International Journal of Data Science and Analytics (2021) 11:141–153
https://doi.org/10.1007/s41060-021-00248-2

REGULAR PAPER

Detection of causally anomalous time-series

Manoj Apte[1] · **Sushodhan Vaishampayan**[1] · **Girish Keshav Palshikar**[1]

Received: 19 August 2020 / Accepted: 15 January 2021 / Published online: 5 February 2021

Abstract
Many complex and important real-life applications, such as surveillance, monitoring and fraud detection, need to identify entire time-series, from a given collection, as anomalous. In this paper, we formulate and propose a solution for this *inter-time-series anomaly detection* problem, which is different from the usual *intra-time-series* anomaly detection, which identifies an anomalous "region" within a given single time-series. We formulate the notion of causally anomalous multi-variate time-series, and propose algorithms to identify them in a given database, using well-established notions of both linear and nonlinear Granger causality. The idea is to use (either domain knowledge or frequently observed) causal relations that hold between the univariate time-series corresponding to individual attributes, and identify those time-series as anomalous where this expected causality is violated. We use the proposed algorithms to detect causally anomalous time-series in several public datasets, in different domains such as economics, engineering, and medicine. Our experiments show that the causally anomalous time-series are not detected by strong baseline algorithms, indicating that this is a new notion of anomaly that complements the more standard formulations of what makes a time-series anomalous. We then present a detailed real-life case-study in a large stock exchange, where these techniques were used to identify agents with suspicious order behavior. We also point out limitations of the proposed notion of causally anomalous time-series.

Keywords Anomaly detection, Time-series · Granger causality · Stock market frauds · Stock market order book surveillance

1 Introduction

Many applications, e.g., such as surveillance, monitoring and fraud detection, collect a database of multi-variate time-series, where each time-series gives the "behavior" of an entity over a period of time, and the time-series share a common "structure." The time-series in a database may or may not be sampled at the same rate. Examples:

- In a data center having N servers, a monitoring program collects information about the servers' utilization (e.g., CPU usage, disk usage and memory usage at every second), resulting in N time-series every day.
- In a stock exchange having N stocks, the corresponding N time-series (collected every day) have data about the price and volume in the trades in that stock.
- In a bank having N accounts, the corresponding N time-series have data about the transactions of these accounts in a given time-period.
- In N systems (e.g., aircraft), each being monitored by a particular type of sensor, the corresponding N time-series have data about the sensor readings.

An important question is: are there any *anomalous* time-series in the given database? In the data center example, an anomalous time-series may indicate a server that is dangerously overloaded, or almost un-utilized or having wild fluctuations in its utilization levels. In the stock market example, an anomalous stock may indicate highly unusual trading patterns. This problem may be called as *inter-time-series* anomaly detection. This is different from the usual *intra-time-series* anomaly detection, where the goal is to identify an anomalous "region" within a given single time-series. For example, a particular kind of anomalous region in a given ECG signal may indicate a specific cardio-vascular disease,

✉ Manoj Apte
manoj.apte@tcs.com

Sushodhan Vaishampayan
sushodhan.sv@tcs.com

Girish Keshav Palshikar
gk.palshikar@tcs.com

[1] TCS Research, Tata Consultancy Services Ltd., Pune 411013, India

or in the sensor data about a chemical process, it may indicate transient faults or sudden changes. In the intra-time-series setting, a region within the single given time-series is anomalous only in the context of the remaining portion of the same time-series. In inter-time-series setting considered in this paper, a time-series in a database may be anomalous as a whole, in the context of the other time-series in the database. We consider, in general, multi-variate time-series.

Clearly, a multi-variate time-series consists of attribute-wise univariate *component* time-series. For example, the multi-variate time-series for a particular server consists of individual univariate time-series, one for each attribute (e.g., CPU utilization, memory usage, disk-space usage). Notions of statistical causality (e.g., Granger causality [1]) between univariate time-series are well-known in statistics. On the basis of domain knowledge, a particular causality relation may be expected to hold among some attributes. Alternatively, one may *observe* that a causality relation holds between two specific attributes in a large majority of time-series in a database. In either scenarios, then, a particular time-series in the database can be called as *causally anomalous* (or just *anomalous*) if its component time-series violate the expected causality relation(s). Purely as an example, one may find that CPU utilization Granger causes memory usage in a majority of time-series in the database. Then, any server whose time-series violates this causal relation is anomalous.

