FIRST YEAR PROGRESS REPORT

EU FET OPEN PROJECT FOC
255987

forecasting financial crises

Deliverable D4.1

"Report on Activity in WP4 - Validation of the forecasting modelling"
WP Leader : Giulia Iori

Objectives

The activity of WP4 consists in empirical analysis of financial data aimed at validating the specific models developed in WP2. The tasks belonging to this work package are:
T4.1 Empirical study of inter-bank markets
T4.2 Validation of interdependence and trend reinforcement.
T4.3 Analysis of the Financial System as a network of sectors
T4.4 Validation of systemic risk indicators.
T4.5 Empirical analysis of query-logs.

Progress Overview

In the first year the activity the activity of WP4 has been focused on Tasks T4.1, T4.2 and T4.5. In T4.1, at CITY we have performed an empirical study of the e-Mid inter-bank market in the period 1999-2009, in order to verify the hypothesis that the dynamics of the cross-sectional interest rate spreads is driven by the strategic behaviour of banks. Moreover, we have been working on the comparison of indices on different markets with the aim of identifying early warning indicator of systemic risk.
At the ETHZ node we have analysed the FED Discount Window dataset and we have characterized the exposure of FED over time during the crisis. The same dataset allows in principle to study also the borrowing patterns of the major financial institutions.

For T4.2 at ETHZ, we have developed a method to detect trend reinforcement and interdependence from co-movements in time series of financial indicators. We have applied this method to the CDS price time series for the major US financial institutions and we have found significant levels of both trend reinforcement and interdependence. The methods is a first step towards a more general approach to the analysis of time series from which we plan in the second and third year to derive some indicators of systemic risk, in line with the objectives of T4.4 (Validation of Systemic Risk Indicators).

Moreover, in line with T3.4 (Simulations, scenario analysis, systemic risk estimation) in WP3 the time evolution of these indicators will be visible from widget on the platform, providing a live monitoring of systemic risk.

In T4.5 we have worked in a collaboration team (BM, CNR, ETHZ) to analyse query-logs in relation to stock markets. We have found that query logs can be used as a proxy of trading volumes and can anticipate, by one day or more, the financial trends. We have also started to work on the analysis of Twitter data (BM, CITY).

Output of the WP

1) V. Hatzopoulos, G. Iori, G. Gabbi, L. Politi, G. Germano (2011) *The microstructure of the European interbank market and its role in the determination of cross-sectional spreads,* working paper available soon.
2) Kapar, Olmo and Iori, *Structural Break Test as early warning indicators,* working paper available soon.
3) Kaushik R., Battiston S., "Trend reinforcement and interdependence in credit default swap networks", Working paper, available on FOC website in October. T4.2, T4.1
4) Puliga M., Battiston S., Tasca P., "Some statistics on FED discount window lending system", Working paper, available on FOC website October, T4.1
5) I. Bordino, S. Battiston, G. Caldarelli, M. Cristelli, A. Ukkonen and I. Weber, Web search queries can predict stock market volumes, *Submitted to PNAS* (2011)

Task T4.1 Empirical study of inter-bank markets (Planned duration Month 3-24)

T4.1.1 Analysis of the e-Mid Market (CITY, CNR)

The aim of this task, carried out by CITY in collaboration with Gabbi from the CNR node, is to perform an empirical study inter-bank markets. This study will eventually inform the theoretical models so that changes in the pattern of lending and borrowing relationships before and during the credit crisis can be reproduced and understood. In the working paper *The microstructure of the European interbank market and its role in the determination of cross-sectional spreads,* Hatzopoulos et al. (2011) analyze the intraday, intra-maintenance period and inter-day dynamics of a number of market variables such as rates, volumes, bid-ask spreads, interest rate volatility, etc. during the years 1999-2009 in the European electronic interbank market of overnight lending (e-MID). The main goal of this study is to explain the observed changes before, during and after the 2007-2008 financial crisis, of the cross-sectional dispersion of lending/borrowing conditions. Unlike previous contributions, focusing on banks' dependent and macro information as explanatory variables of credit spreads, we address the role of market microstructure as a determinant of the credit spreads and assess to what extent banks who experienced better credit condition were those that played more strategically and exploited better market opportunities.

The only electronic market for Interbank Deposits in the Euro Area and US is called e-MID. When the financial crisis started, market players were 246 members from 29 EU countries and the US, of which 30 central banks acting as market observers, 2 Ministries of Finance, 108 domestic banks and 106 international banks.

The database is composed by the records of all overnight transactions registered in the period 01/1999–12/2009. Each line contains a code labelling the quoting bank, i.e. the bank that proposes a transaction, and the aggressor bank, i.e. the bank that accepts a proposed transaction, the volume, time and rate at which the transaction has been executed and the size (for Italian banks only) and nationality of banks. A label indicates the side of the aggressor bank, i.e. whether the latter is lending/selling ("Sell") or borrowing/buying ("Buy") capitals to or from the quoting bank. We do not have any information regarding when and how a particular section of the market is used, i.e. whether some banks prefer to remain anonymous during the negotiation, we do not know whether a transaction is a result of some specific *proposals* and, finally, we do not know the content of the order book, i.e. we do not have complete information on the state of the liquidity, its dynamics and how the banks use this information when acting on the market.

During the 2007-2008 financial crisis a considerable increase in cross-sectional dispersion of credit spreads among banks during the turmoil is observed (see Fig.1).

