Applications of Topological Data Analysis in Economics

Christopher Shultz

March 4, 2023

Abstract

Topological data analysis (TDA) is an emerging method that has gained popularity in recent years, especially in the field of data science. TDA offers a powerful set of tools for analyzing complex and high-dimensional data, making it a promising method for researchers in economics. In this paper, we provide an overview of TDA and explain how it differs from traditional approaches in economics research. We also discuss the different methods TDA provides and how they have been applied in other fields. Finally, we present examples of how TDA has been applied in economics research and discuss the potential for TDA to add value to future research. This paper contributes to the growing literature on the use of TDA in economics and offers a starting point for researchers interested in this approach.

1 Introduction

1.1 Motivation

Topological Data Analysis (TDA) is an emerging multidisciplinary field that has gained popularity in recent years across a number of research areas. TDA provides a powerful set of tools for analyzing complex and high-dimensional data, which makes it a promising method for researchers in economics. However, despite its potential, TDA remains somewhat of a fringe topic in the field of economics. As such, the aim of this paper is to familiarize economists with the basic ideas of TDA and demonstrate some potential applications in the field.

In particular, we aim to achieve the following objectives: first, to provide a comprehensive overview of TDA and explain how it differs from traditional approaches in economics research. Second, to present an overview of the different methods available in TDA and how they have been applied across numerous fields. Third, to describe economic applications in detail and discuss directions for future research.

The motivation for this paper is twofold: First, we aim to raise awareness of TDA among the economics community, as it has significant potential to contribute to the development of new findings. Second, we seek to provide a useful resource for researchers interested in using TDA in their own analyses, but may not know where to start. By providing an overview of the field and demonstrating potential applications in economics, we hope to encourage economists to learn more about TDA and contribute to its growing literature.

1.2 Defining Topological Data Analysis

At the highest level, TDA applies concepts from algebraic topology to identify the shape and structure of data, such as clusters or holes, that may not be apparent using traditional methods. By identifying and tracking such topological features, TDA may help provide insight

Electronic copy available at: https://ssrn.com/abstract=4378151

into the underlying structure of the data, which can help researchers extract features for predictive models and understand patterns and relationships that may not reveal themselves using traditional methods.

TDA is driven by the idea that data has *shape* that can be quantified and understood. This process is driven by the representation of data as a "point cloud" and said point cloud as a simplicial complex. A simplicial complex provides an approximation of the underlying topological object represented by the dataset. Using the ideas of persistent homology, this object contains information about the shape of the data that can be useful in prediction and inference.

TDA has been applied broadly in many fields already, with neuroscience, biology, and computer science at the forefront. It has yielded promising results as a powerful new tool, but remains relatively obscure in the field of economics (outside some applications in the financial sector).

1.3 TDA vs. Traditional Approaches in Economics

Traditional methods employed by economists are often designed to analyze relationships between variables, such as supply and demand or the impact of policy interventions. These methods typically assume that data is linear or describable via some parametric model, such as linear regression. The field has expanded into nonparametric methods and ML in recent years, but still relies heavily on traditional econometric approaches. Many such methods may struggle to capture complex, high-dimensional data with non-linear relationships.

TDA, on the other hand, is designed to analyze data that is complex and high-dimensional, such as social networks, brain imaging data, or multidimensional stock price correlation series. It does not assume a specific parametric model and instead uses mathematical techniques to find patterns and structures within the data. This allows it to find non-linear relationships and provide interesting insights about the underlying structure of the data itself.

Another key difference between TDA and traditional methods in economics is its emphasis on visual representations of data. TDA provides a way to visualize complex data in two or three dimensions, which can make it easier for researchers to identify patterns and relationships. Traditional econometric tools, on the other hand, often rely on statistical tests and measures that may be difficult to interpret or visualize.

Overall, TDA offers a complementary approach to traditional methods and can be especially useful when data is complex or high-dimensional. By identifying patterns and structures that may not be apparent with traditional tools, TDA has potential to contribute new insights and discoveries to the field.

1.4 Overview of this Paper

The remainder of this paper is structured as follows. We begin with a high-level summary of TDA, its core concepts, the methods and modules it provides, and a summary of differences between these tools and traditional approaches. We then survey the literature for applications of TDA across a number of fields, with a particular focus on economics. Finally, we provide a summary of potential use cases for TDA in economics, by method, and provide

Electronic copy available at: https://ssrn.com/abstract=4378151

a narrative on pros/cons and potential challenges. We share insight on existing software implementations of TDA tools available for public use. To conclude, we present a summary of key points and directions for future research.

2 TDA Concepts and Methods

This section provides an overview of the concepts and methods present in topological data analysis. Primarily, we focus on the "traditional" TDA pipeline, which involves representation of data as a point cloud, a point cloud as a simplicial complex, and the extraction of useful features from that simplicial complex or filtration of complexes. We do not explore more advanced topics, including the Ball Mapper algorithm or Persistence Barcodes. For information on those topics, refer to [17] and [26], respectively.

2.1 The TDA Pipeline

While low-dimensional relationships are straightforward to visualize, the visualization of higher-dimensional data is difficult and often requires decomposition into 2- or 3-dimensional space. TDA provides a rigorous framework to examine the "shape" of underlying data [13]. Our aim is to quantify the underlying structure of a multidimensional dataset by representing it in a simplified topological signature. This is a lossy process, meaning that some level of information is lost during the creation of the summary structure. A particular challenge is to put the data into a form that fits standard TDA pipelines while allowing the necessary lossy-ness to reduce dimensionality rather than eliminating important structures [47].

