorithm had some problems finding meaningful extrema for some of the stocks in the period before 2020 (oddly only in the pre-period). Fifth and lastly, recalling the findings of Lien and Wang (2016), the unique information shares do not always produce reliable estimates for the price discovery process and the Hasbrouck method should be preferred.

Recall from section 4 that during the process of calculating the information shares, the Ξ-matrix was derived which displayed the permanent impact of an innovation in a market on the underlying efficient price. Figure 2 shows the distributions of permanent impacts of the Dax companies before and after the cutoff. Appendix 7.3 shows the exact values for each stock. It seems to be the case that the permanent impact of a shock at the Lang & Schwarz exchange overall contributed more to the efficient price

before 2020 than a shock in the Xetra exchange. The differences diminish somewhat after the cutoff and the Xetra exchange catches up. These results are not supported by theory given the importance and size of the two exchanges. A reason could be - again - the usage of EOD-data. A significant shock in the efficient price, i.e. a shock that is *not* attribute to market specific characteristics such as spreads, will most likely be incorporated in the price of the other market within seconds. It is highly unlikely that daily closing prices will reflect these movements in a representative way.
The last result of the analysis of low-frequency data is displayed in table 4. GameStop and AMC attracted the most attention within the last two years in the "battle" of young traders versus established investors. Neobroker like Trade Republic or Robinhood even halted trading with these stocks at some point due to unpredictable volatility[6]. Together with Tesla, these three stocks were on top of the list of Reddit discussions, with Tesla even being the stock most held by Robinhood users (Tokic (2020)). Although for these stocks the majority of the price discovery process probably took place in the US, I felt like I should not miss out at having a look at *the Hot-Stocks* of 2021 when talking about the disruption of traditional markets by neobroker.

However, as displayed in table 4, the results suffer from the same teething problems

	Hasbrouck IS Pre			Pre	Hasbrouck IS Post			Post	Hasbrouck IS	Unique IS
Reference Entity	Lower	Upper	Median	Unique IS	Lower	Upper	Median	Unique IS	Change	Change
AMC	0.0562	0.9972	0.5267	0.4824	0.0087	0.9949	0.5018	0.4931	-0.0249	0.0107
GameStop	0.0006	0.9999	0.5002	0.4367	0.1494	0.9189	0.5342	0.5743	0.0339	0.1376
Tesla	0.0914	0.9725	0.5319	0.8997	0.0133	0.9955	0.5044	0.9588	-0.0275	0.0591

Table 4: Information shares on US-*Hot-Stocks*

as the analysis of the Dax companies before. For AMC and Tesla, the Hasbrouck information shares show a wide range of upper and lower bounds and the change in the median as well as in the unique information share might be negligible. A different result is drawn from the GameStop stock. The Hasbrouck information share could not spot a particular place for price discovery before the cutoff (50.02%) and the unique information share spotted a small tendency towards the Xetra exchange (43.67%). However, the picture changes after the cutoff. I want to note that the "hype" around the financial discussions on Reddit started with GameStop and the stock suddenly started trading at ridiculously high valuations combined with huge volatility. It can be said without too much speculation that no sane long-term investor would have bought this stock in the last two years. However, demand exploded among apps like Robinhood and Trade Republic and if an effect of information advantage on a stock would be expected among these apps, it would be GameStop - because people started to buy the stock simply

[6]For a nice recap of the story around GameStop and the clash of the Reddit community versus short-sellers I recommend the introduction of Welch (2021).

because they liked the company, although fundamentals looked terrible. Anyway, looking at the change in information shares after the cutoff point, both measures of price discovery agree that Lang & Schwarz gained advantage over Xetra. Comparing the Hasbrouck measurement, the lower bound lies at almost 15%, a value to which each of the Dax companies could not even get close to. Also, the shift in the median towards Lang & Schwarz amounts to 3.39 percentage points. Compared to the shifts in the Dax a large value. This result could indicate that German Trade Republic users *teamed up* with American Robinhood users and pushed the price discovery process through the Lang & Schwarz exchange towards the rest of the market. Even with keeping in mind the limited interpretability and no guarantee for causal effects, an interesting result.

