Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica A 355 (2005) 21–33
PHYSICA A
www.elsevier.com/locate/physa

Interest rates hierarchical structure

T. Di Matteo*, T. Aste, S.T. Hyde, S. Ramsden

Applied Mathematics, Research School of Physical Sciences, Australian National University, 0200 Canberra, Australia

Received 31 October 2004; received in revised form 11 January 2005
Available online 4 May 2005

Abstract

We propose a general method to study the hierarchical organization of financial data by embedding the structure of their correlations in metric graphs in multi-dimensional spaces. An application to two different sets of interest rates is discussed by constructing triangular embeddings on the sphere. Three-dimensional representations of these embeddings with the correct metric geometry are constructed and visualized. The resulting graphs contain the minimum spanning tree as a sub-graph and they preserve its hierarchical structure. This produces a clear cluster differentiation and allows us to compute new local and global topological quantities.

Keywords: Interest rates; Data clustering; Correlations; Econophysics

1. Introduction

In this paper we investigate the hierarchical organization of interest rates data and we discuss a general method to characterize the statistical, geometrical and topological properties of correlation matrices in complex systems. A number of physicists have observed that the structure of the correlation coefficients from complex datasets (such as time series from financial markets) can be conveniently studied by mapping the data structures onto graphs [1–7]. Mantegna [1] has studied

*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.2005.02.063

the hierarchical organization of such correlations by retaining only the most relevant correlations which form the minimal spanning tree (MST) (see also [3,5–7]). However, this reduction to a minimal skeleton necessarily results in extreme sensitivity to dynamical variations in the system (small variations in interactions). An extension from trees to graphs has been proposed by Onnela et al. [6], but a general procedure which allows the construction of connected networks with varying degrees of information content and controlled complexity has not been introduced yet. In this framework, we propose to map the interest rates correlations into graphs embedded on manifolds with different topologies in multi-dimensional spaces, both Euclidean and non-Euclidean.

We chose to apply this method to interest rates data, because, in the Econophysics literature, these data have been less investigated in comparison with the investigations performed on stock market prices. Recently, these studies have become very attractive and approached from many different perspectives [8–14].

For several economic reasons, distinct interest rates have very similar statistical behaviors in time, following similar trends. This makes the subject very challenging since one is no more dealing with the statistics of single objects but with the collective dynamics of a whole complex set of highly correlated data.

Our study starts from the analysis of the collective behavior of the stochastic fluctuations of interest rates data by using a clustering linkage procedure which has been proved to be a useful tool to detect differences and analogies among these tangled correlated data [12]. The output of this clustering linkage procedure, gives us the simplest picture of the interest rates hierarchical organization. We show that the same clustering structure naturally emerges from the embedding on the sphere of the graph made by retaining the strongest correlations. We emphasize that the resulting graph contains the MST.

The structure of this paper is as following: Section 2 describes the interest rates data set, Section 3 gives a resumé of the main outcomes obtained from a correlation cluster analysis. The general idea and the specific results are reported in Section 4 where a three-dimensional (3D) visualization of these metric graphs is also shown.

2. Data description

We investigate two data sets: 16 Eurodollars interest rates (Set 1) and 34 interest rates in money and capital markets, referring to government, private, industries securities and commitments (Set 2). Set 1 contains daily values $f_i(t)$, where t is the current date and $3i = \theta$ is the maturity date in the time period 1990–1996 [11]. On this time period we have 16 different time series corresponding to maturity dates ranging from $\theta = 3$ to 48 months with a step of 3 months (reported in Table 1). The behavior of $f_i(t)$ as function of t is shown in Fig. 1 where we use $i(= \theta/3)$ to label the different maturity dates θ. For a better visualization of the plot, in Fig. 1(b) we report only those values corresponding to the following maturity values: $\theta = 3, 15, 30, 48$ months. The interest rates behaviors for all maturity dates follow very similar trends in time, and stay mostly inside the shape traced by the two extreme