At the outset, we emphasize that this notion of causally anomalous time-series is quite limited—it captures only one particular type of anomaly. It fails to identify any time-series in the database as anomalous, if no causal relations are found to hold among the attribute time-series. It only applies to multi-variate time-series in the inter-time-series setting. It identifies an entire time-series as causally anomalous or not; it does not identify which regions in an anomalous time-series make it anomalous.

Our contributions in this paper are as follows. We formalize the novel notion of a causally anomalous multi-variate time-series. We propose unsupervised algorithms (which do not need any training data) to identify causally anomalous time-series in a database, based on both linear and nonlinear Granger causality formalisms (Sect. 4). We use the proposed algorithms to detect causally anomalous time-series in several public datasets, in different domains such as economics, engineering, and medicine (Sect. 5). Our experiments show that the causally anomalous time-series are not detected by strong baseline algorithms, indicating that this is a new notion of anomaly that complements the more standard formulations of what makes a time-series anomalous. We then present a detailed real-life case-study in a large stock exchange, where these techniques were used to identify agents with suspicious order behavior (Sect. 6).

2 Related work

Anomaly detection involves identifying unexpected items or events in a dataset. Most anomaly detection techniques use unsupervised, and do not need any labeled training data. Basic assumptions behind anomaly detection algorithms include rarity of occurrence and significant differences (e.g., value distributions) between anomalous points and "normal" data points. The input for an anomaly detection algorithm can be *point data* (where each data instance is considered as a point in k-dimensional space), or *graph data* (where two data points are joined by edges and points need not be considered to be members of a Euclidean space) or *sequence (or time-series) data* (an ordered list of values with or without a timestamp).

Many anomaly detection algorithms are known for point data; see [2] for an excellent, if dated, survey. In Sect. 5.2, we have already described the basic ideas behind some of these algorithms (RRS, LOF, iForest), which we have used as baselines. Another prominent approach for anomaly detection for point data is based on the idea that anomalous data points occur in less dense regions, whereas normal points occur in highly dense regions. DBSCAN is a density-based spatial clustering algorithm [3] that can be used for marking points in a low density region as outliers. Breunig et al. [4] takes this idea forward and considers the ratio of average density of k nearest neighbors of a data point with that point itself and marks points with local density lower than nearest neighbors as outliers.

We now review some of the work for anomaly detection in time-series data; see [5] for a survey. Most of these techniques work with univariate time-series (e.g., ECG) and identify parts within a time-series as anomalous (the intra-time-series setting described in Sect. 1). They identify either (i) a single value or (ii) a contiguous region in the time-series as anomalous, by comparing with neighboring (pre- and post-) regions in the same time-series. An unusually high (or low) value, or a sudden and sustained change in value are examples of anomalies detected by such techniques. These techniques are not directly usable in our inter-time-series setting (1).

A prominent approach for intra-time-series anomaly detection is proposed in [6,7]. The idea here is to transform a numeric time-series T into a symbolic aggregate approximation (SAX) representation by sliding a window of length n across T: first divide T in equal sized frames and then discretizing T using symbols. Then, they formalize the problem of detecting anomalies in T as the problem of finding discords, i.e., unusual sub-sequences in the SAX representation of T.

Qiu et al. [8] have used Granger graphical models that explore the temporal dependencies between variables by applying L1-regularized learning to Granger causality. Their aim is to find data points in multivariate time-series data

that significantly deviate from the "normal" patterns. Given a reference time-series, they use KL-divergence to compute an anomaly score for a time-series with the reference time-series for each variable, representing whether and how much that variable contributes to the deviation from the reference time-series. Anomalies are reported based on a threshold cutoff. This approach is useful when a suitable reference time-series is available, which is not the often the case.

Some approaches for anomaly detection in time-series are based on clustering. Tatusch et al. [9] clusters data points in various time-series that have the same timestamp, and then analyze behavior of a time-series as a sequence of transitions across these clusters to identify anomalous time-series. The notion of this behavior is based on the distances between the clusters of different points in time. In [10], the authors extend this work by using a different notion of distance between clusters and by using a weighting function. A crucial step in these time-series clustering methods is the so-called *evolutionary clustering*, which is the problem of how to build clusters of data points for different time instants. Several algorithms are known for evolutionary clustering; e.g., see [11].

We note that there is not much work in identifying entire multi-variate time-series in a database as anomalous (our inter-time-series setting described in Sect. 1), using inter-attribute causal relations.