While TDA is an emerging field, many applications rely on some variation of the following pipeline [12].

1. A multidimensional dataset is viewed as a finite set of points in some metric space.
2. A continuous shape is generalized from the data to highlight its underlying geometry. This is usually a simplicial complex or family of simplicial complexes called a filtration that reflect the structure of the data at different scales.
3. Topological information is extracted from these continuous objects using the tools of persistent homology.

The remainder of this section describes the TDA process in detail.

2.2 Data as Point Clouds

Any n-dimensional dataset can be represented as a finite set of points in some metric space. We refer to this representation of data as a *point cloud*. Within the stated space (often Euclidean), some formal concept of distance exists which can describe the relationship between points. Given the often stochastic nature of underlying data, point clouds are viewed as noisy samples taken from some underlying geometric object [10]. The presence of noise is a feature separating TDA from the field of computational geometry [64], of which a classical problem is the topological reconstruction of geometric objects from point clouds.

Textbooks on mathematical analysis (e.g. Rudin [56]) provide a formalization of the metric space, that is a set X in which some distance function ρ is defined alongside the following properties.

Electronic copy available at: https://ssrn.com/abstract=4378151

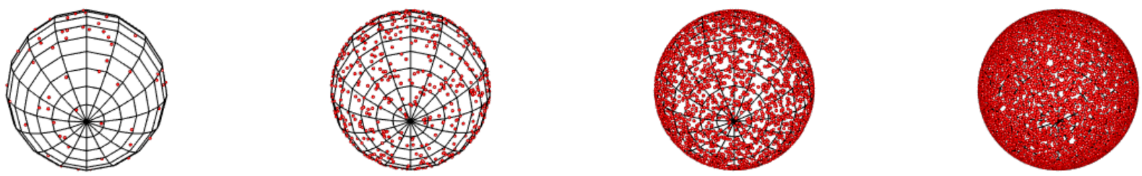
1. $0 \leq \rho(x, y) \leq \infty \ \forall x, y \in X$. This means that the defined distance between two points must exist within the bounds of $[0, \infty)$.

2. $\rho(x, y) = 0$ if and only if $x = y$.

3. $\rho(x, y) = \rho(y, x) \ \forall x, y \in X$. Distance is commutative.

4. $\rho(x, z) \leq \rho(x, y) + \rho(y, z) \ \forall x, y, z \in X$. The triangle inequality applies: the shortest path from any one point to another is a straight line.

A 3-dimensional point cloud is straightforward to visualize. Figure 1 provides a demonstration of discrete points constituting a point cloud with the underlying topological object converging to a sphere as the number of points tends toward infinity. This idea naturally extends to dimensions 4- and above, though they can't be visualized in the same way. Point clouds are usually not uniformly distributed in space, but carry some geometric structure that represents important properties of the systems from which they are generated [11]. The derived shape of a dataset may provide useful information about the underlying phenomena the data represents [33]. To understand the shape and connectivity of the underlying object, we leverage the ideas of persistent homology.



(a) n=100 (b) n=1000 (c) n=5000 (d) n=10000

Figure 1: A 3-d point cloud that converges to a sphere as $n \to \infty$

2.3 Persistent Homology

Persistent homology is an algebraic method for the measurement of topological features of shapes and functions [33]. To understand persistent homology, we outline its basic mathematical preliminaries, the majority of which are derived from algebraic topology, a subject that provides a set of tools for counting and collating holes and other topological features in space [26]. In general, the homology of some topological space X is a set of topological invariants of X represented by its homology groups:

$$H_0(X), H_1(X), H_2(X), \cdots$$

In degree 0, refer to the connectedness of the data; degree 1 describes the presence of holes or tunnels; degree 2 describes voids; and so on [8, 11]. We lose the ability to visualize structures captured for $k \geq 2$, so they aren't commonly used in applications [47]. For example, the image of a circle in $\mathbb{R}^2$ has $H_0 = 1, H_1 = 1$ since there is one connected component and one hole. A torus in $\mathbb{R}^3$ has one connected component, since any two points chosen along the exterior can be connected by drawing a path between them. As a result, $H_0 = 1$. A torus has two holes, the first is the large hole in the center and the second can be seen by cutting the torus in half. Hence, $H_1 = 2$. Finally, it has a single void (inside is hollow), which means that $H_2 = 1$ [46].

Electronic copy available at: https://ssrn.com/abstract=4378151

Before attempting to understand the homology features of a derived shape, we must first understand how to derive a continuous shape from a noisy point cloud. This is accomplished through the use of simplicial complexes [46]. The idea of persistent homology is about the examination of features that *persist* on derived topological objects across various levels of scaling/resolution. Such scaling is defined by the selected distance parameters in constructing the complex.

2.4 Simplicial Complexes

The continuous shape inferred from our discrete point cloud is called a *simplicial complex*, which views our point cloud as the vertices of a combinatorial graph whose edges are determined by some proximity measure [43]. This proximity measure defines the "resolution" of the complex and is often represented as the parameter ϵ.