5.2 High-Frequency Data

One of the major drawbacks mentioned several times already is the behavior of the Hasbrouck information share when dealing with low-frequency data. To overcome this, I also analyze stock market data at 1-minute intervals for a sample of Dax companies as described in section 4.3. As data was scraped from the web just recently, the time component cannot be evaluated and the interpretation of the data can only address the second research question whether the price discovery process of companies that trade at high multiples predominantly takes place at the Lang & Schwarz exchange.

One would expect that the bounds of the information shares now converge and - if an information advantage of one market over the other is present - the median information share tends towards one of the two markets. And indeed, looking at the lower bounds it can be seen that only for Bayer, BASF, and Daimler no significant share of the price discovery process can be attributed to Lang & Schwarz. For the other stocks, minimum information shares start at around 5% (Delivery Hero) and go up to 19% (EON). These values are considerably higher compared to the lower frequency information shares, concluding that the uncertainty of the method diminishes with increasing accuracy of the data. The same accounts for the upper bounds. While for the low frequency data, upper bounds were mostly between 97%-99%, the upper bounds now mainly range from 80% (EON) to 95% (Delivery Hero).

The median information shares are again quite balanced around 50%. Deviations can only be found in the decimals when looking at percentages. However, the correlation between the information shares of the Lang & Schwarz Exchange and the EV/sales-ratio of the stocks lies at 0.3160 this time. Compared to the correlation coefficients before a significant increase, indicating a small positive relation between informed trading at Lang & Schwarz and the valuation of companies. This result gives the first minor confirmation of one of the research questions, namely that Trade Republic user might trade more extensively and informed shares at higher multiples through Lang & Schwarz.

Reference Entity	Hasbrouck IS Lower	Upper	Median	EV/Sales-ratio
Vonovia	0.1756	0.8290	0.5023	11.97
Delivery Hero	0.0505	0.9496	0.5001	9.95
SAP	0.0632	0.9385	0.5009	5.61
Infineon	0.0582	0.9438	0.5010	4.81
Hello Fresh	0.1640	0.8380	0.5010	3.13
Zalando	0.0596	0.9423	0.5010	2.34
Siemens	0.0798	0.9217	0.5008	2.08
Bayer	0.0007	0.9985	0.4996	1.12
Covestro	0.0827	0.9191	0.5009	0.98
BASF	0.0053	0.9995	0.5024	0.96
Deutsche Bank	0.0995	0.9027	0.5011	0.95
EON	0.1926	0.8054	0.4990	0.53
Daimler	0.0146	0.9854	0.5000	0.47

Table 5: Information shares using high frequency data

The results again should not be taken for granted. Although the correlation result seems too large to be a random coincidence, the tendencies in the median information shares are again quite small and might be subject to noise in the data. Second, the same argument from before applies when it comes to markets that closely linked. Identifying a significant information advantage of one German market over the other does not seem plausible. Even though data is sampled at smaller intervals, high-frequency and machine-trading could exploit systematic information advantages within seconds and the effect would probably not be visible anymore in 1-minute intervals. Also, although many datapoints are available for the given stocks, the period observed was rather short and the question of representativeness remains.

6 Conclusion

I evaluated the price discovery process of two German exchanges before and after the market entry of the neobroker Trade Republic. The two hypotheses to test were the following. First, the market venue through which Trade Republic conducts its trades played a larger role in the price discovering process after a new group of young traders joined the market in the beginning of 2020. Second, the user of the app might be subject to another information generating process and rather trade *Hot-Stocks* than *Blue-Chips*. Therefore, the information share of the market of interest should be higher for higher priced stocks, represented by the EV/sales-ratio. The analysis was conducted on two datasets using the Dax companies, one consisting of daily closing prices and one where quotes were sampled at one minute intervals. Hasbrouck information shares and a method to estimate unique information shares introduced by Grammig and Peter (2013) were used as proxies for the price discovery process.
The findings using EOD-data were in line with the literature such that the Hasbrouck method delivers only vague estimates for information shares when dealing with low-frequency data. Intervals were mostly too large to obtain a meaningful metric that could show the information advantage of one market over the other. Both research

questions could neither be confirmed nor completely rejected. Younger traders did not seem to take over a significant market share after accessing the market via the Trade Republic app. I also could not find any relation between highly priced stocks and the information advantage of these investors.

Calculating unique information shares resulted in estimates that were not as centered as the ones before. However, there was also no evidence that the market was disrupted by the neobroker, neither over time nor across certain stocks. The method of estimating unique information shares should be viewed with caution in this context for various reasons.

