Kiel

Working Papers

Kiel Institute for the World Economy

Trading Strategies in the Overnight Money Market: Correlations and Clustering on the e-MID Trading Platform

by Daniel Fricke

No. 1766 | April 2012

Web: www.ifw-kiel.de

Kiel Working Paper No. 1766 | April 2012

Trading Strategies in the Overnight Money Market: Correlations and Clustering on the e-MID Trading Platform

Daniel Fricke

Abstract:
We analyze the correlations in patterns of trading for members of the Italian interbank trading platform e-MID. The trading strategy of a particular member institution is defined as the sequence of (intra-) daily net trading volumes within a certain semester. Based on this definition, we show that there are significant and persistent bilateral correlations between institutions' trading strategies. In most semesters we find two clusters, with positively (negatively) correlated trading strategies within (between) clusters. We show that the two clusters mostly contain continuous net buyers and net sellers of money, respectively, and that cluster memberships of individual banks are highly persistent. Additionally, we highlight some problems related to our definition of trading strategies. Our findings add further evidence on the fact that preferential lending relationships on the micro-level lead to community structure on the macro-level.

Keywords: interbank market, socio-economic networks, community identification

JEL classification: G21, E42

Daniel Fricke
Kiel Institute for the World Economy
24105 Kiel, Germany
Telephone: +49 431 8814-229
E-mail: daniel.fricke@ifw-kiel.de

Trading Strategies in the Overnight Money Market: Correlations and Clustering on the e-MID Trading Platform.[†]

Daniel Fricke[‡§]

This version: April 2012

Abstract

We analyze the correlations in patterns of trading for members of the Italian interbank trading platform e-MID. The trading strategy of a particular member institution is defined as the sequence of (intra-) daily net trading volumes within a certain semester. Based on this definition, we show that there are significant and persistent bilateral correlations between institutions' trading strategies. In most semesters we find two clusters, with positively (negatively) correlated trading strategies within (between) clusters. We show that the two clusters mostly contain continuous net buyers and net sellers of money, respectively, and that cluster memberships of individual banks are highly persistent. Additionally, we highlight some problems related to our definition of trading strategies. Our findings add further evidence on the fact that preferential lending relationships on the micro-level lead to community structure on the macro-level.

JEL-Classification: G21, E42
Keywords: Interbank market, socio-economic networks, community identification

---

[†]The article is part of a research initiative launched by the Leibniz Community. The author is grateful for helpful comments by Thomas Lux, Mattia Montagna, and Giacomo Livan.

[‡]Kiel Institute for the World Economy, Hindenburgufer 66, 24105 Kiel. Correspondence: daniel.fricke@ifw-kiel.de

[§]Department of Economics, University of Kiel, Olshausenstr. 40, 24118 Kiel

1 Introduction and Existing Literature

The global financial crisis (GFC) has shown that the stability of the interbank network is of utmost importance for the macroeconomy. Linkages arising from bilateral lending relationships may lead to systemic risk on the macro-level (see Haldane (2009)). Therefore, regulators and policymakers need to understand the structure and functioning of the interbank network in more detail.

Approaches from the natural sciences have been used to investigate the topology of networks formed by interbank liabilities. Examples for the Italian e-MID (electronic market for interbank deposit) include De Masi *et al.* (2006), Iori *et. al* (2007; 2008), Finger *et al.* (in progress), and Fricke and Lux (2012). Several 'stylized facts' of interbank networks have been identified, among them the finding of disassortative mixing patterns, i.e. high-degree nodes tend to trade with low-degree nodes, and vice versa Cocco *et al.* (2009). Quite recently, Fricke and Lux (2012) argued that this fact may be a key driver for the observation of a hierarchical core-periphery (CP) structure in interbank networks (see also Craig and von Peter (2010)). Thus, preferential lending relationships between individual institutions may lead to community structure at the macro-level. While many authors also note some other form of community structure in the interbank network they analyze (see e.g. Boss *et al.* (2004)), Fricke and Lux (2012) stress that apart from the CP structure, there is no community structure whatsoever in the Italian e-MID network of interbank claims.

Based on these findings, we aim at analyzing the trading strategies of individual institutions in more detail. For this purpose we use a detailed dataset containing all overnight interbank transactions on the e-MID trading platform from January 1999 to December 2010. Splitting the sample into half-yearly subsamples, we define the (intra-) daily net trading volumes of the individual institutions as the trading strategies and analyze the correlations between them.[1] We find evidence for significant and persistent bilateral correlations between institutions' trading strategies. In most semesters we find two anti-correlated clusters of trading strategies, indicating that banks tend to trade preferentially with banks from the other cluster. This finding is both in line with the existence of preferential relationships, and with herding phenomena in banks' trading strategies. Interestingly, the information whether a bank is a continuous net buyer or net seller appears to be the only defining characteristic for its group membership in this anonymous dataset. This is

[1] See Kyriakopoulos *et al.* (2009) for a similar approach. Comparable studies also exist for the patterns of trading members in stock exchanges Zovko and Farmer (2007), and Tumminello *et al.* (2012).

in contrast to the findings of Iori *et. al* (2007) who performed a very similar analysis based on the e-MID data during the sample period 1999-2002, and found two clusters containing mostly large and small banks, respectively. We also highlight some problems related to the definition of trading strategies. Overall, our findings are quite persistent over time and appear to be largely unaffected by the GFC. We add further evidence on the fact that preferential lending relationships on the micro-level lead to community structure on the macro-level.

The remainder of the paper is structured as follows: section 2 briefly introduces the Italian e-MID interbank data, section 3 contains the empirical analysis and section 4 concludes.

2 Dataset

The Italian electronic market for interbank deposits (e-MID) is a screen-based platform for trading unsecured money-market deposits in Euros, US-Dollars, Pound Sterling, and Zloty operating in Milan through e-MID SpA.[2] The market is fully centralized and very liquid; in 2006 e-MID accounted for 17% of total turnover in the unsecured money market in the Euro area and covers essentially the entire domestic overnight deposit market in Italy. Average daily trading volumes were 24.2 bn Euro in 2006, 22.4 bn Euro in 2007 and only 14 bn Euro in 2008.

Available maturities range from overnight up to one year. Most of the transactions (> 80%) are overnight, which is not surprising given that the loans are unsecured. In August 2011, e-MID had 192 members from EU countries and the US, of which 101 were domestic (Italian) and 61 international banks.[3] This composition is, however, not stable over time as can be seen from Figure 1: The left panel shows a clear downward trend in the number of active Italian banks over time, whereas the additional large drop after the onset of the GFC (semester 18) is mainly due to the exit of foreign banks. The right panel shows that the decline of the number of active Italian banks went along with a relatively constant trading volume in this segment until 2008.[4] The overall upward trend of trading volumes was mainly due to the increasing activity of foreign banks until 2008, which virtually faded away after the onset of the crisis. Due to the cyclicality of foreign banks' activity, we will focus on the Italian banks only. Additionally, for reasons

[2]More details can be found on the e-MID website, see http://www.e-mid.it/.

