Finance Research Letters xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

Neighbors matter: Geographical distance and trade timing in the stock market

Kęstutis Baltakys*, Margarita Baltakienė, Hannu Kärkkäinen, Juho Kanniainen

DARE Business Data Research Group, Laboratory of Industrial and Information Management, Tampere University of Technology, Finland

ARTICLE INFO

Keywords:
Investor trading
Geographical distance
Information transfer
Private information
Investor network
Social interactions
Behavioral finance
Behavioral economics
Social networks
Individual investors

JEL classification:
D8
G10

ABSTRACT

The starting point of this paper is that neighboring investors may talk to each other sharing information about their transactions in stock markets, leading to similar trading behavior. We find that pairwise trade timing similarities between investor pairs are negatively associated to geographical distance between corresponding investor pairs. This suggests that local information transfer channels between neighboring individual investors are used in decision making. We also observe that differences in age and language moderate this association. The analysis is conducted using investor level data from different regions of Finland.

1. Introduction

The financial research literature, through numerous empirical studies, has provided evidence of how various behavioral biases affect investors' decisions. One of the less investigated examples is social interaction which is an integral part of the investing process and has an important effect on investor decisions. Specifically, the studies on market participation have found that "social" investors are more likely to invest in the market when the participation rate among their peers is high (Hong et al., 2004; Brown et al., 2008; Heimer, 2014) or when a family member has recently started investing (Li, 2014). Social influence extends beyond the mere choice of whether or not to participate in the stock market. Ivković and Weisbenner (2007) find that an investor's decision to purchase securities in an industry is related to neighborhood purchases of securities in that industry, especially for local stocks. It is also possible to predict individual investor trading using a proxy measure of social contact possibilities based on epidemic models (Shive, 2010).

Direct social interaction is the key underlying construct for social influence, and household investors can use it as a primary channel for acquiring knowledge and information about investments. However, most of the research has investigated the aggregate group behavior influence on individual investors' decisions, while direct investor-to-investor (i2i) communication has not been widely explored. A welcome exception is Ozsoylev et al. (2013), who estimate the empirical investor network to identify information transfer between investors from investor-level transaction data, but they do not use any data or proxies about social links between investors. The aim of this paper is to address this research gap.

* Corresponding author.
E-mail address: kestutis.baltakys@tut.fi (K. Baltakys).

https://doi.org/10.1016/j.frl.2018.11.013
Received 3 July 2018; Received in revised form 6 November 2018; Accepted 17 November 2018

This paper is also related to investor network analysis, which has generated interest in financial studies since the recent financial crisis (Tumminello et al., 2012; Ozsoylev et al., 2013; Emmert-Streib et al., 2018; Baltakys et al., 2018; Ranganathan et al., 2018). Our focus is on the relationship between pairwise trade timing similarity and geographical distance between investors in Finland. In the following, we will loosely refer to such negative relationship as *"local information exchange or transfer"*. We seek to understand whether households that are located closer to each other tend to time their trades in a more similar manner than those that are farther apart. We expect households, as opposed to institutions, to be less sophisticated investors, lacking expertise and seeking to reduce the cost of information search and decision making by word-of-mouth communication (Ivković and Weisbenner, 2007). In this paper, we use a unique account-level transaction data set that has been explored in multiple studies (Grinblatt and Keloharju, 2001; Linnainmaa, 2011; Tumminello et al., 2012; Siikanen et al., 2018).

Our main goal is to seek evidence of local information transfer channels existence and usage between investors living nearby. Our starting point is that geographical proximity enables face-to-face interactions and grants access to the same local sources of information. In this regard, it has been observed that the likelihood and dynamics of friendship decrease as the distance between individuals increases (Backstrom et al., 2010; Preciado et al., 2012).

In this paper, we test if the geographical distance of a pair of investors is associated with their synchronization in trade timing. This research strategy automatically avoids capturing the local bias (Grinblatt and Keloharju, 2001; Zhu, 2003; Ivković and Weisbenner, 2005; Seasholes and Zhu, 2010), because local bias is about the geographical compositions of portfolios, but not about how investors time their transactions. Investors can seek mutual connections; consequently, trade execution timing can be partially explained by investor reactions to received information, which in turn can be private or public. A stronger trade timing similarity at shorter distances could indicate a reaction to private information exchange between neighboring investors. An alternative explanation could be local public information arrivals, but business news on publicly listed companies is widely spread and thus is not supposed to have local effects only. At the same time, due to the small-world effect (Travers and Milgram, 1967), the dynamics of such exchange could lead to a ripple effect of further information dissemination (Ivković and Weisbenner, 2007). Therefore, we do not expect to find extraordinarily strong associations in terms of economic significance. To the best of our knowledge, this is the first investor-level, market-wide study where the geographical distance between investors is linked to their pairwise trade timing similarity.

