Time Series Data Mining with an Application to the Measurement of Underwriting Cycles

Iqbal Owadally[a,*], Feng Zhou[b], Rasaq Otunba[c], Jessica Lin[c], Douglas Wright[a]

[a]*Faculty of Actuarial Science and Insurance, Cass Business School, City, University of London, 106 Bunhill Row, London EC1Y 8TZ, United Kingdom*
[b]*Future of Humanity Institute, University of Oxford, 16/17 St Ebbe's Street, Oxford, OX1 1PT, United Kingdom*
[c]*Department of Computer Science, George Mason University, Nguyen Engineering Building, 4400 University Drive, Fairfax, VA 22030, USA*

Abstract

Underwriting cycles are believed to pose a risk management challenge to property-casualty insurers. The classical statistical methods that are used to model these cycles and to estimate their length assume linearity and give inconclusive results. Instead, we propose to use novel Time Series Data Mining algorithms to detect and estimate periodicity on U.S. property-casualty insurance markets. These algorithms are in increasing use in Data Science and are applied to Big Data. We describe several such algorithms and focus on two periodicity detection schemes. Estimates of cycle periods on industry-wide loss ratios, for all lines combined and for four specific lines, are provided. One of the methods appears to be robust to trends and to outliers.

Keywords: Data science, Algorithms, Big Data, Periodicity, Artificial Intelligence

1. Introduction

Underwriting cycles are cycles in profitability in the property-casualty insurance markets. These cycles appear to be present in different lines of business and in different countries, often independently. They pose a significant risk management challenge, but

*Corresponding author. IO is grateful to CKER, Society of Actuaries, for a research grant. The ideas in this paper were initiated when IO was in receipt of a grant from the Bank of England, to which he also expresses thanks.

Email addresses: `m.i.owadally@city.ac.uk` (Iqbal Owadally), `feng.zhou@philosophy.ox.ac.uk` (Feng Zhou), `rotunba@gmu.edu` (Rasaq Otunba), `jessica@cs.gmu.edu` (Jessica Lin), `dwright@city.ac.uk` (Douglas Wright)

are poorly understood.[1] A comprehensive description of these cycles, along with a summary of the various economic hypotheses explaining their existence, may be found in Harrington, Niehaus and Yu (2014). Recent actuarial studies of insurance cycles include Wang et al. (2011) who use a regime-switching Markov model, Taylor (2008) with a simulation model of the insurance industry, Trufin, Albrecher and Denuit (2009) on risk theory with cycles, and Ingram and Underwood (2010) on behavioral approaches to cycles (see also Ingram and Bush, 2013). Feldblum (2001) gives a very influential and intuitive account of cycles in property-casualty insurance from a practicing actuary's viewpoint.[2]

Classical time series methods are typically used to estimate the length of underwriting cycles. For example, AR(2) models are fitted to loss ratios or premiums (see e.g. Harrington, Niehaus and Yu, 2014). This presumes that cycles have a linear autoregressive character (Boyer and Owadally, 2015). Furthermore, the sample size of available annual insurance data is small. Boyer et al. (2012) and Boyer and Owadally (2015) find that the annual time series data set of loss ratios that is available is not long enough to determine with confidence whether a fitted AR(2) model will be cyclical. Despite using a battery of filters employed in the business cycle literature, Boyer et al. (2012) find that no insurance cycle can be forecast.

Nonlinear time series techniques have also been applied to underwriting cycles. Wang et al. (2011) model the loss ratio using a regime-switching Markov process and find that such a nonlinear Markov model provides a better fit to insurance data than an AR(2) model. Higgins and Thistle (2000) employ a smooth transition regression model. Jawadi et al. (2009) use a switching transition error correction model to examine non-linear cointegration between premiums and other variables.

[1] If insurers do not identify the severity and turning points of market cycles, they may mis-price their products. Cutting prices in response to competition in a 'soft' phase, when profitability and premiums are falling, may lead to steep losses in future years. Conversely, if insurers raise prices excessively in a 'hard' phase, they may lose business and have to abandon certain product lines altogether.

[2] We refer readers to the references cited above for further discussion of the economic theories of underwriting cycles and empirical results supporting these various theories.

These developments suggest that underwriting cycles are not necessarily cycles with a linear autoregressive character. One way of moving beyond the econometric strictures of classical linear time series is to use Time Series Data Mining algorithms developed in the discipline of Computer Science. Data scientists use these algorithms for such purposes as classification, segmentation, indexation and anomaly detection within 'Big Data' from a variety of sources, e.g. financial, medical, marketing etc. It turns out that one can use these algorithms to determine *periodicity* in data, rather than *cyclicality* in a Fourier analysis sense. Testing for cycles turns out to be a joint test of the goodness-of-fit of a statistical model and of the presence of cycles. Techniques like data mining are useful in this context because they enable us to augment the set of models that we can use to test for, and measure, cycles.

The purpose of this paper is to provide an introduction to these algorithms and to apply them to the problem of estimating underwriting cycles. DFA models, such as DYNAMO (CAS, 2013; D'Arch et al., 1997, 1998) and the model of Kaufman et al. (2001), generally use simple exogenous Markov models to capture the effect of underwriting cycles. It has long been recognized that underwriting cycles should be integrated within dynamic financial analysis (Warthen and Somner, 1996, p. 311). Time Series Data Mining should provide improved estimation methods for the length of underwriting cycles and thus improve risk management through better calibrated DFA models.

2. Time Series Data Mining

2.1. Data Mining and Data Science

In Statistics, data mining is often equated with data dredging, which is the practice of testing several hypotheses on a data set in the hope of finding a correlation which spuriously appears to be statistically significant. Rigorous statistical methodology requires that a hypothesis be formulated, an experiment be designed, and data collected. A statistical significance test is then carried out to see if the null hypothesis can be rejected. However, with the information revolution under way, with Big Data and Cloud Computing, businesses have found themselves inundated with data. Big Data refers to large amounts of raw multi-dimensional data collected continually and updated rapidly. How to exploit this data resource is the central concern of *Data Mining* which, in this paper, refers to a sub-discipline of Computer Science.

A more descriptively accurate term for Data Mining is "Knowledge Discovery from Data" (KDD) (Han et al., 2012). Terminology is fluid in the fast-changing world of Computer Science, and Data Mining is also used interchangeably with *Machine Learning*, although the latter tends to emphasize predictive tasks whereas Data Mining includes both predictive and descriptive tasks.

A key difference between Data Mining and Statistics is that Statistics is model-based with hypothesis testing at its core, whereas Data Mining is algorithm-driven with *search* as its principal tool. Statistics has strong theoretical underpinnings whereas Data Mining tends to be practically-oriented and is based on programming practice (Chakrabarti, 2009). However, there is a continuum between Data Mining and Statistics, and computer scientists use statistical tools as much as statisticians use computers, so the distinction is becoming increasingly blurred. Indeed, *Data Science* is now used as an umbrella term for a collection of various disciplines, including Statistics, Data Mining, and Machine Learning, all of which may be applied to Big Data.

Data Mining has been used successfully in a variety of areas, most notably credit assessment, marketing and sales, and biomedical research, so it is likely to find profitable use in Actuarial Science too. One way to understand Data Mining is to consider the functional tasks that are carried out as part of this discipline. These tasks may be broadly classified as predictive or descriptive.

A classic predictive task is *classification*. The closest analogue to this in Statistics is the task of regression. Classification is about discovering, from the data, a model or learning function that can map a data item to one of several predefined classes (Kantardzic, 2011; Han et al., 2012). The model or learning function could be a decision tree or flow chart with if-then rules, implemented as executable code. Neural networks and support vector machines are also used. In such cases, the term *Machine Learning* tends to be used alongside Data Mining. Whereas regression tends to be concerned with numerical data, classification is capable of handling categorical data. Marketers use association rules to uncover what items shoppers are likely to put together in their shopping baskets, so that these items can be located in close proximity in a store, or as suggested purchases on an e-commerce website.

A typical descriptive task in Data Mining is *clustering*. This is very similar to classification except that, whereas classes are predefined in classification (for example, via a training data set that has already been classified by the user), in clustering the groups or *clusters* are unknown at the start and the computer learns how to group the data. Thus, classification is supervised learning whereas clustering is unsupervised (Han et al., 2012). A typical strategy is to use iterative distance-based methods such as k-means clustering where data points are grouped into clusters iteratively until the clusters stabilize (Han et al., 2012; Kantardzic, 2011). Clustering is obviously critical in the era of Big Data and Cloud Computing, where the vast amount and flow of data makes supervised learning difficult. A typical example of clustering is a retail organisation grouping its customers into homogeneous sub-populations according to their shopping habits (Han et al., 2012).