The simplicial complex is built from simple pieces (simplices) identified by the point cloud and chosen resolution parameter ϵ. It was first introduced in 1895 as a triangulation of a manifold [52]. Each observation in the point cloud is a vertex (0-dimensional simplex), an edge (1-dimensional simplex) is defined by its two endpoint vertices, a triangle (2-dimensional simplex) defined by three vertices, and so on. In general, a d-dimensional simplex is defined by $d+1$ vertices [47]. A collection of simplices is a simplicial complex. A variety of complexes exist, but the two most commonly utilized are the Cech Complex and the Vietoris-Rips Complex [26].

While the deduction of a simplicial complex from a particular point cloud and scaling parameter ϵ is useful, it is insufficient for most applications. It's better to understand the *persistence* of topological features at different levels of scaling (ϵ) [26]. This provides a summary of the behavior of the underlying object across all possible resolutions with the idea that features persisting across many values of ϵ are most important [10].

The development of simplicial complexes at various levels of scaling forms a filtration of complexes $\emptyset \subseteq \mathbb{X}_0 \subseteq \mathbb{X}_1 \subseteq \cdots \subseteq \mathbb{X}_n \subseteq \mathbb{X}$. In this context, we consider the "birth" and "death" times of topological features and the range of ϵ values for which they persist [28]. Studies using TDA to examine simplicial complexes directly or studying the homology patterns therein are guaranteed to be robust to perturbations and do not vary under changes in coordinates or deformations of the original samples [52].

Many studies use the Rips complex, primarily because it is viewed as a simpler (and acceptable) approximation to the more complicated Cech complex. While no single Rips complex appropriately approximates a single Cech complex, pairs of Rips complexes "squeeze" the associated Cech complex into a manageable hole [26, 36]. Especially when input data is large, Rips is the common choice [14].

For each instance of the scaling parameter $\epsilon > 0$, the Rips complex $R(X, \epsilon)$ is defined as:

- For each $k = 0, 1, \cdots$, a k-simplex of vertices $x_{i1}, x_{i2}, \cdots x_{ik}$ is part of $R(X, \epsilon)$ if and only if the mutual distance between any pair of its vertices is less than ϵ: $d(x_{ij}, x_{il}) < \epsilon \ \forall x_{ij}, x_{il} \in X$ [28].

This process involves the examination of all pairwise distances between points based on the notion of Euclidean distance. Once the complex is constructed, we can use the tools developed in homology to understand the number and utility of relevant features [46]. The primary tool we use to visualize the persistent homology of our filtration of simplicial complexes is the *persistence diagram*.

Electronic copy available at: https://ssrn.com/abstract=4378151

2.5 Persistence Diagrams and Persistence Entropy

The value of ϵ at which a homology class is first observed is its *birth* and the value at which it disappears is its *death*. This leads to an ordered pair (b, d) associated with each derived homology class that can be visualized via persistence diagram. Classes with a short lifetime exist close to the diagonal, while those that persist over many values of ϵ are further away. Such points are seen as representations of the inherent structure of the data while those close to the diagonal are likely attributable to noise [47]. An example of a persistence diagram is provided in Figure 2.

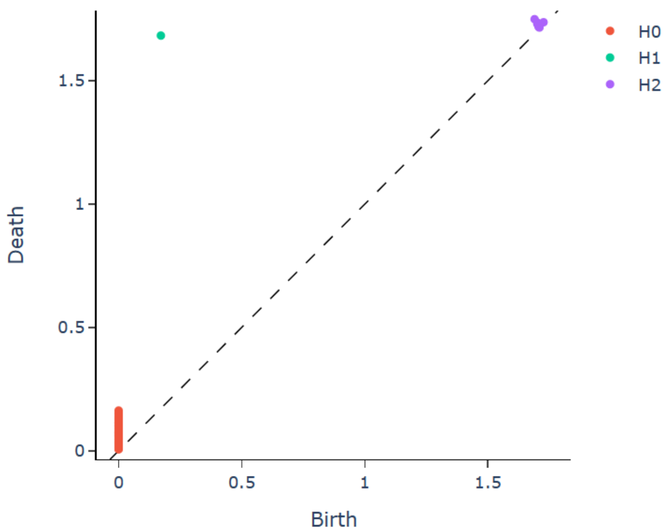


Figure 2: An example of a persistence diagram from the Python *giotto-tda* website.

The stability theorem for persistence is arguable the most important theorem for justifying the use of persistent homology as a tool for data analysis. It says that the distance between two persistence diagrams can't be larger than the distance between the two datasets used to obtain them. This means that even if we're given a dataset infected with noise, the persistence diagram obtained from it is approximately correct because its close enough to the diagram we would have from noise-free data [47].

While the persistence diagram is useful for visualization of the persistence of toplogical features over various resolutions, they aren't suitable for analysis with statistical theory and we therefore must consider alterantive measures [8]. Conceptually, we want to quantify the sequence of birth-death information of the derived homology features and explore their potential utility in prediction and inference. One feasible option is to derive the *persistence entropy* vector described in [48]. This generates a vector of quantified entropy representing the order/disorder of the underlying topological object over time.