Looking at three American *Hot-Stocks* could confirm one of the hypotheses in some parts. Analysis of the much-noticed GameStop stock suggested that traders at Lang & Schwarz became better informed about the company after the market entry of Trade Republic.

The results of higher frequency data confirmed the assumptions about the Hasbrouck method that upper- and lower-bounds converge when using more accurate data. It was also found that the information share of the Lang & Schwarz exchange had a small positive correlation with the valuation of the stocks in the sample, confirming one of the research question in some parts. In general, results of information shares should increase in their interpretability and relevance with more accurate quote data.

Overall it can be said that the expected effects were not found. Younger traders did not systematically disrupt the price discovery process in Germany for the Dax stocks. Reasons for that could be that the amount of trader and the quantities traded are still too small to make a difference. However, if the information advantage would be indeed present for some stock this would contradict the findings of Agarwal et al. (2007) who stated that the size of the market does not matter in the price discovery process. It rather seems that the information advantage is not present, confirming the findings of Pasztor (2021) and Tokic (2020) who see younger retail trader as uninformed market participants and Frijns et al. (2015) who finds that information shares stay at the markets once they are established.

One of the biggest weaknesses of the study is its data. Although I tried collecting low-level datapoints, results were still not as meaningful as hoped. In a recent study, Hasbrouck (2021) uses data down to 10 microseconds, suggesting that the method performs more accurately on very high frequency data.

References

Sumit Agarwal, Chunlin Liu, and S. Ghon Rhee. Where does price discovery occur for stocks traded in multiple markets? evidenve from hong kong and london. *Journal of International Money and Finance*, 26(1):46–63, 2007.

Jeffrey M. Bacidore and George Sofianos. Liquidity provision and specialist trading in nyse-listed non-u.s. stocks. *Journal of Financial Economics*, 63(1):133–158, 2002.

Geir H Bjonnes, Carol L Osler, and Dagfinn Rime. Price discovery in two-tire markets. *International Journal of Finance and Economics*, 26(2):3109–3133, 2021.

Gabriel Frahm, Markus Junker, and Rafael Schmidt. Estimating the tail-dependence coefficient: Properties and pitfalls. *Insurance: Mathematics and Economics*, 37(1): 80–100, 2005.

Henry L. Friedman and Zitong Zeng. Retail investor trading and market reactions to earnings announcements. *Behavioral and Experimental Finance eJournal*, 2021.

Bart Frijns, Aaron Gilbert, and Alireza Tourani-Rad. The determinants of price discovery: Evidence from us-canadian cross-listed shares. *Journal of Banking and Finance*, 59:457–468, 2015.

Alex Frino, Grace Lepone, and Danika Wright. Are paper winners gamblers? evidence from australian retail investors. *Journal of Accounting and Finance*, 59:593–614, 2019.

Roland Fuess, Ferdinand Mager, Michael Stein, and Lu Zhao. Financial crises, price discovery, and information transmission: A high-frequency perspective. *Financial Markets and Portfolio Management*, 32:333–365, 2018.

Imen Ghadhab and Slaheddine Hellara. Price discovery of cross-listed firms. *International Review of Financial Analysis*, 44:177–188, 2016.

Joachim Grammig and Franziska J. Peter. Telltale tails: A new approach to estimating unique information shares. *Journal of Financial and Quantitative Analysis*, 48(2): 459–488, 2013.

Joachim Grammig, Michael Melvin, and Christian Schlag. Internationally cross-listed stock prices during overlapping trading hours: Price discovery and exchange rate effects. *Working Paper Series: Finance and Accounting, No. 78*, 2001.

Joel Hasbrouck. One security, many markets: Determining the contributions to price discovery. *The Journal of Finance*, 50(4):1175–1199, 1995.

Joel Hasbrouck. Price discovery in high resolution. *Journal of Financial Economics*, 19(3):395–430, 2021.

Danqi Hu, Charles M. Jones, Valerie Zhan, and Xiaoyan Zhang. The rise of reddit: How social media affects retail investors and short-sellers' roles in price discovery. *Behavioral and Experimental Finance eJournal*, 2021.

Bruce Lehmann. Some desiderata for the measurement of price discovery. *Journal of Financial Markets*, 5(3):259–276, 2002.

