Interest rates hierarchical structure

T. Di Matteo\textsuperscript{a,}\textsuperscript{*}, T. Aste\textsuperscript{a}, S. T. Hyde\textsuperscript{a} and S. Ramsden\textsuperscript{a}

\textsuperscript{a}Applied Mathematics, Research School of Physical Sciences, Australian National University, 0200 Canberra, Australia.

Abstract

We propose a general method to study the hierarchical organization of financial data. The statistical, geometrical and topological properties of such data are analyzed by embedding the structure of their correlations in metric graphs in multi-dimensional spaces. We show an application to two different sets of interest rates data. In this case we construct triangular embeddings on the sphere. The resulting graph contains the minimum spanning tree as sub-graph and it preserves its hierarchical structure. This results in a clear cluster differentiation and allows to compute new local and global topological quantities. A three dimensional representation of this embedding is constructed together with its projection on a plane by using the Pelting method and a relaxation procedure to converge on the correct metric geometry.

Key words: Interest Rates, Data Clustering, Correlations, Econophysics.

PACS: 89.65.Gh, 89.75.-k, 89.75.Hc, 89.90.+n.

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 financial market data. A number of physicists have observed that the structure of the correlation coefficients associated with complex datasets (such as times series from financial markets) can be conveniently studied by mapping the data-structures onto graphs \[1–7\]. Mantegna \[1\] has studied the hierarchical organization of complex datasets by reduction to the minimal spanning tree (MST) which is a powerful method to investigate correlations in financial systems (see also \[3,5–7\]). However, this reduction to a

\* Corresponding author: Tel: +61 2 61250166, fax: +61 2 61250732.  
Email address: tiziana.dimatteo@anu.edu.au (T. Di Matteo).
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 network 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 in multidimensional spaces, both Euclidean and non-Euclidean.

Why do we choose interest rates data?

In the Econophysics literature the study on interest rates has been less investigated and less predominant in comparison with the investigations performed on stock market prices. Only recently these studies are becoming very attractive and approached from many different perspectives [8–14].

For several economic reasons, 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 motion of a whole complex set of 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. The output of this clustering linkage procedure, gives us the simplest picture of the interest rates hierarchical organization. In this paper we present an original approach that associates the empirical analysis with metric graphs embedded on hyperbolic surfaces in multi-dimensional spaces. These graphs where the genus is the varying parameter controlling the degree of complexity of the system provide natural hierarchies and include the MST in their structure. Multi-dimensional spaces (both Euclidean and non-Euclidean) embedding gives us a locally-planar representation and allows us to introduce novel topological measures to characterize these metric graphs. Here we present an application to interest rates of this novel procedure that allows us to generate graphs. Section 2.1 describes the interest rates data set and Section 2.2 gives a resume of the main outcomes obtained from a correlation cluster analysis. Our idea and results are reported in Section 2.3 where a 3D visualization of these metric graphs is also shown.
Fig. 1. a) Eurodollar interest rates ($f_\theta(t)$) as function of $t$ for $\theta$ ranging from 3 to 48 months; b) Eurodollar interest rates ($f_\theta(t)$) as function of $t$ for $\theta = 3, 15, 30, 48$ months.

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

2 Hierarchical structure

2.1 Data description

We investigate two data sets: 16 Eurodollars interest rates (Set 1) and 34 different kinds of interest rates in money and capital markets, referring to government, private, industries securities and commitments (Set 2). Set 1 contains daily values $f_\theta(t)$, where $t$ is the current date and $\theta$ 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.
Table 1
Data Set 1: Eurodollar interest rates in the time period 1990-1996.