You may, for example, take a series of observations of some metric r over entity i and day j. Thus for each day, you have a vector of observations:

$$X_j = (r_{aj}, r_{bj}, \cdots)$$

Given an arbitrary window size $w = 10$ days, and for each $t \geq w$, we can construct a w-dimensional time series defined as $\mathbb{X} = \{X_{t-9}, X_{t-8}, \cdots X_{t-1}, X_t\}$. We may then apply this sliding window to the dataset and create an array of point clouds to be used in a persistent homology calculation. To convert this series of "frames" into a filtration of simplicial

Electronic copy available at: https://ssrn.com/abstract=4378151

complexes, we can use the Python package *giotto-tda*, and specifically, the *VietorisRipsPersistence* (VRP) object wtihin.

The VRP object takes a set of homology dimensions (e.g. 0, 1, 2) and transforms the vector of point clouds into a set of homology vectors corresponding to the selected homology dimension. You may then extract scalar features of the dataset using the *PersistenceEntropy* object. Derived persistence entropy values can be used directly as inputs to a regression model or visually to examine relationships derived from high-dimensional features. More information on this at the *giotto-tda* website.

2.6 Comparison of TDA to Traditional Methods

Traditional econometric tool which are designed to analyze relationships between variables using linear or parametric methods have been succesful in many cases, but can struggle to capture complex, high-dimensional data that is becoming increasingly common. TDA offers a complementary approach to such traditional methods, focusing on deriving information about the shape and structure of the data.

One key difference between TDA and traditional methods is the way that data is represented. Traditional methods often rely on summary statistics and numerical models, while TDA emphasizes visual representation and the use of topological features to identify patterns in data. This can make it extremely useful for exploring high-dimensional data, as it allows researchers to visualize information that can be difficult to see using other methods.

Another difference is the underlying assumptions of the two approaches. Traditional methods often assume linearity or a specific parametric model, and such assumptions will inevitably be violated. TDA on the other hand, does not make any assumptions about the shape or structure of the data, and can identify non-linear relationships and patterns that can be missed in traditional settings.

A particular challenge in using TDA is that it can be computationally intensive, especially for large datasets. This can make it difficult to apply to certain types of problems, or may lead to the need for specialized hardware or softawre. Additionally, the learning curve is intimidating, though in the opinion of the author, is quite manageable for those familiar with econometrics already.

The important takeaway in comparing TDA with traditional methods is that TDA is very much a *complementary* tool. Two ways value can be derived from this complementary approach are [a] the derivation of visuals that would otherwise be inaccessible; and [b] the retrieval of quantified persistence entropy information from multidimensional time series. Combined with standard approaches, this increases the ability of researchers to identify meaningful trends and uncover previously hidden information.

3 A Survey of Applications in TDA

TDA has been successfully applied to a variety of fields, including biology, neuroscience, computer science, and more recently, economics. It has shown promising results in uncovering new results and discoveries. This section provides a survey of the applications of TDA broadly and then with a targeted focus on economics. As the literature is extensive, we make no attempt to create a comprehensive and rigorous review, nor do we explore the methodological details of every study in depth. Rather, we provide a high-level summary of areas

Electronic copy available at: https://ssrn.com/abstract=4378151

where TDA has been applied with success and focus on its major contributions.

We begin with a review of some key applications in non-economics fields such as biology, neuroscience, and computer science. We provide examples of how TDA has been used in these fields and highlight some key discoveries that it helped produced. Then, we turn our attention to the use of TDA in economics research specifically. Overall, this section highlights the versatility and potential of TDA as a tool for exploring complex, high-dimensional data in a wide range of fields.

3.1 Broader Applications of TDA

We take an intentionally broad and surface-level view of the applications of TDA outside of economics. The approach is to show that TDA has far reaching applicability and broadly demonstrated success.

In neuroscience, TDA has been used to identify topological features within the brain such as cavities [60]; reveal the organization of whole-brain activity maps [58]; explore TBI and spinal cord injury datasets [51]; understand event-related fMRI designs [22]; distill high-dimensional neuroimaging data into simple representations [25]; identify biomarkers of recovery from TBIs [50]; and analyze single-trial EEG signals [63]. Closely related, TDA has also shown success in medical image processing: exploration of hyperspectral imaging datasets [21]; clustering of cardiovascular patients [34]; and classification between healthy patients and diabetics [24].

In the natural sciences, TDA has been used to generate large-scale genomic recombination maps [9]; model the spread of the Zika virus [42]; understand biological aggregation models [62]; analyze immune reconstitution after stem cell transplantation [38]; quantify zebrafish pattern formations [45]; and many more [39, 16, 53, 41, 55].

TDA has been applied in a number of other areas including physics, engineering, signal processing, and operations resarch: analysis of fast radio burst observations [49]; exploration of high-dimensional aviation data [40]; understanding order structure in sodium silicate glasses [61]; identification of cosmic voids in cosmological datasets [65]; improving classification results in movie genre detection [20]; detecting structural information in manufacturing data [32]; mining social network data to identify clusters [2]; study spatial networks [23]; and image tampering detection [3].

In the following section, we review the existing cases of TDA applied to economics, and go in a bit more detail than our assessment of the non-economics publications.

3.2 TDA in Economics

In economics, the primary utilization of TDA has been for the identification of early "warning signals" of market instability [35, 27, 30, 57, 31, 6, 28, 37]. Aside from early warning signals, TDA has shown promise in exploring market dynamics over time [1, 5, 15, 44], examination of economic cycles [19], corporate failure [54]; stock returns [18], investment strategy development [4], financial risk measures [59], and asset allocation [29].