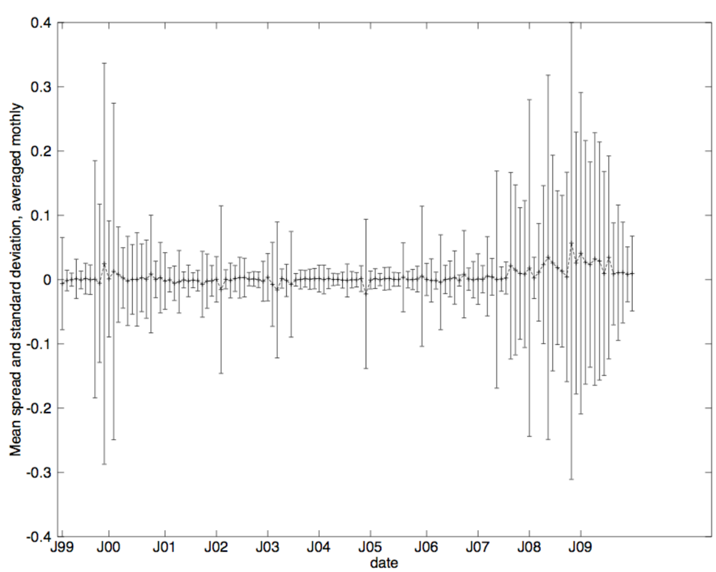

Figure 1. Standard deviation and mean of cross-sectional rates for lenders as a function of time. Unit of averaging time is one month. Labels on the x-axis denote the January of the corresponding year.

In Figure 1, for a given bank b participating in $T_{b,d}$ transactions on a given day d belonging to period p the average bank daily credit spread is calculated as:

$$<c_{bd}> = \frac{1}{T_{bd}} \sum_t r_{bt} - <r_d>, \ \forall d \in p.$$

Our analysis aims to show that the dynamics of the cross-sectional spreads can be (partially) explained in terms the changing market microstructure conditions in the e-Mid and the trading behaviour of banks (contrary to previous studies that indicated size as the only important borrower characteristic to determine spreads).

Opportunities for banks to play strategically in the interbank market have become more significant during the crisis. First of all, there is an intraday term structure effect suggesting that trading in the morning is more expensive than trading in the afternoon. The negative slope of the intraday interest curve has become much steeper during the crisis as shown in Figure 2).

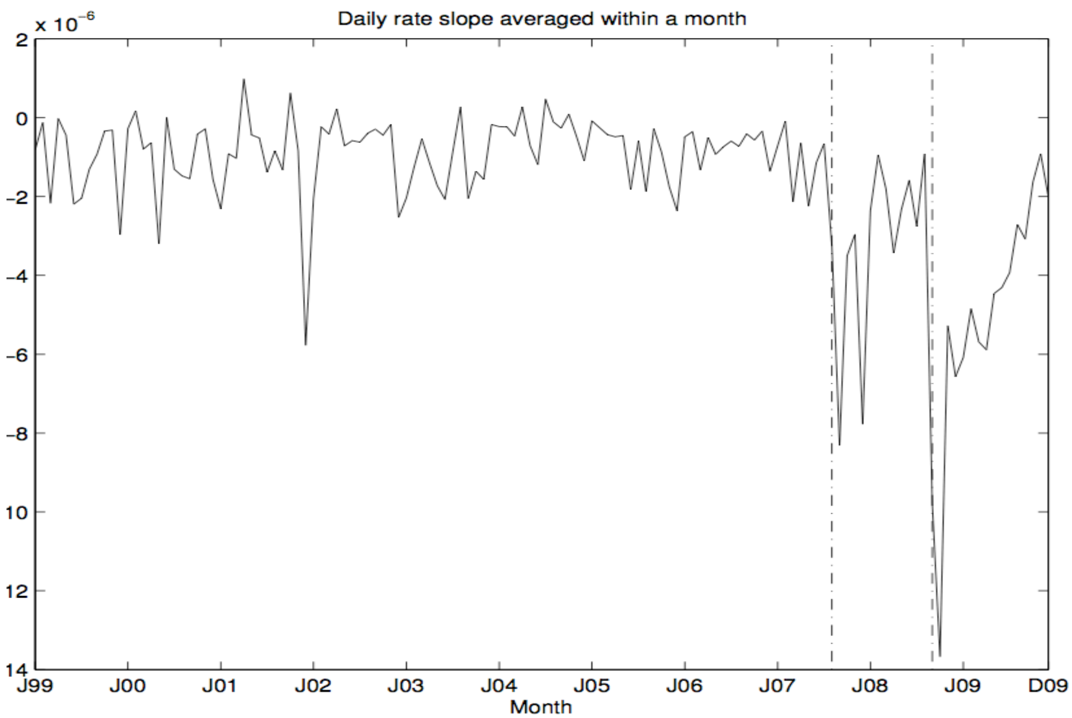

Figure 2. Monthly average of the slope of the intraday interest rate spread (measured respect to daily average). For each day *d* we compute the mean spread value at time windows of size 30 minutes before and after time *t*. Fitting s_r for each day *d* with a simple linear relation , $<s_d t> = m_d t + q_d$, where *t* is in minutes, we estimate the slope of the intraday term structure and its evolution in time.

Second there is a bid-ask spread effect suggesting that banks could trade at a more profitable rate if submitting their orders as quoters and not as aggressor.