[3]Additionally, 29 central banks and 1 ministry of finance acted as market observers.

[4]Thus, the decline decline of active Italian banks appears to be driven by domestic mergers and acquisitions.

explained below, we will only consider the relatively active banks within the respective semesters.

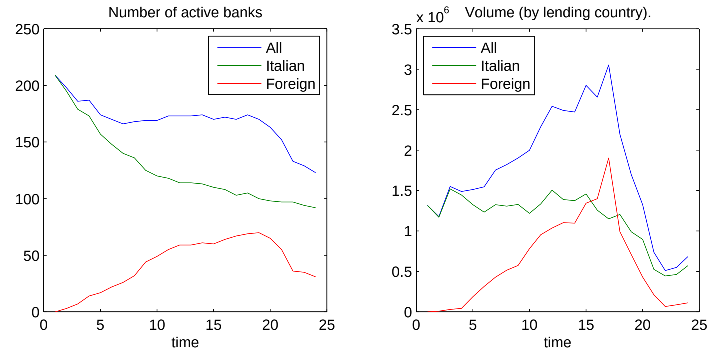

Figure 1: Number of active banks (left) and traded volume (right) over time. We also split the traded volume into money going out from Italian and foreign banks, respectively.

The trading mechanism follows a quote-driven market and is similar to a limit-order-book in a stock market, but without consolidation. The market is transparent in the sense that the quoting banks' IDs are visible to all other banks. Quotes contain the market side (buy or sell money), the volume, the interest rate and the maturity. Trades are registered when a bank (aggressor) actively chooses a quoted order. The minimum quote size is 1.5 million Euros, whereas the minimum trade size is only 50,000 Euros. Thus, aggressors do not have to trade the entire amount quoted. The platform allows for credit line checking before a transaction will be carried out, so trades have to be confirmed by both counterparties. The market also allows direct bilateral trades between counterparties. Contracts are automatically settled through the TARGET2 system.

We have access to all registered trades in Euro in the period from January 1999 to December 2010.[5] For each trade we know the two banks' ID numbers (anonymous to us), their relative position (aggressor and quoter), the maturity and the transaction type (buy or sell). We will focus on all overnight trades conducted on the platform, leaving a total number of 1,317,679 trades. The large sample size allows us to analyze the system's evolution over time. Here we focus on 24 independent semi-annual subsamples.

In the following, we analyze the correlation matrices of the banks' trading strategies. The finding of a core-periphery structure in the interbank network

[5]A detailed description of the dataset can be found in Finger *et al.* (in progress) and Fricke and Lux (2012).

indicates the existence of hierarchical structure. Interestingly, we did not find any further community structure in the network using usual community detecting algorithms. In this paper, we identify clusters of trading strategies in a different type of network based on the (intra-) daily trading patterns of individual banks.

3 Empirical Analysis of Trading Strategies

3.1 Measuring Correlations Between Strategies

Even though the dataset is anonymous, each bank can be identified by a unique ID-number, so we can observe the behavior of all sample banks over the entire sample period. In the following, we will divide the dataset into 24 half-yearly intervals (semesters) which are treated as independent subsamples.[6] In order to define the trading strategies we further divide each semester into different trading sessions. For example, Iori *et. al* (2007) use a Fourier method to use continuous trade data, while Kyriakopoulos *et al.* (2009) use daily data. Here we find that each day in the sample period can be naturally split into two trading sessions, during which a similar fraction of trades occur on average.[7] Figure 2 shows the fraction of trades by the time of the day.[8] For example, we see that roughly 25% of the trades occur between 9 and 10am. This is in line with the findings in Iori *et. al* (2008), and we checked that this observation is in fact stable over the entire sample period.[9] Based on these numbers, we split each day into a morning session (8am-12am) and a afternoon session (12pm-6pm). In total, our sample contains 3,073 days, so we end up with 6,146 trading periods. Due to the heterogeneity of the banks' trading activity, we restrict ourselves to the institutions which were active in a fraction of at least $\theta = 50\%$ of all trading sessions in a particular semester.[10] In this way, we focus on those institutions which were relatively active over a significant part of the sample period. Given the negative trend in the number of active Italian banks, it is unavoidable that the number of banks ending up in the sample is also decreasing over time, see Figure 3.

[6] Later on we will see that the subsamples are in fact not independent.

[7] Note that we also checked that the results are not affected by splitting each day 3 or 4 trading sessions. However, it should be clear that the number of periods without any activity is positively related to the number of subintervals.

[8] The same remarks hold for the transacted volumes, yielding very similar results.

[9] See Iori *et. al* (2007) for an identification of the events that contribute to the intra-day patterns.

[10] We checked the results are qualitatively very similar for other values of θ.

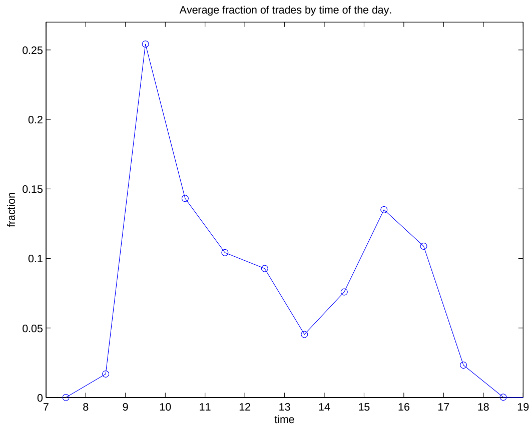

Figure 2: Activity by time of the day, indicated by the fraction of trades occurring at a certain time of the day (hourly intervals).

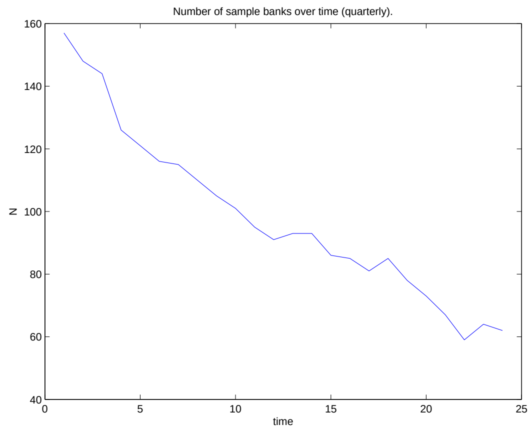

Figure 3: Number of sample banks over time. For each semester, a bank is in the sample if it was active in at least θT trading sessions, with T as the total number of trading sessions in the respective semester.