2.2. Time Series Data Mining

Time Series Data Mining, also known as Temporal Data Mining, is a lesser known but growing sub-field of Data Mining. Time series are, of course, found everywhere, and are an essential part of the Big Data phenomenon. Such data are large, high-dimensional and frequently updated. For example, in the medical field, there are electrocardiograms (ECGs), electroencephalograms (EEGs), gait analysis data etc. One hour of ECG data is about 1 GB large and needs to be analyzed for anomalous rhythms (Aghabozorgi et al., 2015). Another example in engineering concerns telemetry from multiple sensors on board the International Space Station (and the Space Shuttle before its retirement). In finance, high-frequency algorithmic trading consumes financial market data at millisecond intervals. It is not surprising that Data Mining techniques have been re-tooled to deal with time series data.

It is worth noting that Time Series Data Mining is applied not just to time-stamped data but more generally to sequential data. Sequential data are ordered with respect to an index, not necessarily a time index (Kantardzic, 2011). For example, data mining algorithms originally designed for protein sequences and gene expression sequences have been used in fields other than biomedicine.

There are burgeoning applications of Time Series Data Mining in finance and risk management. It has been used in efficient portfolio construction (Guan et al., 2007), for discovering patterns in stock prices (Aghabozorgi and Ying Wah, 2014; Guo et al., 2008), measuring the riskiness of stocks (Stetco, 2013), and constructing hedges (Hsu and Chen, 2014).

Just like general Data Mining, Time Series Data Mining is composed of several different tasks. Ratanamahatana et al. (2010) and Esling and Agon (2012) cite several examples concerned with classification, clustering, as well as many other tasks, all applied to time series data. In *classification*, there may be a large number of time series in a time series database and we seek to assign each time series to a predefined class. For example, Lotte et al. (2007) describe how long electrocardiogram (ECG) time series data for several human subjects could be classified to aid brain-computer interfaces. In *clustering*, time series from a large time series database are grouped together in clusters but the clusters are not known in advance, and the computer 'learns' how to group the time series (or subsequences from the time series) together. This is an important task in DNA analysis (Kerr et al., 2008).

Three particular Time Series Data Mining functionalities are of interest here: *indexing*, *motif discovery* and *periodicity detection*.

Indexing is also known as Query by Content, and is a basic task in any database. For example, one may search for the number of customers who buy both diapers and beer just before the weekend (usually young parents). In this case, an exact match to the query may be found. With a real-valued time series data set, for example ECG data for a large group of heart patients, it is unlikely that one will find an exact match to a subsequence that might represent a heart defect; instead, one performs a *similarity search* (Han et al., 2012). Time series data may also be corrupted by noise, for instance Space Shuttle sensor data, and again the query would involve a search for a data subsequence which closely resembles, but is not necessarily identical to, some other subsequence. Being able to measure similarity (or dissimilarity) is therefore important.

Motifs are repeated patterns in time series data. Discovering motifs is of great interest in DNA analysis in bioinformatics. It is known that overrepresented DNA sequences are of biological significance (Tompa and Buhler, 2001). Caraça-Valente and López-Chavarrías (2000) investigate how patients recover under physiotherapy based on similar patterns, which are essentially motifs. To detect anomalies, for example in ECGs, non-anomalous healthy ECGs are represented using typical repeated patterns, and anomalous heart rhythms may then be detected by comparison with the healthy motifs.

Since motifs are repeated, one can immediately see that they may be relevant to the task of *detecting periodicity*. By periodicity, we mean the regular repetition of a pattern. Periodicity in time series data is ubiquitous. It occurs in electricity consumption, seasonal sales data, ECG data, and economic data; detecting and measuring periodicity can help power generation companies, retailers, medical monitoring teams, and economic forecasters respectively. Underwriting cycles in property-casualty insurance are an illustration of periodicity. Measuring this periodicity has obvious advantages in terms of risk management.

3. Representations of Time Series

In time series analysis, a time series is often represented by its correlogram or periodogram. Likewise, in Time Series Data Mining, time series can be represented in various ways. There are two reasons for doing this. First, the original time series database may be very large and it is easier to search or process a summarized version of the time series. Second, a high-level abstraction means that features such as trends, seasonalities and cycles in the data become more apparent, whilst noisy distortion is removed (Esling and Agon, 2012).

3.1. Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT)

One group of methods for representing time series in Data Mining is transformation-based methods. A classical approach is to use frequency-domain analysis. A Fast Fourier

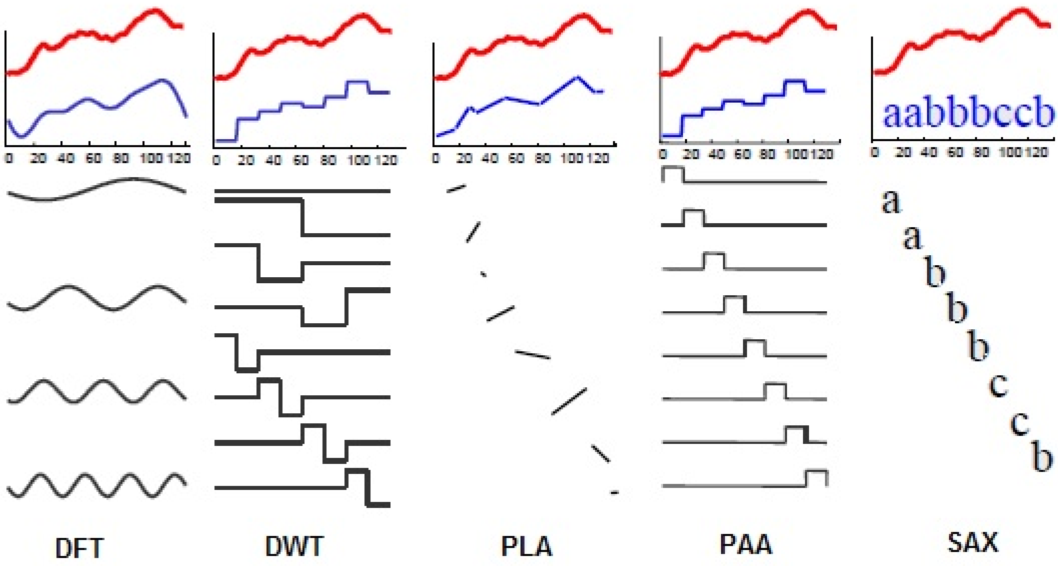


Figure 1: In the time plots at the top, an original time series (top line) and five representations of the time series (bottom) are shown. Constituent parts of the representations appear under the time plots. DFT: Discrete Fourier Transform, composed of sinusoids. DWT: Discrete Wavelet Transform, composed of Haar wavelets. PLA: Piecewise Linear Approximation, composed of linear segments. PAA: Piecewise Aggregate Approximation, composed of box functions. SAX: Symbolic Aggregate Approximation with discretization and symbolic representation. Source: Eamonn Keogh (reproduced with permission).

Transform algorithm is used to calculate the Discrete Fourier Transform (DFT) of a time series. (The function `fft()` can be used with the basic **stats** package in R for this purpose.) The DFT represents a time series by means of a superposition of sinusoids, as shown in the panel labelled DFT in Figure 1.

Another well-known method involves the Discrete Wavelet Transform (DWT) (Kantardzic, 2011). The time series is decomposed into a number of wavelets, whose shape is based on a prototype function or "mother wavelet". This is illustrated in the panel labelled DWT in Figure 1 using 8 Haar wavelets. Whereas the component sinusoids in DFT make a global contribution to the time series, the DWT has the advantage of using wavelets that are localized in time, meaning that finer details of the original time series can be captured. The **wavelets** package in R may be used to implement DWT.

3.2. Piecewise Linear Approximation (PLA) and Piecewise Aggregate Approximation (PAA)

A second group of methods for representing time series uses windowing and piecewise approximations of the original time series (Kantardzic, 2011). One such straightforward method is the Piecewise Linear Approximation (PLA) which simply replaces the time series by a sequence of linear segments. This is depicted in the panel labelled PLA in Figure 1 using 8 linear segments. This method removes noise and reduces the dimensionality of the data, but it is not clear how to determine the number of linear segments to use. The squared error between the PLA and the original data may be minimized but scaling and time axis distortions remain (Kantardzic, 2011).

An improved method is the Piecewise Aggregate Approximation (PAA) (Keogh et al., 2001; Yi and Faloutsos, 2000). PAA can be implemented using function `PAA()` in R package **TSclust**. The original time series is split into several same-length subsequences, and the mean value of each of these subsequences is then used as the height of a box function approximating the segment (Ratanamahatana et al., 2010). This is shown in the panel labelled PAA in Figure 1 using 8 subsequences and therefore 8 box functions.