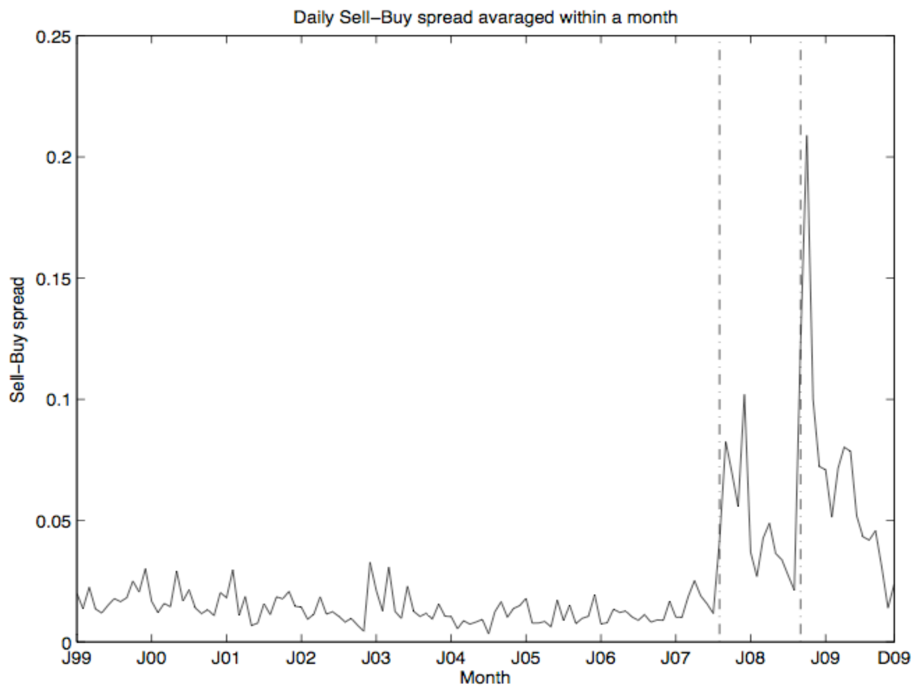

Figure 3. Monthly average of the daily Sell-Buy spread. Two very well defined peaks are clearly present after the crisis milestones.

We calculate the credit spread separately for the lending and borrowing transactions for each bank such that all banks have two separate and coefficients. In a day a bank may participate in both lending and borrowing transactions in which case both coefficients can be defined. As a first step in the analysis we show that while, on average, some banks appear to perform better than other, such good/bad performances are not necessarily consistent over time. As a measure of consistency of bank performance we measure the autocorrelation of lending and borrowing credit spreads, in trading time units and identify the largest lag at which autocorrelations are significantly positive. We normalize the autocorrelation time with the number of days a bank is active, on average, (as a lender or as a borrower) in a maintenance period (credit institutions in the euro area are required to hold minimum reserve balances with NCBs that have to be fulfilled only on average over a one-month maintenance period that runs from the 24th of a month to the 23rd of the following month). The distribution of lags over all banks at which the autocorrelation remains significant (above the confidence interval) is calculated for three periods separately with respect to the borrowing and lending credit spreads.

The main findings are:

- In the early period (2006/7/5-2007/8/9) the great majority of banks autcorrelation decays very rapidly and very few correlations remain significant beyond 2 lags (indicating that bank's performance is mostly driven by good/bad luck).
- In the second period (2007/8/10-2008/9/15) and even more so in the third (2008/9/16-2009/10/21) a coupling of past to future performance appears for a large fraction of the whole system.
- Autocorrelation in performance are more important from the point of view of the lenders than the borrowers.
- It is mainly the poor performing borrowers that experience an increase in their autocorrelation.
- It is the banks that participate more who have an increased tendency to be locked in a performance pattern, good or bad.

To assess the role of strategic trading behaviour we regress the average daily credit spread (separately for lending and borrowing spreads) of each bank versus five variable:

- X_1: (daily volume traded in the morning - daily volume traded in afternoon)/ total volume
- X_2 : (daily volume traded as quoter - daily volume traded as aggressor) / total volume
- X_3 : (daily total borrowing volume - daily total lending volume)/ total volume
- X_4 : time difference between current trade and previous trade
- X_5: daily total volume traded/total volume traded in the market in a day

The number of banks for which we obtain statistically significant positive or negative coefficient for each regressed variable are reported in the tables below. The results show that trading preferentially in the morning and trading preferentially as a quoter correlates, for the majority of banks, with high lending spreads (good perforance) while size and frequency of trades become more often negatively correlated with the lending spreads during the crisis. Trading mostly on the lending side also correlate with good lending performance for a considerable number of banks. On the borrowing side, trading preferentially in the morning is a significant explanatory variable of poor borrowing performance while trading preferentially as a quoter is a significant explanatory variable of good performance for several banks. Only few banks seem to take advantage of the lower borrowing rates available in the afternoon. Trading primarily as a borrower, size and frequency of trades affect different banks in the two opposite directions. Further investigation controlling for bank size and nationality is currently being performed

year	t1	t2	t3	t4	t5	nBanks
1999	12	18	15	8	3	55
2000	5	27	19	3	2	64
2001	12	22	3	6	3	65
2002	18	27	8	8	1	68
2003	36	28	10	5	2	77
2004	14	26	14	4	1	80
2005	32	20	14	8	2	97
2006	36	35	13	6	1	94
2007	39	40	21	11	2	99
2008	45	39	18	15	6	93
2009	48	38	22	2	5	76
all	81	72	36	22	7	110

year	t1	t2	t3	t4	t5	nBanks
1999	0	3	3	4	2	55
2000	2	1	5	3	4	64
2001	2	2	9	5	0	65
2002	3	3	5	2	5	68
2003	0	3	9	5	5	77
2004	0	3	7	4	5	80
2005	1	1	5	3	3	97
2006	0	1	7	5	5	94
2007	3	4	9	14	11	99
2008	0	7	7	9	24	93
2009	3	3	6	5	33	76
all	4	7	17	25	59	110