2. Data

2.1. Postal codes

In this paper, we investigate the household investor pairwise relationships between two network layers: geographical distance layer (Barthélemy, 2011) and trade timing similarity layer aggregated over multiple securities. In the geographical distance layer, nodes are embedded in space using the postal code as a proxy for investor location and the links between investors are actual geographical distances between investor locations. Fig. 1 shows the distribution of Finnish household investors' postal codes in the country. Postal code areas vary from sparsely dense to medium dense with roughly 64% of postal codes having less than 600 investors. We use the postal code locality coordinates to calculate the distance matrix for each pair of investors[1] using the Vincenty's formula (Vincenty, 1975).

2.2. Investor transaction data set

We use a unique investor-level transaction data set obtained from Euroclear Finland Ltd for our analysis.[2] It includes transactions from January 1995 to December 2009 of all domestic investors that traded stocks listed on the Nasdaq OMX Helsinki Exchange. Each transaction also contains meta-data about the investors, such as investor gender, language, and year of birth. We focus on the top 20 most actively traded securities from each market capitalization segment, making a total of 60 securities analyzed.[3] Finnish households typically have very low trading activity, with 47% of investors being active only once throughout the whole observation window of 1995–2009. To conduct a market-wide analysis and have a manageable size data set, we limit the selection of household investors to those who were active in the exchange at least 20 times.[4] We need to have a sufficient number of observations to estimate the trade timing similarity, but at the same time, we expect the inactive investors to be less sophisticated and seek easy channels with a lower information acquisition cost (Siikanen et al., 2018). In addition, we remove the investors with multiple postal code entries.

[1] For investors in the same postal code, or in cases where two or more postal codes share the same coordinates in our database, we measure the distance as 1/4 of the distance to the closest non-zero distance postal code.

[2] The data that support the findings of this study are available from Euroclear Finland Ltd.; however, they cannot be obtained from the authors under the non-disclosure agreement signed with the data provider.

[3] Table A.1 (see Appendix) shows the selected companies, ISINs, number of days the companies securities were traded, number of investors, total number of transactions, and the market capitalization. We have decided to exclude Nokia company, as it is the most widely traded and owned security in Finland.

[4] e.g., an investor who traded 20 securities on one day or an investor who traded one security in 20 days or any combination of both.

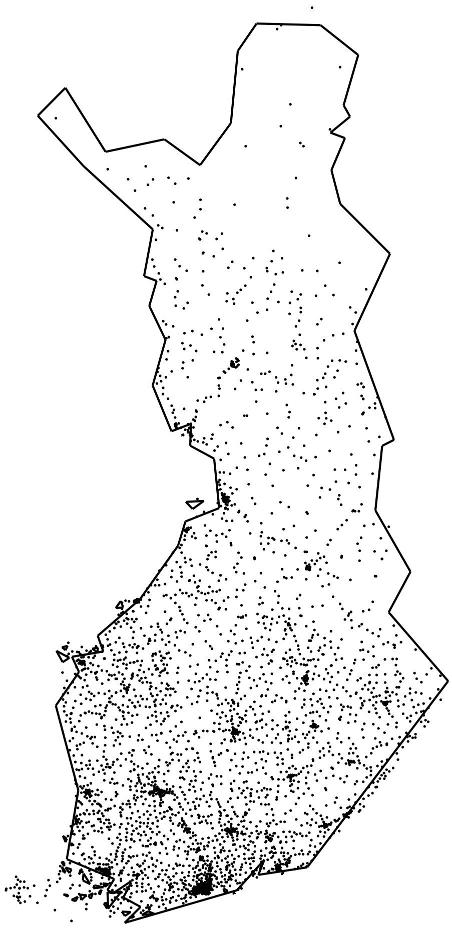

Fig. 1. The distribution of 3621 households postal codes over the map of Finland.