Table 1
Data set 1: Eurodollar interest rates in the time period 1990–1996

i	θ months	i	θ months	i	θ months	i	θ months
1	3	5	15	9	27	13	39
2	6	6	18	10	30	14	42
3	9	7	21	11	33	15	45
4	12	8	24	12	36	16	48

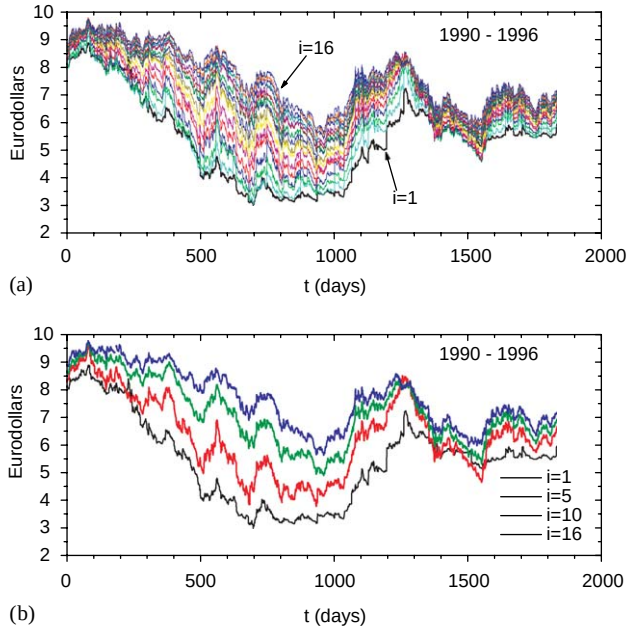

Fig. 1. (a) Eurodollar interest rates ($f_i(t)$) as function of t for i ranging from 1 to 16; (b) Eurodollar interest rates ($f_i(t)$) as function of t for $i = 1, 5, 10, 16$.

maturity values, namely $\theta = 3$ and 48 months. Set 2 contains 34 weekly data for different interest rates during a time period of 16 years between 1982 and 1997 recorded in the Federal Reserve (FR) Statistical Release database [12,15]. We report their main characteristics in Appendix A. 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 2). The behavior in time of these interest rates time series, $f_i(t)$, is shown in Fig. 2, where their average $\bar{f}(t) = \sum_i f_i(t)/34$ is also shown. It is evident from Fig. 2 that all these data follow very similar trends in time and they lie in a rather narrow band around $\bar{f}(t)$.

Table 2
Data set 2: Interest rates in the time period 1982–1997

i	f_i	i	f_i	i	f_i	i	f_i	i	f_i	i	f_i
1	FED	7	FP3	13	CD6	19	TC5Y	25	TBS3M	31	ED6M
2	SLB	8	FP6	14	TC3M	20	TC7Y	26	TBS6M	32	AAA
3	CP1	9	BA3	15	TC6M	21	TC10Y	27	TBS1Y	33	BAA
4	CP3	10	BA6	16	TC1Y	22	TC30Y	28	TC10P	34	CM
5	CP6	11	CD1	17	TC2Y	23	TBA3M	29	ED1M		
6	FP1	12	CD3	18	TC3Y	24	TBA6M	30	ED3M		

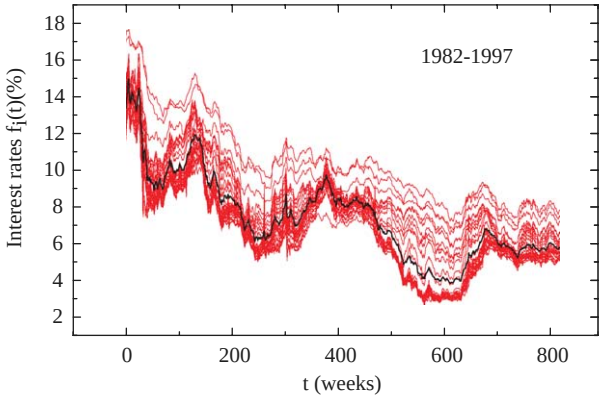

Fig. 2. Interest rates $f_i(t)$ as function of t for $i = 1-34$ (thin gray lines) and their average $\bar{f}(t)$ (thick black line).

