Turkish Economic Review
www.kspjournals.org
Volume 2 March 2015 Issue 1

Estimating Financial Trends by Cubic B-Spline Fitting via Fisher Algorithm

By Mehmet BARAN [a] Sıtkı SÖNMEZER [b †]
& Abdülvahid UÇAR [c]

Abstract. Trends have a crucial role in finance such as setting investment strategies and technical analysis. Determining trend changes in an optimal way is the main aim of this study. The model of this study improves the optimality by cubic b-spline fitting to the equations to reduce the error terms. The results show that cubic b-spline fitting is more efficient compared to the first order Fisher Method and original Fisher Method. This method may be used to determine regime switches as well.
Keywords. Technical Analysis, Trends, Regime Switches, Investment Strategies.
JEL.

1. Introduction

It is often the case that when trying to extract information out of a data stream; researchers face the problem of dividing the financial data into homogeneous parts. When they work with time series, or when they aim to determine trends within a particular data set, it may be essential for them to know how to decompose a sample into sub- samples in which homogeneity is maximized. Intuition or segmenting the data visually may be used for grouping the data, however; these methods lack the efficiency when compared with the statistical methods that are available in the literature.

Various disciplines have the problem of clustering a set of objects as there are different usages, applications and objective functions. Thus, there are many variations of this clustering problem. Maharaj and Inder (1999) and Duncan, Gorr and Szczypula (2001) have given methods of clustering time series data. Eventually, numerous clustering algorithms from different fields of study are present in the literature, however; the merits of these algorithms are dubious (Gonzales 1985).

The method proposed in this study is a continuation of the line of research first established by Fisher (1958), who proposed a novel method for segmenting time series data. This method is then futher developed by Baran and Sonmezer (2013). It has to be noted that Fisher has divided the grouping problem into two parts; first one is introduced as the “unrestricted problem” and second one is as the “restricted problem”. This paper deals with the restricted problem, which arises in many cases in finance where certain conditions, on the basis of prior information, theory or for convenience, are imposed most of the time and offers an improved version that gives smaller error terms. The first part of the study provides literature survey

[a †] Marmara University, Faculty of Engineering, Computer Engineering, İstanbul, Turkey. ☏. +90216-348-0909/251. ✉. mehmet.baran@marmara.edu.tr
[b] Beykent University, FEAS, Department of Business, İstanbul, Turkey. ☏. +90212-867-51-40/5231. ✉. sitkisonmezer@beykent.edu.tr

regarding the method used in finance related sectors. In the second part, Fisher method and the proposed improvements is presented. The final part presents an example showing the improvement.

2. Literature Review

There is a huge amount of publication devoted to the problem of dividing a time series into homogenous segments, mostly by statisticians and economists. In this paper, only the articles that have utilized methods that determine homogeneous grouping with financial data are reviewed. We also limited ourselves to review only the puplications which have some relation to Fisher's method.

Boginski, Butenko & Partelos (2005) have examined market graph with cliques and independent sets by using correlations; cliques are sets of interconnected vertices and independent sets are sets of vertices without connections. Similar object clusters are assumed to be reflected by cliques whereas, objects that are not similar are represented by independent sets. They claim that a suitable similarity criterion shall be determined before dividing a data set into parts and there is hardship in the fact that the number of clusters are not known a priori.

Huang & Zhang (2012) have investigated structural changes in Singapore's private housing market and in particular the impact of government policies on housing price determination by utilizing Fisher method to find "1) the specification of the number of changes in the model 2) the detection of the change point, or the boundaries of intervals over which each of the model pieces applies 3) the estimation of the model parameters within each subdomain."

Kumar & Patel (2010) have come up with a new method derived from Fisher method which is based on the tradeoff between decreased variance and increased bias arising from combining. With this modified version, it is possible to determine when it is beneficial to combine or not. Adams and Lim (2011) also utilized the Fisher method in forming sub groups for their data to minimize sum of squared deviations for growth polarity.

Baran & Sonmezer (2013) used linear regression (instead of constant levels as proposed by Fisher) in the original Fisher Algorithm and their method of clustering financial data reduced the errors substantially compared to Fisher (1958). A review of this algorithm is presented here below. The method presented below is a continuation of the line of research initiated by this article.