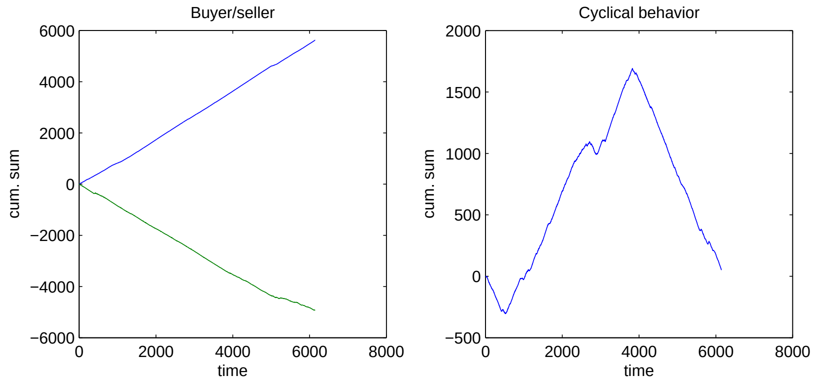

Figure 4: Examples of banks' trading strategies in the Italian interbank market. Here we plot cumulative sums of particular signed strategies (1,-1,0) over the entire sample period. The examples were not chosen randomly, but to illustrate different trading styles. The left panel shows a constant buyer/seller, and the right panel shows a highly cyclical trading strategy.

For each trading period and for each institution, we define the trading strategy as the net traded volume in Euros. Net volume is the total buy volume (money borrowed from other banks) minus total sell volume (money lent to other banks). Thus, a bank with a positive (negative) net volume is a net buyer (seller) of money. Banks with a zero strategy were either inactive in the trading session or ended up with a flat position after buying and selling equal amounts.[11] Figure 4 shows examples of cumulative (signed) trading strategies for three institutions which were active during most of the sample period. We see a substantial level of heterogeneity in the employed strategies, with banks acting as continuous buyers/sellers (left panel) and more complicated, possibly cyclical, patterns (right panel).

In the end, we obtain a strategy matrix $\mathbf{S}$ of dimension $N \times T$, where T denotes the number of trading sessions in a particular semester and N the number of sample banks in this semester. Note that in each semester the strategy matrix only includes transactions between the sample banks, thus ignoring transactions with other banks that might act relatively infrequently.[12]

[11] We also defined signed strategies (similar to the approach in Zovko and Farmer (2007)), where we assign a +1, -1 or a 0 describing the strategies in a simpler form. However, given that our network is closed, we work with the valued version since it contains more information. A technical reason is that it may happen that a bank has the same strategy in each trading session. In such a case, we cannot compute the correlation of this strategy with other strategies.

[12] In this sense, we work with a closed network of banks.

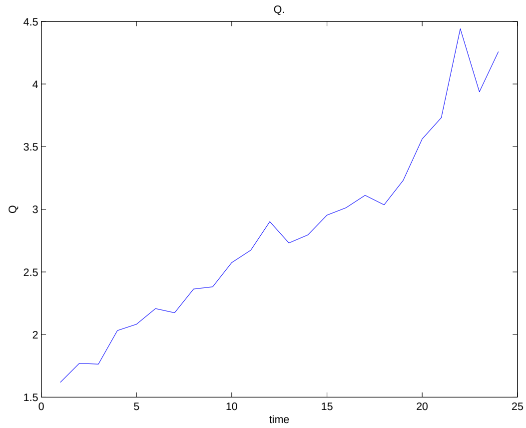

Figure 5: Time evolution of $Q = T/N$.

In everything that follows, we will work with the normalized bank-specific time-series (mean equal to 0 and variance equal to 1). For our sample period, $T \sim 240$, while N varies over time as shown above. In total, $N_T = 218$ different institutions are in the final sample, but obviously not each bank is active in each semester. Therefore, the ratio $Q = T/N$ is not constant over time as can be seen from Figure 5. We see that this ratio exceeds 1 for the complete sample period, so we can reliably estimate all correlations.[13]

From the strategy matrix $\mathbf{S}_t$ we can construct the $N \times N$ semi-annual correlation matrices $\mathbf{C}_t$ between the institutions' strategies. Due to the normalization of the strategies the correlation matrix can be simply calculated based on the Pearson estimator (see Laloux *et al.* (1999))

$$\mathbf{C}_t = \frac{1}{T-1}\mathbf{S}_t\mathbf{S}_t^T, \tag{1}$$

with $()^T$ denoting the transpose of a matrix.

A color example of such a correlation matrix for the first half of 2006 can be seen in Figure 6. Dark colors represent high absolute correlations, with warm and cold colors indicating positive and negative correlations, respectively. As will be explained in more detail below, the left panel shows the raw correlation matrix after applying the clustering approach and the right panel shows the filtered correlation matrix with identical ordering. In both cases, the two (anti-correlated) clusters are readily visible, with the level of

[13]For a set of N time series of length T, the correlation matrix contains $\frac{N(N-1)}{2}$ entries which have to be determined by NT observations. If T is not very large compared to N, the empirical correlation matrix is to a large extent noisy. We should also note that many results from random matrix theory require $Q \geq 1$, see below.

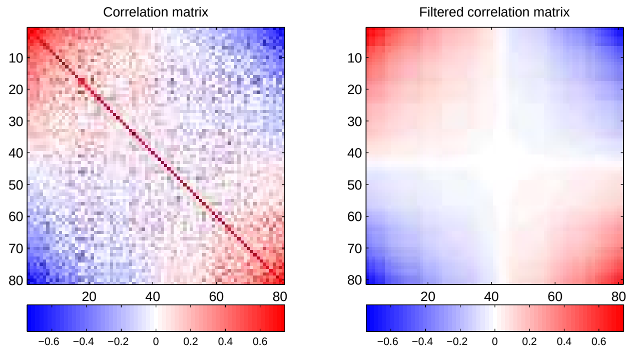

Figure 6: Example of a correlation matrix after applying the clustering approach for the first half of 2006. Left: raw correlation matrix. Right: filtered matrix. Warm colors represent positive correlations, cold colors negative ones. White indicates uncorrelated strategies.

noisy correlations significantly reduced in the filtered matrix. We will show below that similar observations hold for most semesters.

As a first rough formal test for the significance of the computed correlations, we use a t-test assuming normally distributed disturbances. Later on, we will provide more thorough significance tests, but even this simple approach is already quite suggestive. Figure 7 shows the fraction of significant correlations at the 5% level, among all correlations, over time. Obviously, these values are quite large, with an average value of close to 50% which clearly exceeds the 5% we would expect randomly with a 5% acceptance level of the test. In the following, we will provide a more detailed analysis of the correlation matrices, mostly in terms of their eigenvalues.

3.2 Structure in the Correlation Matrices

In this section we investigate the eigenvalue spectrum of the observed correlation matrices. Note that since the correlation matrices are symmetric, all eigenvalues are positive real numbers (see Bouchaud and Potters (2003)). We denote v_i as the (normalized) eigenvector corresponding to the ith eigenvalue λ_i,[14] where the eigenvalues are sorted from largest to smallest, such that $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_N$. By definition, an eigenvector is a linear combination of the different components $j = 1, \cdots, N$ such that $\mathbf{C}v_i = \lambda_i v_i$. Thus, λ_i is

[14]We drop the time-indices in most of the following to save notation. It should be clear that we have a correlation matrix for each semester, and thus time varying eigenvalues and eigenvectors.