3. Correlation cluster formation

We study the hierarchical structure arising from the correlations between the interest rates fluctuations $\Delta f_i(t) = f_i(t + \Delta t) - f_i(t)$ with $\Delta t = 1$ day for Set 1 and $\Delta t = 1$ week for Set 2. To this end, we compute the metric distance $d_{i,j}$ between the series Δf_i and Δf_j (which is defined in [16] and used for financial time series in [1]): $d_{i,j} = \sqrt{2(1 - c_{i,j})}$ with $c_{i,j}$ the correlation coefficient among the i,j interest rates fluctuations:

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

where the symbol $\langle \ldots \rangle$ denotes a time average performed over the investigated time period an σ_i is the standard deviation defined as

$$\sigma_i = \sqrt{\frac{1}{T_2 - T_1} \sum_{t=T_1}^{T_2} (\Delta f_i(t) - \langle \Delta f \rangle)^2} , \tag{2}$$

where T_1 and T_2 delimit the range of t. The correlation coefficients are computed between all the pairs of indices labeling our interest series. Therefore we have a 16×16 (for Set 1) and 4×34 (for Set 2) symmetric matrix with $c_{i,i} = 1$ on the diagonal. By definition, $c_{i,j}$ is equal to zero if the interest rates 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. In order to single out the clustering structure, 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 defined, one can introduce several ultra-metric distances. Mantegna et al. [2–4] have used the 'subdominant ultra-metric', obtained by calculating the minimum spanning tree connecting several financial time series. As the correlations between interest rates are strong in any part of the analyzed period, we have instead considered a different ultra-metric space that emphasizes the cluster structure of the data [11,12]. The result of this procedure, for the 16 Eurodollars interest rates (Set 1) in the whole time period 1990–1996, tells us that the data set is gathered into 3 main clusters: $Cls_1 = \{3, 6\}$, $Cls_2 = \{9, 12, 15, 18, 21\}$, $Cls_3 = \{24, 27, \ldots, 45, 48\}$. The first cluster($Cls_1$) gathers together interest rates with maturity shorter than 1 year; Cls_2 contains those with maturity dates between 1 and 2 years; whereas Cls_3 includes those with maturity dates which are larger than 2 years. For the other 34 interest rates (Set 2) the same procedure yields to a separation in the following clusters:

- all the interest rates with maturities equal to 1 month (CP1, FP1, CD1, ED1M);
- all the interest rates with maturities 3 and 6 months (CP3, CP6, FP3, FP6, BA3, BA6, CD3, CD6, ED3M, ED6M);
- Treasury securities at 'constant maturity' (TC), and Treasury bill secondary market rates (TBS) with maturities 3 and 6 months (TC3M, TC6M, TBS3M, TBS6M);
- Treasury bill rates (TBA) with maturities 3 and 6 months (TBA3M, TBA6M);
- all the interest rates with maturities between 1 and 3 years (TC1Y, TC2Y, TC3Y, TBS1Y);
- all the interest rates with maturities larger than 3 years (BAA, AAA, TC5Y, TC7Y, TC10Y, TC30Y, TC10P).

Finally, there are also three isolated elements, namely FED, SLB and CM. In the next section, we show how the same cluster structure obtained with the ultrametric/linkage procedure spontaneously emerges from the planar sub-graph made by the most correlated links.