3. Methods

3.1. Measuring trading similarity

To compare the trading position taken by an investor on a given day, irrespective of the absolute volume traded, a categorical variable is calculated as in Tumminello et al. (2012) that describes the investor's trading activity. For each investor i and each trading day t we take the volume sold $V_s(i, t)$ and the volume bought $V_b(i, t)$ for a certain security. Then the scaled net volume ratio (Eq. (1)) can be calculated:

$$r(i, t) = \frac{V_b(i, t) - V_s(i, t)}{V_b(i, t) + V_s(i, t)} \tag{1}$$

The trading state of day t can be assigned for an investor, for a chosen threshold θ (we set $\theta = 0.01$):

$$\begin{cases} b - \text{primarily buying state, when} \quad r(i, t) > \theta \\ s - \text{primarily selling state, when} \quad r(i, t) < -\theta \end{cases}$$

In fact, the use of $\theta \neq 0$ automatically excludes day traders (investors who close their positions at the end of the day). In this paper, we focus on the investor pairs trading simultaneously in the same direction.

Subsequently, to measure the trade timing similarities over all 60 securities and both buying and selling behaviors, Jaccard coefficients for each investor pair i, j are calculated:

$$J_{ij} = \frac{\sum_{z,d} M_{11}^{(z,d;i,j)}}{\sum_{z,d} [M_{01}^{(z,d;i,j)} + M_{10}^{(z,d;i,j)} + M_{11}^{(z,d;i,j)}]} \tag{2}$$

where $M_{11}^{(z,d;i,j)}$ is the total number of trading days where both i and j are in the state $d \in \{b, s\}$ in security z, $M_{01}^{(z,d;i,j)}$ is the total number of trading days where i is not and j is in the state $d \in \{b, s\}$ in security z, and $M_{10}^{(z,d;i,j)}$ is the total number of trading days where i is in the state $d \in \{b, s\}$ in security z and j is not. The Jaccard coefficient is chosen for its straightforward interpretation as the fraction of common choices of trading activity.

3.2. Regression model

Using a linear regression model, we seek to explain how the distance between Finnish households is associated with the trade timing. The dependent variable in our model is the Jaccard similarity J_{ij}, ranging between [0,1], with 0 having no similarity and 1 being identical in the choice of trade timing. The explanatory variable is the distance D_{ij} (in km) between investors i and j.

We expect the age, language, and gender of an investor pair to alter the friendship and communication and possibly have an effect on the *local information exchange*. Thus the following variables obtained from the data set attributes are used to examine the moderating effects:

- Dummy age variable A_{ij}, equal to 1 if absolute age difference between i and j is more than 10 years, 0 otherwise.
- Language variable L_{ij}, equal to 1 if investors i and j speak different languages, 0 otherwise.
- Female pair dummy variable FF_{ij}, equal to 1 if investors i and j are both female, 0 otherwise.
- Mixed gender pair dummy variable MF_{ij}, equal to 1 if investors i is male (female) and j is female (male), 0 otherwise.

In addition, these dummies are used as control variables. To summarize, we define the regression model as follows:

$$\begin{aligned} J_{ij} = \alpha_0 \; &+ \; \alpha_1 \cdot D_{ij} + (\beta_{12} \cdot A_{ij} + \beta_{13} \cdot L_{ij} + \beta_{14} \cdot FF_{ij} + \beta_{15} \cdot MF_{ij}) \cdot D_{ij} \\ &+ \; \alpha_2 \cdot A_{ij} + \alpha_3 \cdot L_{ij} + \alpha_4 \cdot FF_{ij} + \alpha_5 \cdot MF_{ij}. \end{aligned} \tag{3}$$

The baseline in the regression is that in a pair, investors are approximately of the same age, they speak the same language, and both of them are male. Thus, we expect that α_1 is negative (longer distance, less communication, and thus fewer synchronized transactions). Differences from the baseline settings are handled by the dummy variables described above.

Age has been found to impact communication and communication patters in many ways. For instance, according to Leskovec and Horvitz (2008), especially in the online context, people tend to communicate more with each other when they are similar with respect to age, language, and location. Moreover, generation-related differences and biases have been found to negatively affect tacit knowledge transfer (Liebowitz et al., 2007). In addition, compared to the older generations, generation Y (usually defined as those born between the early 1980s and mid-late 1990s) has been found to rely heavily on technology to communicate with each other (Bolton et al., 2013), while the earlier generations commonly value more face-to-face communication (Venter, 2017). Consequently, we expect β_{12} and β_{13} to be positive, making the association between distance and trading similarity less negative.