What follows in this section is a review of these applications, a discussion of common methods, themes, and outcomes, and an attempt to draw conclusions about where TDA has been proven successful so far, and where that may lead for future research questions.

Electronic copy available at: https://ssrn.com/abstract=4378151

Early Warning Signals

The primary application in finance and economics has been in the identification of early signatures of market instabilites. These are often framed as "warning signals." The potential is demonstrated repeatedly across multiple studes assets, and time frames.

The major applications of such techniques began with Marian Gidea's 2017 paper [28]. Within, a TDA-based method for the detection of early warning signals for critical transitions in financial data is developed. The time series of multiple stock prices are used to build time-dependent correlation networks, which are envisioned as topological structures in a process similar to the TDA pipeline described in section 2. An important note from this analysis is in the discussion of a *critical transition*, and what that term really means.

In terms of a complex financial system, a critical transition refers to an abrupt change in the behavior of the system arising due to changes in external conditions, which makes the system switch from one steady state to another [28]. The common example of a critical transition in financial systems is the overall market crash (e.g. 2007-2009 financial crisis). Many other possible applications of "critical transitions" exist in areas studied by economists, and this is one area potentially fruitful for further study.

In a similar paper (2018), Gidea and Katz [28] provide a similar analysis of four major US stock market indices during the technology crash of 2000 and the financial crisis of 2007-2009. The authors use a sliding window approach to extract time-dependent point clouds, derive a filtration of simplicial complexes, and extract relevant topological information therefrom. Temporal changes in persistence landscapes are quantified via L^p-norms. In the times prior to financial meltdowns, these norms exhibit strong growth, potentially creating an "early warning signal" of changes to the complex system underlying the derived simplicial complexes. Basu and Li [5] provide a similar application, which utilizes TDA to integrate with a standard machine learning framework, thus boosting its predictive power, especially in regard to anticipating crashes. Other studies in this realm include [30, 57, 35, 66, 67, 1].

Exploration of Market Dynamics Over Time

Moving away from the idea of developing early warning signals, another application is the examination of market dynamics. de Carufel et al [15] use analytical topology to characterize financial price series. The authors refer to a method implemented by Brooks [7], which consists of decomposing data into components of its *total variation*. The de Carufel analysis takes this methodology and applies it to price data, wherein the topological decomposition of total variation is used to demonstrate the scaling property of financial instruments.

Basu and Dlotko [5] introduce a new technique for analysis of the evolution of correlations for multiple time series and apply TDA to gain insights about commodity futures markets over a 20 year period. Such analyses may provide an understanding of the evolution of (and changes in) correlation structures in a variety of time series problems.

Majumdar and Laha [44] develop an approach for time series classification and clustering based on the ideas in TDA, e.g. persistent homology. The clustering method they propose examines the topological similarities and dissimilarities of some well-known time series models in finance. The classification method examines whether topological features can be used to distinguish between time series models using simulated data. Both perform well and lend further legitimacy to the use of TDA in the analysis of time series, particularly the dynamics

Electronic copy available at: https://ssrn.com/abstract=4378151

of complex economic systems.

Other Stated Applications
Aside from these two areas, many other applications have been published. Sato [**?**] investigated the relationships between TDA barcodes and traditional financial risk measures, such as growth rate, volatility, and correlation coefficients. Results support the effectiveness of TDA as a valid risk measure, especially in detecting rapid changes in a short period of time. This suggests the potential for the use of TDA as a supplementary tool in financial risk management. Another risk-related application is presented in Baitinger and Flegel [4]. Dlotko et al. [18] use TDA to explore the association between seven of the most commonly studied financial ratios and stock returns, using the TA Ball Mapper algorithm to identify interdependencies between factors. Their results show significant opportunity for investors to exploit the knowledge attainable through the use of TDA-derived information.

In a separate paper, Dlotko et al. [19] use TDA to "topologically map the macroeconomy," which offers a new representation of data to inform policy and deepen understanding of the underlying mechanisms fueling the economy. 2-d snapshots of multi-dimensional space capture non-monotonic relationships, and specific examples show how some countries have returned to Great Depression levels, and reappraise the links between real private capital growth and the performance of the economy. The specific point here is that TDA reveals information about non-monotonicity where monotonicity may otherwise have been assumed.

Qiu et al. [54] use TDA to examine whether failing firms from the US organize neatly along the five predictors of default proposed by the Z-score models. Each firm represented as a point in a 5-d point cloud, each dimension being one of the five predictors. Visualizing that cloud using Ball Mapper reveals failing firms are not always located in similar regionso f the point cloud, i.e. they aren't concentrated in an easily separable area of the space. The authors recommend understanding where in parameter space failure often occurs and to drive risk-averse lenders toward spaces where failure typically does not occur.

4 Guidance on Using TDA in Economics

The summary of public literature from section 3.2 yields an overview of the core areas that TDA has added value in economics to date. This section presents an overview of recommendations for exploring the utility of TDA in the broad field of economics.

- **Financial Network Analysis** - the most dominant area of TDA research applied to economics, TDA can be used to generate "early warning signals" of structural changes to complex systems, most notably the financial markets. This may be used for systemic risk monitoring, preventative alerts for banks or insurers, etc.