4. Results from a 2D embedding and discussion

Here the general idea is the construction and characterization of metric graphs (networks of specific topology and geometry) that encode relevant information

concerning the hierarchical organization, interactions and dynamical properties of these systems. For a set of n interest rates we can associate a point in a multi-dimensional space to each of the n interest rates. To all pairs (i, j) a metric distance d_{ij} is associated and the resulting network is an nth order 'complete graph' (K_n). In this construction, the length of each edge is equal to the metric distance between the two interest rates increments in the multi-dimensional space and short distances are associated with highly correlated rates values. The problem we are addressing is to extract maximal information both qualitative (visual) and quantitative by topological and geometric simplification of the complete graph, without excessive information loss. Such simplification can be obtained by starting from the set of n unconnected nodes and for a given genus g by connecting iteratively two nodes if and only if the resulting graph can be embedded on an orientable surface of genus g: S_g [17]. This process will end with a maximally connected graph compatibly with the surface genus. Here we present an application of this genus-dependent procedure to the two sets of interest rates data (Sets 1 and 2) for the genus $g = 0$ case (the sphere). In Figs. 3, 4 their 3D representations, respectively for Sets 1 and 2, are shown. In both figures we can observe that the resulting graph on S_0 is a triangulation of a topological sphere and we can visualize the hierarchical organization of the whole system [18]. Each node represents an interest rates and the length of each edge is the metric distance $d_{i,j}$ introduced in the previous section. Different colors (on line version) have been chosen to distinguish different clusters. We have relaxed the resulting network numerically [19,20] seeking to make all vertex angles as equal as possible, consistently with the imposition of edge-lengths equal to $d_{i,j}$. A detailed description of this relaxation procedure is given in Appendix B. Note that both graphs in Figs. 3, 4 contain as sub-graph the minimum spanning tree (MST) shown

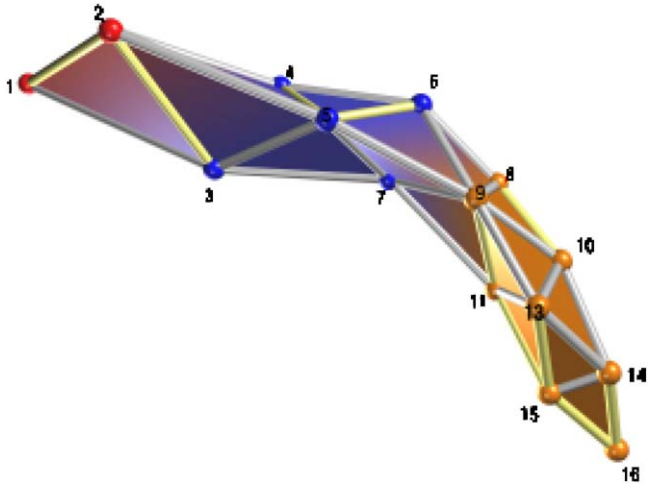

Fig. 3. Three-dimensional representation of the embedding on S_0 of the correlation structure of the 16 Eurodollar interest rates (Set 1). Each edge-length corresponds to the metric distance $d_{i,j}$.

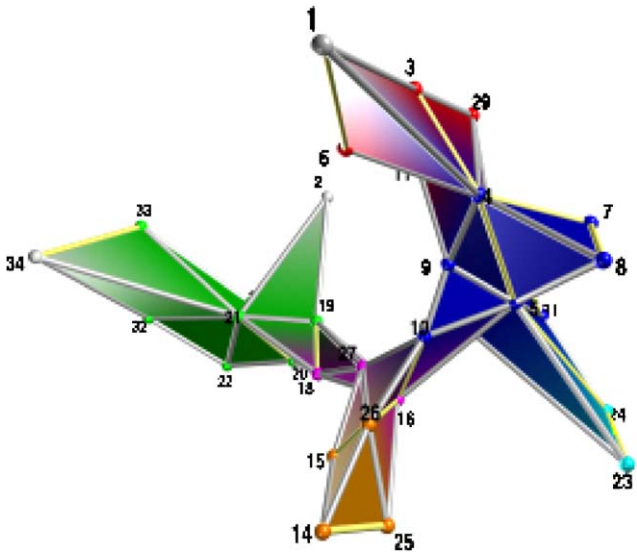

Fig. 4. Three- dimensional representation of the embedding on S_0 of the correlation structure of the 34 interest rates (Set 2). Each edge-length corresponds to the metric distance $d_{i,j}$.