With the remaining two variables, we want to contrast different gender combinations. Studies show that females are more risk averse (Estes and Hosseini, 1988; Embrey and Fox, 1997; Fehr-Duda et al., 2006) and show a stronger desire to use a financial adviser compared to men (Stinerock et al., 1991). Therefore, we would expect the coefficients β_{14} and β_{15} to be negative.

Geographical distances can have very different meanings in rural and urban areas. For example, in Helsinki, the population density is over $3000/\text{km}^2$, but in Lapland, it is around $1.8/\text{km}^2$ and neighbors can be kilometers apart from each other; thus, people are used to traveling long distances on a daily basis. Therefore, we analyze the trade timing similarity association to the distance between investors separately for different metropolitan and rural areas. The separate analyses on different regions can also be justified by the finding of (Gilbert et al., 2008) that people in urban areas, as opposed to people in rural areas, have more friends, on average, and they are situated closer, and yet they tend to have more friends scattered throughout the country than those in rural areas do. In total, we run 14 regression analyses for different geographical areas. This also acts as a robustness check for the results.[5]

4. Main results

The results in Tables 1 and 2 confirm our expectations and show a consistent significant, negative relationship between distance and trade timing similarity across all regions, except for one positive and one negative insignificant relationship. Our findings are consistent regarding the distance and trade timing similarity association.

In 13 regions, an investor pair with an age difference larger than 10 years, on average, has a smaller trade timing similarity than a pair of investors who are similar in age. Taking into account the age difference moderator, in most cases, we observe a reduced negative relationship between distance and trade timing similarity for investor pairs with a larger age difference. These finding are in line with the literature (Liebowitz et al., 2007; Leskovec and Horvitz, 2008) stating that similarity in age is important for better communication.

We also observe a smaller trade timing similarity for investors who speak different languages. In nine regions, investors who share the same main language are more similar in trade timing than investor pairs who speak different languages. The language moderator also weakens the negative effect on distance association to trade timing, except for Turku, Southeastern Finland, and Northern

[5] The set of metropolises $\mathcal{M}$ is defined as follows: $\mathcal{M}$= {Helsinki, Tampere, Turku, Oulu}, where in our definition the Helsinki area also includes investors from neighboring Espoo and Vantaa municipalities. The set of rural regions $\mathcal{R}$ is defined as follows: $\mathcal{R}$= {Uusimaa, Eastern Tavastia, Southwestern Finland, Western Tavastia, Central Finland, Southeastern Finland, Ostrobothnia, Northern Savonia, Eastern Finland, Northern Finland}

Table 1
Linear regression estimates for four metropolitan areas with control variables and moderators. Pairwise relationships are estimated over 60 securities and buying and selling behaviors. Standard errors are given in parentheses () and economic significance is given in curly brackets {}. Economic significance is normed by Jaccard coefficient standard deviation.

