Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Physica A 339 (2004) 181–188

PHYSICA A

www.elsevier.com/locate/physa

# An interest rates cluster analysis

T. Di Matteo[a,*], T. Aste[a], R.N. Mantegna[b]

[a] *Applied Mathematics, Research School of Physical Sciences, Australian National University, Canberra 0200, Australia*

[b] *INFM, Dipartimento di Fisica e Tecnologie Relative, Università di Palermo, viale delle Scienze, 90128 Palermo, Italy*

**Abstract**

An empirical analysis of interest rates in money and capital markets is performed. We investigate a set of 34 different weekly interest rate time series during a time period of 16 years between 1982 and 1997. Our study is focused on the collective behavior of the stochastic fluctuations of these time series which is investigated by using a clustering linkage procedure. Without any a priori assumption, we individuate a meaningful separation in 6 main clusters organized in a hierarchical structure.


*PACS:* 89.65.Gh; 05.45.Tp; 89.90.+n

*Keywords:* Interest rates; Data clustering; Correlations; Econophysics

## 1. Introduction

Since long time financial data have been widely studied by economists, mathematicians and, more recently, by physicists [1–3]. The variations of these financial time series can be seen as stochastic processes where a set of financial quantities is varying in time as a consequence of underlying economic changes. The present availability of enormous sets of financial data allows to get increasingly important insights on the complex behavior of these systems starting from empirical studies. These investigations are leading to more and more accurate results on risk assessment and search for market imperfections. One of the important points in these analyses is to individuate

* Corresponding author. Tel.: +61-2-61250166; fax: +61-2-61250732.
*E-mail address:* tiziana.dimatteo@anu.edu.au (T. Di Matteo).

0378-4371/$ - see front matter 
doi:10.1016/j.physa.2004.03.041

similarities and specificities among the analyzed financial time series. This search has been widely exploited for stocks price changes whereas interest rates have been less investigated [4–9]. For several economic reasons, interest rates and bonds have very similar statistical behavior in time or, in other words, they are all highly correlated. Their multivariate dynamics have been studied with a correlation-based clustering procedure in a set of US treasury securities where an underlying hierarchical structure has been detected [7]. Here we investigated a partially different and less homogeneous set to evaluate the degree of hierarchal organization among different time series in a diversified group of bonds.

## 2. An empirical analysis on interest rates

We investigate weekly data for 34 selected interest rate time series recorded in the Federal Reserve (FR) Statistical Release database [10]. In the following we will indicate these time series with the symbol $f_i(t)$, where $t$ is the current date and $i$ is a number which labels the different time series (see Table 1). The different interest rate time series analyzed are: The Federal funds rate (FED); State & local bonds (SLB); Commercial Paper (CP); Finance Paper placed directly (FP); Bankers acceptances (BA); The rate on certificates of deposit (CD); (Note that in these cases the numbers 1, 3 and 6 stand for maturity dates of 1, 3 and 6 months.); The yields on Treasury securities at 'constant maturity' (TC) (in particular the TC at 3 and 6 months (TC3M, TC6M)) and $1, 2, 3, 5, 7, 10$, and 30 years (TC1Y-TC30Y) maturities; The Treasury bill rates (TBA) with maturities of 3 and 6 months (TBA3M, TBA6M, TBS3M, TBS6M) and 1 year (TBS1Y); The Treasury long-term bond yield (TC10P); The Eurodollar interbank interest rates (ED) with maturity dates $1, 3$ and 6 months (ED1M, ED3M, ED6M), respectively; The Corporate bonds Moody's seasoned rates (AAA, BAA) and The Conventional mortgages rates (CM). Their characteristics can be found in Ref. [10]. Unless differently stated, we report weekly data obtained from unweighted averages of daily data ending on Friday.