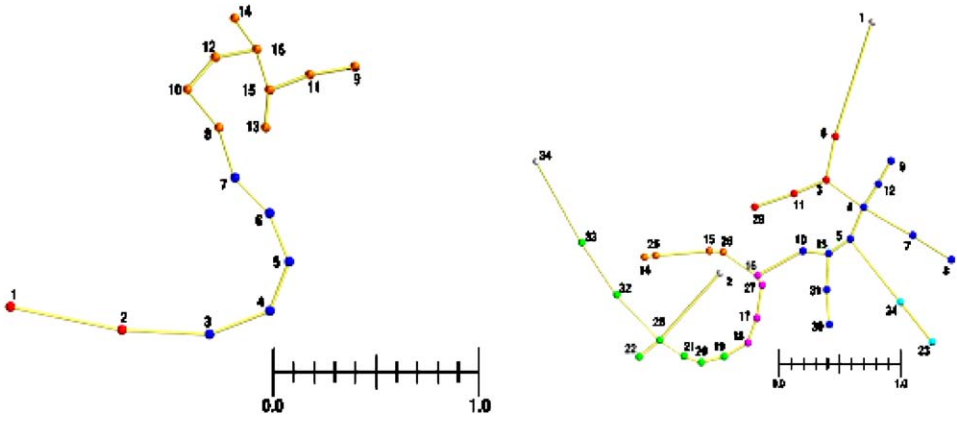

Fig. 5. Two-dimensional representations of the minimum spanning tree (MST) with edge-lengths equal to $d_{i,j}$. (Left) MST for Set 1. (Right) MST for Set 2.

in Fig. 5. In these MST representations the systemhas also been relaxed to approach the real distances using the same procedure as for the 3D case. The planar graphs of Figs. 3, 4 are a natural further step from the construction of the MST. In Ref. [21], we prove that such graphs preserve the hierarchical organization of the MST and allow us to compute new local and global topological quantities. The embedding on

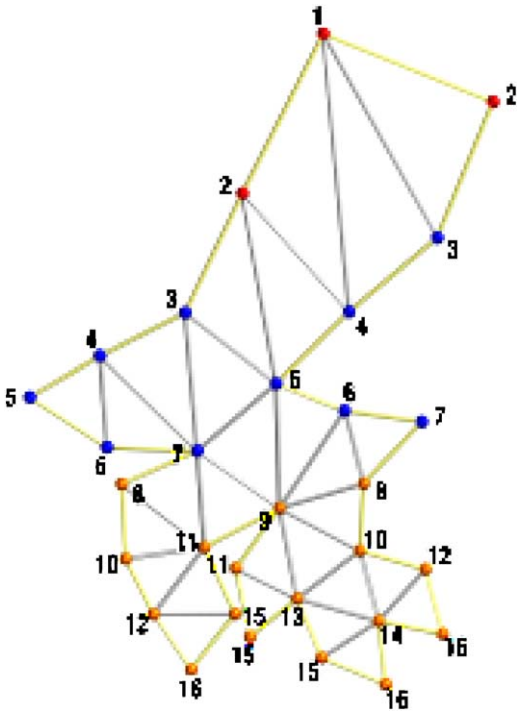

Fig. 6. Two-dimensional Pelting representation of the graph in Fig. 3 which opens it into a topological disk on the plane.

S_0 gives us a clear clustering differentiation, we can see from Figs. 3 and 4 that the clusters described in Section 3 naturally emerge from this construction. Once we have the embedding on S_0 we can project the 3D graph on the plane for a convenient alternative visualization. This has been done by using the Pelting Surface Operator [22] (inside the 3D Houdini package). In our case the Pelting is constructed by cutting the surface along the MST. The result is a topological disk, with every edge of the MST opening out into edges of the boundary. Every edge is treated as a spring, with every point on the boundary connected to a surrounding circular frame. The spring network is relaxed approaching as closely as possible to metric distances to create a disc-like mesh with no overlaps. Some deviations from actual metric distances are inevitable. The results are shown in Figs. 6 and 7.