3 Overview of Granger causality

3.1 Linear Granger causality

Interdependence between two time-series can be computed using cross-correlation (in time domain) or coherence (in frequency domain) functions. However, this does not capture *causal* interdependence between two time-series. *Granger causality (GC)* techniques provide a better understanding of the *causal* relationship among two time-series. A univariate time-series Y *Granger-causes* another univariate time-series X if using Y values allows us to get better predictions for X values. A simple approach (the Wald test) to test whether Y Granger causes X is as follows. As an example, using lag = 2, build an auto-regressive (AR) model for X, which predicts the value X_t at time t using previous two values X_{t-1} and X_{t-2}:

$$X_t = \alpha + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \epsilon_t \quad (1)$$

We can use the actual data in the given time-series X to compute the estimated values $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2$ of the coefficients. The R^2 value for this *restricted model* (denoted $R^2_{\text{restricted}}$) gives its predictive accuracy. Next, build a new AR model, called *full model*, which predicts X_t using X_{t-1}, X_{t-2} as well as 2 new variables corresponding to lagged values Y_{t-1}, Y_{t-2}:

$$X_t = \alpha + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 Y_{t-1} + \beta_4 Y_{t-2} + \epsilon_t \quad (2)$$

Again, estimate the coefficients and obtain a new R^2 value (denoted R^2_{full}) for this model. Now, use the F-test to check if the new model has better accuracy than the old model. Compute the F-value as follows:

$$F = \frac{((R^2_{\text{full}} - R^2_{\text{restricted}})/m)}{((1 - R^2_{\text{full}})/((n-2)\cdot(\ell-1)))} \quad (3)$$

Here, m denotes the number of additional variables in the full model ($m = 2$ in our example) and n denotes the total number of elements in time-series X (Y is assumed to have the same number of elements as X) and ℓ is the lag ($\ell = 2$ in our example). Next, compare this computed F-value to the critical value $F^*_{\alpha,m,2\ell}$ of the F-distribution with m degrees of freedom (DoF) in numerator and $(n-2)(\ell-1)$ DoF in the denominator (this value is obtained from tables). Here, α is the level of significance; typically, $\alpha = 0.05$. If computed F-value $> F^*$ then we reject the null hypothesis H_0 (both models have the same accuracy) and accept the alternative hypothesis H_1 that adding ℓ lag values of Y to the AR model improves prediction accuracy of X, i.e., Y Granger-causes X. In general, we may have to try different lag (ℓ) values, and we say Y *Granger-causes* X if this statement is true for at least one lag ℓ value. One issue is that we have tacitly assumed that X and Y have the same sampling rate, i.e., have the same timestamps at which the values are measured. If this is not the case, then one possibility is to re-sample both the time-series to have the same timestamps.

We need to check some conditions on time-series X, Y, before we apply any Granger causality test to check if X Granger-causes Y. We now summarize these restrictions. Roughly, a time-series is not stationary if it contains trends, seasonality, and other time-dependent effects such as the mean, variance, auto-correlation vary with time. Formally, a time-series X_t is *stationary* if the unconditional joint probability distribution of d random variables (positions) within the time-series does not depend on time. For example, suppose $d = 2$ and consider the joint PDF for values occurring at two positions X_{11} and X_{18} within the time-series. Then for a stationary time-series, the PDF is the same for say X_{12} and X_{19}, X_{13} and X_{20}, X_{35} etc., as long as these two positions are at a distance 7 from each other. All stationary time-series are also *integrated with order 0* (denoted $I(0)$), but not all $I(0)$ time-series are stationary. Informally, a time-series is integrated with order 0 if its auto-covariance "quickly" decays to 0. If time-series X, Y are both $I(0)$ (or at least stationary) then a Granger causality test can be applied to check if one time-series Granger-causes another time-series. Statistical tests, such as augmented Dicky-Fuller (ADF) test, can check whether or not the given time-series is stationary.

If the time-series X is not stationary, then we can take the first difference of the time-series to create another time-series ($Z_t = X_{t+1} - X_t$, for every $1 \leq t < |X|$), and then test the new time-series Z for stationarity. If Z is stationary then X is called *integrated with order 1* (denoted $I(1)$). Suppose we are given two time-series X and Y. Suppose you are able to find a constant value β such that $Y_t - \beta X_t$ is relatively constant, for every $1 \leq t < |X|$, i.e., this new time-series is stationary $I(0)$. Then, X and Y are said to be *co-integrated*. There are statistical tests—such as Engle-Granger test or Johansen test—which check whether the given two time-series are co-integrated. Thus if X (or Y) is not $I(0)$ (or stationary), then we can still test them for Granger causality if X is $I(1)$ (or Y is $I(1)$), or X, Y are co-integrated.