Table 1
Interest rates and standard deviations in the time period 1982–1997

| $i$ | $f_i$ | $\sigma_i$ | $i$ | $f_i$ | $\sigma_i$ | $i$ | $f_i$ | $\sigma_i$ | $i$ | $f_i$ | $\sigma_i$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | FED | 0.30935 | 10 | BA6 | 0.17225 | 19 | TC5 | 0.15715 | 28 | TC10P | 0.13608 |
| 2 | SLB | 0.13001 | 11 | CD1 | 0.21291 | 20 | TC7Y | 0.15363 | 29 | ED1M | 0.2166 |
| 3 | CP1 | 0.22257 | 12 | CD3 | 0.1901 | 21 | TC10Y | 0.14863 | 30 | ED3M | 0.19926 |
| 4 | CP3 | 0.19011 | 13 | CD6 | 0.19299 | 22 | TC30Y | 0.13288 | 31 | ED6M | 0.20104 |
| 5 | CP6 | 0.17951 | 14 | TC3M | 0.1925 | 23 | TBA3M | 0.21838 | 32 | AAA | 0.10695 |
| 6 | FP1 | 0.21418 | 15 | TC6M | 0.18271 | 24 | TBA6M | 0.20186 | 33 | BAA | 0.09411 |
| 7 | FP3 | 0.15407 | 16 | TC1Y | 0.16993 | 25 | TBS3M | 0.17672 | 34 | CM | 0.11556 |
| 8 | FP6 | 0.13842 | 17 | TC2Y | 0.16347 | 26 | TBS6M | 0.16262 | | | |
| 9 | BA3 | 0.17838 | 18 | TC3Y | 0.16227 | 27 | TBS1Y | 0.14789 | | | |

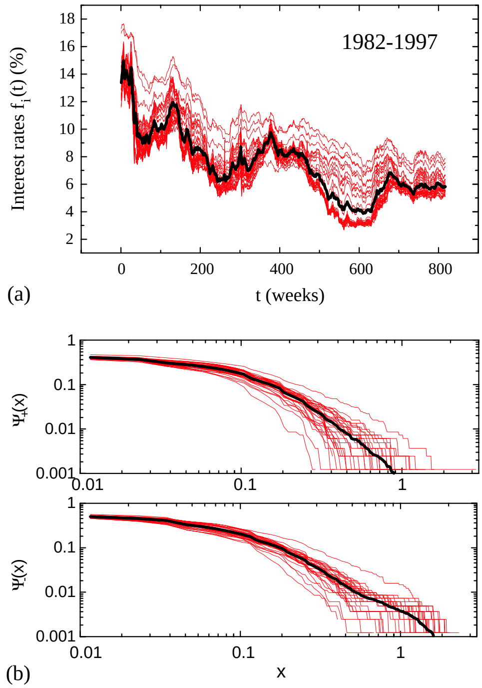



Fig. 1. (a) Interest rates $f_i(t)$ as function of $t$ for $i=1$–34 (gray lines) and their average $\bar{f}(t)$ (black line). (b) Cumulative distributions for the probabilities of the interest rates fluctuations. $\Psi_+(x)$ and $\Psi_-(x)$. (The black lines are their averages.)

## 3. Fluctuations

The interest rate time series, $f_i(t)$ vs. $t$ are shown in Fig. 1, where their average $\bar{f}(t)=\sum_i f_i(t)/34$ is also shown. It is evident from Fig. 1 that all these data follow very similar trends in time and they lay in a narrow band around $\bar{f}(t)$. The interest rate fluctuations are analyzed by studying the changes in their values from one week to the following week: $\Delta f_i(t)=f_i(t+\Delta t)-f_i(t)$, where $\Delta t=1$ week. The quantities $\Delta f_i(t)$ show stochastic fluctuations around the zero with similar behaviors for all the interest rates. These fluctuations are analyzed in the 'tails' region by computing the cumulative distributions $\Psi_\pm(\pm x)$ $(x>0)$, a quantity which tells us the probability to find a weekly change which is *larger than* $x$ $(+)$, or *smaller than* $-x$ $(-)$. It is defined as: $\Psi_+(x)=1-\int_{-\infty}^{x} p(\xi)\,\mathrm{d}\xi$ and $\Psi_-(x)=\int_{-\infty}^{-x} p(\xi)\,\mathrm{d}\xi$ with $p(\xi)$ being the probability density distributions of $\Delta f_i(t)$ (see Fig. 1 (b)). These distributions are highly leptokurtic and are characterized by non-Gaussian profiles. The standard deviation of $\Delta f_i(t)$ is defined as: $\sigma_i=\sqrt{1/(T_2-T_1)\sum_{t=T_1}^{T_2}(\Delta f_i(t)-\langle\Delta f\rangle)^2}$, where $T_1$ and $T_2$ delimit the range of $t$, and $\langle\Delta f\rangle$ is the average over time of $\Delta f_i(t)$ (which tends to zero for $T_2-T_1\to\infty$). We compute the standard deviations of $\Delta f_i$ for each