<table>
<thead>
<tr>
<th>i</th>
<th>θ months</th>
<th>i</th>
<th>θ months</th>
<th>i</th>
<th>θ months</th>
<th>i</th>
<th>θ months</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>5</td>
<td></td>
<td>15</td>
<td></td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>6</td>
<td></td>
<td>6</td>
<td></td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>9</td>
<td></td>
<td>7</td>
<td></td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>12</td>
<td></td>
<td>8</td>
<td></td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>i</th>
<th>f_i</th>
<th>i</th>
<th>f_i</th>
<th>i</th>
<th>f_i</th>
<th>i</th>
<th>f_i</th>
<th>i</th>
<th>f_i</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FED</td>
<td>7</td>
<td>FP3</td>
<td>13</td>
<td>CD6</td>
<td>19</td>
<td>TC5Y</td>
<td>25</td>
<td>TBS3M</td>
</tr>
<tr>
<td>2</td>
<td>SLB</td>
<td>8</td>
<td>FP6</td>
<td>14</td>
<td>TC3M</td>
<td>20</td>
<td>TC7Y</td>
<td>26</td>
<td>TBS6M</td>
</tr>
<tr>
<td>3</td>
<td>CP1</td>
<td>9</td>
<td>BA3</td>
<td>15</td>
<td>TC6M</td>
<td>21</td>
<td>TC10Y</td>
<td>27</td>
<td>TBS1Y</td>
</tr>
<tr>
<td>4</td>
<td>CP3</td>
<td>10</td>
<td>BA6</td>
<td>16</td>
<td>TC1Y</td>
<td>22</td>
<td>TC30Y</td>
<td>28</td>
<td>TC10P</td>
</tr>
<tr>
<td>5</td>
<td>CP6</td>
<td>11</td>
<td>CD1</td>
<td>17</td>
<td>TC2Y</td>
<td>23</td>
<td>TBA3M</td>
<td>29</td>
<td>ED1M</td>
</tr>
<tr>
<td>6</td>
<td>FP1</td>
<td>12</td>
<td>CD3</td>
<td>18</td>
<td>TC3Y</td>
<td>24</td>
<td>TBA6M</td>
<td>30</td>
<td>ED3M</td>
</tr>
</tbody>
</table>

with a step of three months (reported in Tab. 1). The behavior of $f_\theta(t)$ as function of $t$ is shown in Fig.1 where we use $i(=\theta)$ to label the different maturity dates $\theta$. 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 maturity values, namely $\theta = 3$ and $\theta = 48$ months. Set 2 contains 34 different weekly interest rate time series during a time period of 16 years between 1982 and 1997 recorded in the Federal Reserve (FR) Statistical Release database [12,15]. In appendix A are reported their main characteristics. 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 Tab.2). The interest rate time series, $f_i(t)$ v.s. $t$ are 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 lay in a narrow band around $\bar{f}(t)$.

### 2.2 Correlation cluster formation

We study the hierarchical structure arising from the correlations between the interest rate 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 have computed the metric distance $d_{i,j}$ between the series $\Delta f_i$ and $\Delta f_j$ 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 correlations among the \( i, j \) interest rates fluctuations:

\[ c_{i,j} = \frac{\langle \Delta f_i \Delta f_j \rangle - \langle \Delta f_i \rangle \langle \Delta f_j \rangle}{\sigma_i \sigma_j} \]  

(1)

where the symbol \( \langle ... \rangle \) denotes a time average performed over the investigated time period and \( \sigma_i \) is the standard deviation defined as:

\[ \sigma_i = \sqrt{\frac{1}{T_2 - T_1} \sum_{t=T_1}^{T_2} (\Delta f_i(t) - < \Delta f >)^2} \]  

(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 \( \times \) 16 (for Set 1) and 34 \( \times \) 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 to 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: \( \hat{d}_{i,j} \leq \max \{ \hat{d}_{i,k}, \hat{d}_{k,j} \} \), called ‘ultra-metric inequality’. Once the metric distance \( d_{i,j} \) is defined, one can introduce several ultra-metric distances. Mantegna et al. have used the ‘subdominant ultra-metric’, obtained by calculating the minimum spanning tree connecting several financial time series [2–4]. Being the correlations between interest rates 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: \( Cls_1 = \{3, 6\} \), \( Cls_2 = \{9, 12, 15, 18, 21\} \), \( Cls_3 = \{24, 27, ..., 45, 48\} \). The first cluster \( (Cls_1) \) gathers together \( \Delta f_\theta(t) \) with maturity shorter than 1 year; \( Cls_2 \) contains those with maturity dates between 1 year 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 several clusters organized in the following hierarchical structure:

- all the interest rates with maturities equal to 1 months (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).
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 $d_{i,j}$.

- 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.

Interest rates data are highly correlated and their analysis by studying the MST is not enough to extract a meaningful hierarchical structure. Therefore in the next section we are proposing a different approach to deal with this peculiar type of data.

2.3 Results from a 2D embedding and discussion

Here the key 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 $n$ interest rates we can associate a point in a multi-dimensional space with each of the $n$ interest rates. To all pairs $(i, j)$ a metric distances $d_{ij}$ is associated and the resulting network is an $n$-th 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 high-dimensional space and short distances are associated with highly correlated rates values.
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 $d_{i,j}$.