- **Evolution and Correlation of Time Series** - similar to the prior point, but not limited to financial markets, TDA provides a mechanism for quantifying and tracking otherwise "hidden" information about multiple time series, their correlations, and their evolution through time.

- **Enancement of Predictive Modeling Frameworks** - TDA may be useful in enhancing the quality / accuracy of predictive modeling frameworks via the generation of new features and the uncovering of previously "hidden" information about the structure of the underlying system. This is especially true in time series, wherein "topological

Electronic copy available at: https://ssrn.com/abstract=4378151

data analysis on time series (TDA-TS)" is used to analyze time series data by identifying patterns that can help predict future trends. This has been used to analyze stock prices and GDP growth.

- **Topological Clustering** - TDA can be used to analyze the distribution of economic activity in multidimensional space. This can include the identification of clusters of economic activity, measuring the distance between clusters, and identifying factors that contribute to cluster membership.

- **Visualization of High-Dimensional Data** - The Mapper algorithm (not covered extensively in this paper) can be used to visualize high-dimensional data by reducing it to a lower-dimensional space. Examples include the examination of consumer preference data to identify market segments, labor market data to identify areas of skill concentration, etc.

- **Dataset Similarity Comparison** - the Bottleneck distance can be used to compare and analyze the similarity between datasets with different resolutions. E.g. it has been used to compare economic time series data with different levels of granularity to identify patterns and relationships in time.

TDA is in the early stages of development and shows great potential to add value to economics research. This is true in that it provides a number of new and unique methods as well as their broad potential for application across a number of economics sub-fields.

4.1 Software Implementation of TDA Logic

A number of open source software exists to provide researchers with easy-to-use tools for applying TDA methods to various problems. Here are some common examples.

- **GUDHI** is an open source library written in C++ that provides a wide range of tools, including persistent homology, cohomology, and Mapper. It can be accessed via Python or C++.

- **TDAstats** is an R package that provides many TDA tools, including persistent homology, Mapper, etc.

- **Giotto-TDA** is a Python package that includes Mapper, persistent homology, and clustering.

Many others exist, including Ripser, Dionysus, and Ayasdi (commercial). The open source packages are simple to easy, publicly available, and well-documented.

4.2 Potential Challenges of Using TDA in Economics Research

Topological Data Analysis (TDA) has great potential in the field of economics, but it also presents several challenges to researchers. One of the main challenges is ensuring data quality and preparation. TDA requires high-quality data that is free from noise and errors, which can require specialized domain-specific knowledge and expertise. Another challenge is computational complexity, as TDA algorithms can be computationally intensive, especially for large and complex data sets. The choice of parameters, such as the scale parameter or the resolution parameter, is another challenge in TDA, which can significantly affect the results and require domain-specific knowledge. In the case of sliding window-based time series approaches, the choice of the size of the sliding window is another relevant "hyperparameter."

Electronic copy available at: https://ssrn.com/abstract=4378151

Interpreting results is also challenging, as TDA generates topological features that may be difficult to relate to economic concepts. Reproducibility may be difficult, as TDA algorithms can be sensitive to the choice of software implementation, parameter settings, and data preparation. The theoretical foundations of TDA are still being developed, making it challenging to explain and justify TDA results to non-experts. Finally, integrating TDA with existing methods in economics may also be challenging.

To address these challenges, researchers need to have a strong mathematical grounding to arrive at an understanding of TDA and its applications in economics, as well as the domain-specific knowledge and expertise required to analyze economic data. Furthermore, researchers need to adopt a rigorous and transparent approach to data preparation, algorithm selection, and interpretation of results, and to engage in open and reproducible research practices.

5 Conclusion

In conclusion, Topological Data Analysis (TDA) offers a powerful tool for analyzing complex and high-dimensional data in economics research. TDA's potential to unveil underlying structure and patterns in data can help advance our understanding of complex economic systems. While TDA is a relatively new method, it has gained popularity in recent years and offers a promising approach for researchers who seek to uncover hidden insights in large and complex data sets.

One of the key advantages of TDA is its ability to reveal hidden information that traditional statistical methods may overlook. By mapping data to a geometric space, TDA can extract key features and relationships between variables that may be difficult to discern using traditional statistical methods. TDA can provide a new way of looking at the same data, allowing researchers to see patterns and trends that might otherwise be missed.

Despite the challenges of data preparation, computational complexity, parameter selection, and interpretation of results, TDA has a relatively low learning curve, making it accessible to researchers with a wide range of expertise. TDA algorithms are readily available in a range of open-source software packages, which can be used to analyze complex data sets quickly and efficiently. Furthermore, TDA software packages are constantly being updated and improved, providing researchers with access to cutting-edge tools and techniques.

In economics research, TDA has already been applied to a range of areas, including financial markets, labor markets, natural resource economics, and environmental economics. TDA has been used to analyze the structure and connectivity of financial networks, identify market segments in consumer preference data, and assess the spatial distribution and connectivity of ecosystem services. TDA has also been used to identify clusters of related industries and to analyze the spatial and temporal patterns of land use and land cover change.

By moving away from traditional financial markets, TDA offers a new direction of research in economics, with applications in areas such as labor markets, where TDA has been used to analyze the spatial distribution of job opportunities and identify areas of skills concentration. TDA can also be applied to natural resource economics and environmental economics to identify priority areas for conservation and restoration, and to analyze the effects of habitat fragmentation on biodiversity.