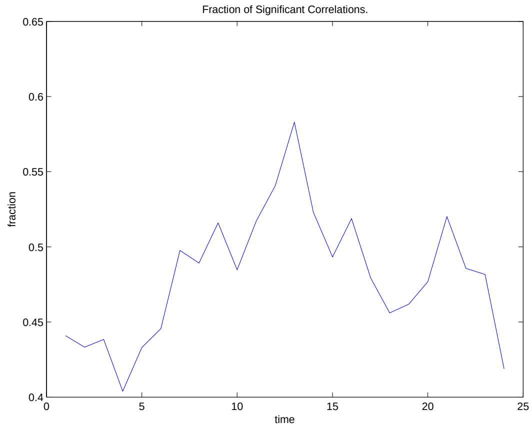

Figure 7: Fraction of significant correlations over time, based on a standard t-test testing the significance of the correlation coefficient at the 5% level.

the variance of the correlation matrix captured by the weighting based on the corresponding eigenvector.

3.2.1 Time-Persistence of Correlations

As a first step, we investigate whether the observed correlations are stable over time. This is helpful for the dynamic analysis, since we expect trading strategies, and thus the corresponding clusters, to be quite persistent. In this way, the time-persistence of individual correlation pairs offers a first test for spuriousness, since spurious correlations are not likely to persist in time. In contrast, if the correlations are persistent, then the clusters of institutions are also likely to persist in time.

Since we can track individual institutions over time, we can test whether the (time-varying) correlations between institutions' trading strategies are significantly different from zero. In total, we have $N_T = 218$ institutions in our sample. Therefore, we have to evaluate the significance of $(N_T^2 - N_T)/2 = 23,653$ relationships. Given that two sample banks may be active in different semesters, several of these correlations cannot be evaluated. Nevertheless, for all institutions jointly belonging to the sample in at least two semesters, we can check whether the observed correlations are significantly different from zero using a simple t-test assuming Gaussian residuals.[15]

Doing this, we find 3,203 significant combinations at the 5% significance

[15]Note that this is a joint test also on the autocorrelation of observed correlations, since highly volatile values are unlikely to be significant.

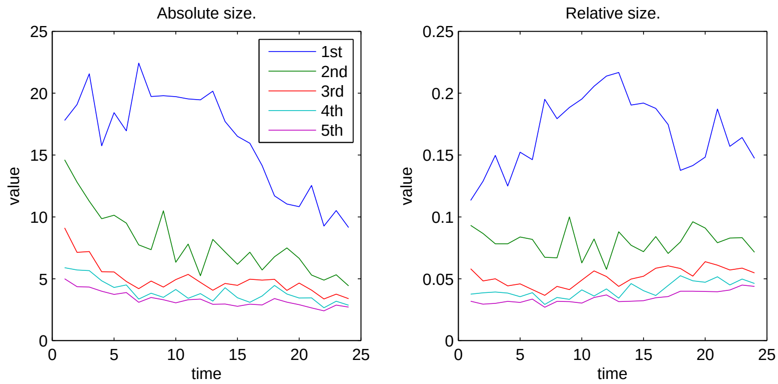

Figure 8: Absolute (left) and relative (right) size of the first 5 Eigenvalues of **C** over time. The relative size normalizes the observed eigenvalues by their total sum N.

level, i.e. a fraction of roughly 10% of all correlation pairs is significantly different from zero. This value is way above the 5% level we would expect to occur in a purely random fashion.[16] Hence, a large number of the correlations between individual banks' trading strategies are significantly different from zero over the sample period. We can therefore expect any identified clusters to persist over time.

3.2.2 Time Evolution of the Largest Eigenvalues

For a set of infinite length uncorrelated time series, **C** would correspond to the identity matrix with all eigenvalues equal to 1. If the time series are of finite length, however, the eigenvalues will never exactly equal 1 (see Zovko and Farmer (2007)). In fact, the distribution of the eigenvalues in the case of iid Gaussian time-series is known. Below, we will use this information in order to detect significant clustering by finding eigenvalues significantly larger than the predicted values. As a first step, however, we focus on the absolute and relative size of the eigenvalues over time. Since $\sum_i \lambda_i = N$, we can standardize the observed eigenvalues to make the values comparable over time. Figure 8 shows the dynamics of the five largest eigenvalues, both in absolute terms (top panel) and standardized (bottom panel).

Note that the time evolution of the largest eigenvalue is, in qualitative terms, very similar to the dynamics of the fraction of significant correlations in Figure 7. Thus, we might expect at least the largest eigenvalue to carry

[16]Here we disregard those correlation pairs between banks that were never jointly active. Taking this into account, increases the fraction of significant correlations to roughly 20%.

important information on the structure of the correlation matrix. Except for the beginning and the end of the sample period, its absolute and relative importance is quite strong. In contrast, the (relative) sizes of the second to fifth eigenvalues are far more stable over time.[17]

While the eigenvalues give us a macroscopic description of our network in terms of the variance, investigating the leading eigenvectors allows us to understand how much a particular principal component affects the original variables. Below, we will use information from the leading eigenvector to define the trading clusters.[18]

3.2.3 Random Matrix Theory and Eigenvalue Density

Here we test the significance of the observed eigenvalues based on results from random matrix theory (RMT). It is well documented that the eigenvalues of a correlation matrix can be used to separate true information from noise. The null hypothesis of uncorrelated strategies, translates itself to the assumption of iid random elements in the strategy matrices, which gives the so-called random Wishart matrices or Laguerre ensemble of RMT.[19].

The basic idea, is to compare the observed eigenvalue frequency distribution with a hypothetical one obtained from the correlation matrix for N iid Gaussian random variables of length T. More formally, for N random uncorrelated variables of length T, in the (thermodynamic) limit $T, N \longrightarrow \infty$, with $Q = T/N \geq 1$, the density of the eigenvalues $p(\lambda)$ is given by the functional form

$$p(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda^{\max} - \lambda)(\lambda - \lambda_{\min})}}{\lambda}, \tag{2}$$

with

$$\lambda_{\min}^{\max} = \sigma^2(1 + 1/Q \pm 2\sqrt{1/Q}), \tag{3}$$

where σ^2 being the variance of the time series, which is equal to 1 due to the standardization, and $\lambda \in [\lambda_{\min}, \lambda^{\max}]$. Equations (2) and (3) define the well-known Marčenko-Pastur (MP) distribution of eigenvalues from iid Gaussian random variables Marčenko and Pastur (1967). Empirical values falling outside the interval defined by Eq. (3) are likely to carry important information

[17]Note that the first eigenvalue monitors information about the primary cluster in the correlation matrix, both in terms of size and its average correlation. The subsequent eigenvalues carry information on subsequent clusters, see Friedman and Weisberg (1981).