5. Conclusion

We have investigated the hierarchical structure of two sets of interest rates by reducing their correlation matrices to a sub-set of relevant interactions which can be

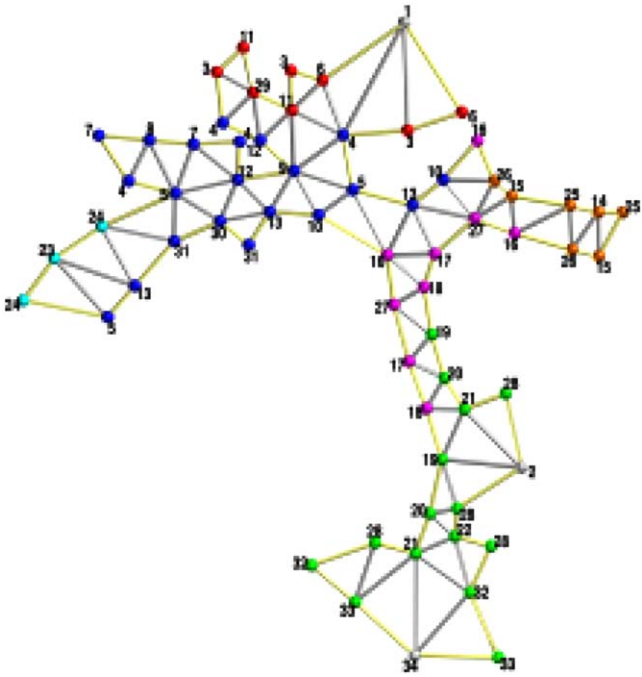

Fig. 7. Two-dimensional Pelting representation of the graph in Fig. 4 which opens it into a topological disk on the plane.

mapped into a metric graph topologically embedable on a sphere. We show that such a procedure yields structures which are naturally organized accordingly to the same clustering structure which can be extracted from a linkage procedure over an ultrametric distance. These graphs contain the minimum spanning tree and they can be considered as a natural extension of the MST approach; our graphs share the same hierarchical structure but contain more information. We note that the simplest embedding (genus 0) reveals relations not observable with the MST technique. This embedding procedure can be extended to surfaces of higher genus constructing in this way networks with different degree of complexity and tunable information content. This will introduce new investigation tools (e.g. genus versus information content) and pose new challenges for their visualizations.

Acknowledgements

We thank Rosario Mantegna and Michele Tumminello for many helpful and fruitful discussions. T. Di Matteo also benefited from discussion with participants at the COST P10 'Physics of Risk' meeting in Nyborg (DK), April 2004. We acknowledge partial support from ARC Discovery Project DP0344004 (2003),

Australian partnership for advanced computing national facilities (APAC) and M.I.U.R.-F.I.S.R. Project 'Ultra-high frequency dynamics of financial markets'.

Appendix A. Data description

Hereafter we list the different interest rates time series analyzed and their main characteristics.

