Topological Data Analysis of Treasury Yield Curve Rates

Christopher Shultz

February 1, 2021

Abstract

We explore the use of Topological Data Analysis (TDA) to study the evolution of US Treasury Yield Curve Rates from 2015 - 2020. We particularly focus on the use of TDA to understand the shape and structure of our multidimensional rates series and whether useful economic inference can be derived from topological tools.

1 Introduction

Topological Data Analysis (TDA) is an emerging multidisciplinary field that uses tools from topology, statistics, and scientific computing to extract insights about complex data. Many existing papers introduce the technique to an unfamiliar reader [8, 10, 21, 26]. From a practical perspective, TDA can be used primarily to explore the structure of multidimensional data with regard to shape and connectivity.

TDA has shown significant promise in a variety of fields, including cosmology [11, 12, 25, 24], image analysis [7, 19, 9, 23, 1], finance [16, 15, 17, 20, 22, 5, 13, 18], and neuroscience [2, 3, 4, 6], among many others. This paper does not attempt to present a comprehensive overview of all existing applications of TDA in the scientific literature. However, across all examined papers, the primary theme involves the extraction of insights about the shape and structure of complex multidimensional datasets.

With this paper, I aim to demonstrate the utility of TDA for exploring economic and financial questions, with a specific focus on treasury yield curve rates. A similar exploration takes place in [16] with attention to the equity markets, and we follow their example in our study of rates. Their findings are particularly interesting as they present one of few existing analyses utilizing TDA to explore economic and financial research. Gidea and Katz find a strong rising trend for trading days prior to financial downturns, which implies that TDA may be useful as a new econometric analysis to complement traditional methods.

This analysis is meaningful for two main reasons. First, it is (to our knowledge) the first implementation of TDA to examine the shape and structure of treasury yield curve rates. Second, it provides an introduction of TDA to the economics community, who may not have been previously exposed. Based on our survey of the literature, applications of TDA are common in finance, but not much was found in the economics space. Given its utilization across a wide number of fields and the increasing complexity of economic data, TDA has the potential to add significant value to researchers in economics and finance.

To complete our analysis, we utilize daily treasury yield curve rates data from 2015 - 2020 across multiple maturity periods. ***description of method here. description of results here. description of why this matters here***.

The remainder of this paper is structured as follows: section 2 describes the mathematical and theoretical details of TDA, including a detailed view of some existing applications; section 3 provides a summary of the existing literature on yield curve rates and the methodologies that have been used to date. Sections 4 and 5 explore our methodology and dataset, respectively, providing a detailed outline of the analysis we undertake and the nuances of the underlying dataset. Finally, sections 6 and 7 present our results, conclusions, and thoughts on future research.

2   Topological Data Analysis

TDA allows for the extraction of new information from high-dimensional, incomplete, and/or complex datasets. It shows promise as an emerging method for providing empirically sound methods for analysis and understanding of the underlying data that can be represented as point clouds in Euclidean space. TDA packages are implemented in a number of languages, including C++, Python, and R.

Chazel and Michel [10] provide a generalized TDA pipeline that most studies implementing TDA employ:

1. The input data is a finite set of points with some distance or similarity metric. In many cases, including this paper, the distance can be induced by the underlying metric space. Typically, data is represented as a series of points in a Euclidean n-dimensional space $\mathbb{E}^n$ [14].

2. Using the point cloud from [1], a "continuous" shape is built on top of the underlying dataset to highlight its shape. This typically takes place through the use of simplicial complexes.

3. Topological or geometric information is inferred from the structures built on the data.

4. The extracted information provides new families of features and descriptions of the data that can be used to better understand the data or be combined with other existing features for modeling.

The standard topological descriptors are the barcode and the persistence diagram, which provide a multiscale representation of the homology of the geometric construction. Homology in degree 0 describes the connectedness of the data; degree 1 detects holes or tunnels; degree 2 captures voids, etc. [8]. Features that *persist* when the resolution changes are referred to as persistent homology.

With our collection of data in $\mathbb{E}$, we use the point cloud as the vertices of a combinatorial graph whose edges are determined by proximity. While this approach allows for useful clustering, a number of additional higher order features are ignored under this structure. Additional features can be accurately derived by thinking of the graph as a scaffold for a higher-dimensional object. Specifically, one completes the graph to a simplicial complex - a space built from simple pieces (simplicies) identified combinatorially along faces. One way to fill the higher dimensional simplicies of the proximity graph is below:

Definition 1: Given a collection of points $x_\alpha \in \mathbb{E}^n$, the **Rips Complex**, $\mathcal{R}_\epsilon$ is the abstract simplicial complex whoes k-simplices correspond to unordered $(k+1)$-tuples of points $x_\alpha \binom{k}{0}$, which are pairwise within distance ϵ.

2.1   Simplicial Complexes

2.2   Barcodes

After conversion of our data set into a family of simplicial complexes, we view these topological objects via a theory of persistent homology that is encoded in the form of a Betti number: a **barcode** [14].