The standard formulation of Granger causality has several limitations. The main limitation is that it is a linear notion. That is why we now discuss one formulation of nonlinear Granger causality and use it in our experiments. There are other well-known limitations of Granger causality. First, philosophically speaking, it is really a probabilistic notion of causality and does not match with Hume's notation of causality. More importantly, it is seriously affected by confounding and when X and Y are both affected by a third process and may not always yield correct results in such scenarios. Finally, it applies only to pairs of variables and in general, we may need to use multi-variate extensions of Granger causality. Selecting such pairs of variables to test for causal relation can be tricky, and for this reason, we have used domain knowledge to identify them.

3.2 Nonlinear Granger causality

Linear Granger causality is unable to detect causal relations if they are nonlinear; see [12] for an example. Thus, we need a test to check for nonlinear causality between two time-series. We first summarize the statistical test given by Baek and Brock [13] for testing whether a time-series Y *nonlinearly Granger causes* another time-series X. We assume that X, Y are strictly stationary, weakly dependent (i.e., $I(0)$) and satisfy some "mixing" conditions given in [14]. Let $L, M \geq 1$ denote given lag values for time-series X and Y, respectively, and let m denote the given *lead value* for X. For example, if $m = 3$, then the *lead vector* for X at time index t is $X_t^m = (X_t, X_{t+1}, X_{t+2})$. Similarly, the *lag vectors* are: $X_{t-L}^L = (X_{t-L}, X_{t-L+1}, \ldots, X_{t-1})$ and $Y_{t-M}^M = (Y_{t-M}, Y_{t-M+1}, \ldots, Y_{t-1})$. The null hypothesis H_0 is that Y does not nonlinearly Granger cause X. The test procedure first fits a full linear model (using L lag values of X and M lag values of Y) to the data, and obtains the residuals for X, i.e., difference between predicted and actual values of X. Similarly for Y. Any remaining predictive power of these residuals time-series can be considered as nonlinear causal relation [13]. We now summarize the statistical hypothesis test developed by Baek and Brock [13] for this purpose.

Let $\epsilon > 0$ be a given small positive number. Let t, s be any two time indexes. Quantities $C1, C2, C3, C4$ are defined as given below. $C1$ is the joint probability that (i) the distance between lag vectors X_{t-L}^{m+L} and X_{s-L}^{m+L} at time indexes t and s is $< \epsilon$; as well as (ii) the distance between lag vectors Y_{t-M}^M and Y_{s-M}^M at time indexes t and s is $< \epsilon$. $\|\cdot\|$ denotes distance, for which the max norm is used. $C2, C3, C4$ are understood similarly.

$$C1(m+L, M, \epsilon) = \mathbf{Pr}(\|X_{t-L}^{m+L} - X_{s-L}^{m+L}\| < \epsilon, \|Y_{t-M}^M - Y_{s-M}^M\| < \epsilon) \quad (4)$$

$$C2(L, M, \epsilon) = \mathbf{Pr}(\|X_{t-L}^{L} - X_{s-L}^{L}\| < \epsilon, \|Y_{t-M}^M - Y_{s-M}^M\| < \epsilon) \quad (5)$$

$$C3(m+L, \epsilon) = \mathbf{Pr}(\|X_{t-L}^{m+L} - X_{s-L}^{m+L}\| < \epsilon) \quad (6)$$

$$C4(L, \epsilon) = \mathbf{Pr}(\|X_{t-L}^{L} - X_{s-L}^{L}\| < \epsilon) \quad (7)$$

Then, [13] give correlation integral-based estimators $\widehat{C1}, \widehat{C2}, \widehat{C3}, \widehat{C4}$ for the above quantities, which can be computed from the data, i.e., the actual realizations of the time-series X and Y. They show that under the null hypothesis and other conditions on X, Y given above, the test statistic given on the left follows the Normal distribution with mean 0 and variance σ^2 (which depends on m, L, M, ϵ):

$$\sqrt{n}\left(\frac{\widehat{C1}}{\widehat{C2}} - \frac{\widehat{C3}}{\widehat{C4}}\right) \sim N(0, \sigma^2) \quad (8)$$

They give an estimator for σ^2 that can be computed from the data. Here, $n = n_0 + 1 - m - \max(L, M)$ and n_0 is the length of the given realization of time-series X (and Y). The hypothesis test procedure now simply computes the value of the test statistic from given data and if it is more than the critical value obtained from the normal distribution on the right side (for, the significance level of, say, 0.05), then H_0 is rejected and the alternative hypothesis that Y nonlinear Granger causes X is accepted. We actually use a modified form of this Baek and Brock test, as given by Hiemstra and Jones [15], where they have given a better estimator for σ^2. Note that the choice of values for lags L, M, lead m and ϵ are important.