3.3. Symbolic Aggregate Approximation (SAX)

A third method, known as Symbolic Aggregate Approximation (SAX), has been developed by Lin et al. (2003). It supplies a symbolic representation of a real-valued time series, as pictured in the panel labelled SAX in Figure 1. The SAX conversion can be achieved using the function `convert.to.SAX.symbol()` in the **TSclust** package in R. Matlab code for SAX is also made available by Lin (2016).

The procedure for SAX is to z-normalize the original time series, and then discretize it as in the PAA method. The heights of the box functions thus created have a standardized Normal distribution, the area under which is then split into a number of equi-probable regions, each of which is assigned an alphabetical symbol. If the box function height lies in the nth part of the distribution, it is assigned the nth symbol. The time series is therefore converted into a symbol string where every symbol is equally likely to appear. This is illustrated in Figure 2. Further details about SAX are given by Lin et al. (2003).

SAX removes noise but retains general features of the data such as cycles. Despite its simplicity, it enjoys significant success in Time Series Data Mining tasks such as clustering and classification (Ratanamahatana et al., 2010). SAX is a crucial part of one of the periodicity detection algorithms that we use on underwriting cycles in a subsequent section.

4. Dissimilarity Measures

4.1. Uses of Dissimilarity Measures

Dissimilarity or distance measures are essential to many Time Series Data Mining tasks. Just like a cross-correlation tells us about the relationship between two time series in time series analysis, the dissimilarity measure tells us how distant or different one time series is from another. Dissimilarity measures are sometimes used explicitly in time series data mining tasks such as clustering, and are often used implicitly within algorithms for many other tasks.

For instance, we might wish to perform clustering of electroencephalogram (EEG) time series data for multiple individuals to group different types of brain activity. The

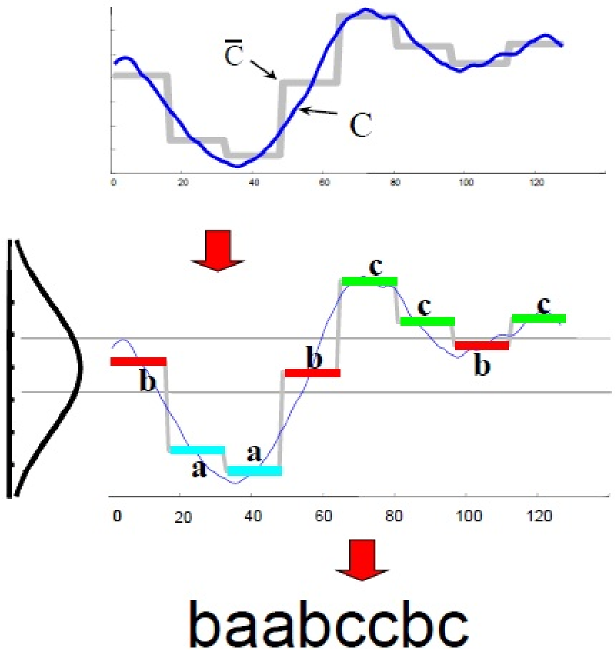


Figure 2: Symbolic Aggregate Approximation (SAX). An original real-valued time series C (top) is discretized using the Piecewise Aggregate Approximation (PAA) algorithm resulting in $\overline{C}$. The discrete values are assigned a symbol such that every symbol is equally likely to occur according to the standard Normal distribution. C is converted to a symbol string `baabccbc`. Source: Eamonn Keogh (reproduced with permission).

computer should 'learn' to cluster or group the EEGs by itself, by combining the EEGs that are similar in the same cluster. A measure of dissimilarity or distance is therefore required.

Dissimilarity measures are also used in periodicity detection. Several dissimilarity measures exist but we shall discuss only a few that are are relevant to our discussion of periodicity measurement on underwriting cycles.

Consider two time series $X = \{x_t : t \in T, x_t \in S\}$ and $Y = \{y_t : t \in T, y_t \in S\}$. Without loss of generality, the discrete time set T can be assumed to be $\{0, 1, \ldots, n-1\}$. If the series are real-valued, then the state space $S = \mathbb{R}$. If the series are symbolic sequences, then S represents an alphabet of symbols.

4.2. Minkowski and Euclidean Distances

Metrics in normed vector spaces can be used as dissimilarity or distance functions for numeric time series. The general Minkowski distance based on the L_p norm,

$$\left(\sum_{t \in T} |x_t - y_t|^p \right)^{1/p} \tag{1}$$

may be evaluated with the `minkowskiDistance()` function in R's **TSdist** package, and is most commonly used in its Euclidean ($p = 2$) or Manhattan ($p = 1$) form. Functions `euclideanDistance()` and `diss.EUCL()` may be used in packages **TSdist** and **TSclust** respectively to output the Euclidean distance directly.

These metrics are easy to calculate and are parameter-free but suffer from the disadvantage of not being robust. Dissimilarity measures should be robust to noise, outliers, scaling (amplitude variations) and warping (temporal variations) in the two time series being compared (Esling and Agon, 2012). L_p-based distances do not satisfy these requirements. The time series can be z-normalized first to cope with amplitude variations (in the vertical axis), but since these measures operate in lock-step (i.e. all corresponding time points are compared in a one-to-one fashion), they suffer from being too rigid (Ratanamahatana et al., 2010). For example, underwriting cycles may start later or earlier at certain points during some decades, but these metrics do not allow for any local time-shifting.

4.3. Hamming Distance

If X and Y are symbolic sequences (e.g. $S = \{a, b, c, d\}$), then a Hamming distance may be used: $\sum_{t \in T} \mathbb{I}_{x_t \neq y_t}$, where $\mathbb{I}_A$ is an indicator function taking value 1 if A is true and 0 otherwise. For example, the Hamming distance between *babdb* and *aaddd* is 2. The Hamming distance is also an inflexible, lock-step dissimilarity measure.

4.4. Dynamic Time Warping (DTW)

DTW is in widespread use in speech recognition and was introduced by Berndt and Clifford (1994, 1996) to the Artificial Intelligence community. Unlike the Hamming distance and the Minkowski distance, DTW is intended to be resilient to noisy distortions, such as may occur when speech is recorded, digitized and transferred over networks. DTW can be computed in R using packages **TSdist**, **TSclust** and **dtw**.

Consider the two time series X and Y, defined in section 4.1, of finite length n. The DTW algorithm starts by calculating a dissimilarity matrix. This is a symmetric matrix containing a dissimilarity value for every pair of data points in X and Y:

$$\begin{pmatrix} d(0,0) & d(0,1) & d(0,2) & \dots & d(0,n-1) \\ d(1,0) & d(1,1) & d(1,2) & \dots & d(1,n-1) \\ d(2,0) & d(2,1) & d(2,2) & \dots & d(2,n-1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d(n-1,0) & d(n-1,1) & d(n-2,2) & \dots & d(n-1,n-1) \end{pmatrix}$$

where $d(i,j)$ represents a distance such as in the Euclidean metric, in which case $d(i,j) = |x_i - y_j|^2$, or as in the Hamming distance in which case $d(i,j) = \mathbb{I}_{x_i \neq y_j}$.

Whereas the Euclidean and Hamming distances add the terms in the leading diagonal, DTW seeks to warp the time axis locally so as achieve an optimal alignment of the time series X and Y. It does this by adding matrix elements starting from cell $(0,0)$ and ending at cell $(n-1, n-1)$ but proceeding along a warping path which may deviate from the leading diagonal. A warping path is a contiguous path from cell $(0,0)$ to $(n-1, n-1)$. The minimum warping path associated with the smallest cumulative distance from cell $(0,0)$ to $(n-1, n-1)$ may be found by dynamic programming, just like a shortest path

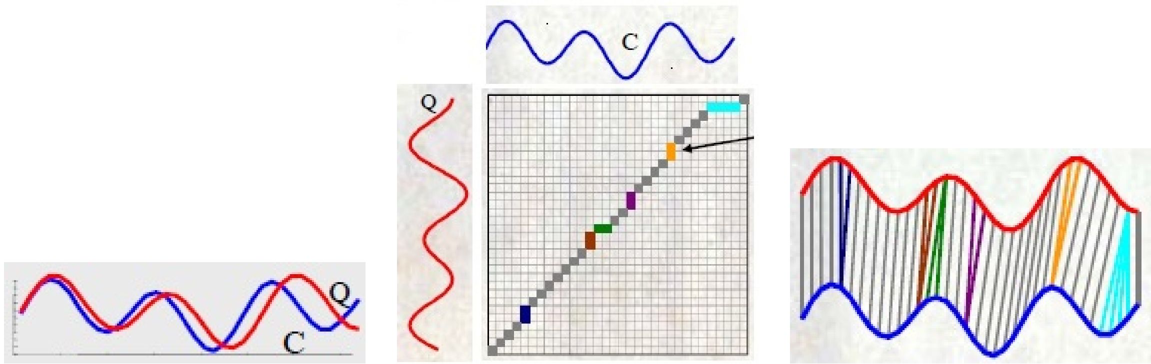


Figure 3: Dynamic Time Warping (DTW). Left panel: two time series C and Q are similar but misaligned. Middle panel: a warping path that minimizes the distance between C and Q is highlighted with an arrow. Right panel: the effect is to warp the time axis to align C and Q more closely. Source: Eamonn Keogh (reproduced with permission).

problem. There are several strategies used to speed up this optimization, most notably the Keogh (2002) lower bounding technique.