We believe that TDA has the potential to add value to economics research by enabling researchers to see patterns and trends that might otherwise be missed. By providing a powerful

Electronic copy available at: https://ssrn.com/abstract=4378151

tool for analyzing complex and high-dimensional data, TDA offers new directions of research in economics and can help advance our understanding of complex economic systems. As TDA continues to evolve and improve, it will likely become an increasingly important tool for researchers in the field of economics.

References

[1] A. Aguilar and K. Ensor. Topology data analysis using mean persistence landscapes in financial crashes. *Journal of Mathematical Finance*, 10(4), 2020.

[2] K. Almgren, M. Kim, and J. Lee. Extracting knowledge from the geometric shape of social network data using topological data analysis. *Entropy*, 19(7), 2017.

[3] A. Asaad and S. Jassim. Topological data analysis for image tampering detection. *International Workshop on Digital Watermarking*, 2020.

[4] E. Baitinger and S. Flegel. The better turbulence index? forecasting adverse financial market regimes with persistent homology. *Expert Systems with Applications*, (156), 2020.

[5] D. Basu and P. Dlotko. The four seasons of commodity futures: Insights from topological data analysis. *SSRN*, (3506780), 2019.

[6] D. Basu and T. Li. A machine-learning-based early warning system boosted by topological data analysis. *SSRN*, (3394704), 2019.

[7] M. Brooks. Varilets: Additive decomposition, topological total variation, and filtering of scalar fields. *arXiv*, (1503.04867v3), 2016.

[8] P. Bubenik. Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research*, 16(1):77–102, 2015.

[9] P. Camara. Topological data analysis generates high-resolution, genome-wide maps of human recombination. *Cell Systems*, 3(1):83–94, 2016.

[10] G. Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

[11] F. Chazal, M. Glisse, and C. Labruere. Convergence rates for persistence diagram estimation in topological data analysis. *Journal of Machine Learning Research*, 16:3603–3635, 2015.

[12] F. Chazal and B. Michel. An introduction to topological data analysis: Fundamental and practical aspects for data scientists. *arXiv*, (1710.04019), 2017.

[13] J. Curry. Topological data analysis and cosheaves. *Japan Journal of Industrial and Applied Mathematics*, 32(2):333–371, 2015.

[14] S. Dantchev and I. Ivrissimtzis. Efficient construction of the cech complex. *Computers and Graphics*, 36(6):708–712, 2012.

[15] J. de Carufel. A topological approach to scaling in financial data. *arXiv*, (1710.08860), 2017.

[16] M. Dindin, Y. Umeda, and F. Chazal. Topological data analysis for arrythmia detection through modular neural networks. *Canadian Conference on Artificial Intelligence*, 2020.

Electronic copy available at: https://ssrn.com/abstract=4378151

[17] P. Dlotko. Ball mapper: A shape summary for topological data analysis. *arXiv*, (1901.07410), 2019.

[18] P. Dlotko, W. Qiu, and S. Rudkin. Financial ratios and stock returns reappraised through a topological data analysis lens. *arXiv*, (1911.10297), 2019.

[19] P. Dlotko, S. Rudkin, and W. Qiu. Topologically mapping the macroeconomy. *arXiv*, (1911.10476), 2019.

[20] P. Doshi and W. Zadrozny. Movie genre detection using topological data analysis. *International Conference on Statistical Language and Speech Processing*, 2018.

[21] L. Duponchel. Exploring hyperspectral imaging data sets with topological data analysis. *Analytica Chemica Acta*, 1000(13):123–131, 2018.

[22] C. Ellis. Feasibility of topological data analysis for event-related fmri. *Network Neuroscience*, 3(3):695–706, 2019.

[23] M. Fent and M. Porter. Spatial applications of topological data analysis: Cities, snowflakes, random structures, and spiders spinning under the influence. *Physical Review Research*, 2(3), 2020.

[24] K. Garside, R. Henderson, I. Makarenko, and C. Masoller. Topological data analysis of high resolution diabetic retinopathy images. *PloS one*, 14(5), 2019.

[25] C. Geniesse. Generating dynamical neuroimaging spatiotemporal representations (dyneusr) using topological data analysis. *Network Neuroscience*, 3(3):763–778, 2019.

[26] R. Ghrist. Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2008.

[27] M. Gidea. Topological data analysis of critical transitions in financial networks. *International Conference and School of Network Science*, 2017.

[28] M. Gidea and Y. Katz. Topological data analysis of financial time series: Landscapes of crashes. *Physica A: Statistical Mechanics and Its Applications*, 491.

[29] A. Goel, P. Pasricha, and A. Mehra. Topological data analysis in investment decisions. *Expert Systems with Applications*, 147, 2020.

[30] H. Guo. Empirical study of financial crises based on topological data analysis. *Physica A: Statistical Mechanics and Its Applications*, 558, 2020.

[31] W. Guo. Analysis of global stock markets' connections with emphasis on the impact of covid-19. *Phsyica A: Statistical Mechanics and Its Applications*, 569, 2021.

[32] W. Guo and A. Banerjee. Identification of key features using topological data analysis for accurate prediction of manufacturing system outputs. *Journal of Manufacturing Systems*, 43, 2017.