Donald Lien and Zijun Wang. Estimation of market information shares: A comparison. *The Journal of Futures Markets*, 36(11):1108–1124, 2016.

Seema Narayan and Russel Smyth. The financial econometrics of price discovery and predictability. *International Review of Financial Analysis*, 42:380–393, 2015.

Alexander Osipovich. Individual-invester boom reshapes u.s. stock market. *The Wall Street Journal*, 2020. URL `https://www.wsj.com/articles/individual-investor-boom-reshapes-u-s-stock-market-11598866200`.

Michael S. Pagano, John Sedunov, and Raisa Velthuis. How did retail investors respond to the covid-19 pandemic? the effect of robinhood brokerage customers on market quality. *Finance Research Letters*, 43, 2021.

James Pasztor. Robinhood - hero or villain? *Journal of Financial Service Professionals*, 75(5):18–22, 2021.

Roberto Stein. The top 5 predictable effects of new entries in robinhood's "100 most popular" list. 2020. URL `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3694588`.

Damir Tokic. Robinhoods and the nasdaq whale: The makings of the 2020 big-teck bubble. *Journal of Corporate Accounting and Finance*, 31:9–14, 2020.

Ivo Welch. The wisdom of the robinhood crowd. *Journal of Finance, Forthcoming*, 2021.

7 Appendix

7.1 Empirical Cointegrating Vectors

Table 6 and 7 show the result of the β-coefficient when regressing the returns of the Lang & Schwarz exchange on the returns of Xetra. Recall, to match the theoretical cointegrating relation $\boldsymbol{\beta}' = (1 - 1)$ the β-coefficient is supposed to be $\beta = 1$.

Reference Entity	β-coefficient
Adidas	1.000631
Airbus	1.000544
Allianz	0.996074
BASF	0.999499
Bayer	0.999840
Beiersdorf	0.996760
BMW	0.999499
Brenntag	1.001934
Continental	1.000777
Covestro	1.001351
Daimler	0.999173
Delivery Hero	1.000457
Deutsche Bank	0.999760
Deutsche Boerse	0.999698
Deutsche Post	1.000514
Deutsche Telekom	0.995606
EON	0.995180
Fresenius	1.000059
Fresenius Medical Care	0.999712
Heidelberg Cement	1.000106
Hello Fresh	1.000229
Henkel	0.999689
Infineon	1.001516
Linde	0.998635
Merck	1.001496
MTU	1.000439
MunichRe	0.999197
Porsche	1.003813
Puma	1.000851
Qiagen	1.000950
RWE	1.000626
SAP	0.997106
Sartorius	1.002642
Siemens	0.999799
Symrise	1.001517
Volkswagen	1.002179
Vonovia	0.999900
Zalando	1.001993
AMC	0.988077
GameStop	1.000595
Tesla	0.999364

Table 6: Empirical cointegrating vectors using low frequency data

Reference Entity	β-coefficient
BASF	0.867620
Bayer	1.086935
Covestro	0.914394
Daimler	0.980810
Delivery Hero	0.934089
Deutsche Bank	0.854185
EON	0.799233
Hello Fresh	0.846582
Infineon	0.911000
SAP	0.926948
Siemens	0.913744
Vonovia	0.837908
Zalando	0.921970

Table 7: Empirical cointegrating vectors using high frequency data

7.2 Model Selection

Table 8 shows the Akaike criterion for different specifications of the vector error correction model to identify the number of lags used for the final specification to obtain the residuals. Lang & Schwarz and Xetra indicate which return series was used as the independent variable.