The overall effect is akin to warping or distorting the time axis of both time series to align them more closely before calculating the distance using, for example, the Euclidean distance. This is illustrated in Figure 3. If underwriting cycles do not have a precisely regular cyclical structure, then the warping effect achieved in DTW can compensate and can uncover periodicity.

The DTW dissimilarity can be calculated directly in R using functions `dtwDistance()` and `diss.DTWARP()` in packages **TSdist** and **TSclust** respectively. The dissimilarity matrix and warping path may also be computed using the functions `dtwDist()` and `dtw()` in package **dtw** of R. See Giorgino (2009) for more details about DTW.

5. Algorithms for Periodicity Detection

5.1. Periodicity Detection in Data Mining

Data scientists use a number of tools when investigating periodicity in a time series database. They use classical time series analysis (autocorrelations and partial autocorrelations) and frequency domain analysis (spectral density estimates). They also use algorithms that are based on time series representations such as SAX (section 3) and dissimilarity measures such as DTW (section 4). In this section, we consider two popular and successful periodicity detection algorithms, WARP and MBPD.

5.2. Warping for Periodicity (WARP)

The WARP algorithm was developed by Elfeky et al. (2005) with the purpose of being resilient to noise. It utilizes the DTW dissimilarity measure (see section 4). The key idea in WARP is to measure the DTW dissimilarity between a time series and a locally time-shifted version of the time series. In other words, we measure the distance between the data and a lagged, time-warped version of itself to cater for any noisy distortion. If the DTW distance is locally minimized at lags of τ, 2τ, 3τ etc. (measured in relevant time units), then this suggests a period of τ.

Because of noise and other distortions, the local minima in DTW are unlikely to occur at lags which are exact multiples of the true period (if it exists). An averaging scheme may then be used. For example, if the local minima occur at lags of τ_1, τ_2 and τ_3, then the period is estimated to be $\frac{1}{3}\left(\tau_1 + \frac{1}{2}\tau_2 + \frac{1}{3}\tau_3\right)$.

WARP is easily implemented in R using a shifting function and the DTW dissimilarity function `dtwDistance()` in the **TSdist** package.

WARP is potentially a useful tool for discovering and measuring periodicity on insurance underwriting cycles because of its resilience to noise and its ability to stretch or compress the time axis. It is possible that insurance markets exhibit periodicity but that hard and soft market periods are sometimes longer and sometimes shorter than they are at other times. Classical time series and frequency-domain methods look for cycles of a precise length, whereas periodicity may be more fluid.

5.3. Motif-Based Periodicity Detection (MBPD)

The Motif-Based Periodicity Detection (MBPD) scheme was designed by Otunba, Lin and Senin (2014), based on a motif discovery algorithm, GrammarViz, by Li, Lin and Oates (2012). In turn, GrammarViz employs SAX (see section 3.3) as well as an Artificial Intelligence tool, Sequitur, developed by Nevill-Manning and Witten (1997). Java code

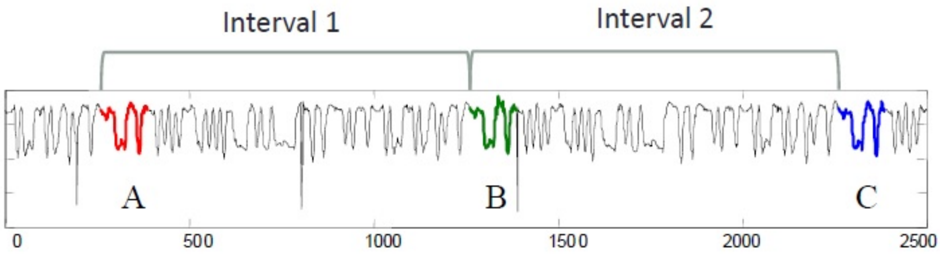


Figure 4: A motif or repeated subsequence which recurs three times, in positions A, B and C, in time series data. Source: Eamonn Keogh (reproduced with permission).

for GrammarViz is made available by Senin (2016). The code for Sequitur is also freely available (Nevill-Manning and Witten, 2016).

MBPD is based on motif discovery, a key task in Data Mining as discussed earlier in section 2. Motifs are repeated patterns in time series data. More precisely, they are subsequences which are very similar, according to some threshold value of a dissimilarity measure, and which recur in the time series (Li, Lin and Oates, 2012). Figure 4 shows three occurrences of a motif. In insurance loss ratio data, a motif could be a crest or trough indicating a soft or hard insurance market.

Suppose that the motif discovery algorithm discovers n motifs in the time series data. For the ith motif $(i = 1, \ldots, n)$, we record the intervals between each occurrence of the motif, $\{\tau_i(1), \tau_i(2), \ldots\}$. Two such inter-motif intervals are shown in Figure 4 for a motif, and we require that there be at least three occurrences of a repeated subsequence for it to be regarded as a significant motif. We then calculate the mean $\overline{\tau}_i$ of the intervals between the occurrences of the ith motif, as well as their standard deviation s_i. MBPD posits that:

$$\text{estimate of period} \ = \ \overline{\tau}_j, \quad \text{where } j = \arg\inf_i s_i. \tag{2}$$

In other words, all motifs are discovered, we then identify the motif which repeats itself with the most regularity, and take the average inter-motif interval for this motif as the approximate period of the time series data.

This is a simple enough scheme, but it relies on the motif discovery algorithm, GrammarViz, which itself relies on the AI software, Sequitur. We sketch how these algorithms work here and refer the reader to Li, Lin and Oates (2012) and Nevill-Manning and Witten (1997, 2016) for more details.

GrammarViz first converts a real-valued time series into a symbolic string using SAX (see section 3). It then calls on Sequitur for grammar induction on this string.

In Computer Science and Linguistics, a grammar is a set of rules that describe how all possible strings can be constructed from an alphabet. (An alphabet is simply a set of allowable symbols, and a string is a sequence of these symbols.) Grammars are therefore compact generative representations of strings. Conversely, a grammar can be viewed as a compressed version of a string because, given the grammar rules, one can reconstitute the string.

For example, consider the symbolic sequence $X = abcdabcdeab$, which may be the SAX representation of a real-valued time series. Sequitur performs grammar induction and converts this into three grammar rules (Li, Lin and Oates, 2012):

$$X \to BBeA \qquad A \to ab \qquad B \to Acd \tag{3}$$

Sequitur works in an online fashion and builds the rules as it goes through string X one symbol at a time until it reaches the last symbol. Repeated subsequences (for example, ab) are merged and assigned a new symbol (for example, $A \to ab$) as Sequitur progresses through the string. The usefulness of grammar induction to periodicity detection is immediately clear: a compact summarization of data takes place where recurring patterns are detected and identified. The recurring pattern in X is seen, from the repetition of B in the first grammar rule, to be $abcd$.

GrammarViz maps all grammar rules back on to the original real-valued time series data. Each rule maps to a motif in the data. Rules can be ranked by their length and frequency, and may be visualized in a time plot of the time series. Li, Lin and Oates (2012) describe this in greater detail. For each rule or motif i of interest, the mean $\overline{\tau}_i$ and standard deviation s_i of the intervals between each occurrence of the motif can be computed, as described earlier. The period of the time series data is then estimated as in equation (2).

Otunba, Lin and Senin (2014) test MBPD on several synthetic pseudo-randomly simulated time series and demonstrate that MBPD detects and estimates the period more accurately than using peaks in spectral density estimates obtained by a Fast Fourier Transform. They also test MBPD on real public data sets, including sunspot data and electrocardiograms (ECG). These data sets have been analyzed at length in the literature using a variety of statistical methods, including classical time series methods. Otunba, Lin and Senin (2014) show that MBPD estimates a period close to the consensus estimates from these other methods.

6. Application to Underwriting Cycles

6.1. Data

We calculate the loss ratio, which is the ratio of losses plus loss adjustment expenses to premium earned, from data collated by A.M. Best Co. on U.S. property-casualty insurance for every year over nearly six decades from the early 1950s. We create time series of loss ratios, one for all lines combined, and one for each of automobile, fire, homeowners (multiple peril) and commercial (multiple peril) insurance. The all-lines time series starts in 1951, homeowners insurance starts in 1955, and the others start in 1954. In years prior to 1982, the data set is separated by mutual and stock companies, so we merge them.