3. Grouping (Segmenttation) Algorithms

A grouping algorithm divides a time series consisting of N data points $x(i)$, $1<=i<=N$, into K contiguous and mutually exclusive segments:

$$[1,N_1]\cup[N_1+1,N_2]\cup\ldots\cup[N_{K-2}+1,N_{K-1}]\cup[N_{K-1}+1,N] \qquad (1)$$

where $[a,b]$ denotes a segment which begins at a and ends at b. The division is done in such a way that the "homogeneity" within the segments is minimized, while the differences between the segments are maximized. When such a grouping is found, the segments are said to represent the "trends" in the data and the segment boundaries represent the transition regions between the trends.

Grouping algorithms can be classified by their criterion of "homogeneity". Below, we will define three grouping algorithms: Fisher (1958), Baran & Sönmezer (2013), and the grouping algorithm proposed in this article.

4. Fisher Grouping

Define the mean of the j^{th} segment by:

$$M_j = \frac{1}{N_j - N_{j-1}} \sum_{i=N_{j-1}}^{N_j} x(i) \quad (2)$$

where $N_0 = 1$and $N_k = N$.

Define the mean-square error of the *j* th segment by

$$e_j = \sum_{i=N_{j-1}}^{N_j} \left(x(i) - M_j\right)^2 \quad (3)$$

Define the mean-square error of the whole time series by

$$E = \sum_{i=0}^{K-1} e_i \quad (4)$$

In 1954, Fisher proposed an algorithm that computed the segment boundaries $N_1, N_2, ... , N_{K-1}$in Eq. (1) in such a way that the MS error E defined in Eq. (4) is minimized. The algorithm is based on dynamic programming, whose time complexity is $O(KN^2)$. For details of fisher algorithm, see Fisher (1954), Baran & Sönmezer (2013).

6. Problems of Fisher grouping in Segmenting Financial Data

Fisher grouping is perfect for time series where the data remains in near-constant trends for long durations, and the transition regions from one trend to the neighboring trend are relatively narrow. This is one reason why Fisher algorithm found particular favor with hydrologists: Most of the lakes and rivers have near-constant depth for long periods, with very rapid change of depth during short intervals of time (ie, spring rains) between them. Unfortunately, financial data does not fit into this pattern. "Trends" in financial data are rarely defined by approximately constant levels with quick changes between them. Therefore, Fisher algorithm in its classical form is largely unsuitable for discovering trends in financial data, and boundaries between the neighboring trends.

7. First Order Grouping

In order to alleviate the deficiencies of the Fisher grouping algorithm, the first order grouping algorithm is proposed by Baran & Sönmezer (2013). In this algorithm, everything is the same with the Fisher grouping algorithm, with the exception that the definition of homogeneity for a segment (ie, equations (2) and (3)) is replaced by

$$e_j = \sum_{i=N_{j-1}}^{N_j} \left(x(i) - M_j\right)^2 \quad (3)$$

For every segment j, least mean squares (LMS) algorithm is used to compute a_j and b_j. Then, equation (4) is used to calculate the segment error e_j. After the e_j's are computed for all possible segments, the rest of the algorithm is exactly the same with Fisher's original algorithm: Employing dynamic programming to find optimal trends and trend boundaries. Time complexity remains the same, even though each step is more expensive because of the LMS algorithm.

A possible improvement of this algorithm, not mentioned in Baran & Sönmezer (2013), is to use least median squares rather than least mean squares (Rousseeuw, 1984). Least median squares is more expensive than LMS, but it brings a degree of insulation against outliers.

8. Problems of First Order Grouping in Segmenting Financial Data

First order grouping algorithm is good for time series composed of trends with linear growth or decay, and regions of quick transition between them. This is a much better assumption to use when working in financial data. Indeed, compared with the Fischer algorithm, first order grouping algorithm finds a much improved segmentation with much lower error rates for the same number of segments.

But, trends are frequently neither constant, nor linear. Usually they do not have a recognizable shape. Hence, an algorithm with a more malleable homogeneity criterion is required.