[18]We also performed several bootstrap-tests on the significance of the largest eigenvalues. For example, we randomly reshuffled buy and sell periods for the institutions, constructed the correlation matrix and compared the resulting eigenvalues with those from the observed correlation matrices. The results are comparable with those based on RMT as detailed below, and are available upon request.

[19]See Laloux *et al.* (1999)

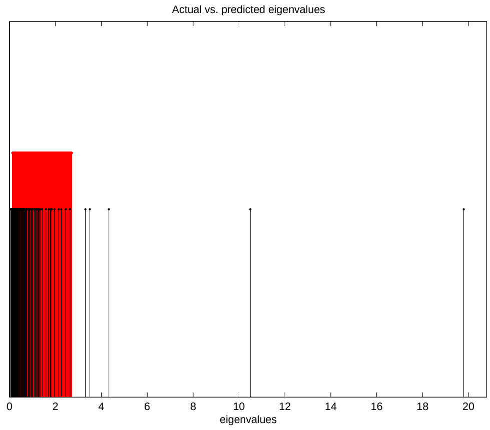

Figure 9: Comparison of observed and predicted eigenvalues from the MP distribution, for semester 9. Black bars indicate observed eigenvalues, the red block shows the range of the predicted distribution, as defined in Eq. 3.

about the system. Note that the closed-form solution is only a limiting result and, as shown above, the parameter Q changes over time. Consequently, the predicted eigenvalue density under the null changes from semester to semester. Therefore, we run separate tests for each semester, where we compare the observed eigenvalue densities to the range of the eigenvalues under the null, which can be calculated by using the observed Q in Eq. (3).[20]

Figure 9 shows an illustrative example of the results of our analysis. There we compare the actual eigenvalues (black bars) with the range predicted by RMT (red area). We see that the bulk of the eigenvalues lies within the range of the MP distribution. However, in this particular semester, 5 eigenvalues are outside the upper limit of the MP distribution, with one very large eigenvalue, one of medium size and the remaining eigenvalues of comparable size. The results are comparable for most periods, usually containing 5 or 6 eigenvalues significantly larger than those predicted under the null.[21] This is further evidence for the non-random structure of the correlations between trading strategies, even though we have to bear in mind that both N and T are quite small, such that there may be small-sample biases.[22] Nevertheless,

[20]Note that σ and Q could also be treated as free parameters that could be estimated from the data, in order to find the optimal fit of the MP distribution.

[21]Note that there are often also values smaller than the predicted $\lambda_{\min}$, which we will not comment on in the following.

[22]We found comparable results for other aggregation periods, e.g. yearly data, and other selected samples, e.g. other values for θ. Thus, we are quite confident that the results do not depend crucially on these issues.

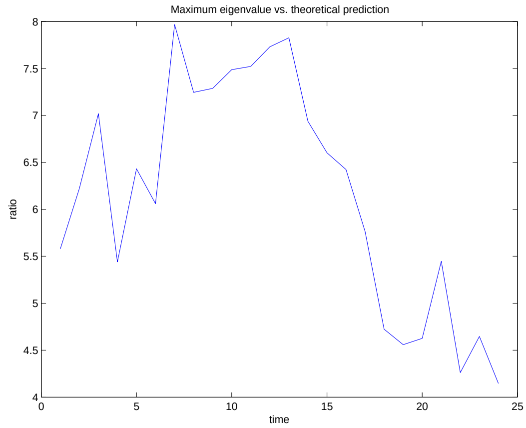

Figure 10: Relative size of the largest eigenvalue compared to $\lambda_{\max}$, as defined in Eq. 3.

the largest eigenvalue is usually substantially larger than the theoretical prediction, cf. Figure 10. Thus, in most of what follows, we will focus on the largest eigenvalue and the corresponding leading eigenvector.

In the next section we filter out the noisy component of the observed correlation matrices. This will allow us to identify the clusters of trading strategies more convincingly.

3.2.4 Filtering the Correlation Matrices

The presence of a well-defined bulk of eigenvalues in agreement with the MP distribution suggests that the contents of **C** are highly affected by noise. In other words, those eigenvalues outside the predicted range are likely to contain substantial information about the underlying community structure, i.e. those components we are actually interested in.[23] In the following, we will therefore filter out the noisy component from the correlation matrices, using the approach of Kim and Jeong (2005).[24]

With the complete set of eigenvalues and eigenvectors, we can decompose the correlation matrix **C** as

$$\mathbf{C} = \sum_{i=1}^{N} \lambda_i v_i v_i', \qquad (4)$$

[23] In the future, it would be very interesting to use the approach of Livan *et al.* (2011) and set up a factor model that might be able to reproduce the observed eigenvalue spectra.

[24] Similar approaches for cross-correlation matrices of stock returns can be found in Laloux *et al.* (1999), Plerou *et al.* (2002), and Livan *et al.* (2011).

where v_i and λ_i defined as above. Because only the eigenvectors corresponding to the few largest eigenvalues contain the information on significantly non-random structure, we can identify a filtered correlation matrix for these groups by choosing a partial sum of Eq. (4), denoted as $\mathbf{C}^g$. Thus, we posit that the eigenvalue spectrum of the correlation matrix is organized into a group part defined by the largest eigenvalues, and a random part containing the bulk of small eigenvalues consistent with the MP distribution.[25] Thus, we can write the correlation matrix as

$$\mathbf{C} = \mathbf{C}^g + \mathbf{C}^r = \sum_{i=1}^{N_g} \lambda_i v_i v_i' + \sum_{i=N_g+1}^{N} \lambda_i v_i v_i'. \tag{5}$$

Note that the calculation of $\mathbf{C}^g$ involves a choice on the number of leading eigenvalues relevant for its calculation. Here we simply use the leading eigenvector only ($N_g = 1$), but the results do not depend on this assumption.[26] In the following, we will always show the results for the raw and the filtered correlation matrix.

3.3 Clustering of Trading Behavior

We have both shown that the correlations between banks' trading strategies are quite persistent and that a number of eigenvalues of the correlation matrices significantly exceed those predicted in case of completely uncorrelated strategies. Thus, there is some evidence in favor of clustering, i.e. non-random structure, in the correlation matrices. However, we have also seen that there is a substantial level of noise in the correlation matrices, which is why we constructed filtered correlation matrices, based on the significant eigenvalues and their corresponding eigenvectors only.

In this section we will show that the components of the leading eigenvector allow us to identify clusters of trading behavior.[27] We usually find two

[25] Note that for correlations of stocks returns this is slightly different, since there the first eigenvalue is usually an order of magnitude larger than the rest. The corresponding eigenvector is simply the 'market' itself, containing roughly equal components on all stocks with almost all components having the same sign. See for example Plerou *et al.* (2002). We checked that such a 'market' vector does not exist in our case; the distribution is more uniform with roughly equal numbers of large and small eigenvalues. Note that this uniform distribution also deviates from the standard Gaussian distribution for the eigenvector components of a random correlation matrix, see Guhr *et al.* (1998).