Table 1. Number of positive significant (left) and negative significant (right) t-values for lenders.

year	t1	t2	t3	t4	t5	nBanks
1999	7	0	3	4	4	50
2000	9	0	4	2	2	54
2001	6	3	9	8	6	57
2002	11	2	11	5	3	69
2003	16	2	6	5	11	77
2004	11	5	14	10	4	83
2005	21	6	9	5	4	92
2006	24	9	10	5	14	94
2007	22	4	11	5	16	94
2008	22	7	15	12	14	88
2009	27	6	14	11	11	72
all	49	12	19	15	24	100

year	t1	t2	t3	t4	t5	nBanks
1999	1	11	12	3	1	50
2000	3	17	17	1	2	54
2001	1	9	2	6	0	57
2002	2	16	4	1	0	69
2003	3	16	6	5	0	77
2004	5	21	7	4	4	83
2005	6	13	11	5	3	92
2006	2	30	11	3	4	94
2007	4	20	16	12	7	94
2008	1	24	20	6	6	88
2009	5	27	12	5	6	72
all	6	61	32	18	18	100

Table 2. As **Table 1** above but for borrowers.

Another dimensions over which banks can play strategically is in the choice of counterparties. Some preliminary analysis shows that preferential lending, measured by the participation ratio, occurs both before and during the crisis. Further analysis in this direction is in progress.

Finally we are testing the hypothesis of adverse selection as an explanation for the decreased trading activity during the crisis, and the impact of the crisis on market geographical segmentation (Cassola, Holthausen and Lo Duca (2008)).

References

V. Hatzopoulos, G. Iori, G. Gabbi, L. Politi, G. Germano (2011) *The microstructure of the European interbank market and its role in the determination of cross-sectional spreads,* available soon.

T4.1.2 Structural Break Test as early warning indicators (CITY)

In a separate study Kapar, Olmo and Iori at CITY have analyzed indices on different markets with the aim of identifying early warning indicator of systemic risk of the recent financial crisis. We concentrate on the variables that are proxy for several key global market conditions. In particular, S&P 500 is used as a proxy for the general conditions in the stock markets; VIX index as an implied volatility measure and market uncertainty; iTraxx Europe index as a credit risk measure and JP Morgan EMBI Index as an indicator of the movements in the emerging markets. Besides these indices, we also examine TED and LIBOR-OIS spreads. Ted spread is measured as the difference between the LIBOR rate and the U.S. treasury bill rate while Libor-OIS spread is the difference between Libor and the overnight indexed swap rate. For the analysis, we use monthly data in the period from July 2004 to May 2011. Finally we consider the spread between the e-MID O/N Euro and the ECB main refinancing rate in the period from July 2004 to December 2009.

Although July 2007 is accepted as the beginning date of the financial crisis, using structural break tests, we show that break dates in the indices related to the interbank market appears already at the end of 2006 and in the early months of 2007. During the same period we observed a rapid decline in structured credit mortgage-backed instruments measured by the ABX indices1 The prices of the riskier tranches started moving down in late 2006 and in February 2007 (when New Century Financial and HSBC, the number-three and number-two subprime lenders, announced problems) when the lower-rated indexes saw a dramatic drop of 29 percent in only one month. Early signals of the propagation of the crisis to other markets around February 2007 are also reported by Gonzales-Hermosillo and Hesse (2009) which find, using Markov-regime switching techniques, that VIX, forex swap and TED spread moved into high volatility regimes in February 2007.

We apply the Bai and Perron (1998,2003) structural break test to determine the location of potential multiple break points in the variables. This test is aiming to find out the break points in a linear regression which consist of stationary regressors like the model below with m breaks and (m+1) regimes:

$$r_t = z_t\beta^j + et, t = T_{j-1} + 1,..., T_J$$

Breakpoints are treated as unknown and the equation and are estimated using least squares methodology. In order to apply structural break test, we firstly check stationarity of the variables with Augmented Dickey Fuller Test and Zivot Andrews unit root test. The result of the test are presented in the table below:

financial variable	Break 1	Break 2	Break 3
VIX Index	May 2007	October 2008	-
S and P 500 index	October 2007	February 2009	-
iTraxx index	May 2007	November 2008	February 2010
EMBI index	May 2007	November 2008	-
TED Spread	February 2007	September 2008	December 2009
Libor-OIS Spread	July 2007	October 2008	January 2010
Euribor-OIS Spread	June 2007	September 2008	September 2009
eMID-ECB Base Rate	December 2006	September 2008	-

Table 5: The occurence in time of break points.

According to the test results, while most of the indices show breaks around May-July 2007 the TED spread and eMID (both measures related to the interbank market) show signs of stress much earlier. Given that, these results suggest that the crisis propagated first to the interbank markets, due to liquidity squeeze, and spilled over the other markets only later.

Besides retrospective detection of structural changes, we also test the break dates in eMID interbank rate with sequential test where break dates are computed sequentially when new observations arrive (Chu et al.,1996;Leisch et al.,2000). Depending on the historical period chosen, this test picks up the break point between February 2007 and June 2007 in eMID rate. In conclusion,this preliminary findings highlight the existence of signals of crisis before common belief, and other econometric methodologies and financial markets will be investigated to assess the robustness of these findings.