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. In [17] it has been proposed to map $K_n$ on a 2D hyperbolic surface. Such reduction can be obtained also with an opposite procedure: it starts from the set of $n$ unconnected nodes and for a given genus $g$ it connects iteratively two nodes if and only if the resulting graph can be embedded on a 2D hyperbolic surface $S_g$. This process will end with a triangular embedding. Here we present an application of this genus-dependent procedure to the two sets of interest rates data (Set 1 and Set 2) for the genus $g = 0$ case (the sphere). In Figs. 3, 4 their 3D representations, respectively for Set 1 and Set 2, are shown. In both figures we can observe that the resulting graph on $S_0$ is a triangulation and we can visualize the hierarchical organization of the whole system. Each node represents an interest rate and the length of each edge is the metric distance $d_{i,j}$ introduced in the previous section. Different colours (on line version) have been chosen to distinguish different clusters. Note also that our embedding gives no edges crossing. We have relaxed the resulting network numerically ([18,19]) seeking to make all vertex angles as equal as possible, consistent 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 in Fig. 5. For these MST representations the system has also been relaxed to the real distances using the same procedure as for the 3D case. The previous results are a natural further step from the construction of the MST which preserve its hierarchical organization and allow us to compute new local and global topological quantities [20].
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.

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

Embedding on $S_0$ gives us a clear clustering differentiation, we can see from Figs. 3 and 4 the natural formation of the clusters described in Section 2.2. Once we have the embedding on $S_0$ we can project the $3D$ graph on the plane. This has been done by using the Pelting Surface Operator [21] (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 to create a $2D$ mesh with no overlaps. Distances, however are not respected. The results are shown in Figs. 6 and 7.
Fig. 7. Two dimensional Pelting representation of the graph in Fig. 4 which opens it into a topological disk on the plane.

A further development of the present analysis will consist in the investigation of embeddings in surfaces with higher genus. This will introduce new investigation tools (e.g. genus versus information content) and pose new challenges for their visualizations.

Acknowledgements

Many thanks to Rosario Mantegna and Michele Tumminello for many helpful and fruitful discussions. T. Di Matteo benefited from discussions with the participants to the COST P10 ‘Physics of Risk’ meeting in Nyborg (DK), April 2004. We acknowledge partially financial support from ARC Discovery project DP0344004 (2003).

A Data description

Hereafter we list the different interest rate 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) [22] 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 U.S. 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 [23]. 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 U. S. Government securities market for each business day. They have maturities of 3 and 6 months (TBA3M, TBA6M, TBS3M, TBS6M) and 1 year (TBS1Y) [24].

• 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 month, 3 months 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 obtained from unweighted averages of daily data ending on Friday.

B Networks relaxation

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 \((x_i, y_i, z_i)\) [19,25]. 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:

\[
E = E_{\text{angle}} + E_{\text{length}}
\]  

(B.1)

with:

\[
E_{\text{angle}} = k_b \sum_{i,j,k=1}^{n(n-1)} (\pi - \theta_{ijk})^2
\]  

(B.2)

and

\[
E_{\text{length}} = k_s \sum_{i,j=1}^{n} (\delta_{ij} - d_{i,j})^2
\]  

(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. \(\theta_{ijk}\) denotes the angle (centered on vertex \(i\)) subtended by the three (edge-linked) vertices \(i, j, k\); of magnitude:

\[
\theta_{ijk} = \arccos\left(\frac{\delta_{ij}^2 + \delta_{ik}^2 - \delta_{jk}^2}{2\delta_{ij}\delta_{ik}}\right)
\]  

(B.4)

where \(\delta_{ij}\) denotes the distance of the vector joining vertices \(i\) and \(j\):

\[
\delta_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2}
\]  

(B.5)
The force acting on each $n$-connected vertex is the gradient of $E$ respect to $x_i, y_i, z_i$:

$$F_{x_i} = -\frac{dE}{dx_i}; \quad F_{y_i} = -\frac{dE}{dy_i}; \quad F_{z_i} = -\frac{dE}{dz_i}. \quad (B.6)$$

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

$$dx_i \propto F_{x_i}; \quad dy_i \propto F_{y_i}; \quad dz_i \propto F_{z_i}.$$ 

(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


[22] These series end on August 29, 1997. On September 2, 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, August 29, 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.

[23] On and after October 28, 1998, data are stop yields from multiple-price auctions (prior to October 28 average yield was used).

[24] 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.