4 Causally anomalous time-series

Let $D = \{T_1, T_2, \ldots, T_N\}$ be a given database of N time-series. Each T_i in D is a multivariate time-series of the form: $\langle \mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \ldots, \mathbf{x}_{m_i}^{(i)} \rangle$; here, m_i = length of time-series T_i. We assume that each element $\mathbf{x}_j^{(i)}$ of each time-series in D has $k \geq 2$ attributes, i.e., each element is a k-vector. Our task

is to find out all time-series in D, each of which is causally anomalous with respect to a majority of the time-series in D. We now formalize the notion of causally anomalous time-series.

Let $T^{[a]}$ denote the univariate time-series obtained from T by removing all attributes, except a-th attribute. The univariate time-series $T^{[b]}$ is obtained from T by removing all attributes except b-th attribute. If time-series $T^{[a]}$ Granger causes time-series $T^{[b]}$, then we denote it as $a \Rightarrow b$; if $T^{[a]}$ does not Granger cause $T^{[b]}$, then we denote it as $a \not\Rightarrow b$.

In some applications, we may have *a priori* domain knowledge that $a \Rightarrow b$ is expected to hold for some particular attributes a and b. In such a case, any time-series in D for which $a \not\Rightarrow b$ is called *causally anomalous*. Alternatively, suppose we do not have any such domain knowledge. In that case, for a given pair of attributes a and b, we examine each time-series in D, and check whether $a \Rightarrow b$. Now, there are several possibilities: (i) For a "majority" of time-series in D, it is true that $a \Rightarrow b$. In that case, any time-series for which $a \not\Rightarrow b$ is called *causally anomalous*; (ii) for a "majority" of time-series in D, it is true that $a \not\Rightarrow b$. In that case, any time-series for which $a \Rightarrow b$ is called *causally anomalous*. (iii) Neither the relation $a \Rightarrow b$ nor $a \not\Rightarrow b$ holds for a "majority" of time-series in D. In that case, our approach fails to identify any causally anomalous time-series in D.

We now only need to formalize the notion of "majority" mentioned above. Rather than a Boolean check against a threshold value, we use a simple statistical test of proportions [16] to check whether the observed proportion is equal or greater than the user-specified threshold constant h_0. Let $p_{\text{obs}} = count\_pos/N$ be the observed proportion of time-series for which $a \Rightarrow b$ holds (N is the number of time-series in the database). We use the right-tailed test where the null hypothesis is $H_0 : p_{\text{obs}} = h_0$ against the alternative hypothesis $H_1 : p_{\text{obs}} > h_0$. The test statistic is: $t = \frac{p_{\text{obs}} - h_0}{\sqrt{\frac{h_0(1-h_0)}{N}}}$. Note that t is a random variable, as it is based on p_{obs}, which is computed from the given set of time-series and varies if the sample is changed. By a well-known result, assuming H_0 is true and sample size N is large, t has Standard Normal distribution $t \sim N(0, 1)$. We reject H_0 if the computed value of t is "too large" i.e., if the area under the standard Normal PDF to the right of t (as computed by the subroutine $pnorm$ in the algorithm) is less than the required level of significance, say 0.05.

The check whether $T^{[a]}$ Granger causes time-series $T^{[b]}$ can be implemented through a subroutine $check\_Causality(T^{[a]}, T^{[b]}, a, b, c\_type)$, which returns $TRUE$ if time-series $T^{[a]}$ Granger causes time-series $T^{[b]}$ at some lag m (in the range [0, lagmax], where lagmax is a user-specified constant) and returns $FALSE$ when $T^{[a]}$ does not Granger cause $T^{[b]}$ at any lag in the range [0, lagmax]. For nonlinear causality, along with lagmax, another parameter leadmax is also required. c_type specifies the type of causality to be tested by the routine (0 for linear and 1 for non-linear). This routine encapsulates checks for both linear and nonlinear Granger causality in a single abstraction.