6.2. Time Series Analysis over Nearly 6 Decades of Data

Figure 5 shows, in graphical form, the loss ratios for all lines combined of U.S. property-casualty insurance over the period 1951–2011, their spectrum, autocorrelations and partial autocorrelations. From the time plot in the top left panel of Figure 5, we observe that the loss ratios trend upwards: insurance markets have become increasingly competitive over time. The loss ratios also have an appearance of cyclicality, with successive crests and troughs in evidence, marking soft and hard markets. When the data is de-trended by means of linear regression, the periodogram (top right panel of Figure 5) slopes downwards with several local peaks but these cannot be distinguished from noise

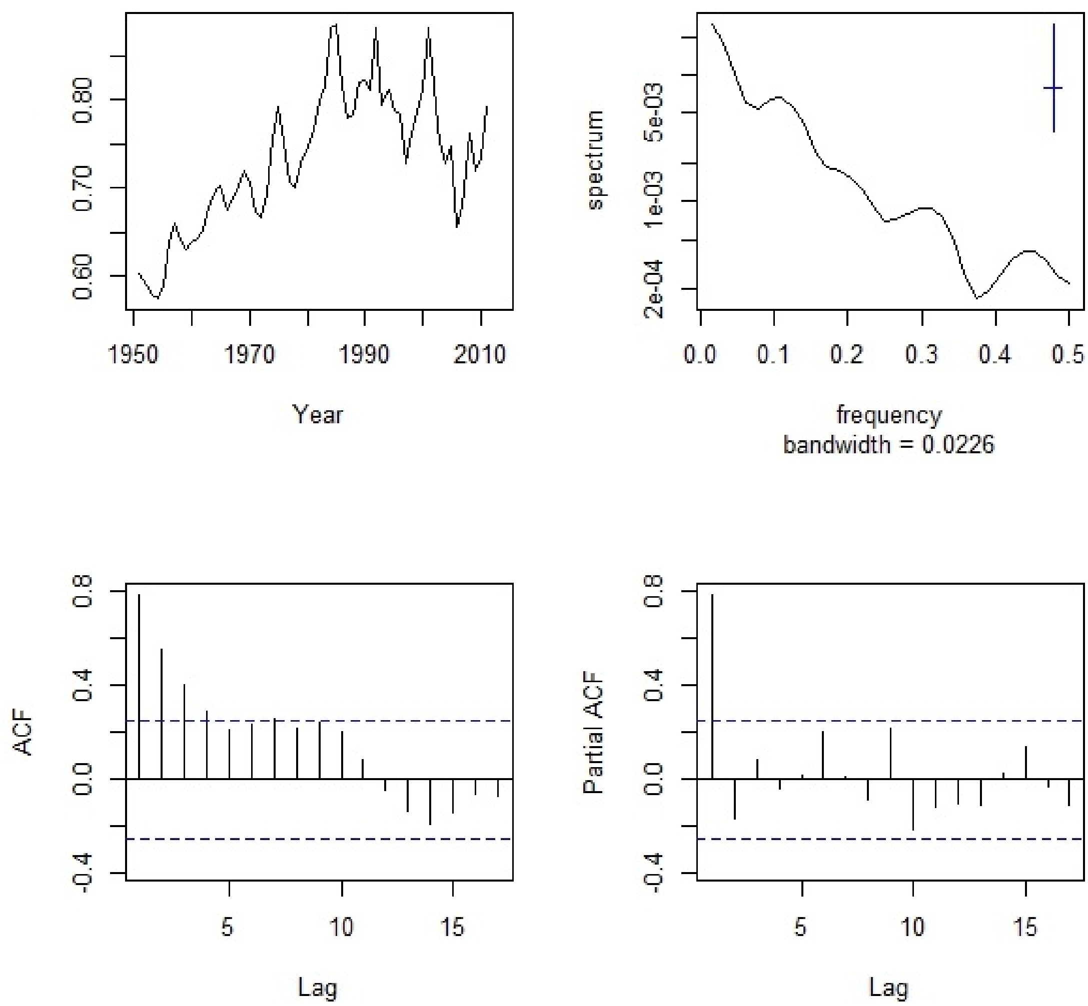


Figure 5: Loss ratios for U.S. property-casualty insurance, all lines combined, 1951–2011. Top left: time plot. Top right, bottom left and bottom right: periodogram, correlogram and partial correlogram of de-trended loss ratios.

in the data.[3] The correlogram (bottom left panel of Figure 5) shows decaying autocorrelations, not significantly different from zero at large lags. The decaying correlogram and barely significant partial autocorrelations after lag 0 (bottom right panel of Figure 5) suggest a parsimonious AR(p) model with small p.

Before any time series estimation is undertaken, the trend-stationarity or difference-stationarity of the data must be investigated. Harrington and Yu (2003) provide a comprehensive study of the trend-stationarity of loss, expense and combined ratios in the U.S., and this is updated by Harrington, Niehaus and Yu (2014). Based on this, we fit a succession of ARMA(p, q) models to the time series of loss ratios:

$$\left(1 - \sum_{j=0}^{p} \phi_j B^j\right)(X_t - \mu) \;=\; \phi_0 t \;+\; \left(1 + \sum_{j=0}^{q} \theta_j B^j\right)\epsilon_t, \tag{4}$$

where $\epsilon_t \sim$ i.i.d. Normal random variables and B is the backward shift operator. In equation (4), we include the time trend as a regressor following Harrington, Niehaus and Yu (2014) and others, so that the trend is estimated simultaneously with the ARMA coefficients.

Table 1 lists the bias-corrected Akaike Information Criterion (AICc) (Brockwell and Davis, 2006, p. 301) for the various ARMA models when fitted to the data. According to the AICc, the AR(2) model provides a good parsimonious fit to the data, but it is bettered by both the AR(1) and ARMA(1,1). (Similar results are obtained with the Akaike and Bayes Information Criteria.)

In most of the insurance cycle literature (e.g. Trufin, Albrecher and Denuit, 2009; Lamm-Tennant and Weiss, 1997), cycles are identified and estimated using an AR(2) model. It is well known that, provided $\phi_1^2 + 4\phi_2 < 0$, complex roots occur in the characteristic equation of the AR(2) process, and the cycle period τ may then be estimated using (Hamilton, 1994; Sargent, 1987):

$$\tau \;=\; \frac{2\pi}{\arccos(\phi_1/2\sqrt{-\phi_2})}. \tag{5}$$

[3]The periodogram is smoothed with suitable Daniell filter weights to reduce the confidence interval on the spectral density estimate.

	$p=0$	$p=1$	$p=2$	$p=3$
$q=0$	-167.67	**-223.56**	**-222.42**	-220.50
$q=1$	-206.06	**-222.75**	-220.52	-218.07
$q=2$	-217.23	-220.52	-218.26	-222.14
$q=3$	-216.11	-218.27	-215.71	-219.43

Table 1: Values of bias-corrected Akaike Information Criterion (AICc) for various ARMA models fitted to loss ratios for U.S. property-casualty insurance, all lines combined, 1951–2011. Highlighted: the AR(1), AR(1,1) and AR(2) models give the best fit in descending order.

When we fit an AR(2) model to the loss ratios for all lines combined of U.S. property-casualty insurance, we find that $\widehat{\phi}_1 = 0.8928$ (0.1333), $\widehat{\phi}_2 = -0.1442$ (0.1627) (standard errors are in parentheses). The model may be validated by considering the residuals, shown in Figure 6: the residuals are serially uncorrelated with large Ljung-Box p-values. If we accept the AR(2) model, then we are immediately confronted with the fact that these AR(2) coefficients are such that $\widehat{\phi}_1^2 + 4\widehat{\phi}_2 = 0.2203 > 0$, which violates the criterion for AR(2) cycles. In other words, we cannot identify underwriting cycles of an autoregressive nature in the all-lines data.

We repeat this analysis on each of the automobile, fire, homeowners and commercial lines. Figure 7 contains time plots of loss ratios on these lines, displayed together with all-lines loss ratios for comparison. The high peak in the homeowners insurance loss ratio in 1992 is the result of Hurricane Andrew. We find again that the AR(2) model does not provide the best fit to the data, although the AICc ranks it in the top 3 models. The parameter and cycle period estimates, when an AR(2) model is parameterized using the data, are listed in Table 2. Sensible estimates of the cycle period are obtained for the automobile and fire lines. However, for homeowners insurance, no cycle is identified,[4] and a spurious period of 54 years is obtained for the commercial insurance line ($\widehat{\phi}_1^2 + 4\widehat{\phi}_2 =$

[4]We also controlled for the outlier due to Hurricane Andrew by creating a dummy variable for 1992 in equation (4). The resulting AR(2) model then estimated a suspiciously long cycle of 21.4 years in the homeowners line.