[33] Edelsbrunner H. and J. Harer. Persistent homology: A survey. *Contemporary Mathematics*, 432:257–282, 2008.

[34] D. Hwang. Topological data analysis of coronary plaques demonstrates the natural history of coronary atherosclerosis. *JACC: Cardiovascular Imaging*, 2021.

[35] Y. Katz and A. Biem. Time-resolved topological data analysis of market instabilities. *Physica A: Statistical Mechanics and Its Applications*, 571, 2021.

Electronic copy available at: https://ssrn.com/abstract=4378151

[36] M. Kerber and R. Sharathkumar. Approximate cech complex in low and high dimensions. *International Symposium on Algorithms and Computation*, pages 666–676, 2013.

[37] W. Kim. Investigation of flash crash via topological data analysis. *Topology and Its Applications*, 2020.

[38] T. Lakshmikanth. Mass cytometry and topological data analysis reveal immune parameters associated with complications after alogeneic stem cell transplantation. *Cell Reports*, 20(9), 2017.

[39] M. Li and B. Park. Modeling the spread of the zika virus using topological data analysis. *PloS one*, 13(2), 2018.

[40] M. Li, M. Ryerson, and H. Balakrishnan. Topological data analysis for aviation applications. *Transportation Research Part E: Logistics and Transportation Review*, 128, 2019.

[41] T. Liao. tmap: An integrative framework based on topological data analysis for population-scale microbiome stratification and association studies. *Genome Biology*, 20(1), 2019.

[42] D. Lo and B. Park. Modeling the spread of the zika virus using topological data analysis. *PloS one*, 14(5), 2018.

[43] P. Lum, G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson, and G. Carlsson. Extracting insights from the shape of complex data using topology. *Scientific Reports*, 3(1):1–8, 2013.

[44] S. Majumdar and A. Laha. Clustering and classification of time series using topological data analysis. *Expert Systems with Applications*, 162, 2020.

[45] M. McGuirl, A. Volkening, and B. Sandstede. Topological data analysis of zebrafish patterns. *Proceedings of the National Academy of Sciences*, 117(10), 2020.

[46] P. Medina and R. Doerge. Statistical methods in topological data analysis for complex, high-dimensional data. *arXiv*, (1607.05150), 2016.

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

[48] E. Munch, A. Myers, and F. Khasawneh. Persistent homology of complex networks for dynamic state detection. *Physical Review E*, 100, 2019.

[49] J. Murugan and D. Robertson. An introduction to topological data analysis for physicists. *arXiv*, (1904.11044), 2019.

[50] J. Nielson. Uncovering precision phenotype-biomarker associations in traumatic brain injury using topological data analysis. *PloS one*, 12(3), 2017.

[51] J. Nielson, J. Paquette, and A. Liu. Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury. *Nature Communications*, 6, 2015.

[52] A. Patania, F. Vaccarino, and G. Petri. Topological analysis of data. *EPJ Data Science*, 6(1):1–6, 2017.

[53] J. Pike. Topological data analysis quantifies biological nano-structure from single molecule localization microscopy. *Bioinformatics*, 36(5), 2020.

Electronic copy available at: https://ssrn.com/abstract=4378151

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

[55] R. Rabadan and A. Blumberg. Topological data analysis for genomics and evolution: Topology in biology. *Cambridge University Press*, 2019.

[56] W. Rudin. Real and complex analysis. *McGraw-Hill International Editions*, 1987.

[57] P. Saengduean, S. Noisagool, and F. Chamchod. Topological data analysis for identifying critical transitions in cryptocurrency time series. *2020 IEEE International Conference*, 2020.

[58] M. Saggar, O. Sporns, and J. Gonzalez-Castillo. Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nature Communications*, 9, 2018.

[59] M. Sato. Can tda be a new risk measure? an application to finance of persistent homology. *SSRN*, (2710275), 2016.

[60] A. Sizemore, J. Phillips-Cremins, R. Ghrist, and D. Bassett. The importance of the whole: Topological data analysis for the network neuroscientist. *Network Neuroscience*, 3(3):656–673, 2019.

[61] S. Sorensen. Revealing hidden medium-range order in amorphous materials using topological data analysis. *Science Advances*, 6(37), 2020.

[62] C. Topaz, L. Ziegelmeier, and T. Halverson. Topological data analysis of biological aggregation models. *PloS one*, 10(5), 2015.

[63] Y. Wang, H. Ombao, and M. Chung. Topological data analysis of single-trial electroencephalographic signals. *The Annals of Applied Statistics*, 12(3), 2018.

[64] L. Wasserman. Topological data analysis. *Annual Review of Statistics and Its Application*, 5:501–532, 2018.

[65] X. Xu. Finding cosmic voids and filament loops using topological data analysis. *Astronomy and Computing*, 27, 2019.

[66] P. Yen and S. Cheong. Using topological data analysis (tda) and persistent homology to analyze the stock markets in singapore and taiwan. *Frontiers in Physics*, 99, 2021.

[67] P. Yen, K. Xia, and S. Cheong. Understanding changes in the topology and geometry of financial market correlations during a market crash. *Entropy*, 23(9), 2021.

6 Statements & Declarations

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Electronic copy available at: https://ssrn.com/abstract=4378151

Author Contributions

Not Applicable: Single Author

Electronic copy available at: https://ssrn.com/abstract=4378151