Algorithm $GC\_Anomaly$ uses these ideas to identify causally anomalous time-series in a given database D with respect to a given pair of attributes. Basically, for given attributes a, b, it just counts (and records in array $flag$) in how many time-series in D the causality $a \Rightarrow b$ holds and in how many $a \not\Rightarrow b$ holds. It then uses a statistical hypothesis test of proportions (as explained above) to check if either of these counts is "close" or above the user-specified threshold h_0, and returns the indexes of time-series which violate the direction (dir) of the majority causal relation that holds in D (these are the causally anomalous time-series). Running this algorithm in a loop over all possible attribute pairs will identify causally anomalous time-series with respect to any pair of attributes. Note that it is possible that neither $a \Rightarrow b$ nor $a \not\Rightarrow b$ may hold for majority of time-series in D, in which case the algorithm does not find any causally anomalous time-series in D.

5 Experiments

5.1 Datasets

To our knowledge, there are no public-domain multivariate time-series datasets containing explicitly marked causal relations among its attributes. Hence, we use several public-domain time-series datasets, for which there is domain knowledge which gives an idea of the causal relations that are expected to hold between some pairs of attributes (Table 1). The Electric Motor Temperature Dataset[1] contains 52 time-series, each corresponding to a session where the permanent magnet synchronous motors (PMSM) are monitored using sensors. The measurement sessions range between 1 and 6 h. The causal relations expected to hold in this dataset are taken from [17], which intuitively correspond to *voltage causes current*, *voltage causes angular motor speed* and *temperature causes (changes to) torque*.

The World Development Indicators dataset from World Bank[2] contains data related to the development indicators for various countries from the years 1960 up to 2018. Following are some of the attributes, along with the number of countries for which the attribute is available: births attended by skilled health staff (% of total) (96 countries), immunization, measles (% of children ages 12–23 months) (191 countries), prevalence of underweight, weight for age (% of children

[1] https://www.kaggle.com/wkirgsn/electric-motor-temperature.

[2] https://databank.worldbank.org/source/world-development-indicators/.

Algorithm 1: Algorithm *GC_Anomaly*

input : $D = \{T_1, T_2, \ldots, T_N\}$ // N time-series, each element is k-dimensional vector
input : $A = \{1, 2, \ldots, k\}$; $a, b \in A$ // attribute names
input : $h_0 \in [0, 1]$ // user-given min. proportion reqd. for majority
output: $O \subseteq \{1, 2, \ldots, N\}$ // indexes of anomalous time-series
$count\_pos := 0$; $count\_neg := 0$; $O := \emptyset$;
for i = *1 to* N **do**
  $(flag[i]) := check\_Causality(T_i^{[a]}, T_i^{[b]}, a, b, c\_type)$;
  if $flag[i] == TRUE$ **then**
    count_pos++;
  end
  else
    count_neg++;
  end
end
$t_+ = \frac{(count\_pos/N) - h_0}{\sqrt{\frac{h_0(1-h_0)}{N}}}$;
$t_- = \frac{(count\_neg/N) - h_0}{\sqrt{\frac{h_0(1-h_0)}{N}}}$;
// area under Std. Normal prob. density to the right of obs. value t_+
if $pnorm(t_+) < 0.05$ **then**
  // $a \Rightarrow b$ holds in majority of time-series in D
  $dir := TRUE$
end
else
  if $pnorm(t_-) < 0.05$ **then**
    // $a \not\Rightarrow b$ holds in majority of time-series in D
    $dir := FALSE$
  end
  else
    // Neither $a \Rightarrow b$ nor $a \not\Rightarrow b$ holds in majority of time-series
    // in D (no statistically significant difference in their counts)
    return$((\emptyset, NULL))$;
  end
end
for i = *1 to* N **do**
  if $dir == TRUE \wedge flag[i] == FALSE$ **then**
    $O := O \cup \{i\}$;
  end
  if $dir == FALSE \wedge flag[i] == TRUE$ **then**
    $O := O \cup \{i\}$;
  end
end
return$((O, dir))$;

under 5) (18 countries) with mortality rate under-5 (per 1000 live births) and contraceptive prevalence any methods (% of women ages 15-49) with population growth (annual %) (34 countries). The number of countries for each causal relation are different due to non-availability of different attributes for all countries.

The Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet Data Set is handwriting dataset (PDSST) [18] consisting of 25 People with Parkinson's Disease and 15 healthy individuals. The dataset consists of handwriting recordings of three types. We have considered the static spiral test (SST) data. SST is used for clinical research in literature for different purposes like determining motor performance, measuring tremor and diagnosing Parkinson disease. Since people with Parkinson's are not able to draw a spiral properly, we can consider a causal relation between the x and y coordinates for such people (since the number of individuals with Parkinson's disease is more).