Panel A: Distance				
	Helsinki	Tampere	Turku	Oulu
Distance	−6.23e−06 ***	−8.24e−06 *	8.02e−07	−1.64e−05 ***
	(4.47e−07)	(3.52e−06)	(6.01e−06)	(3.02e−06)
	{−0.006}	{−0.002}	{1.82e−04}	{−0.008}
Panel B: Moderators				
Age diff. ≥ 10	2.72e−06***	1.49e−05***	1.49e−05*	1.53e−05***
	(4.84e−07)	(4.16e−06)	(6.73e−06)	(3.59e−06)
	{0.002}	{0.005}	{0.004}	{0.008}
Different language	1.49e−06*	2.23e−05	−3.90e−05 **	1.18e−04***
	(5.81e−07)	(1.53e−05)	(1.18e−05)	(2.04e−05)
	{6.40e−04}	{0.002}	{−0.005}	{0.009}
Female-female	3.88e−06***	1.15e−05	−4.01e−05 **	−5.64e−06
	(7.80e−07)	(1.23e−05)	(1.51e−05)	(1.06e−05)
	{0.001}	{0.001}	{−0.004}	{−7.91e−04}
Male-female	3.43e−06***	−7.84e−06	−2.52e−05 ***	−1.36e−05 ***
	(4.68e−07)	(4.56e−06)	(6.97e−06)	(3.93e−06)
	{0.002}	{−0.002}	{−0.006}	{−0.006}
Panel C: Constant and control variables				
Constant	0.003***	0.003***	0.003***	0.003***
	(5.60e−06)	(3.10e−05)	(3.77e−05)	(4.03e−05)
	{0}	{0}	{0}	{0}
Age diff. ≥ 10	−9.38e−05 ***	−3.69e−04 ***	−2.65e−04 ***	−3.08e−04 ***
	(6.17e−06)	(3.62e−05)	(4.21e−05)	(4.81e−05)
	{−0.004}	{−0.012}	{−0.009}	{−0.010}
Different language	−3.13e−04 ***	−6.92e−04 ***	−5.15e−05	−0.001 ***
	(8.30e−06)	(1.25e−04)	(6.87e−05)	(2.40e−04)
	{−0.009}	{−0.006}	{−0.001}	{−0.008}
Female-female	−1.20e−04 ***	6.95e−05	1.85e−04	8.27e−06
	(1.23e−05)	(9.73e−05)	(9.58e−05)	(1.44e−04)
	{−0.002}	{8.12e−04}	{0.003}	{8.52e−05}
Male-female	−1.40e−04 ***	3.39e−05	6.60e−05	9.02e−05
	(6.24e−06)	(3.87e−05)	(4.35e−05)	(5.29e−05)
	{−0.005}	{0.001}	{0.002}	{0.003}
Panel D: Sample size and Jaccard standard deviation				
N	26,953,337	3,039,121	1,805,007	1,253,352
Jaccard std. dev.	0.013	0.015	0.014	0.015

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Finland, which have a strengthening language moderating effect. This might be related to the different language distributions (especially Finnish and Swedish) across regions.

The gender control variables do not suggest a universal rule of effects on the trade timing similarity. However, the moderators hint at a stronger negative relationship between trade timing similarity and distance in the rural areas if one investor is female, which is in line with gender research studies.

The main results show high statistical significance. This can partially be explained by the very high number of observations. For a robustness check, we perform a bootstrap analysis. For each of the 1000 iterations, we sample 10% of the relationships, run the same regression analysis, and observe the signs of the coefficients. These results also confirm that our findings are robust (see Tables C.2 and C.3 in the Appendix). At the same time, the economic significance is quite limited, but this is expected, as we are analyzing all the relationships in the regions, while only a fraction of them may have actual social connections.

5. Additional robustness checks

In addition to the main analysis, we run a number of additional robustness checks. First, we conduct three analyses: for aggregated urban and rural areas (see Table C.4), an analysis without moderators, and an analysis with the logarithmic distance transformation $ln(D_{ij})$. All results also confirm the negative association between distance and trade timing, and are available on request. Fourth, we reason that investors might not converse about all of their traded securities but about a subset of special interest companies. To capture this relationship, instead of measuring the similarity over all securities, we estimate the trade timing similarity to be the maximum of the similarities observed for different securities (see eq. (B.1) in the online Appendix). The regression results are consistent with the main findings and can be found in the online Appendix (see tables C.5 , C.6 and C.7).

Table 2
Linear regression estimates for 10 rural regions with control variables and moderators. Pairwise relationships are estimated over 60 securities and buying and selling behaviors. Standard errors are given in parentheses () and economic significance is given in curly brackets {}. Economic significance is normed by Jaccard coefficient standard deviation.