	Lang & Schwarz			Xetra			Minimizing Lag	
Reference Entity	2 lags	3 lags	4 lags	2 lags	3 lags	4 lags	LuS	Xetra
Adidas	15166.7	15162.8	15163.8	15103.3	15102.8	15103.3	2	2
Airbus	5372.9	5374.3	5371.1	5362.6	5363.7	5359.3	3	3
Allianz	6055.5	6041.8	6027.7	6048.7	6036.1	6024.3	3	3
BASF	3579.9	3583.5	3578.2	3542.9	3544.6	3540.9	3	3
Bayer	4171.7	4172.3	4174.0	4063.7	4066.2	4066.4	1	1
Beiersdorf	3922.1	3921.0	3918.8	3859.3	3861.4	3861.0	3	1
BMW	3965.3	3961.0	3958.3	3921.5	3917.2	3914.0	3	3
Brenntag	3102.4	3099.9	3094.4	2928.0	2926.4	2921.2	3	3
Continental	6032.7	6034.4	6036.1	6007.8	6009.3	6012.1	1	1
Covestro	3871.7	3874.7	3874.1	3740.2	3743.4	3741.1	1	1
Daimler	3073.4	3076.3	3077.6	3039.7	3040.7	3041.6	1	1
Delivery Hero	4859.9	4863.6	4866.5	4800.0	4803.0	4804.9	1	1
Deutsche Bank	-105.7	-107.4	-110.1	-157.3	-163.8	-165.3	3	3
Deutsche Boerse	5082.3	5063.9	5063.6	4961.9	4947.4	4948.2	3	2
Deutsche Post	2003.6	2005.9	2006.9	1990.9	1990.9	1992.9	1	2
Deutsche Telekom	-720.7	-734.2	-734.5	-788.5	-804.5	-804.6	3	3
EON	-1517.2	-1521.1	-1531.9	-1612.1	-1612.3	-1618.3	3	3
Fresenius	3251.9	3253.9	3249.7	3253.7	3256.3	3252.9	3	3
Fresenius Medical Care	3927.3	3924.5	3927.8	3862.7	3862.9	3863.8	2	1
Heidelberg Cement	3794.2	3795.0	3792.5	3724.8	3726.0	3725.2	3	1
HelloFresh	3644.1	3646.8	3634.7	3571.3	3573.6	3560.0	3	3
Henkel	4053.3	4053.1	4054.2	3963.0	3964.7	3966.7	2	1
Infineon	1950.9	1952.3	1949.7	1875.9	1877.2	1874.7	3	3
Linde	4049.2	4043.5	4041.5	3924.1	3916.4	3916.1	3	3
Merck	5001.9	4999.2	5000.9	4949.4	4950.4	4952.8	2	1
MTU	7005.2	7008.9	7010.3	6992.9	6995.6	6997.1	1	1
MunichRe	6689.9	6690.1	6688.3	6651.3	6651.3	6653.5	3	1
Porsche	4181.3	4179.0	4180.9	4105.0	4104.9	4108.6	2	2
Puma	4197.6	4201.5	4205.0	4088.6	4089.9	4092.6	1	1
Qiagen	2235.2	2238.5	2241.6	1993.0	1996.2	1999.6	1	1
RWE	1731.4	1717.0	1706.0	1674.2	1663.7	1656.8	3	3
SAP	5097.3	5092.5	5089.3	5009.7	5003.1	5000.0	3	3
Sartorius	8875.5	8874.5	8875.8	8823.4	8825.9	8829.1	2	1
Siemens	4939.2	4942.0	4937.4	4959.7	4961.7	4959.3	3	3
Symrise	4072.2	4067.1	4069.0	3897.5	3899.3	3899.1	2	1
Volkswagen	6548.9	6551.4	6553.7	6487.6	6491.2	6493.7	1	1
Vonovia	2346.8	2341.1	2333.7	2307.3	2305.6	2299.7	3	3
Zalando	4277.8	4280.2	4284.1	4193.1	4193.3	4196.9	1	1
AMC	3112.3	3115.6	3113.7	2838.1	2838.6	2826.9	1	3
GameStop	8837.1	8748.5	8710.5	7810.1	7746.1	7695.5	3	3
Tesla	9947.3	9942.0	9920.5	9344.0	9343.5	9325.3	3	3
						Average:	2.24	2.02

Table 8: Akaike criterion on different specifications of the VECM

7.3 Permanent Impact Coefficients

Table 9 and 10 display the elements of the Ξ-matrix matrix for each company in the pre- and post-period. The elements of are ordered by column, i.e. the first two values represent the first column of the matrix, the last two values the second column.