5.2 Baselines

As mentioned earlier, our notion of causally anomalous time-series is quite restricted, applies only to multivariate time-series and identifies entire time-series in a database as causally anomalous or not (inter-time-series anomaly setting). This makes a direct comparison with other time-series anomaly detection techniques somewhat difficult. For example, it is not useful to compare with non-causality-based techniques that analyze univariate time-series to identify intra-time-series regions as anomalous.

Our first baselines "condense" each time-series to a point in a multi-dimensional space and then use standard anomaly detection techniques to identify anomalous points. Here, we used the well-known outlier detection algorithms RRS [19], LOF [4], iForest [20] algorithms, along with a simple statistical outlier detection algorithm based on Mahalanobis distance, as baselines. These algorithms do not directly work on time-series data; instead they identify anomalous points in a given set of p-dimensional data points. Hence, we map each of pair of our univariate time-series X and Y to a single vector containing summary statistics for each of the two component time-series: min, max, average, standard deviation, quartiles, auto-correlations, etc. In Table 1, we take all possible attribute pairs, and transform the associated two univariate time-series datasets in this manner to a single datapoint, in order to apply these baseline algorithms to them.

The RRS algorithm [19] uses the distance of each point from its kth nearest neighbor (NN) as the measure to decide the outlier. The more the distant a point is from its kth neighbor, the more likely is the point to be an outlier. Once the distances from the kth nearest neighbor are computed for each point, then we mark top-m data points as anomalous (we use $k = 5, m = 5$). The LOF algorithm [4] computes local outlier factor as the ratio of average density of k nearest neighbors of the data point and the average density of the data point itself. The idea is that the average density of a normal point must be close to the average density of its neighbor. Thus, a point having a marginal difference between the average density of its k nearest neighbors and its own average density is likely to be an outlier. This ratio is converted to an outlier score. We mark top-m data points as anomalous (we use $k = 5, m = 5$).

The iForest algorithm [20] builds a *random partition tree* as follows: first randomly select an attribute A, then randomly select a split value between the maximum and minimum

Table 1 Public-domain time-series datasets used

Dataset	#Time-series	#Attributes	Expected causal relations
Electric motors	52	12	$pm \Rightarrow torque$,
			$u\_d \Rightarrow i\_d, u\_q \Rightarrow i\_q$
			$u\_d \Rightarrow motor\_speed$
			$u\_q \Rightarrow motor\_speed$
World dev. ind.	96	2	$birth\_skilled\_staff \Rightarrow mortality\_under\_5$
World dev. ind.	191	2	$immunization\_measles \Rightarrow mortality\_under\_5$
World dev. ind.	18	2	$prevalence\_underweight \Rightarrow mortality\_under\_5$
World dev. ind.	34	2	$contraceptive\_prevalence \Rightarrow population\_growth$
PDSST	40	2	$x \Rightarrow y$

values of A, and then partition the instances in the dataset accordingly. This partitioning of instances is repeated recursively until all instances are *isolated* (i.e., become the only entry in a partition). Many such random partition trees are constructed for the same dataset, and points which have a short path length in many of these trees are identified as anomalies.

The *Mahalanobis distance* between two p-dimensional points $\mathbf{x}$ and $\mathbf{y}$ in a dataset is given by:

$$R(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T \cdot C^{-1} \cdot (\mathbf{x} - \mathbf{y}) \tag{9}$$

where C is the $p \times p$ sample co-variance matrix. This way, we get distance of each point from the mean of both attributes. These distances are sorted in descending order and the points that are very far from the mean are marked as anomalous.

In addition, we combine dynamic time warping [21,22] and the RRS algorithm to design another simple baseline algorithm for detection of anomalous time-series; this algorithm works directly on time-series data. Dynamic time warping (DTW) [21,22] is a well-known technique for measuring similarity of the *shapes* of two time-series. We use DTW to construct a similarity matrix between a set of time-series and then use RRS algorithm (which uses this matrix) to identify anomalous time-series. We call this baseline algorithm as RRS_DTW. Note that this method works directly on the time-series, without mapping each time-series to a single vector, and identifies an entire time-series as anomalous.