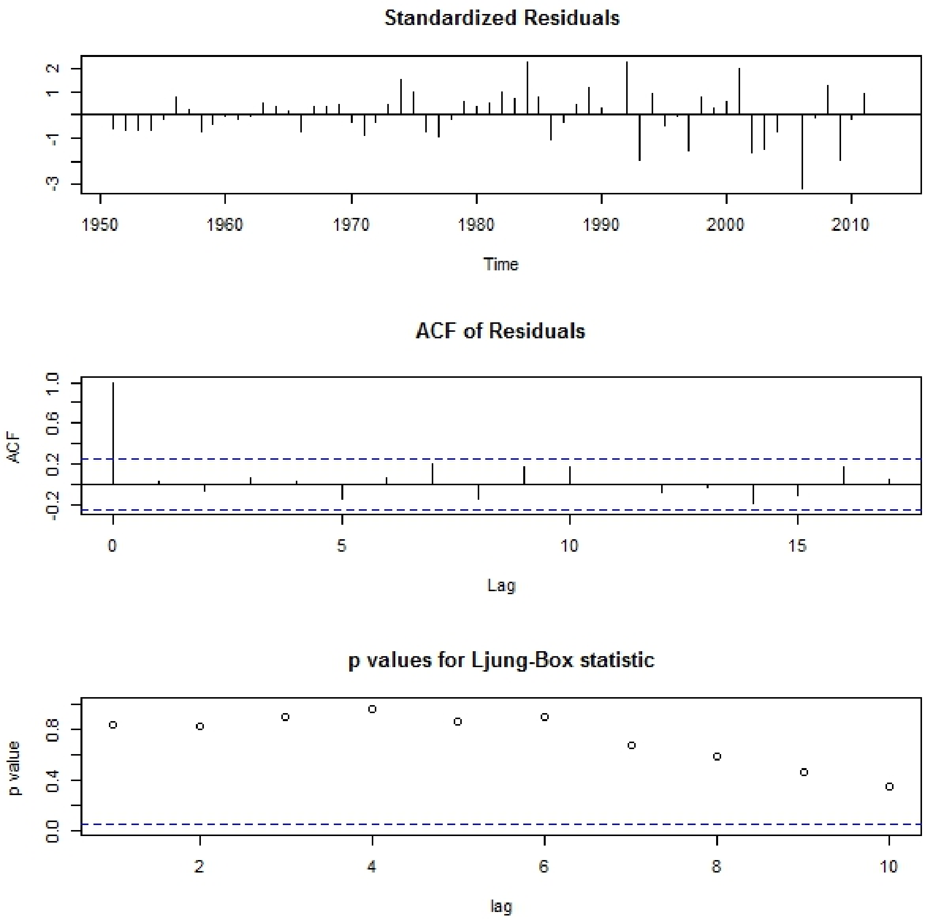


Figure 6: Residuals from the AR(2) model fitted to loss ratios for U.S. property-casualty insurance, all lines combined, 1951–2011.

	$\widehat{\phi}_1$	$\widehat{\phi}_2$	$\widehat{\mu}$	$\widehat{\phi}_0$	$\widehat{\tau}$
All lines	0.8928	-0.1442	-5.3387	0.0031	n/a
1951–2011	(0.1333)	(0.1627)	(1.8118)	(0.0009)	
Automobile	1.1933	-0.4641	-2.8469	0.0018	12.5
1954–2011	(0.1172)	(0.0559)	(1.5657)	(0.0008)	
Fire	0.6896	-0.1741	-1.9588	0.0013	10.5
1954–2011	(0.1381)	(0.1934)	(1.8189)	(0.0009)	
Homeowners	0.3689	0.1022	-8.0819	0.0044	n/a
1955–2011	(0.1326)	(0.1328)	(2.6320)	(0.0013)	
Commercial	0.8943	-0.2026	-4.0305	0.0024	54
1954–2011	(0.1312)	(0.1465)	(3.2365)	(0.0016)	

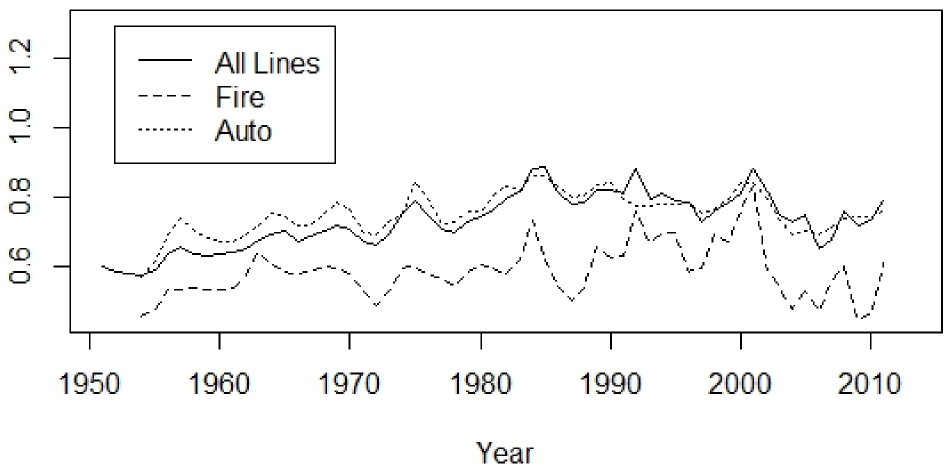
Table 2: Estimates for the AR(2) model fitted to nearly 6 decades of loss ratios for U.S. property-casualty insurance for all lines combined and for four specific lines. (Standard errors are in parentheses.) $\widehat{\tau}$ is the estimated cycle period in years and n/a means that no AR(2) cycle is identified.

$-0.0106 < 0$, so the cyclicality criterion holds but only just).

6.3. Time Series Analysis over 3 Decades of Data

The argument can be made that insurance markets have changed considerably since the early 1950s, both in terms of product and regulation. Consequently, the length of underwriting cycles will change over time, and an attempt to model insurance losses over such a long period is bound to be inconsistent. In their recent survey, Harrington, Niehaus and Yu (2014) estimate cycle periods in seven 25-year overlapping sub-periods, beginning in the 1950s, to be between 4.4 and 6.1 years (for all lines combined). However, these shorter sub-periods capture only 4–5 full cycles (assuming that cycles are indeed present), so the confidence interval on the period estimate is very large. This point is made by Boyer et al. (2012) and Boyer and Owadally (2015), particularly in view of the nonlinearity relating τ to the AR(2) coefficients in equation (5).

We repeat the earlier analysis but for a 30-year sub-period, 1982–2011. Just as for the de-trended all-lines data, the spectra exhibit no significant peak, suggesting that there is



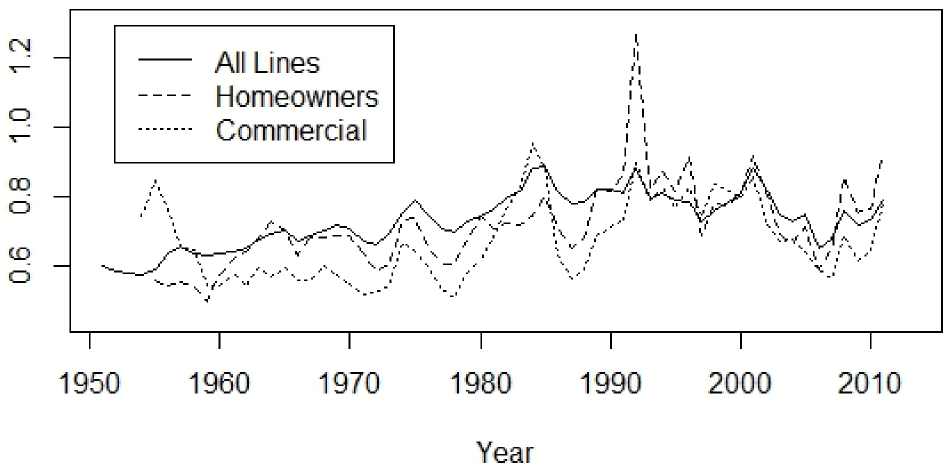


Figure 7: Time plots of loss ratios for U.S. property-casualty insurance, 1951–2011.