Panel A: Distance	Uusimaa	Eastern Tavastia	South-Western Finalnd	Western Tavastia	Central Finalnd	South-Eastern Finalnd	Ostrobothnia	Northern Savonia	Eastern Finland	Northern Finland
distance	−3.84e−07	−6.64e−06 ***	−8.45e−06 ***	−3.00e−06 ***	−2.46e−06 ***	−4.54e−06 ***	−9.05e−07 ***	−2.67e−06 ***	−2.82e−06 ***	−5.71e−07 ***
	(2.66e−07)	(4.67e−07)	(1.51e−07)	(3.29e−07)	(1.25e−07)	(4.86e−07)	(1.68e−07)	(4.76e−07)	(1.58e−07)	(1.71e−07)
	{−0.001}	{−0.014}	{−0.032}	{−0.008}	{−0.014}	{−0.011}	{−0.003}	{−0.007}	{−0.020}	{−0.004}
Panel B: Moderators										
age diff. ≥ 10	5.11e−07	1.50e−06**	2.57e−06***	9.83e−07*	1.22e−06***	−2.17e−07	2.51e−07	7.39e−07	6.81e−07***	4.24e−07*
	(2.97e−07)	(5.54e−07)	(1.57e−07)	(3.94e−07)	(1.52e−07)	(5.88e−07)	(1.83e−07)	(5.72e−07)	(1.94e−07)	(2.09e−07)
	{0.001}	{0.003}	{0.011}	{0.003}	{0.007}	{−6.21e−04}	{9.55e−04}	{0.002}	{0.005}	{0.003}
different language	7.64e−07*	2.43e−06	7.85e−06***	1.48e−05***	6.77e−08	−4.09e−07	3.90e−07*	8.49e−06	1.53e−06	−7.23e−06 **
	(3.49e−07)	(2.81e−06)	(2.17e−07)	(3.04e−06)	(9.13e−07)	(3.54e−06)	(1.90e−07)	(4.66e−06)	(1.77e−06)	(2.50e−06)
	{0.002}	{9.89e−04}	{0.036}	{0.005}	{5.55e−05}	{−1.55e−04}	{0.001}	{0.002}	{9.73e−04}	{−0.004}
female-female	−3.60e−06 ***	−8.36e−06 ***	−6.26e−06 ***	−7.97e−07	6.13e−07	−2.44e−06	−1.34e−06 *	1.28e−07	−1.61e−06 *	3.23e−07
	(8.60e−07)	(1.56e−06)	(4.03e−07)	(1.20e−06)	(4.73e−07)	(1.83e−06)	(5.29e−07)	(1.66e−06)	(6.29e−07)	(6.69e−07)
	{−0.003}	{−0.006}	{−0.009}	{−7.08e−04}	{9.49e−04}	{−0.002}	{−0.002}	{1.04e−04}	{−0.003}	{6.18e−04}
male−female	−1.63e−06 ***	−4.63e−06 ***	−2.26e−06 ***	−1.56e−06 ***	−6.44e−08	3.79e−07	−8.83e−07 ***	−1.46e−06 *	−9.24e−07 ***	−4.71e−07 *
	(3.27e−07)	(6.01e−07)	(1.68e−07)	(4.36e−07)	(1.69e−07)	(6.57e−07)	(2.01e−07)	(6.29e−07)	(2.19e−07)	(2.36e−07)
	{−0.004}	{−0.009}	{−0.008}	{−0.004}	{−2.95e−04}	{8.91e−04}	{−0.003}	{−0.003}	{−0.005}	{−0.003}
Panel C: Constant and control variables										
constant	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.004***	0.003***
	(1.60e−05)	(2.56e−05)	(1.45e−05)	(2.45e−05)	(1.77e−05)	(3.84e−05)	(1.51e−05)	(3.73e−05)	(3.15e−05)	(2.98e−05)
	{0}	{0}	{0}	{0}	{0}	{0}	{0}	{0}	{0}	{0}
age diff. ≥ 10	−7.88e−05 ***	−2.08e−04 ***	−4.24e−04 ***	−2.91e−04 ***	−3.18e−04 ***	2.32e−05	−1.90e−04 ***	−2.88e−04***	−3.13e−04 ***	−2.96e−04 ***
	(1.88e−05)	(3.03e−05)	(1.65e−05)	(2.93e−05)	(2.14e−05)	(4.63e−05)	(1.66e−05)	(4.45e−05)	(3.87e−05)	(3.63e−05)
	{−0.003}	{−0.007}	{−0.014}	{−0.010}	{−0.011}	{7.38e−04}	{−0.007}	{−0.009}	{−0.010}	{−0.010}
different language	−1.90e−04 ***	−0.002 ***	−9.69e−04 ***	−0.002 ***	−8.74e−04 ***	2.33e−04	−1.94e−04 ***	−5.43e−04	8.19e−05	9.05e−04*
	(2.73e−05)	(1.75e−04)	(2.94e−05)	(2.52e−04)	(1.43e−04)	(2.57e−04)	(1.75e−05)	(3.07e−04)	(3.31e−04)	(3.91e−04)
	{−0.006}	{−0.011}	{−0.029}	{−0.007}	{−0.005}	{0.001}	{−0.007}	{−0.002}	{2.78e−04}	{0.003}
female−female	−1.63e−04 **	4.57e−04***	6.64e−04***	−2.72e−04 **	−1.63e−04 *	−2.62e−04	6.92e−05	2.87e−04*	2.33e−04	−5.35e−05
	(5.51e−05)	(8.36e−05)	(4.43e−05)	(9.00e−05)	(6.78e−05)	(1.44e−04)	(4.78e−05)	(1.29e−04)	(1.24e−04)	(1.11e−04)
	{−0.002}	{0.006}	{0.009}	{−0.003}	{−0.002}	{−0.003}	{8.81e−04}	{0.003}	{0.002}	{−6.18e−04}
male−female	−1.05e−04 ***	2.17e−04***	1.29e−04***	−2.14e−05	−1.12e−04 ***	−1.82e−04 ***	−4.59e−05 *	2.14e−04***	1.26e−04**	−1.62e−05
	(2.07e−05)	(3.28e−05)	(1.78e−05)	(3.25e−05)	(2.41e−05)	(5.16e−05)	(1.82e−05)	(4.88e−05)	(4.37e−05)	(4.05e−05)
	{−0.004}	{0.007}	{0.004}	{−6.94e−04}	{−0.003}	{−0.005}	{−0.002}	{0.006}	{0.004}	{−5.13e−04}
Panel D: Sample size and Jaccard standard deviation										
N	5,935,341	3,154,892	11,648,346	3,983,213	5,301,202	1,922,045	12,072,544	1,858,117	2,246,499	2,128,835
Jaccard std. dev.	0.013	0.014	0.014	0.014	0.015	0.016	0.013	0.016	0.016	0.014