[26] We found very similar results using all eigenvalues exceeding the theoretical upper limit by a certain threshold.

[27] We also use hierarchical clustering techniques, which yielded very similar results. One important disadvantage of these techniques, however, is that they remain silent about the optimal number of clusters.

clusters of banks, with highly correlated strategies within the clusters, but anti-correlated strategies between the clusters. This finding suggests that the banks in one cluster tend to trade in similar directions and their counterparties are usually from the other cluster. At first sight, this result may seem trivial, because in the money market there is always one buyer and one seller, so some banks will necessarily have correlated strategies. But what is surprising is that the clusters are rather large (compared to the number of sample banks) and persist over time. Thus, banks appear to have preferred counterparties in the market, which they turn to repeatedly when they are willing to trade. We will discuss the meaning of the clusters in more detail below.

3.3.1 Clustering Based on the Leading Eigenvectors

Our clustering approach consists of dividing the leading eigenvector into positive and negative components. Components with equal sign end up in the same cluster,[28] i.e. the indicator for the group membership of bank j (g_j) can be defined as

$$g_j = \begin{cases} 1 & \text{if } v_1^j > 0, \\ -1 & \text{if } v_1^j < 0, \\ 0 & \text{if } j \text{ was inactive.} \end{cases} \tag{6}$$

We do this for each semester individually, and the results for one particular semester were already shown above in Figure 6. For most periods we observe similar patterns, with two relatively large groups with anti-correlated strategies, and a third group containing the remaining banks which are mostly uncorrelated with the rest and therefore have no clear strategy in this particular semester. For some semesters, the usefulness of the filtering approach is more obvious than for the example above, cf. Figure 11 for the first half of 1999. In this period, it is hard to identify any structure in the raw correlation matrix, even after reordering the indices according to the leading eigenvector. In contrast, the filtered correlation matrix shows a similar structure as the example above, however, with one of the groups being rather small.

3.4 Time-Persistence of Clusters

So far we have only looked at the structure of the correlation matrices for each semester individually, finding that there is in fact evidence for non-

[28] In principle, we could divide the clusters even further by using the next eigenvectors. However, we find that the usage of the leading eigenvector is already sufficient to identify the large clusters.

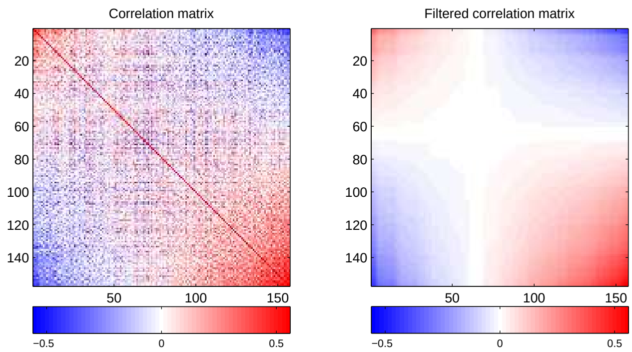

Figure 11: Example of a correlation matrix after applying the clustering approach for the first half of 1999. Left: raw correlation matrix. Right: filtered matrix. Warm colors represent positive correlations, cold colors negative ones. White indicates uncorrelated strategies.

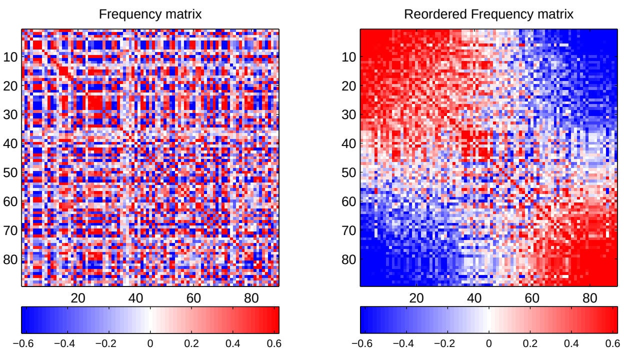

Figure 12: Frequency matrix $\mathbf{F}$, as defined in the text. Left: raw matrix. Right: after reordering.

random structure in the matrices. Here we aim at assessing the persistence of the identified clusters over the sample period. We already discussed the time-persistence of individual correlation pairs in the trading strategies. Here we check whether two sample banks tend to appear in the same clusters over time. If so, this would be evidence for the non-spuriousness of the identified clusters and therefore suggest that the identified clusters are meaningful. In order to shed light on this issue, we construct a frequency matrix $\mathbf{F}$, where each element $f_{ij} \in [0, 1]$ contains the number of semesters that two banks were in the same cluster minus the number of semesters that they were in different clusters, relative to the total number of semesters that the two banks were jointly active.

	B_t	S_t	E_t
B_{t-1}	.6832	.2118	.1050
S_{t-1}	.1807	.6996	.1197
E_{t-1}	.0272	.0331	.9397

Table 1: Transition matrix: B, S and E stand for buyer, seller, and exit, respectively.

Figure 12 shows the results of this exercise. The left panel shows the raw matrix, the right panel shows the same matrix, but after reordering the indices according to the size of the components of the leading eigenvector of this matrix. The results are striking: the two clusters appear even for the complete sample period since there are many large entries (in absolute terms) in the frequency matrix. Hence, there is not only stability in the individual correlations, but also in the group memberships of individual banks. Note that this is strong evidence for the non-randomness of the half-yearly matrices, so the identified clusters are indeed significant. If group membership was completely random, it would be impossible to observe this phenomenon.

Another way to investigate the persistence of the cluster memberships is to construct a transition matrix of going from the one to the other cluster. We find that the two clusters contain those banks which mainly bought and sold money (see below), respectively, so we label the clusters B and S. Taking the possibility of exit into account (E), Table 1 shows for example, that there is a probability of 68.32% that a bank which was in the buy cluster in the last period (B_{t-1}) will remain in this cluster in the next period. This is way above 50%, the level we would expect if cluster membership was completely random (ignoring the possibility of exit). Note that the diagonal elements in this matrix are largest, so there is a significant level of persistence in the cluster memberships. This is also in line with the findings of Fricke and Lux (2012), who show that there is a substantial level of persistence in the banks' strategies (in terms of being intermediaries, lenders, or borrowers) based on quarterly aggregates. We should also stress that, since our clustering approach is only based on the sign of the components of the leading eigenvectors, there are a number of elements in these which are close to zero i.e. insignificant. Therefore, it comes as no surprise that the probability of switching from the buy to the sell cluster (or vice versa) is also quite high, with a value of 21.18%. This is driven by banks with less clear trading strategies, which might put them randomly in one or the other cluster. On average, this effect should vanish since the switching probabilities are of comparable magnitude.