Early signals of financial crisis around February 2007 detected in this paper is in line with Gonzales-Hermosillo and Hesse (2009) which find that VIX, forex swap and TED spread move into high volatility regime in February 2007 by using Markov-regime switching technique. As a result, we provide evidence that the variables related to interbank market reflect earlier episodes of distress in February 2007 which coincide with the crash of Shangai stock exchange and rapid decline in structured credit mortgage-backed instruments measured by the ABS indices (ABX). In conclusion,this preliminary findings highlight the existence of signals of crisis before common belief, and other econometric methodologies and financial markets will be investigated to assess the robustness of these findings.

T4.1.3 The FED Discount Window Market (ETHZ)

The FED Discount Window Lending system is the ordinary mechanism for the financial institutions operating in the US to cope with temporary and small shortages of liquidity. However, the FED is also the last resort lender for the cases when the liquidity shortage is serious. Financial institutions in distress can apply for emergency loans from the FED even if they can only provide commercial papers as a collateral. The dataset, as described in T1.1 of WP1 contains the amount, data and maturity of all the loans granted by FED in a 2-year window (March 2008 - March 2010). At ETHZ we have carried out a preliminary analysis which is reported in a working paper that will be available on the FOC website by October 3, 2011. In that report, we provide the most relevant statistics at the aggregate level for the row data. This analysis constitutes a precondition for any type of investigation of this dataset, which we are happy to pursue in collaboration with other nodes of FOC.

At the microscopic level, we start from the individual loans and we find a broad distribution of loan size (which ranges from 1'000 USD to 61 Billions USD), but no signature of power law. There seems to be a characteristic size of the individual loan, in log scale. Over time, the number of loans granted shows a marked increase since October 2008.

By aggregating by institution, we study the liability across single institutions and we find a distribution with an exponential right tail, indicating a characteristic scale of the liability around few million USD. This number should not be underestimated, since few millions of liquidity are already a significant amount even for medium/large institutions.

We then compute an estimate of FED's credit exposure during the crisis. In particular, we compute the total outstanding lending of FED to all the institutions benefiting from the DW, taking into account the duration of each loan. We then characterize the total FED exposure in terms of:

- Temporal dynamics (how much was FED exposure over time)
- Breakdown by type of collateral provided by the borrower. These greatly varyin terms of risk (e.g., asset backed commercial papers, Primary Dealer Credit Facility, etc.)
- Concentration of the risk. We measure on how many borrowers was the exposure concentrated, using the non-linear Herrfindhal concentration index (akin to the order parameter in statistical mechanics).
- Probability distribution of occurrence. We aim to provide an estimate of how likely it is that FED had to face exposure higher than a certain level.

We find that during the months January- August 2009 there was a peak of risk concentration quickly followed by a peak in the total exposure, as shown in Figure 5. We also find an increase in the fraction of risky collaterals represented by ABCP, as shown in Figure 6.

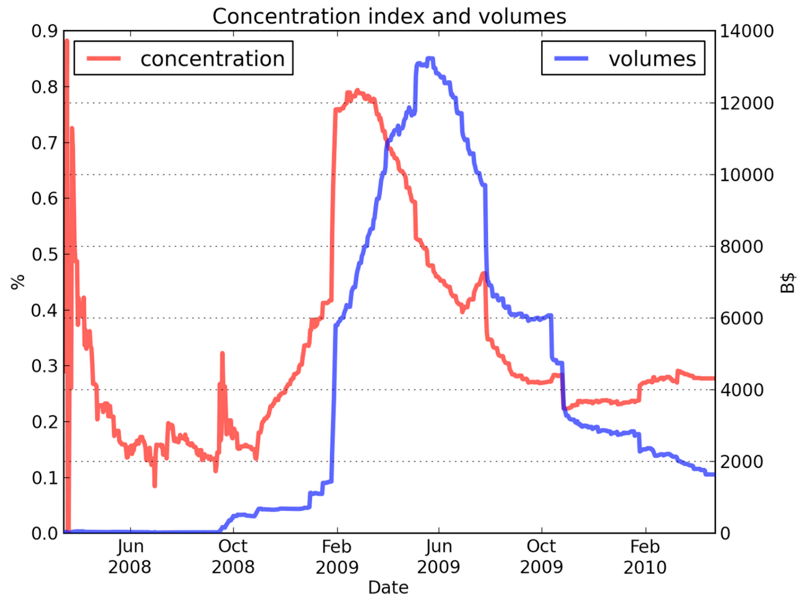

Figure 5. Total exposure of FED over time and its concentration index (index =1 for highest concentration).

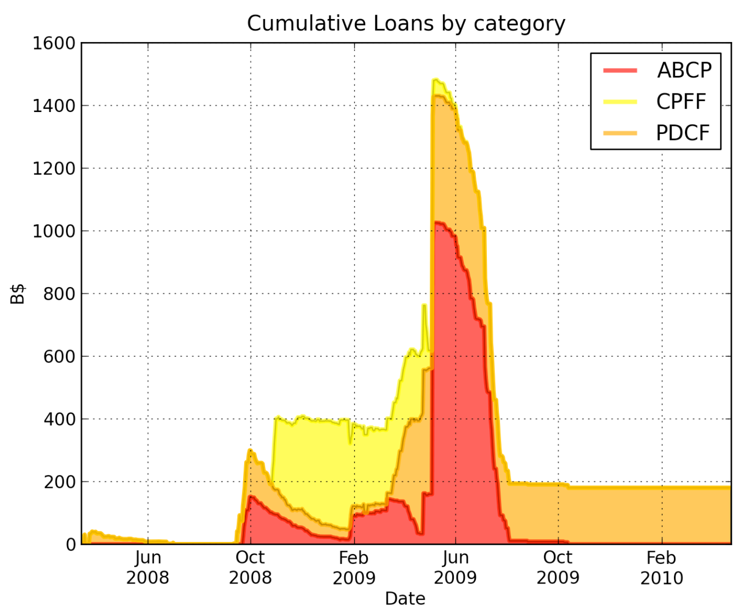

Figure 6. Total exposure of FED , breakdown by asset type. ABCP are the most risky.