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

As a yet additional robustness check, to control for increase in information flows due to internet and online trading in the 2000s, we estimated the trade timing similarities over different years and five-year periods (see Table C.8 and C.9 in the Appendix) and ran regression analysis with period dummies. The results concerning the distance coefficients, age difference and languages are consistent with the main findings.

Sixth, to investigate if the results are not driven only by active investors we ran the main regression analysis excluding 5% of the most active investors. The results confirm our main findings and can be found in the Appendix (see Table C.10).

Finally, as a reference to our main within- regional analysis we performed a regression analysis for across regional investor pairs. We ran 1000 bootstrap iterations for each urban region, sampling 500,000 investor pairs within an urban region, and the same amount of investor pairs where one of the investors in the pair originates from the urban area and the other comes from anywhere outside the analyzed urban region. For both the within-regional and across-regional sample we ran the regression models and observed the resulting coefficient distributions (see resulting distributions of α_1 coefficients in Fig. C.1 in the Appendix). We find that parameter estimates of within-regional regressions are systematically more negative compared to corresponding parameters of across-regional regressions. The fact that the within-regional coefficients are systematically lower than the across-regional coefficients suggests that the distance between investors is more important at closer distances.

6. Conclusions

This study contributes to the literature on financial information transfer channels in that the findings confirm the expected negative association between the trade timing similarities and distances between investors. The results are robust and show that the closer investors are to each other, the higher their trade timing similarity is. The observed negative association suggests the existence of local information diffusion channels between household investors. Investor age and language are important attributes that systemically moderate the observed association.

Acknowledgments

The research project leading to these results received funding from the EU Research and Innovation Programme Horizon 2020 under grant agreement No. 675044 (BigDataFinance) and Tampere University of Technology doctoral school. M.B. is grateful for the grants received from The Finnish Foundation for Technology Promotion, The Foundation for Advancement of Finnish Securities Market and The Finnish Foundation for Share Promotion. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.frl.2018.11.013