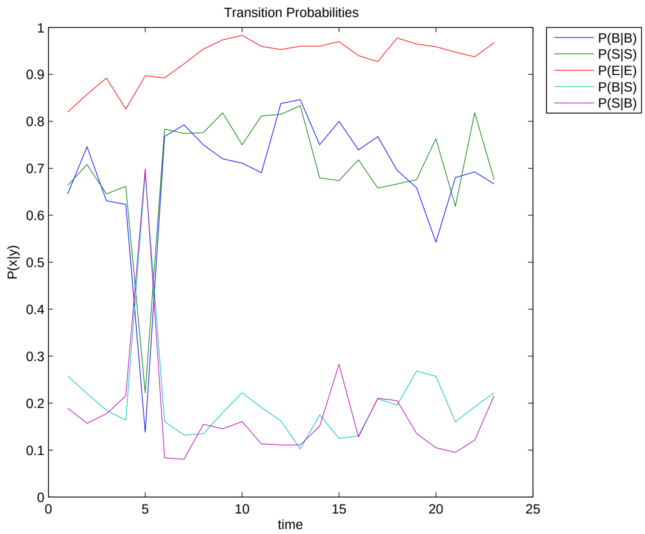

Figure 13: Transition probabilities over time. $P(x|y)$ shows the probability of going from status y to x. B, S and E stand for buyer, seller, and exit, respectively.

We checked that these results are very stable over time, cf. Figure 13. Interestingly, similar to the remarks in Fricke and Lux (2012), there appears to be a substantial reversal in the trading strategies in the second half of 2001 (semester 5), which Fricke and Lux (2012) identified as a significant structural break in many time-series of the e-MID dataset. Interestingly, the GFC (starting around semester 18) is hardly visible, which is somewhat surprising given that core banks tended to reverse their strategies during this time (cf. Fricke and Lux (2012)).

3.5 A Closer Look at the Clusters

In this section we want to take a closer look at the individual clusters. The usual interpretation of the finding of two anti-correlated clusters is that banks in the same cluster have similar strategies (i.e. tend to trade in the same direction), trading with banks in the other cluster. For example, Iori *et. al* (2007) state that the two clusters contain large and small banks, respectively, which tend to trade with each other. At first sight, this interpretation makes intuitive sense, since two strategies can only be positively correlated, if the two banks tend to trade in the same directions during the same trading sessions. Note that the statement *during the same trading sessions* is important, a simple example illustrates this point: suppose bank A is a continuous

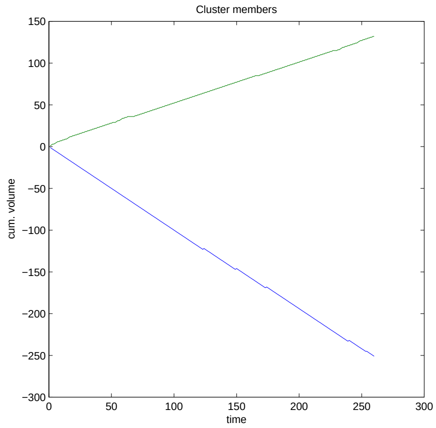

Figure 14: Most anti-correlated (signed) trading strategies in semester 3.

buyer during a single semester. With equal probability it buys 1 Euro on the market or is inactive in a particular trading session. The cumulative trading volume of bank A will therefore be a monotonically increasing function of time. Now suppose that bank B is also a continuous buyer of money, always buying 1 Euro when bank A is inactive. Again bank B's cumulative trading volume will be a monotonically increasing function of time. What does the correlation of the banks' trading strategies tell us about their relationship? By definition, the correlation between the trading strategies of the two banks will be -1. Observed negative correlations may therefore be spurious, in the sense that two strategies may turn out to be identical in terms of the cumulative trading volumes, but not in terms of the individual trading sessions. Quite interestingly, this fact has not been mentioned before and we know of no paper that investigated the trading strategies in this detail after identifying the clusters.[29]

To illustrate these points, consider Figures 14 and 15, where we show the cumulative trading volumes (signed strategies) of the two most anti-correlated banks in semester 3 and 10, respectively. Figure 15 shows the usual pattern, namely one bank continuously buys money in the market, whereas the other bank continuously sells money in the market, and the banks tend to trade in the same trading sessions. In contrast, Figure 14 shows the case where banks with, in terms of the sign of the cumulative trading volume, very similar strategies, end up in different clusters simply because they tend to trade asynchronously.

All in all, this indicates that we might not really measure what we would

[29]Note that the continuous approach of Iori *et. al* (2007) also suffers from this problem.

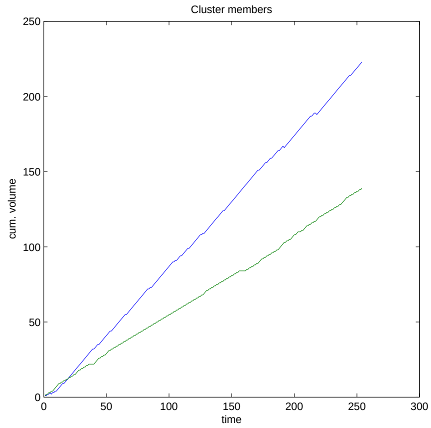

Figure 15: Most anti-correlated (signed) trading strategies in semester 10.

like to measure using our definition of trading strategies. Since we focus only on the relatively active banks, the effect is likely to be small for relatively large values of θ. In order to investigate this in more detail, we checked that one cluster indeed contains banks that tend to buy most of the time, while banks in the other cluster tend to sell most of the time. As another robustness check, we calculated the correlation between trading strategies, using only those trading periods where both banks were jointly active. While this affects particularly those cases where banks tend to trade asynchronously, the general patterns observed in the baseline correlation matrices remain unaffected. We also checked that the total trading volumes in the two clusters are roughly equal, so the clusters indeed tend to trade with each other. Nevertheless, the examples show, that our approach may at times not be able to identify meaningful clusters. Summing up, the story is much simpler than expected: the two clusters contain those banks with clear trading strategies over time, i.e. mostly buying or selling money during a particular semester.

We also checked the results with respect to the CP model in Fricke and Lux (2012), since it might be that one cluster contains the core banks (which are similar across many dimensions and possibly even in terms of their trading strategies), and the other cluster contains the periphery periphery banks. However, we do not find any evidence in this regard.[30] Figure 16 shows an example to highlight that there is apparently no relationship between the coreness of banks in one or the other cluster. There we see the same

[30]Note that this is in contrast to the findings of Iori *et. al* (2007), since the coreness of individual banks was found to be highly correlated with transaction volumes (as proxy for bank size) of individual banks.