	$\widehat{\phi}_1$	$\widehat{\phi}_2$	$\widehat{\mu}$	$\widehat{\phi}_0$	$\widehat{\tau}$
All lines	0.4870	-0.2491	8.211	-0.0037	5.9
1982–2011	(0.1744)	(0.1809)	(2.2634)	(0.0011)	
Automobile	1.0802	-0.6300	9.0571	0.0041	7.7
1982–2011	(0.1347)	(0.1308)	(0.4648)	(0.0008)	
Fire	0.5810	-0.1787	7.8661	-0.0036	7.7
1982–2011	(0.1958)	(0.1962)	(6.5018)	(0.0032)	
Homeowners	0.2616	0.1055	0.1751	0.0003	n/a
1982–2011	(0.1902)	(0.1829)	(9.6359)	(0.0050)	
Commercial	0.6946	-0.1999	6.9532	-0.0031	9.2
1982–2011	(0.2080)	(0.1888)	(11.7350)	(0.0062)	

Table 3: Estimates for the AR(2) model fitted to 3 decades of loss ratios for U.S. property-casualty insurance (1982–2011) for all lines combined and for four specific lines. (Standard errors are in parentheses.) $\widehat{\tau}$ is the estimated cycle period in years and n/a means that no AR(2) cycle is identified.

no cycle, and correlograms tail off, suggesting stationarity. According to the AICc, the AR(2) model is the best model for automobile insurance loss ratios, but ranks as only the second or third best model for the loss ratios on all lines combined and on the other three lines. Parameter and cycle period estimates appear in Table 3.[5]

Standard errors on parameter estimates in Table 3 are higher than in Table 2, reflecting the smaller sample size. There is greater uncertainty on the period estimates when loss ratio data are restricted to the shorter sub-period of 1982–2011. Boyer et al. (2012) and Boyer and Owadally (2015) show that, for shorter data series and larger parameter standard errors, there is a larger probability that the cyclicality criterion $\phi_1^2 + 4\phi_2 < 0$ does not hold, and that the presence of autoregressive cycles is not statistically significant.

[5]No cycle is detected on the homeowners insurance line, irrespective of whether a dummy variable is used in 1992 to control for the outlier represented by Hurricane Andrew.

6.4. Time Series Data Mining: WARP

As demonstrated above, classical time series analysis is not very conclusive about the presence of underwriting cycles on insurance markets, and the AR(2) model is not necessarily the best fitting model to loss ratios. When older data points are removed, cyclical AR(2) models can be fitted, but estimates of cycle lengths suffer from a large confidence interval. Time series methods are also not robust to outliers. Can Time Series Data Mining fare better?

The WARP algorithm (section 5.2) was run on the loss ratios for all lines combined and for each of the four lines described in section 6.1. The results from WARP were inconclusive and it could not detect any periodicity, either over the full length of data (nearly 6 decades), or over the last 3 decades. No periodicity was detected when de-trended loss ratios were input to the WARP algorithm.

Recall that WARP compares the time series with a time-shifted version of itself, with time accelerated locally to minimize dissimilarity. This time-elasticity feature should have enabled WARP to match the recurring peaks and valleys of insurance profitability, evident in Figure 7, and therefore spot and measure periodicity. It seems, however, that the noisiness of the data overwhelmed the algorithm.

6.5. Time Series Data Mining: MBPD

The MBPD algorithm works by detecting *motifs* (repeated patterns) in the data (section 5.3). In the insurance loss ratio data depicted in Figure 7, this is likely to be crests and troughs when soft and hard markets occur. MBPD was able to detect periodicity in the full-length data, on the all-lines loss ratios as well as on the four individual lines. The periods that it estimated are listed in Tables 4 and 5, which also summarize the results from the other methods that were used. When it detected a period of n years, it also tended to highlight periods of approximately multiples of n years, which is perhaps reassuring.

A number of key points are worth highlighting here about the results with MBPD.

1. MBPD detects periodicity on all the series in Tables 4 and 5, even when the AR

	AR	WARP	MBPD
All lines 1951–2011	n/a	n/a	8.5
Automobile 1954–2011	12.5	n/a	8.0
Fire 1954–2011	10.5	n/a	8.0
Homeowners 1955–2011	n/a	n/a	13.0
Commercial 1954–2011	54	n/a	9.0

Table 4: Estimates of cycle periods of underwriting cycles for U.S. property-casualty insurance over nearly 6 decades using three different methods.

	AR	WARP	MBPD
All lines 1982–2011	5.9	n/a	9.5
Automobile 1982–2011	7.7	n/a	7.5
Fire 1982–2011	7.7	n/a	7.7
Homeowners 1982–2011	n/a	n/a	6.0
Commercial 1982–2011	9.2	n/a	16.0

Table 5: Estimates of cycle periods of underwriting cycles for U.S. property-casualty insurance over 3 decades to 2011 using three different methods.

(autoregressive) method fails to detect cycles.

2. The cycle period estimates from MBPD are comparable to those cited in earlier time series studies: see for example Boyer and Owadally (2015), Boyer et al. (2012), Lamm-Tennant and Weiss (1997) and Cummins and Outreville (1987). The cycle periods from MBPD for the auto and fire insurance lines in Table 5 are similar to the ones from the 30-year autoregressive modeling.

3. MBPD detects periodicity in the homeowners insurance line, despite the presence of an outlier caused by Hurricane Andrew in 1992. The algorithm therefore appears to be robust to outliers in financial and economic data.

4. MBPD detects and measures periodicity without any de-trending being applied to the data. (It obtained almost the same results when de-trended time series were passed to it.) The literature on business cycles shows that the de-trending method or filter that is used when pre-processing data has an effect on the subsequent modeling and on measurement of cycles (Nelson and Plosser, 1982; Rudebusch, 1993). MBPD can potentially bypass such problems.

5. The fact that MBPD detects periodicity in insurance loss ratios suggests that underwriting cycles are not necessarily cycles in a Fourier analysis sense at all, and that they may not be predictable in a classical econometric framework, as argued by Boyer et al. (2012) and Boyer and Owadally (2015). The underwriting cycles that are reported by insurance industry professionals, are better described using the concept of periodicity employed in Time Series Data Mining.

6. MBPD has the Artificial Intelligence algorithm Sequitur at its heart. This helps it detect repeated patterns in a way that is much more similar (albeit not identical) to the way that human beings recognise patterns and perceive the world. The insurance market is made up of actuaries, underwriters, loss adjusters, managers, and their perception of underwriting cycles is very real. MBPD is able to capture their reality of underwriting cycles.

7. Conclusion

Underwriting cycles are poorly understood but have significant effects on the property-casualty insurance industry. Classical statistical methods fail either to detect cycles or to estimate their lengths with reasonable confidence. We propose Time Series Data Mining algorithms as an alternative. Various applications of Time Series Data Mining are described, particularly classification, clustering, indexing, motif discovery and periodicity detection. These tasks are unified in their use of time series representations, dissimilarity measures and search techniques. Different representations of time series are described, particularly using the Discrete Fourier and Wavelet Transforms (DFT, DWT), as well as methods using windowing and piecewise approximations such as Symbolic Aggregate Approximation (SAX).

Dissimilarity measures, including particularly Dynamic Time Warping (DTW), are described. DTW was developed in the field of Artificial Intelligence and is used extensively in speech recognition. DTW is central to the WARP periodicity detection algorithm. MBPD is the second periodicity detection algorithm that we consider. It is based on motif discovery software, GrammarViz, which itself depends on SAX and on another Artificial Intelligence tool, Sequitur, for grammar induction. Throughout, we reference software tools in R, Matlab and Java that may be used to implement these algorithms.

We use industry-wide loss ratios for U.S. property-casualty insurance, on all lines combined as well as on four specific lines (automobile, fire, homeowners and commercial). We demonstrate that time series methods are inconclusive. If the time series data set is too long and the underlying data generating process is not time-homogeneous, or if the time series has outliers, time series methods may fail to detect cycles. If the time series data set is too short, the confidence interval in the estimate of autoregressive parameters widens to the point that cyclicality may not be statistically significant.

The first periodicity detection scheme, WARP, does not detect any periodicity. Its keynote feature of stretching and compressing the time series in the time axis cannot compensate for the noisiness of the data and its lack of a regular cyclical structure. MBPD is much more successful. The estimates that it furnishes are in line with previous studies, whether based on the full 6 decades length of the data or the last 3 decades. MBPD appears to be robust to the presence of an outlier and does not require de-trending. The pattern recognition that MBPD employs bears similarity to human pattern recognition, and the underwriting cycles described by insurance professionals are therefore detectable and measurable.

References

Aghabozorgi, S. and Ying Wah, T. (2014). Stock market co-movement assessment using a three-phase clustering method. *Expert Systems with Applications*, 41, 4, 1301–1314.

Aghabozorgi, S., Shirkhorshidi, A.S, and Ying Wah, T. (2015). Time-series clustering—a decade review. *Information Systems*, 53, 16–38.

Berndt, D.J. and Clifford, J. (1994). Using dynamic time warping to find patterns in time series. Proceedings of the AAAI-94 Workshop on Knowledge Discovery in Databases (KDD 1994), 359–370.