9. Cubic B-Spline Grouping

In this article, a more developed version of First order grouping algorithm is proposed, in which the deficiencies mentioned in the above paragraphs are taken into account. Fisher's original algorithm is "zero order", as it approximates segments with constants, which are zero order curves. First order algorithm will approximate segments with first order curves, i.e. lines, which are least mean squares approximations of the data in the segment. Newly proposed algorithm, in contrast, will approximate segments with cubic b-splines. As cubic b-splines are completely elastic and malleable, they have more degrees of freedom fit to the underlying trends. As they are smooth, they will not overfit and they will reject all noise.

In cubic b-spline grouping algorithm, each segment $[N_i, N_j]$ is divided into three contiguous subsegments and a spline curve formed from the weighted sum of six cubic splines are defined on these subsegments (a maximum of six cubic splines on a line divided to three). The weights of these cubic splines are chosen in such a way to minimize the L2 distance e_j between the spline curve and the data, by the least mean squares algorithm. Once the errors e_j are computed, the trend boundaries are found by following exactly the same steps as described in Baran and Sönmezer (2013), using dynamic programming.

If the data contains large number of outliers, least median fit rather than least mean fit may be preferred.

10. Practical Considerations

For least-squares spline fitting, MATLAB's spap2 function is used, with l=3 (each segment is divided into three subsegments) and k=4 (cubic splines are used), as described in Mathworks documentation. With these parameters, spap2 internally divides each segment into three by calling aptknt function of MATLAB.

11. Results

The three algorithms (original fisher algorithm, first order fisher algorithm and cubic b-spline grouping algorithm) are compared by running them on seven different time series: Namely, they are; 10 year US treasury returns, Volatility Index, 2030 Turkish Eurobonds, Silver, USD, Euro and Crude Oil prices. Table 1 indicates that when data is segmented into nine parts, Spline fitting outperforms the remaining models by reducing the error term significantly for each variable. These results are valid for other segmentation counts as well. Segment count of nine is chosen arbitrarily.

Table 1: *Error Term reduction by Methods*

Variable	Segment count	Original Fisher	First Order Fisher	Spline Fit

9

10 yr US treasury(YUST10)	2.50e-02	2.37e-02	2.13e-02
VIX	5.07e-00	4.73e-00	4.31e-00
TURKEY CDS	1.25e-00	1.14e-00	9.57e-01
SILVER	6.27e-01	6.01e-01	5.17e-01
USD	2.06e-01	1.94e-01	1.70e-01
EURO	1.75e-01	1.66e-01	1.46e-01
CRUDE OIL	5.99e-01	5.53e-01	4.90e-01

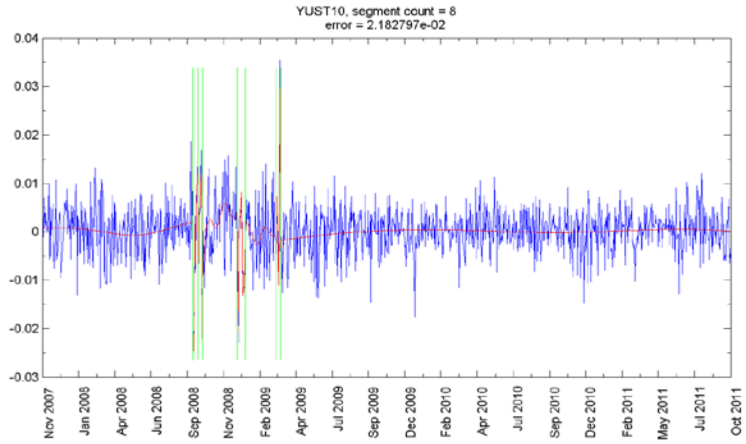

Figure 1: *Segmentation with cubic b-spline fitting for 10 yr US Treasury Bonds*

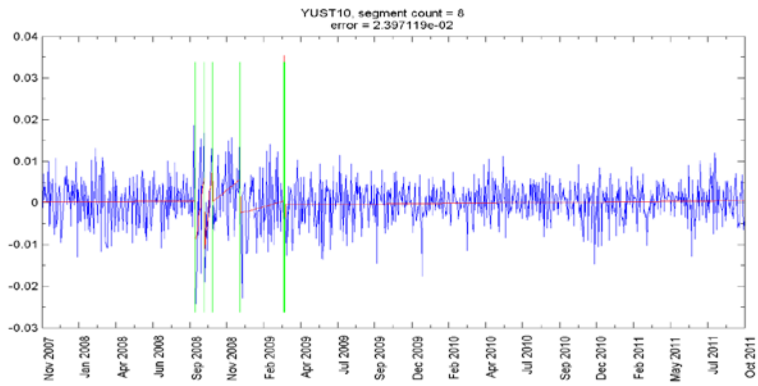

Figure 2: *Segmentation with first order Fisher Method for 10 yr US Treasury Bonds*