T4.2 Interdependence and trend reinforcement (ETHZ) (Planned duration Month 1-24)
The goal of this activity is to try and detect in empirical data the presence of two mechanisms, namely interdependence and trend reinforcement, which play an important role in several of the models developed in WP2. Interdependence, refers to the situation in which a variation in the financial robustness of institution i is likely to be followed by a variation in the financial robustness of institution j at the same time or at a later time.

Similarly, trend reinforcement refers to the situation in which a variation in the financial indicator is likely to be followed by a variation in the same direction at a later time t. Cross-correlation and auto-correlation analysis is not well suited to detect trend interdependence and reinforcement due to some important drawbacks (e.g, absence of correlation does not imply absence of interdependence). For this reason, we have adopted a conditional probability approach and we have developed a method based on the detection of the so-called ε-drawdowns and ε-drawups. Essentially, these are movements in the time series of the indicator for which the rebound exceeds a threshold of magnitude ε.

We have focused our effort on daily time series of Credit Default Swaps (CDS) spreads for the major US financial institution, which we gathered from the Bloomberg platform. In the case of CDS spreads, a drawup signifies an increase in the default probability of the reference entity as perceived by the market (although the default probability is not a simple function of the spread). We count both the drawups (also drawdowns, although not mentioned hereafter) in a given security and the joint drawups (i.e., in the same day) in pairs of securities. We take the frequency of drawups as an estimate of the probability P_i that security i has a drawup. Similarly, we take the frequency of joint drawups as an estimate of the probability P_{ij} that both security i and j have an drawup. We also account for possible time lags τ between the drawups. The expected probability of joint drawups in the case of two statistically independent time series is $P_{ij} = P_i P_j$.

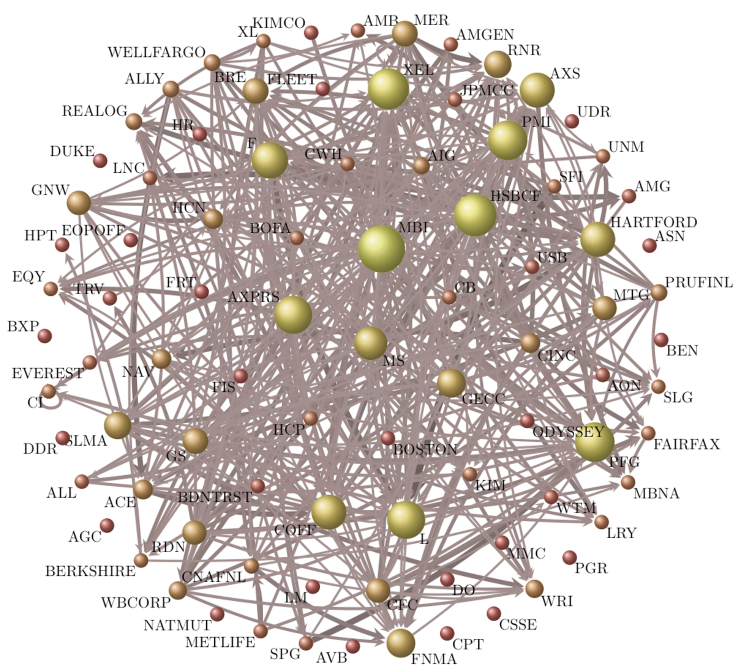

Figure. Network of interdependencies among references entities, estimated from CDS spreads time series.

Therefore, as a possible estimate of the dependence between two financial institutions we can take the deviation of the probability of joint drawups from the case of independent series: $\Delta_{ij}=P_{ij} - P_iP_j$. We also consider the possibility of time lags between the drowups. Analogously, we take as an estimate of the trend reinforcement of security i the deviation from the case of independent series' of the probability of joint drawups in i's time series at a given time lag. In order to assess the level of trend reinforcement in the CDS data, we measure the distribution of values Δ_{ii} across securities and we compare it to the distribution obtained with a control data set of simulated time series of (1) random walks. A statistical analysis based on q-q plots shows that a significant fraction of securities (e.g. about 43 % in the second period and 31% in the first period) tend to show in their spread evolution a level of persistence that cannot be explained by random walks alone. Similarly, we find that there are a number of pairs (e.g. 1.3\% in period 1, 11.2\% in period 2) of securities that exhibits a significant level of interdependence

Finally from the filtered values of Δ_{ii} , we obtain a directed matrix, inwhich we inteepret as a network of dependencies among institutions (see Figure) and which we further investigate in terms of centrality.