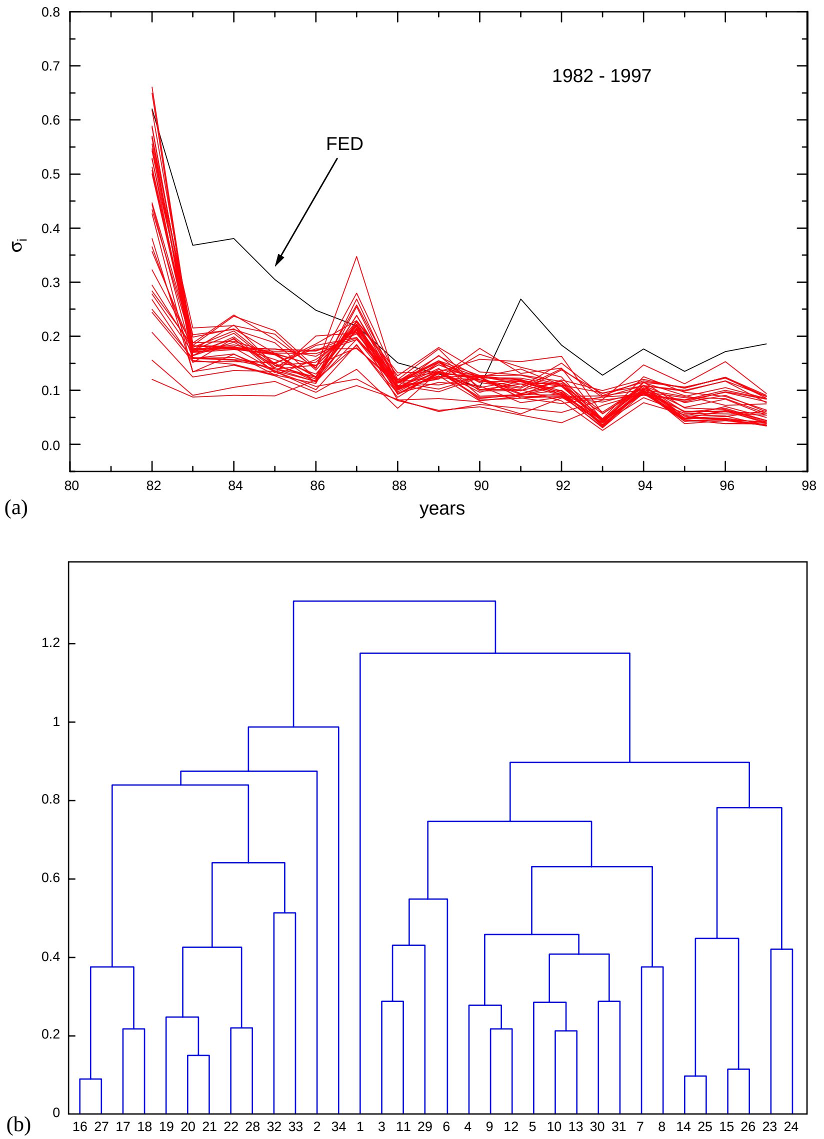



Fig. 2. (a) Standard deviations of $\Delta f_i(t)$ for all the rates analyzed ($i = 1..34$) as function of the years between 1982 and 1997. (b) Hierarchical tree obtained from the correlation coefficients of the 34 interest rates fluctuations time series $\Delta f_i(t)$ in the time period 1982–1997. (On the $x$-axis are reported the $i$ values and on the $y$-axis the ultra-metric distances.)

interest rate series for the whole period 1982–1997 (see Table 1) and for each year (see Fig. 2 (a)). We can observe an overall decreasing trend of $\sigma_i$ in the time period 1982–1997 with similar fluctuations for all the interest rates series except for FED.