	Matrix Elements			
Reference Entity	[1,1]	[2,1]	[1,2]	[2,2]
Adidas	1.1149	1.1149	-0.0126	-0.0126
Airbus	1.0582	1.0582	-0.0382	-0.0382
Allianz	0.9745	0.9745	0.0576	0.0576
BASF	0.5174	0.5174	0.3933	0.3933
Bayer	0.9384	0.9384	-0.0161	-0.0161
Beiersdorf	1.1789	1.1789	-0.1916	-0.1916
BMW	0.5858	0.5858	0.3164	0.3164
Brenntag	0.4927	0.4927	0.5254	0.5254
Continental	0.2126	0.2126	0.7595	0.7595
Covestro	0.8089	0.8089	0.2171	0.2171
Daimler	0.3244	0.3244	0.5482	0.5482
Delivery Hero	0.5520	0.5520	0.3432	0.3432
Deutsche Bank	0.8574	0.8574	0.1096	0.1096
Deutsche Boerse	0.7774	0.7774	0.2755	0.2755
Deutsche Post	0.9291	0.9291	0.0514	0.0514
Deutsche Telekom	0.6269	0.6269	0.3794	0.3794
EON	0.6050	0.6050	0.3401	0.3401
Fresenius	1.1993	1.1993	-0.2155	-0.2155
Fresenius Medical Care	0.6913	0.6913	0.2863	0.2863
Heidelberg Cement	0.5301	0.5301	0.4265	0.4265
HelloFresh	0.7431	0.7431	0.1651	0.1651
Henkel	0.5501	0.5501	0.4184	0.4184
Infineon	0.5613	0.5613	0.4002	0.4002
Linde	1.6776	1.6776	-0.5625	-0.5625
Merck	1.1651	1.1651	-0.1461	-0.1461
MTU	1.1480	1.1480	-0.0483	-0.0483
MunichRe	0.9200	0.9200	0.1540	0.1540
Porsche	0.2924	0.2924	0.6184	0.6184
Puma	0.5788	0.5788	0.4621	0.4621
Qiagen	1.3716	1.3716	-0.4059	-0.4059
RWE	0.3889	0.3889	0.5932	0.5932
SAP	0.8734	0.8734	0.1588	0.1588
Sartorius	0.8587	0.8587	0.2060	0.2060
Siemens	0.6630	0.6630	0.2805	0.2805
Symrise	0.5403	0.5403	0.4670	0.4670
Volkswagen	0.3151	0.3151	0.6319	0.6319
Vonovia	0.3668	0.3668	0.6360	0.6360
Zalando	1.1006	1.1006	-0.1345	-0.1345
AMC	1.0474	1.0474	-0.0434	-0.0434
GameStop	1.2151	1.2151	-0.1496	-0.1496
Tesla	1.6224	1.6224	-0.4725	-0.4725

Table 9: Elements of the Ξ-matrix 2018-2019

Reference Entity	Matrix Elements [1,1]	[2,1]	[1,2]	[2,2]
Adidas	0.5964	0.5964	0.3416	0.3416
Airbus	0.5765	0.5765	0.3512	0.3512
Allianz	0.4403	0.4403	0.4856	0.4856
BASF	0.5180	0.5180	0.4716	0.4716
Bayer	0.5838	0.5838	0.4461	0.4461
Beiersdorf	0.5271	0.5271	0.4189	0.4189
BMW	0.2065	0.2065	0.7873	0.7873
Brenntag	0.3367	0.3367	0.7045	0.7045
Continental	0.2914	0.2914	0.6887	0.6887
Covestro	0.0117	0.0117	1.0020	1.0020
Daimler	0.3251	0.3251	0.6393	0.6393
Delivery Hero	0.2966	0.2966	0.6488	0.6488
Deutsche Bank	0.8625	0.8625	0.1502	0.1502
Deutsche Boerse	0.6941	0.6941	0.3492	0.3492
Deutsche Post	0.8355	0.8355	0.2907	0.2907
Deutsche Telekom	0.8497	0.8497	0.2483	0.2483
EON	0.7817	0.7817	0.2111	0.2111
Fresenius	0.0497	0.0497	0.8707	0.8707
Fresenius Medical Care	0.8793	0.8793	0.1436	0.1436
Heidelberg Cement	0.3795	0.3795	0.5783	0.5783
HelloFresh	0.7540	0.7540	0.1884	0.1884
Henkel	0.6742	0.6742	0.4105	0.4105
Infineon	0.6213	0.6213	0.4823	0.4823
Linde	0.8954	0.8954	0.2804	0.2804
Merck	0.3683	0.3683	0.7014	0.7014
MTU	0.8368	0.8368	0.1167	0.1167
MunichRe	0.7528	0.7528	0.2512	0.2512
Porsche	0.3393	0.3393	0.5945	0.5945
Puma	0.8123	0.8123	0.2007	0.2007
Qiagen	0.7300	0.7300	0.2550	0.2550
RWE	0.3734	0.3734	0.6596	0.6596
SAP	0.7000	0.7000	0.4018	0.4018
Sartorius	-0.2144	-0.2144	1.1855	1.1855
Siemens	0.5437	0.5437	0.4748	0.4748
Symrise	0.3724	0.3724	0.7032	0.7032
Volkswagen	0.4610	0.4610	0.4890	0.4890
Vonovia	0.8473	0.8473	0.2679	0.2679
Zalando	0.5910	0.5910	0.4636	0.4636
AMC	0.9818	0.9818	0.2952	0.2952
GameStop	0.7353	0.7353	0.2280	0.2280
Tesla	1.0334	1.0334	-0.0490	-0.0490