2.3   Persistence Diagrams

3   Treasury Yield Curve Rates

4   Methodology

5   Data

The dataset analyzed in this study was retrieved from the US Treasury [1] and contains daily yield curve rates data from January 2, 2015 to December 31, 2020 for a number of different maturities: 1

---

[1] https://www.treasury.gov/resource-center/data-chart-center/interest-rates

mo, 3 mo, 6 mo, 1 year, 2 yr, 3 yr, 5 yr, 7 yr, 10 yr, 20 yr, and 30 yr. The time series are visualized below in Figure 1 , with our shortest and longest maturities highlighted.
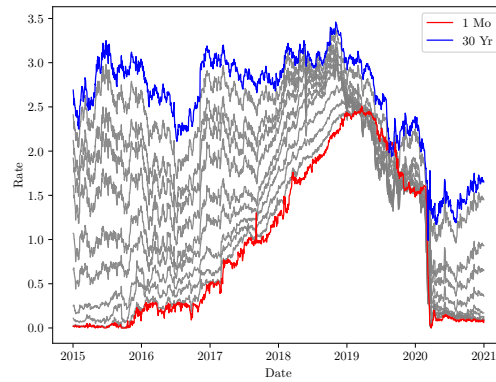

Figure 1: Treasury Yield Curve Rates 2015-2020

6   Results

7   Conclusion

References

[1] Rubin D. Adcock, A. and G. Carlsson. Classification of hepatic lesions using the matching metric. 121:35–42, 2014.

[2] Brandt V. Arai, M. and Y. Dabaghian. The effects of theta precession on spatial learning and simplicial complex dynamics in a topological model of the hippocampal spatial map. 10(6), 2014.

[3] A. Babichev and Y. Dabaghian. Persistent memories in transient networks, 2016.

[4] Arai M. Basso, E. and Y. Dabaghian. Gamma synchronization of the hippocampal spatial map - topological model, 2016.

[5] D. Basu and P. Dlotko. The four seasons of commodity futures: insights from topological data analysis, 2019.

[6] Marron J. Miller E. Pieloch A. Bendich, P. and S. Skwerer. Persistent homology analysis of brain artery trees. 10:198–218, 2016.

[7] Ovsjanikov M. Oudot S. Bonis, T. and F. Chazal. Persistence-based pooling for shape pose recognition. New York: Springer:19–29, 2016.

[8] P. Bubenik. Statistical topological data analysis using persistence landscapes. 16:77–102, 2015.

[9] Oudot S. Carriere, M. and M. Ovsjanikov. Stable topological signatures for points on 3d shapes. 34:1:12, 2015.

[10] F. Chazal and B. Michel. An introduction to topological data analysis: fundamental and practical aspects for data scientists, 2017.

[11] Y. Chen, S. Ho, P. Freeman, C. Genovese, and L. Wasserman. Cosmic web reconstruction through density ridges: method and algorithms. 454:1140–1156, 2015.

[12] R. Van de Weygaert, E. Platen, G Vegter, B. Eldering, and N. Kruithof. Alpha shape topology of the cosmic web. New York: IEEE:224–234, 2010.

[13] Qiu W. Dlotko, P. and S. Rudkin. Financial ratios and stock returns reappraised through a topological data analysis lens, 2019.

[14] R. Ghrist. Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society*, 45:61–75, 2007.

[15] M. Gidea. Topological data analysis of critical transitions in financial networks. Springer Proceedings in Complexity, 2017.

[16] M. Gidea and Y. Katz. Topological data analysis of financial time series: Landscapes of crashes. 491:820–834, 2018.

[17] Xia S. An Q. Zhang X. Sun W. Guo, H. and X. Zhao. Empirical study of financial crises based on topological data analysis. 558, 2020.

[18] Kim Y. Lee G. Kim, W. and W. Kook. Investigation of flash crash via topological data analysis. 2020.

[19] Ovsjanikov M. Li, C. and F. Chazal. Persistence-based structural recognition. New York: IEEE:2003–2010, 2014.

[20] S. Majumdar and A. Laha. Clustering and classification of time series using topological data analysis with applications to finance. 162, 2020.

[21] E. Munch. A user's guide to topological data analysis. 4(2):47–61, 2017.

[22] Rudkin S. Qiu, W. and P. Dlotko. Refining understanding of corporate failure through a topological data analysis mapping of altman's z-score model. 156, 2020.

[23] Couture H. Marron J. Perou C. Singh, N. and M. Niethammer. Topological descriptors of histology images. New York: Springer:231–239, 2014.

[24] T. Sousbie. The persistent cosmic web and its filamentary structure-i: Theory and implementation. 414:350–383, 2011.

[25] Vegter G. Edelsbrunner H. Jones B. Pranav P. Gavrilova M. Tan C. Van de Weygaert, R. and M. Mostafavi. Alpha, betti, and the megasparsec universe: on the topology of the cosmic web. Heidelberg, GER.: Springer-Verlag:60–101, 2011.

[26] L. Wasserman. Topological data analysis. 5:501–532, 2018.