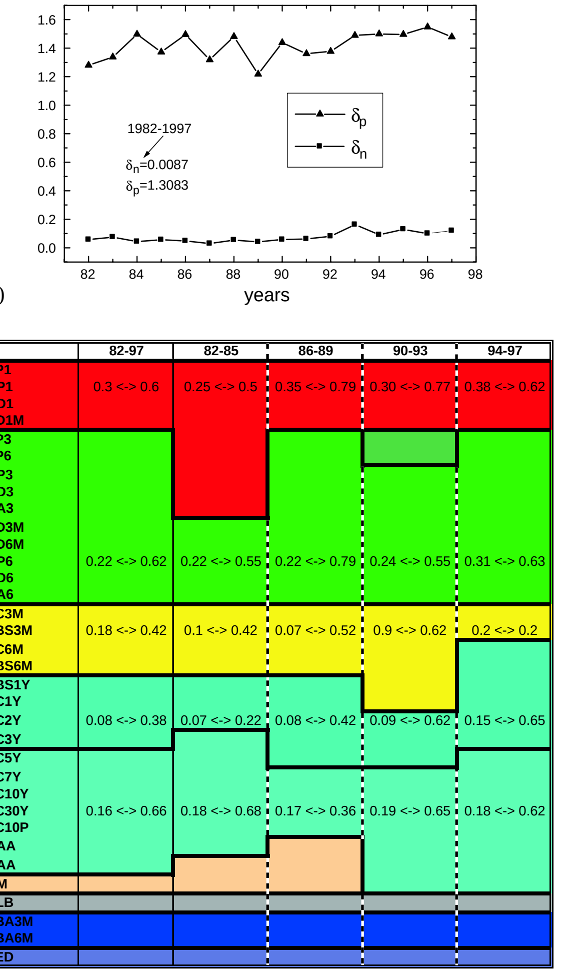
## 4. Cluster analysis and discussion

To understand the geometrical and topological structure of the correlation coefficients, we use the metric distance $d_{i,j}$ between the series $\Delta f_i$ and $\Delta f_j$ which is defined in Ref. [11] and used for financial time series in Ref. [12]: $d_{i,j} = \sqrt{2(1 - c_{i,j})}$ with $c_{i,j}$ the correlations among the $i, j$ interest rates weekly changes

$$c_{i,j} = \frac{\langle \Delta f_i \Delta f_j \rangle - \langle \Delta f_i \rangle \langle \Delta f_j \rangle}{\sigma_i \sigma_j} \,, \tag{1}$$

where the symbol $\langle \cdots \rangle$ denotes a time average performed over the investigated time period. The correlation coefficients are computed between all the pairs of indices labeling our interest series. Therefore we have a $34 \times 34$ symmetric matrix with $c_{i,i} = 1$ on the diagonal. By definition, $c_{i,j}$ is equal to zero if the interest rate series $i$ and $j$ are totally uncorrelated, whereas $c_{i,j} = \pm 1$ in the case of perfect correlation/anti-correlation. Therefore, $d_{i,j}$ can vary between 0 and 2. We determine an ultra-metric distance $\hat{d}_{i,j}$ which satisfies the first two properties of the metric distance and replaces the triangular inequality with the stronger condition: $\hat{d}_{i,j} \leqslant \max[\hat{d}_{i,k}, \hat{d}_{k,j}]$, called 'ultra-metric inequality'. Once the metric distance $d_{i,j}$ is used, one can introduce several ultra-metric distances. Mantegna et al. have used the 'subdominant ultra-metric', obtained by calculating the minimum spanning tree connecting several financial time series [13–16]. Here, we consider a different ultra-metric space that emphasizes the cluster-structure of the data. In our case, a 'cluster' is a set of elements with relative distances $d_{i,j}$ which are smaller than a given threshold distance $\bar{\delta}$, whereas disjoined clusters have some elements which are at distances larger than $\bar{\delta}$. We define the *ultra-metric distance* $\hat{d}_{i,j}$ between two distinct elements $i, j$ belonging to two different clusters as the maximum metric distance between all the couples of elements in the two clusters [8]. The linkage procedure yields to a hierarchical graph as shown in Fig. 2 (b), which refers to the 34 time series in the whole period 1982–1997. The clustering process starts with a nucleation between TC1Y and TBS1Y at the ultra-metric distance $\delta_n = 0.087$. The clustering ends when all the series merge in a unique large cluster at the ultra-metric distance $\delta_p = 1.3083$. At this distance all the interest rates with maturity dates smaller or equal than 6 months (already merged with the FED at $\hat{d} = 1.18$) make a single cluster with another cluster composed of interest rates with maturity dates larger or equal to 1 year. The same analysis performed on each year, gives comparable $\delta_n$ and $\delta_p$ values which are plotted in Fig. 3 (a). Let us now consider the intermediate region by analyzing the cluster evolution at the threshold distance $\bar{\delta} = \frac{1}{\sqrt{2}} = 0.707\ldots$, which is half the way between completely uncorrelated series ($c_{i,j} = 0$ and $d_{i,j} = \sqrt{2}$) and completely correlated ones ($c_{i,j} = 1$ and $d_{i,j} = 0$). At this threshold distance, the cluster analysis on the whole data set (1982–1997) leads to 6 clusters and 3 isolated elements, as one can see from Fig. 2 (b).