Berndt, D.J. and Clifford, J. (1996). Finding patterns in time series: a dynamic programming approach, in: Fayyad, U.M., Pialetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (Eds.), *Advances in Knowledge Discovery and Data Mining*. American Association for Artificial Intelligence, Menlo Park, CA, pp. 229–248.

Boyer, M.M., and Owadally, I. (2015). Underwriting apophenia and cryptids: are cycles statistical figments of our imagination? *Geneva Papers on Risk and Insurance Issues and Practice*, 40, 2, 232–255.

Boyer, M.M., Jacquier, E. and Van Norden, S. (2012). Are underwriting cycles real and forecastable? *Journal of Risk and Insurance*, 79, 4, 995–1015.

Brockwell, P.J. and Davis, R.A. (2006). *Time Series: Theory and Methods*, 2nd ed. Springer Series in Statistics. Springer, New York.

Caraça-Valente, J.P. and López-Chavarrías, I. (2000). Discovering similar patterns in time series. Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2000), 497–505

CAS (2013). Public access DFA model (DYNAMO). Casualty Actuarial Society. URL: `http://www.casact.org/research/index.cfm?fa=dynamo`

Chakrabarti, S. (ed.). (2009). *Data Mining: Know It All.* Elsevier/Morgan Kaufmann Publishers, Burlington, MA.

Cummins, J.D. and Outreville, J.F. (1987). An international analysis of underwriting cycles in property-liability insurance. *Journal of Risk and Insurance*, 54, 246–262.

D'Arch, S., Gorvett, R., Herbers, J., Hettinger, T., Lehmann, S. and Miller, M. (1997). Building a public access PC-based Dynamic Financial Analysis model. *Casualty Actuarial Society Forum.*

D'Arch, S., Gorvett, R., Hettinger, T., and Walling, R. (1998). Using the public access Dynamic Financial Analysis model: a case study. *Casualty Actuarial Society Forum.*

Elfeky, M.G., Aref, W.G. and Elmagarmid, A.K (2005). WARP: time warping for periodicity detection, in: Fifth IEEE International Conference on Data Mining (ICDM 2005), IEEE.

Esling, P., and Agon, C. (2012). Time-series data mining. *ACM Computing Surveys*, 45(1), 1–34.

Felblum, S. (2001). Underwriting cycles and business strategies. *Proceedings of the Casualty Actuarial Society*, 38, 175–235.

Giorgino, T. (2009). Computing and visualizing Dynamic Time Warping alignments in R: the dtw package. *Journal of Statistical Software*, 31, 7, 1–24.

Guan, H. and Jiang, Q. (2007). Cluster financial time series for portfolio. Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition, 851–856.

Guo, C., Jia, H. and Zhang, N. (2008). Time series clustering based on ICA for stock data analysis. Proceedings of 4th International Conference on Wireless Communications, Networking and Mobile Computing 2008 (WiCOM 2008), 1–4.

Hamilton, J.D. (1994). *Time Series Analysis.* Princeton University Press, Princeton, N.J.

Han, J., Kamber, M., and Pei, J. (2012). *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann Series in data management systems. Morgan Kaufman/Elsevier, Waltham, MA.

Harrington, S.E., Niehaus, G. and Yu, T. (2014). Volatility and underwriting cycles. In: Dionne, G. (ed.) (2014). *Handbook of Insurance*, 2nd ed. Springer, New York.

Harrington, S.E. and Yu, T. (2003). Do property-casualty insurance underwriting margins have unit roots? *Journal of Risk and Insurance*, 70, 4, 715–733.

Higgins, M., and Thistle, P.D. (2000). Capacity constraints and the dynamics of underwriting profits. *Economic Inquiry*, 38, 3, 442–457.

Hsu, Y.-C., Chen, A.-P. (2014). A clustering time series model for the optimal hedge ratio decision making. *Neurocomputing*, 138, 358–370.

Ingram, D. and Bush, E. (2013). Collective approaches to risk in business: an introduction to Plural Rationality Theory. *North American Actuarial Journal*, 17, 297–305.

Ingram, D. and Underwood, A. (2010). The Human Dynamics of the Insurance Cycle and Implications for Insurers: An Introduction to the Theory of Plural Rationalities. Monograph, Society of Actuaries, Schaumburg, Illinois.

Jawadi, F., Bruneau, C. and Sghaier, N. (2009). Nonlinear cointegration relationships between non-life insurance premiums and financial markets. *Journal of Risk and Insurance*, 76, 3, 753–783.

Kantardzic, M. (2011). *Data Mining: Concepts, Models, Methods, and Algorithms*, 2nd ed. John Wiley and IEEE Press, Hoboken, NJ.

Kaufman, R., Gadmer, A. and Klett, R. (2001). Introduction to dynamic financial analysis. *ASTIN Bulletin*, 31, 1, 213–249.

Keogh, E. (2002). Exact indexing of dynamic time warping. Proceedings of 28th International Conference on Very Large Databases, Hong Kong, 406–417.

Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S. (2001). Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and Information Systems*, 3, 263–286.

Kerr, G.,Ruskin, H., Crane, M., and Doolan, P. (2008). Techniques for clustering gene expression data. *Computers in Biology and Medicine*, 38, 3, 283–293.

Lamm-Tennant, J. and Weiss, M.A. (1997). International insurance cycles: rational expectations/institutional intervention. *Journal of Risk and Insurance*, 64, 415–439.

Li, Y., Lin, J. and Oates, T. (2012). Visualizing Variable-Length Time Series Motifs. Proceedings of the 2012 SIAM International Conference on Data Mining. SIAM, Anaheim, CA, 895–906.

Lin, J. (2016). SAX. URL: `http://cs.gmu.edu/~jessica/sax.htm`

Lin, J., Keogh, E., Lonardi, S., Chiu, B. (2003). A Symbolic Representation of Time Series, with Implications for Streaming Algorithms. Workshop on Research Issues in Data Mining and Knowledge Discovery. Proceedings of 8th ACM International Conference on Management of Data (SIGMOD), 2–11.

Lotte, F., Congedo, M., L'Ecuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 4, 1–13.

Nelson, C.R. and Plosser, C.I. (1982). Trends and random walks in macroeconomic time series: some Evidence and implications. *Journal of Monetary Economics*, 10, 139–162.

Nevill-Manning, C.G., Witten, I.H. (1997). Identifying hierarchical structure in sequences: a linear-time algorithm. *Journal of Artificial Intelligence Research*, 7, 67–82.

Nevill-Manning, C., Witten, I. (2016). Sequitur. URL: `http://www.sequitur.info/`

Otunba, R., Lin, J. and Senin, P. (2014). MBPD: Motif-based period detection. Proceedings of the 1st International Workshop on Pattern Mining and Application of Big Data. Presented at the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2014), Tainan, Taiwan, Springer, 793–804.

Ratanamahatana, C.A., Lin, J., Gunopulos, D., Keogh, E., Vlachos, M. and Das, G. (2010). Mining Time Series Data. In: Maimon, O., Rokach, L. (Eds.), *Data Mining and Knowledge Discovery Handbook 2010*. Springer US, Boston, MA, pp. 1049–1077.

Rudebusch, G. (1993). The uncertain unit root in real GNP, American Economic Review, 83, 264–272.

Sargent, T. (1987). *Macroeconomic Theory*, 2nd ed. Academic Press, Boston, MA.

Senin, P. (2016). GrammarViz 3.0. URL: `http://grammarviz2.github.io/grammarviz2_site/`

Stetco, A., Zeng, X. and Keane, J. (2013). Fuzzy cluster analysis of financial time series and their volatility assessment. Proceedings of 2013 IEEE International Conference on Systems, Man, and Cybernetics, 91–96.

Taylor, G. (2008). A simple model of insurance market dynamics. *North American Actuarial Journal*, 12, 242–262.

Tompa, M. and Buhler, J. (2001). Finding motifs using random projections. Proceedings of the 5th International Conference on Computational Molecular Biology, Montreal, Canada, 67–74.

Trufin, J., Albrecher, H. and Denuit, M. (2009). Impact of underwriting cycles on the solvency of an insurance company. *North American Actuarial Journal*, 13, 3, 385–403.

Wang, S.S., Major, J.A., Pan, H.C., Leong, J.W.K. (2011). U.S. Property-Casualty: underwriting cycle modeling and risk benchmarks. *Variance*, 5, 2, 91–114.

Warthen, T.V. and Somner, D.B. (1996). Dynamic financial modeling—issues and approaches. *CAS Forum* (Casualty Actuarial Society), Spring 1996, 291–328.

Yi, B. and Faloutsos, C. (2000). Fast time sequence indexing for arbitrary L_p norms. Proceedings of the 26th International Conference on Very Large Databases, 385–394.