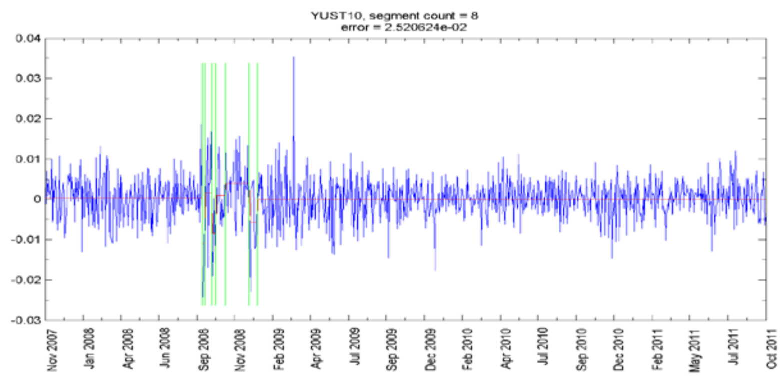

Figure 3: *Segmentation with original Fisher Method for 10 yr US Treasury Bonds*

Figure 1 depicts the results of the Fisher algorithm with cubic b-spline fit with 8 trends. The resulting error is 2.18e-02. Green vertical lines show the trend boundaries and red lines show the fitted curves. Figure 2 is the same but first order Fisher method is used. The resulting error is 2.397e-02. Figure 3 depicts the original Fisher method finding 8 trends in the same time series. It is clear that cubic b-spline fit reduces error terms significantly.

12. Conclusion

It is concluded that Fisher method can be improved via cubic b-spline fitting. Various time series are analyzed and in each of them cubic b-spline fitting outperformed original Fisher method and line fitting Fisher methods. Cubic b-spline fitting may be preferred by the parties interested in setting and finding trends with the least possible error terms. These interested parties may be technical analysts and momentum investors who may want to time the market by finding trends.

References

Adams-Kane, J. & Lim, J. J. (2011). Global Growth Poles in a Multipolar World Economy, Worldbank

Baran. M, & Sonmezer, S. (2013). How to Group Fınancıal Data wıth Maximum Homogeneıty**?,** *Emerging Markets Journal*, 3(1), 13-19.

Boginski, V., Butenko, S., & Pardalos, P. M. (2005). Statistical analysis of Financial networks, *Computational Statistics & Data Analysis*, 48, 431-443.

Duncan, G., Gorr, W. L., & Szczypula, J. (2001). Forecasting Analogous Time Series, In J. S. Armstrong (Ed.), Principles of Forecasting: A handbook for researchers and practitioners, Dordrecht: Kluwer, 195–213.

Fisher, W. D. (1958). On grouping for maximum homogeneity, *Journal of the American Statistical Association*, 53, 789-798.

Gonzales, T. F. (1985). Clustering to Minimize the Maximum Intercluster Distance, *Theoretical Computer Science*, 38, 293-306.

Huang, W., & Zhang, Y. (2012). Structural Change Modeling of Singapore Private Housing Price in Simultaneous Equation Model, *Journal of Financial Risk Management*, 1(2), 7-14.

Maharaj, E. A., & Inder, B. A. (1999). Forecasting time series from clusters, (Monash Econometrics and Business Statistics Working Papers). Monash University, Department of Econometrics and Business Statistics, 9/99.

Kumar, M, & Nitin R. P. (2010). Using clustering to improve sales forecasts in retail merchandising, *Annals of Operations Research*, 174(1), 33-46.

Rousseeuw, P. J. (1984). Least Median of Squares Regression, *Journal of the American Statistical Association*, 79, 871-880.