The corresponding interest rates associated with these clusters are summarized in first column (1982–1997) of Fig. 3 (b) where the ultra-metric distances at which the nucleation process starts and ends for each cluster are also indicated. As one can see, the empirical analysis allow us to distinguish several different clusters that gather




(a)

(b)

Fig. 3. (a) Ultra-metric distances $\delta_n$ and $\delta_p$ at which the clustering process begins and ends, as function of the years between 1982 and 1997. (b) Cluster-structure persistence in the period 1982–1997. In the first column the interest rates are indicated. In the second column, the gray tones distinguish the different clusters as resulting from the analysis over the whole time period. The other columns refer to the cluster analysis over the four selected time periods, namely 82–85, 86–89, 90–93, 94–97. The numbers inside each cluster refer to the ultra-metric distances at which each cluster starts and ends its clusterization.

together meaningful quantities:

- all the interest rates with maturities equal to 1 month;
- all the interest rates with maturities 3 and 6 months with distinctions for the Treasury securities at ‘constant maturity’ (TC), Treasury bill rates (TBA) and Treasury bill secondary market rates (TBS);

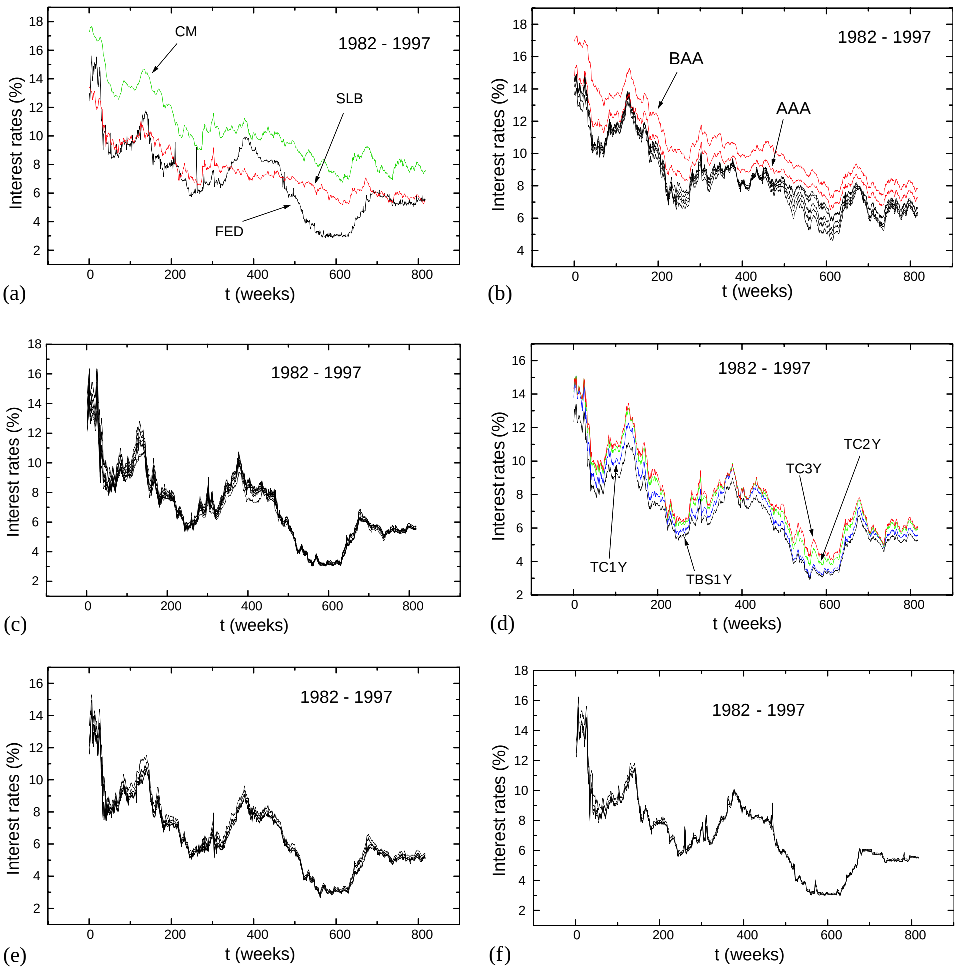



Fig. 4. Interest rates behaviors in the period 1982–1997. The figures refer to the data sets gathered into the clusters obtained from the linkage procedure. (a) FED, SLB and CM; (b) BAA, AAA, TC5Y, TC7Y, TC10Y, TC30Y, TC10P; (c) CP3, CP6, FP3, FP6, BA3, BA6, CD3, CD6, ED3M, ED6M; (d) TC1Y, TC2Y, TC3Y, TBS1Y; (e) TC3M, TC6M, TBA3M, TBA6M, TBS3M, TBS6M; (f) CP1, FP1, CD1, ED1M.

- all the interest rates with maturities between 1 and 3 years;
- all the interest rates with maturities larger than 3 years.

Fig. 4 reports the plot for the interest rates time series $f_i(t)$ grouped into the different sets retrieved from the cluster analysis described above. For some of the series the data-collapse is impressive, indicating that the correlations inside the clusters are strong in any part of the analyzed period. It is therefore interesting to investigate whether a cluster structure, similar to the one obtained