- *The Federal funds rate* (*FED*) is the cost of borrowing immediately available funds, primarily for one day. The effective rate is a weighted average of the reported rates at which different amounts of the day's trading through New York brokers occurs. The weekly data are unweighted averages of 7 calendar days ending on Wednesday of the current week.
- *The State & local bonds* (*SLB*) consists of 20-year tax-exempt bonds, primary market general obligation, 20 Bonds in index mixed quality. We report weekly data ending on Thursday.
- *The Commercial Paper* (*CP*) and the *Finance Paper placed directly* (*FP*) [23] are unweighted averages of offering rates, reported each business day to the FR Bank of New York, on commercial paper placed by several leading dealers for firms whose bond rating is AA or the equivalent and on paper directly placed by finance companies. The symbols CP1, CP3, CP6 stand for maturity dates of 1, 3 and 6 months.
- *The Bankers acceptances* (*BA*) rates are representative of the closing yields for each business day as obtained from dealers by the FR Bank of New York. They are short-term negotiable time drafts or bill of exchange drawn on and accepted by a bank on behalf of its customers. The BA3 rates refer to a maturity date equal to 3 months and the BA6 to a maturity date equal to 6 months. These last are trading rates for the best rated money center banks.
- *The rate on certificates of deposit* (*CD*) is a simple average of dealer rates on negotiable certificates of deposit nationally traded in the secondary market. These rates CD1 (maturity date = 1 month), CD3 (maturity date = 3 months) and CD6 (maturity date = 6 months) are determined for each business day.
- *The yields on Treasury securities at* '*constant maturity*' (*TC*) are interpolated by the US Treasury from the daily yield curve. This curve, which relates the yield on a security to its time to maturity, is based on the closing market bid yields on actively traded Treasury securities in the over-the-counter market. These market yields are calculated from composites of quotations obtained by the FD Bank of New York. The constant maturity yield values are read from the yield curve at fixed maturities, currently 3 and 6 months (TC3M, TC6M) and 1, 2, 3, 5, 7, 10, and 30 years (TC1Y-TC30Y).
- *The Treasury bill rates* (*TBA*) are weekly averages computed on an issue-date basis [24]. *The Treasury bill secondary market rates* (*TBS*) are the averages of the bid rates quoted on a bank discount basis by a sample of primary dealers who report to the FR Bank of New York. The rates reported are based on quotes at the

official close of the US Government securities market for each business day. They have maturities of 3 and 6 months (TBA3M, TBA6M, TBS3M, TBS6M) and 1 year (TBS1Y) [25].

- *The Treasury long-term bond yield* (*TC10P*) are the unweighted average of yields on all issues of bonds outstanding which are neither due nor callable in less than 10 years. It represents yield on US Treasury bonds with maturity over 10 years.
- *The Eurodollar interbank interest rates* (*ED*) are bid rates with maturity dates 1, 3 and 6 months (ED1M, ED3M, ED6M), respectively.
- *The Corporate bonds Moody's seasoned rates* (*AAA*, *BAA*) are average yield to maturity on selected long-term bonds.
- *The Conventional mortgages rates* (*CM*) are contract interest rates on commitments for fixed-rate first mortgages.

Unless differently stated, we report weekly data ending on Friday, obtained from unweighted averages of daily data.

Appendix B. Network relaxation procedure

The numerical code we use to relax the generated networks runs as follows. The initial network geometry consists of a set of vertices placed at random in Cartesian space (x_i, y_i, z_i) [20,26]. That initial structure is then 'relaxed' by motion under the influence of a vector force on each n-connected vertex. Those forces are calculated by the gradient of the (elastic) energy function. We adopt the following form for the energy:

$$E = E_{angle} + E_{length} \tag{B.1}$$

with:

$$E_{angle} = k_b \sum_{i,j,k=1}^{n(n-1)/2} (\pi - \theta_{ijk})^2 \tag{B.2}$$

and

$$E_{length} = k_s \sum_{i,j=1}^{n} (\delta_{ij} - d_{i,j})^2 \,, \tag{B.3}$$

where k_b, k_s denote the elastic moduli for equalizing angles and edges respectively and $d_{i,j}$ denotes the rest spring length. The indices i, j, k label the vertices. θ_{ijk} denotes the angle (centered on vertex i) subtended by the three (edge-linked) vertices i, j, k of magnitude:

$$\theta_{ijk} = \arccos\left(\frac{\delta_{ij}^2 + \delta_{ik}^2 - \delta_{jk}^2}{2\delta_{ij}\delta_{ik}}\right), \tag{B.4}$$

where δ_{ij} denotes the distance of the vector joining vertices i and j:

$$\delta_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2} \,. \tag{B.5}$$

The force acting on each n-connected vertex is the gradient of E respect to x_i, y_i, z_i:

$$F_{x_i} = -\frac{\mathrm{d}E}{\mathrm{d}x_i}; \quad F_{y_i} = -\frac{\mathrm{d}E}{\mathrm{d}y_i}; \quad F_{z_i} = -\frac{\mathrm{d}E}{\mathrm{d}z_i} \,. \tag{B.6}$$

In order to minimize the energy, the position of the vertices changes by an amount proportional to these forces:

$$\mathrm{d}x_i \propto F_{x_i}; \quad \mathrm{d}y_i \propto F_{y_i}; \quad \mathrm{d}z_i \propto F_{z_i} \,. \tag{B.7}$$

In practice, the magnitudes of the elastic moduli are tuned to ensure convergence to a final configuration with all edges of length equal to $d_{i,j}$ and angles as nearly equal as possible.