T4.5 Empirical analysis of query-logs (Planned duration Month 24-42)

T4.5.1 Correlation between query-logs and stock market trends(BM,CNR,ETHZ)
Recent investigations showed that Web search volumes can be used to accurately track several social phenomena. One of the most successful results in this direction, concerns the epidemic spreading of influenza virus among people in the USA. It has been shown that the activity of people querying search engines for keywords related to influenza and its treatment allows to anticipate the actual spreading as measured by official data on contagion collected by Health Care Agencies. We show that trading volumes of stocks traded in NASDAQ-100 are correlated with the volumes of queries related to the same stocks. In particular, query volumes anticipate in many cases peaks of trading by one day or more. Our analysis is carried out on a unique dataset of queries, submitted to an important web search engine, which enable us to investigate also the user behaviour. We show that the query volume dynamics emerges from the collective but seemingly uncoordinated activity of many users. Fig. 1 (top panel) shows the time evolution of the query volume of the ticker ``NVDA'' and the trading
volume of the corresponding company stock ``NVIDIA Corporation''

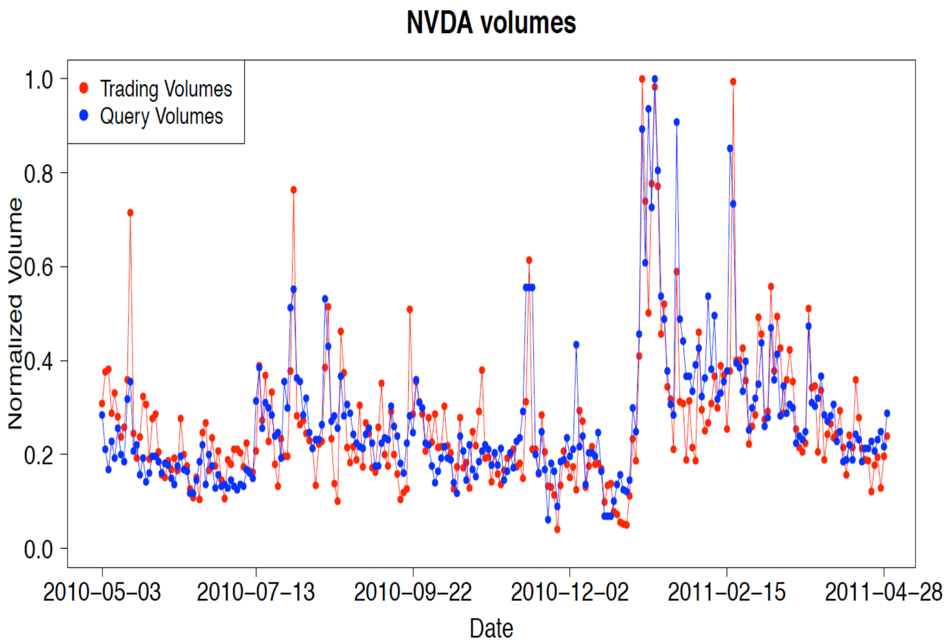

Figure 7. Example of time evolution of trading volume and query log volume related to Apple.com

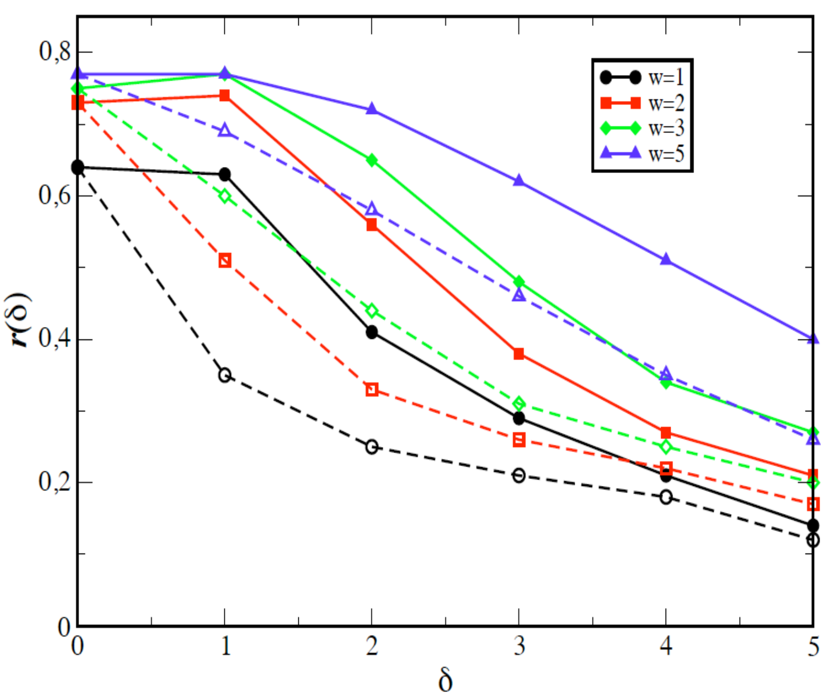

Figure 8. The time-lagged Pearson cross correlation $r(\delta)$ as a function of the time lag.

The time-lagged Pearson cross correlation $r(\delta)$ coefficient between two time series Xi and Yi is:

$$r(\delta) = \frac{\sum_{i=1,N} (Xi - <X>)(Yi - <Y>)}{\sqrt{\sum_{i=1,N} (Xi - <X>)^2} \sqrt{)\sum_{i=1,N} (Yi - <Y>)^2}}$$

where <X>, <Y> are the sample average of the two time series and δ is a time lag. The coefficient ranges from -1 (anticorrelation) to 1 (correlation). To reduce short-term fluctuations, for each time series, we also average the signal at a given time t with the w-1 previous points. We use windows of width w =1,2,3,5.