References

Backstrom, L., Sun, E., Marlow, C., 2010. Find me if you can: improving geographical prediction with social and spatial proximity. Proceedings of the 19th International Conference on World Wide Web. ACM, pp. 61–70.
Baltakys, K., Kanniainen, J., Emmert-Streib, F., 2018. Multilayer aggregation with statistical validation: application to investor networks. Sci. Rep. 8 (1), 8198.
Barthélemy, M., 2011. Spatial networks. Phys. Rep. 499 (1–3), 1–101.
Bolton, R.N., Parasuraman, A., Hoefnagels, A., Migchels, N., Kabadayi, S., Gruber, T., Komarova Loureiro, Y., Solnet, D., 2013. Understanding generation y and their use of social media: a review and research agenda. J. Serv. Manag. 24 (3), 245–267.
Brown, J.R., Ivković, Z., Smith, P.A., Weisbenner, S., 2008. Neighbors matter: causal community effects and stock market participation. J. Finance 63 (3), 1509–1531.
Embrey, L.L., Fox, J.J., 1997. Gender differences in the investment decision-making process. J. Financ. Counsel. Plan. 8 (2), 33.
Emmert-Streib, F., Musa, A., Baltakys, K., Kanniainen, J., Tripathi, S., Yli-Harja, O., Jodlbauer, H., Dehmer, M., 2018. Computational analysis of the structural properties of economic and financial networks. Journal of Network Theory in Finance 4 (3), 1–32.
Estes, R., Hosseini, J., 1988. The gender gap on wall street: an empirical analysis of confidence in investment decision making. J. Psychol. 122 (6), 577–590.
Fehr-Duda, H., De Gennaro, M., Schubert, R., 2006. Gender, financial risk, and probability weights. Theory Decis. 60 (2–3), 283–313.
Gilbert, E., Karahalios, K., Sandvig, C., 2008. The network in the garden: an empirical analysis of social media in rural life. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, pp. 1603–1612.
Grinblatt, M., Keloharju, M., 2001. How distance, language, and culture influence stockholdings and trades. J. Finance 56 (3), 1053–1073.
Heimer, R.Z., 2014. Friends do let friends buy stocks actively. J. Econ. Behav. Organ. 107, 527–540.
Hong, H., Kubik, J.D., Stein, J.C., 2004. Social interaction and stock-market participation. J. Finance 59 (1), 137–163.
Ivković, Z., Weisbenner, S., 2005. Local does as local is: information content of the geography of individual investors' common stock investments. J. Finance 60 (1), 267–306.
Ivković, Z., Weisbenner, S., 2007. Information diffusion effects in individual investors' common stock purchases: covet thy neighbors' investment choices. Rev. Financ. Stud. 20 (4), 1327–1357.
Leskovec, J., Horvitz, E., 2008. Planetary-scale views on a large instant-messaging network. Proceedings of the 17th International Conference on World Wide Web. ACM, pp. 915–924.
Li, G., 2014. Information sharing and stock market participation: evidence from extended families. Rev. Econ. Stat. 96 (1), 151–160.
Liebowitz, J., Ayyavoo, N., Nguyen, H., Carran, D., Simien, J., 2007. Cross-generational knowledge flows in edge organizations. Ind. Manag. Data Syst. 107 (8), 1123–1153.
Linnainmaa, J.T., 2011. Why do (some) households trade so much? Rev. Financ. Stud. 24 (5), 1630–1666.
Ozsoylev, H.N., Walden, J., Yavuz, M.D., Bildik, R., 2013. Investor networks in the stock market. Rev. Financ. Stud. 27 (5), 1323–1366.
Preciado, P., Snijders, T.A., Burk, W.J., Stattin, H., Kerr, M., 2012. Does proximity matter? distance dependence of adolescent friendships. Soc. Netw. 34 (1), 18–31.
Ranganathan, S., Kivelä, M., Kanniainen, J., 2018. Dynamics of investor spanning trees around dot-com bubble. PLoS ONE 13 (6), e0198807.
Seasholes, M.S., Zhu, N., 2010. Individual investors and local bias. J. Finance 65 (5), 1987–2010.

Shive, S., 2010. An epidemic model of investor behavior. J. Financ. Quant. Anal. 45 (1), 169–198.
Siikanen, M., Baltakys, K., Kanniainen, J., Vatrapu, R., Mukkamala, R., Hussain, A., 2018. Facebook drives behavior of passive households in stock markets. Finance Res. Lett.
Stinerock, R.N., Stern, B.B., Solomon, M.R., 1991. Gender differences in the use: of surrogate consumers for financial decision-making. J. Prof. Serv. Mark. 7 (2), 167–182.
Travers, J., Milgram, S., 1967. The small world problem. Phychol. Today 1 (1), 61–67.
Tumminello, M., Lillo, F., Piilo, J., Mantegna, R.N., 2012. Identification of clusters of investors from their real trading activity in a financial market. New J. Phys. 14 (1), 013041.
Venter, E., 2017. Bridging the communication gap between generation y and the baby Boomer generation. Int. J. Adolesc. Youth 22 (4), 497–507.
Vincenty, T., 1975. Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. Survey Rev. 23 (176), 88–93.
Zhu, N., 2003, The local bias of individual investors, Working paper, Yale University.