for the period 1982–1997, could be retrieved from an analysis on shorter periods. This is of course a delicate point since the fragmentation of the data sets will increase the fluctuations due to the noise. We choose to divide the whole period 1982–1997 in four smaller periods of 4 years. The results are reported in Fig. 3 (b) where it is evident how the cluster structure is mostly conserved (and partially modified) in this 4-years period analysis. In conclusion, from the analysis of different kinds of interest rates in money and capital markets, referring to government, private, industries securities and commitments, we have shown how the used clustering linkage procedure is useful to detect differences and analogies among these tangled correlated data.

## References

[1] R.N. Mantegna, H.E. Stanley, An Introduction to Econophysics, Cambridge University Press, Cambridge, 2000.
[2] J.P. Bouchaud, M. Potters, Theory of Financial Risks, Cambridge University Press, Cambridge, 2000.
[3] M.M. Dacorogna, R. Gencay, U.A. Muller, R. Olsen, O.V. Pictet, An Introduction to High-Frequency Finance, Academic Press, New York, 2001.
[4] A.R. Pagan, A.D. Hall, V. Martin, Modeling the Term Structure, Handbook of Statistics, vol. 14, Elsevier Science B. V., Amsterdam, 1997.
[5] R. Rebonato, Interest-Rate Option Models, Wiley, New York, 1998.
[6] J.P. Bouchaud, N. Sagna, R. Cont, N. EL-Karoui, M. Potters, cond-mat/9712164 (1997), Appl. Math. Finan. 6 (1999) 209–232.
[7] M. Bernaschi, L. Grilli, D. Vergni, Physica A 308 (2002) 381–390.
[8] T. Di Matteo, T. Aste, J. Theor. Appl. Finan. 5 (2002) 122–127, cond-mat(0101009).
[9] T. Alderweireld, J. Nuyts, Physica A 331 (2004) 602–616.
[10] http://www.federalreserve.gov/releases/h15/data.htm.
[11] J.C. Gower, Biometrika 53 (1966) 325–338.
[12] R.N. Mantegna, Eur. Phys. J. B 25 (1999) 193–197.
[13] G. Bonanno, N. Vandewalle, R.N. Mantegna, Phys. Rev. E 62 (2000) R7615–R7618.
[14] G. Bonanno, F. Lillo, R.N. Mantegna, Quant. Finan. 1 (2001) 96–104.
[15] S. Miccichè, G. Bonanno, F. Lillo, R.N. Mantegna, Physica A 324 (2003) 66–73.
[16] G. Bonanno, G. Caldarelli, F. Lillo, R.N. Mantegna, Topology of correlation based minimal spanning trees in real and model markets, Phys. Rev. E 68 (2003) 046130.