T4.5.2 Twitter data and stock market movements (BM, CITY)

Twitter (www.twitter.com) is a social networking and messaging service that allows users to post messages of at most 140 characters, called "tweets", to other Twitter users. Contrary to email or SMS messages, tweets are not intended to be a private communications medium. A tweet is received by every user who (in Twitter terminology) "follows" the sender. Unlike in most other social networking sites, connections in Twitter are directed, meaning that any user can become a follower of any other user without having to be confirmed as a "friend". Addressing a tweet to a restricted subset of the recipients is not possible. Moreover, it is common for interesting tweets to be re-sent ("re-tweeted") more or less verbatim by their recipients to "spread the word" to their own followers. Most of the tweets are also openly visible to anybody e.g. through a web-based search interface. Yahoo! has access to all tweets posted in the past 18 months, including metadata (number of followers, total number of tweets, etc) about the users at the time the tweet was posted.

In finance data from Twitter has already been used to try and predict stock market movements from trader sentiments (Bollen et al, 2011, Sprenger and Welpe, 2010) .

Yahoo! and City have started analyzing twitter data with the goal to study how financial information propagates within the trader community (for example study the emergence of gurus and herding behaviour) and how the information flow correlates to the unfolding a financial crisis.

Identifying the trader community is a challenge in itself. There are websites with topic-specific lists of Twitter users. For instance: http://wefollow.com/twitter/trading shows a list of people hat define themselves as "traders". at the time of writing, over 1,000 Twitter users had added themselves under the topic "trading" at wefollow.com. Of course not all of these are professionals, but at least it is a starting point. Also, simply by considering those users who frequently use stock ticker symbols in their tweets are potential candidates for the "trader" community. Furthermore traders have their own jargon (for example they use the $ symbol in front of the stock tickers). We took the tickers from sp-500 and nasdaq-100 indices, and filtered a list of users who have at least 50 tweets (sent in 2011) that contain one of the tickers prefixed with the $-sign. (Using only the tickers, i.e. without the $-sign, leads to a lot of noise created by tweets in languages other than

English that contain words matching the tickers). In an initial experiment we extracted a set of approximately 3500 users that seem suitable candidates as traders for the analysis.

Supplementary Analysis

To complement the analysis done on the robustness of national economies we also considered the network of countries and products from UN data on country production. We define the country-country and product-product networks and we introduce a novel method of community detection based on elements similarity. As a result we find that country clustering reveals unexpected socio-geographic links among the most competing countries. On the same footings the products clustering can be efficiently used for a bottom-up classification of produced goods. Furthermore we define a procedure to rank different countries and their products over the global market. These analyses are a good proxy of country GDP and therefore could be possibly used to determine the robustness of a country economy. We introduce the new concept of competition space and determine its topological structure. By introducing the new concept of Minimal Spanning Forest, we analyse in detail the competition space and uncover the geographical correlations in country competition

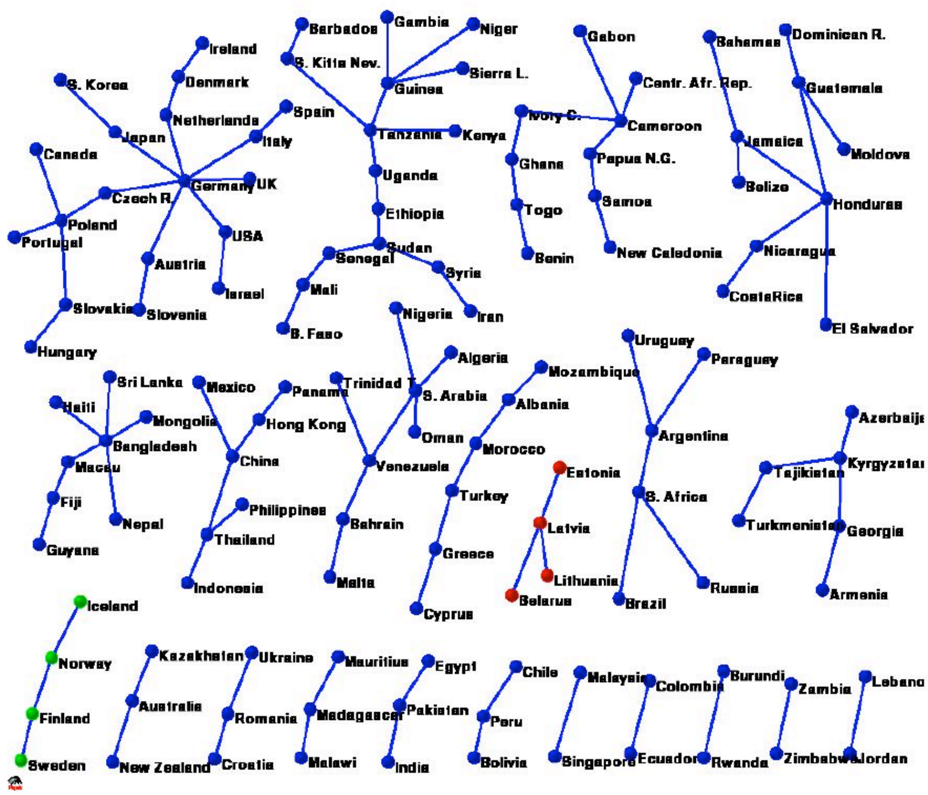

Figure 9. Minimal Spanning Forest showing the communities of countries more strictly competing on the same products.

Bibliography:

Sprenger, Timm O. and Welpe, Isabell M., Tweets and Trades: The Information Content of Stock Microblogs (November 1, 2010). Available at SSRN: http://ssrn.com/abstract=1702854

Johan Bollen, Huina Mao, Xiao-Jun Zeng: Twitter mood predicts the stock market. Journal of Computational Science, 2(1), March 2011, 1-8.