References

[1] R.N. Mantegna, Eur. Phys. J. B 25 (1999) 193–197.
[2] G. Bonanno, N. Vandewalle, R.N. Mantegna, Phys. Rev. E 62 (2000) R7615–R7618.
[3] G. Bonanno, F. Lillo, R.N. Mantegna, Quantitative Finance 1 (2001) 96–104.
[4] G. Bonanno, G. Caldarelli, F. Lillo, R.N. Mantegna, Phys. Rev. E 68 (2003) 046130.
[5] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, Eur. Phys. J. B 30 (2002) 285.
[6] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, A. Kanto, Phys. Rev. E 68 (2003) 056110.
[7] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, Physica A 324 (2003) 247.
[8] A.R. Pagan, A.D. Hall, V. Martin, Modeling the Term Structure, in: HandBook of Statistics, vol. 14, Elsevier Science, New York, 1997.
[9] R. Rebonato, Interest-Rate Option Models, Wiley, New York, 1998.
[10] J.P. Bouchaud, N. Sagna, R. Cont, N. EL-Karoui, M. Potters, cond-mat/9712164 (1997), Appl. Math. Finance 6 (1999) 209–232.
[11] T. Di Matteo, T. Aste, J. Theoret. Appl. Finance 5 (2002) 122–127 cond-mat/0101009.
[12] T. Di Matteo, T. Aste, R.N. Mantegna, Physica A 339 (2004) 181 cond-mat/0401443.
[13] T. Di Matteo, M. Airoldi, E. Scalas, Physica A 339 (2004) 189 cond-mat/0401445.
[14] T. Alderweireld, J. Nuyts, Physica A 331 (2004) 602–616.
[15] http://www.federalreserve.gov/releases/h15/data.htm
[16] J.C. Gower, Biometrika 53 (1966) 325–338.
[17] T. Aste, T. Di Matteo, S.T. Hyde, Physica A 346 (2005) 20–26 cond-mat/0408443.
[18] See additional materials at: http://wwwrsphysse.anu.edu.au/~tas110/ econophysics.html
[19] M. O'Keeffe, Z. Kristallogr. 196 (1991) 21.
[20] S.T. Hyde, S. Ramsden, T. Di Matteo, J.J. Longdell, Solid State Sci. 5 (2003) 35–45.
[21] M. Tumminello, T. Aste, T. Di Matteo, R.N. Mantegna, A new tool for filtering information in complex systems, submitted 2005.
[22] D. Piponi, G. Borshukov, Seamless texture mapping of subdivision surfaces by model pelting and texture blending, Siggraph Proceedings, 2000, pp. 471–476.
[23] These series end on 29 August 1997. On 2 September 1997, the FR Board has began to calculate its statistics on Commercial paper from data transmitted electronically to the FR Board from the Depository Trust Company (DTC) of New York City. The preceding Friday, 29 August 1997, will mark the as-of date for the last releases of data on CP market activity collected at the FR Bank of New York through mail and telephone surveys of market participants.

[24] On and after 28 October 1998, data are stop yields from multiple-price auctions (prior to 28 October average yield was used).
[25] The FR uses data on CP outstanding to track sources and uses of external corporate funding, measure the aggregate flow of funds, and determine the composition of short-term financing in credit markets. With information on CP interest rates, the FR monitors short-term financial markets and gauges the current cost of funds to businesses. Relationships between CP interest rates and those on Treasury bills shed light on investor perceptions about the risk in short-term business lending.
[26] S.T. Hyde, S. Ramsden, Eur. Phys. J. B 3178 (2003) 273–284.