Table 10: Elements of the Ξ-matrix 2020-2021

7.4 Webscraping

As described in section 4.3, stock quotes at 1-minute intervalls are obtained by scraping the websites of finanzen.net and ls-tc.de. By scraping content from other websites, the owner of the website does not usually give official permission to do so. Therefore, webscraper can sometimes enter a gray-area when it comes to property and ownership rights. Websites usually tell potential crawlers what parts of the websites are allowed to be read systematically by machines and which are not via a *robots.txt*-file. The asterisk tells the user what is allowed and what is not.

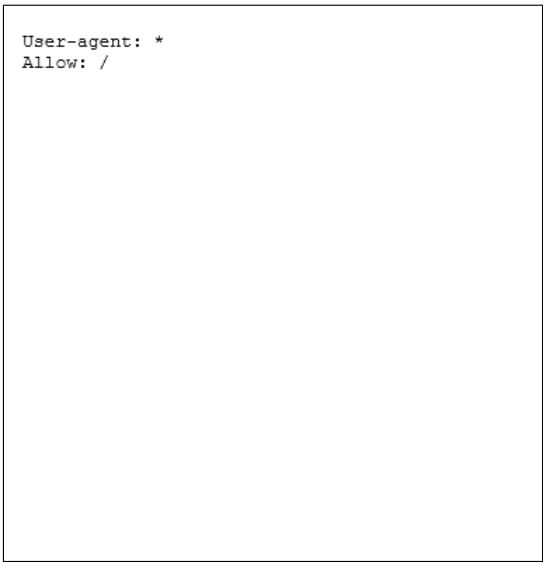

```
User-agent: *
Allow: /
```

Figure 3: *robots.txt*-file of the Lang & Schwarz webpage

The *robots.txt*-file of the Lang & Schwarz website can be accessed via https://www.ls-tc.de/robots.txt. The snapshot shown in Fig. 3 was taken on 23.01.2022. It can easily by seen that the owner explicitly allows all contents to be scraped without exceptions, therefore the scraping process was in line with all ownership guidelines.
Fig. 7 displays the *robots.txt*-file of finanzen.net which can be accessed via the url https://www.finanzen.net/robots.txt. Here, it seems somewhat harder to determine which access the owner grants and which not. During the data collection process, data was obtained from https://finanzen.net/timesandsales/* which would be forbidden for example for "zertifikate", "anleihen", and "optionsscheine", but apparently not for "aktien". The snapshot was also taken at the end of the data generating process on 23.01.2022, concluding that all data generation took place in accordance with the owner.

```
User-agent: Mediapartners-Google
Disallow:

User-agent: Flamingo_SearchEngine
Disallow: /

User-agent: *
Disallow: /*@nasdaq
Disallow: /*@nyse
Disallow: /*@inpopup
Disallow: /*@inRating
Disallow: /*@intpagenr
Disallow: /*@intPersonNr
Disallow: /*@lus
Disallow: /*@pkId
Disallow: /*@print
Disallow: /*@stBoerse
Disallow: /1154235/
Disallow: /adpreview/
Disallow: /ajax/
Disallow: /datenschutz
Disallow: /depot/
Disallow: /export4g/
Disallow: /fonds/suche?shareid
Disallow: /hinweise/mynews
Disallow: /holedaten.asp
Disallow: /impressum
Disallow: /kontakt
Disallow: /kurse/kurse_kursliste_detail.asp
Disallow: /mynews/
Disallow: /news/news_suchergebnis.asp
Disallow: /pdf/
Disallow: /problem-melden
Disallow: /realtime_stuttgart/
Disallow: /seite-bewerten
Disallow: /stockchart/export.asp?stSId
Disallow: /suchergebnis.asp
Disallow: /umfragen/
Disallow: /user/
Disallow: /webresource.axd
Disallow: /werben
Disallow: /wikifolio/
Disallow: /*print=true
Disallow: /*pageMode=Print
Disallow: /aktien/aktien_detail.asp?
Disallow: /*@Quelle
Disallow: /*@intRating
Disallow: /*@intLandNr
Disallow: /*@pkZeit
Disallow: /*@sortKurs
Disallow: /*@sortArt
Disallow: /*@sortDatum
Disallow: /*@sortAnzahl
Disallow: /*@sortMelde
Disallow: /fonds/*/BER
Disallow: /fonds/*/ber
Disallow: /fonds/*/HAM
Disallow: /fonds/*/ham
Disallow: /fonds/*/XQTX
Disallow: /fonds/*/xqtx
```