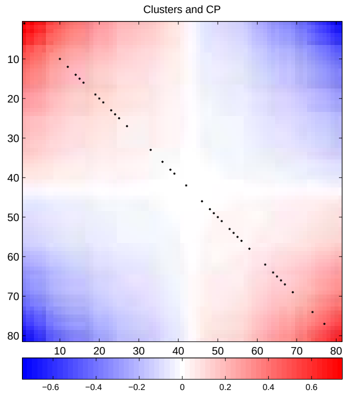

Figure 16: Relation between clusters and core-periphery model. Filtered correlation matrix for the first half of 2006, with core banks highlighted by black dots on the main diagonal.

(filtered and reordered) correlation matrix as in Figure 6, but with core banks explicitly marked by a black dot on the main diagonal. There seems to be no relationship between the core-periphery model and the trading clusters, since the core banks are spread somewhat evenly over the network. However, given that core banks tend to act on both market sides, it should be clear that the core banks usually do not appear among the most anti-correlated (buy and sell) strategies. Similar results hold for other sample periods. Thus, the structure of the correlation matrices is not driven by the core membership of individual banks. This is not surprising, given that most core banks act as intermediaries and tend to distribute money across the complete system, rather than building up large positions in a continuous way. Therefore, core banks are spread evenly across the correlation network.

It would, of course, be very interesting to relate our findings to external information about the sample banks, e.g. using balance sheet data. However, the anonymity of our dataset makes such an exercise impossible. It would be interesting to carry out such an analysis in the future based on different non-anonymous datasets. In this way, the aim would be to model the behavior of individual banks based on observable characteristics in order to come up with realistic agent-based models of the interbank market.

4 Conclusions

In this paper, we analyzed the correlations in patterns of trading for members of the Italian interbank trading platform e-MID. We showed that there are significant and persistent bilateral correlations between institutions' trading strategies, and in most semesters we find evidence for the existence of two anti-correlated clusters. The two clusters mostly contain continuous net buyers and net sellers of money, respectively. The clusters, in terms of the individual cluster memberships, are highly persistent. However, we have also seen that there are certain problems related to our definition of trading strategies. Additionally, we highlight some problems related to our definition of trading strategies, since the observed negative correlations may be spurious. The reason lies in the fact that two strategies may be identical in terms of the cumulative trading volumes, but not in terms of the trading time.

Our findings add further evidence on the fact that preferential lending relationships on the micro-level lead to community structure on the macro-level. Furthermore, despite finding no evidence of community structure in the usual network of interbank liabilities apart from the CP structure, see Fricke and Lux (2012), trading on the e-MID platform appears to be a relatively structured process in terms of trading strategies. Given that each trade involves two counterparties, i.e. a buying and a selling side, it may appear trivial that we identify these clusters. However, the high level of persistence shows that most banks tend to have rather stable trading strategies over time, since they tend to appear in the same cluster over time, trading with the same counterparties.

In the future, we need to explore the trading behavior of individual banks in more detail, to understand the evolution of the interbank network in more detail. The main aim is to build an artificial banking system, that allows to test the effects of different regulatory measures in a laboratory setting. We are still at the beginning of understanding the complexity of actual banking networks.

Bibliography

BOSS, M., H. ELSINGER, M. SUMMER, AND S. THURNER (2004): "Network topology of the interbank market," *Quantitative Finance*, 4(6), 677–684.

BOUCHAUD, J.-P., AND M. POTTERS (2003): *Theory of Financial Risk and Derivative Pricing - From Statistical Physics to Risk Management (2nd Edition)*. Cambridge University Press.

COCCO, J. F., F. J. GOMES, AND N. C. MARTINS (2009): "Lending relationships in the interbank market," *Journal of Financial Intermediation*, 18(1), 24–48.

CRAIG, B., AND G. VON PETER (2010): "Interbank tiering and money center banks," Discussion Paper, Series 2: Banking and Financial Studies 12/2010, Deutsche Bundesbank.

DE MASI, G., G. IORI, AND G. CALDARELLI (2006): "Fitness model for the Italian interbank money market," *Phys. Rev. E*, 74(6), 66112.

FINGER, K., D. FRICKE, AND T. LUX (in progress): "The Italian Interbank Network: An Overview," In progress., Kiel Insitute for the World Economy.

FRICKE, D., AND T. LUX (2012): "Core-Periphery Structure in the Overnight Money Market: Evidence from the e-MID Trading Platform," Kiel Working Papers 1759, Kiel Institute for the World Economy.

FRIEDMAN, S., AND H. F. WEISBERG (1981): "Interpreting the first eigenvalue of a correlation matrix," *Educational and Psychological Measurement*, 41(1), pp. 11–21.

GUHR, T., A. MÜLLER-GROELING, AND H. A. WEIDENMÜLLER (1998): "Random Matrix Theories in Quantum Physics: Common Concepts," *Phys. Rep.*, 299, 190–425.

HALDANE, A. (2009): "Rethinking the financial network," Speech delivered in April 2009 at the Financial Student Association, Amsterdam. URL: http://www.bankofengland.co.uk/publications/speeches/2009/speech386.pdf.

IORI, G., G. DE MASI, O. V. PRECUP, G. GABBI, AND G. CALDARELLI (2008): "A network analysis of the Italian overnight money market," *Journal of Economic Dynamics and Control*, 32(1), 259–278.

Iori, G., R. Renò, G. D. Masi, and G. Caldarelli (2007): "Trading strategies in the Italian interbank market," *Physica A: Statistical Mechanics and its Applications*, 376(0), 467 – 479.

Kim, D.-H., and H. Jeong (2005): "Systematic analysis of group identification in stock markets," *Phys. Rev. E*, 72, 046133.

Kyriakopoulos, F., S. Thurner, C. Puhr, and S. W. Schmitz (2009): "Network and eigenvalue analysis of financial transaction networks," *The European Physical Journal B - Condensed Matter and Complex Systems*, 71(4), 523–531.

Laloux, L., P. Cizeau, J.-P. Bouchaud, and M. Potters (1999): "Noise Dressing of Financial Correlation Matrices," *Phys. Rev. Lett.*, 83, 1467–1470.

Livan, G., S. Alfarano, and E. Scalas (2011): "Fine structure of spectral properties for random correlation matrices: An application to financial markets," *Phys. Rev. E*, 84, 016113.

Marčenko, V., and L. A. Pastur (1967): "Distribution of eigenvalues for some sets of random matrices," *Mat. Sb. (N.S.)*, 72(4), 507–536.

Plerou, V., P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley (2002): "Random matrix approach to cross correlations in financial data," *Phys. Rev. E*, 65, 066126.

Tumminello, M., F. Lillo, J. Piilo, and R. N. Mantegna (2012): "Identification of clusters of investors from their real trading activity in a financial market," *New Journal of Physics*, 14(1), 013041.

Zovko, I. I., and J. D. Farmer (2007): "Correlations and clustering in the trading of members of the London Stock Exchange," Quantitative Finance Papers 0709.3261, arXiv.org.