(a) Part 1

```
Disallow: /fonds/*/FII
Disallow: /fonds/*/fii
Disallow: /fonds/*/MUN
Disallow: /fonds/*/mun
Disallow: /fonds/*/BAE
Disallow: /fonds/*/bae
Disallow: /fonds/*/FSE
Disallow: /fonds/*/fse
Disallow: /fonds/*/TGT
Disallow: /fonds/*/tgt
Disallow: /fonds/*/DUS
Disallow: /fonds/*/dus
Disallow: /fonds/*/STU
Disallow: /fonds/*/stu
Disallow: /anleihen/suche?
Disallow: /fonds/suche?
Disallow: /etf/suche?
Disallow: /zertifikate/suche#
Disallow: /zertifikate/suche?
Disallow: /zertifikate/suche_
Disallow: /optionsscheine/suche#
Disallow: /optionsscheine/suche?
Disallow: /optionsscheine/suche_
Disallow: /knockouts/suche#
Disallow: /knockouts/suche?
Disallow: /knockouts/suche_
Disallow: /zertifikate/*/timesandsales/
Disallow: /zertifikate/*/boersenplaetze/
Disallow: /zertifikate/*/historisch/
Disallow: /zertifikate/*/chart/
Disallow: /zertifikate/*/realtimechart/
Disallow: /chart/zertifikate/
Disallow: /anleihen/timesandsales/
Disallow: /anleihen/boersenplaetze/
Disallow: /anleihen/historisch/
Disallow: /anleihen/charttool/
Disallow: /anleihen/kuponhistorie/
Disallow: /optionsscheine/*/timesandsales/
Disallow: /optionsscheine/*/boersenplaetze/
Disallow: /optionsscheine/*/historisch/
Disallow: /optionsscheine/*/chart/
Disallow: /optionsscheine/*/realtimechart/
Disallow: /knockouts/*/timesandsales/
Disallow: /knockouts/*/boersenplaetze/
Disallow: /knockouts/*/historisch/
Disallow: /knockouts/*/chart/
Disallow: /knockouts/*/realtimechart/

Sitemap: https://www.finanzen.net/xml/google-sitemaps/google_sitemap.xml
Sitemap: https://www.finanzen.net/ratgeber/sitemap.xml
Sitemap: https://www.finanzen.net/sitemap/sitemap.xml
Sitemap: https://www.finanzen.net/xml/google-sitemaps/google_sitemap_videos.xml
Sitemap: https://www.finanzen.net/xml/google-sitemaps/google_sitemap_news_all.xml
Sitemap: https://www.finanzen.net/xml/google-sitemaps/google_sitemap_analysen_all.xml
Sitemap: https://www.finanzen.net/xml/google-sitemaps/google_sitemap_fokus.xml
Sitemap: https://www.finanzen.net/xml/google-sitemaps/extra-sitemap.xml
Sitemap: https://www.finanzen.net/xml/google-sitemaps/internal-etf-links-sitemap.xml
```

(b) Part 2

Figure 4: *robots.txt*-file of the finanzen.net webpage