For our $GC\_Anomaly(linear)$ algorithm, we use $h_0 = 0.55$ and we iterate lag_value from 1 to 4. Here, lag_value is the number of past values considered for building regression models (both full and restricted) and h_0 is the threshold above which the majority is considered a a true majority. For $GC\_Anomaly(non\text{-}linear)$ algorithm, we use $h_0 = 0.55$ and we iterate lag_value from 1 to 4. Specifically for the nonlinear setting, we have to give additional parameter $lead\_value$ which is taken as $lead\_value = lag\_value + 1$.

5.3 Results

Table 2 shows the results of all baseline algorithms, as well as our algorithms, on various public-domain datasets which are explained in Table 1. The public-domain datasets do not provide a ground truth for anomaly detection. So in Table 2, we have given the results of our algorithm $GC\_Anomaly(linear)$, and in the columns for the baselines, we have given the no. of anomalies reported by each baseline algorithm and common anomalies with $GC\_Anomaly(linear)$ are given in bracket; e.g., for the dataset Electric Motors ($u\_d \Rightarrow i\_d$) results for the column LOF show that LOF algorithm has reported five anomalies out of which one is common with $GC\_Anomaly(linear)$. We observe that the causally anomalous time-series are not captured well by the baselines, i.e., they have low overlap with outputs of our algorithms. This shows that our algorithms are capturing genuinely new types of anomalies.

The next question is: how well is the domain knowledge (i.e., expected causal relations among attributes in Table 1) reflected in these datasets? We found that the expected causal relations among attributes do not always hold in the datasets when we use linear Granger causality. As an example, in the electric motors dataset, the domain knowledge expects that the causal relation $pm \Rightarrow torque$ should hold; however, we found that in majority of the time-series in this dataset, this causal relation was found not to hold (using linear Granger causality). Similarly, in the World Development Indicators dataset, we found that some of the expected causal relations were found not to hold (using linear Granger causality). Interestingly, in the Electric Motors dataset, the modified Baek and Brock test ($GC\_Anomaly(non\text{-}linear)$) captured all the causal relations as expected from the domain knowledge and no causally anomalous time-series were detected by $GC\_Anomaly(non\text{-}linear)$. Similarly, in the World Development Indicators dataset also, all the expected causal relations were detected by the modified Baek and Brock test ($GC\_Anomaly(non\text{-}linear)$).

Table 2 Results on public-domain datasets

Dataset with expected causality	GC_anomaly (linear)	RRS	LOF	iForest	Mahalanobis	RRS_DTW
Elec. motors ($u\_d \Rightarrow i\_d$)	4	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)
Elec. motors ($u\_d \Rightarrow i\_q$)	2	5 (0)	5 (1)	5 (0)	5 (1)	5 (0)
Elec. motors ($u\_d \Rightarrow motor\_speed$)	8	5 (0)	5 (0)	5 (0)	5 (1)	5 (0)
Elec. motors ($u\_q \Rightarrow i\_d$)	14	5 (1)	5 (1)	5 (1)	5 (2)	5 (3)
Elec. motors ($u\_q \Rightarrow i\_q$)	11	5 (0)	5 (0)	5 (0)	5 (2)	5 (0)
Elec. motors ($u\_q \Rightarrow motor\_speed$)	2	5 (0)	5 (0)	5 (0)	5 (1)	5 (0)
Elec. motors ($pm \Rightarrow torque$)	11	5 (2)	5 (2)	5 (2)	5 (1)	5 (3)
World. dev. indicators ($birth\_skilled\_staff \Rightarrow mortality\_under\_5$)	13	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)
World. dev. indicators ($contraceptive\_prevalance \Rightarrow population\_growth$)	2	5 (0)	5 (0)	5 (0)	5 (0)	5 (1)
World. dev. indicators ($prevalence\_underweight \Rightarrow mortality\_under\_5$)	1	5 (0)	5 (0)	5 (0)	5 (0)	5 (1)
World. dev. indicators ($immunization\_measles \Rightarrow mortality\_under\_5$)	31	5 (1)	5 (1)	5 (0)	5 (0)	5 (1)
Parkinson's disease (PDSST)	10	5 (0)	5 (1)	5 (1)	5 (3)	5 (1)

It is interesting to compare the anomalies detected by $GC\_Anomaly(linear)$ with those detected by $GC\_Anomaly(non\text{-}linear)$. For example, in the World Development Indicator dataset, one expected causal relation is prevalence_underweight $\Rightarrow$ mortality_under_5, for which we do not know the causal form: linear or non-linear. In this case, $GC\_Anomaly(linear)$ reports BFA, KWT and PER as anomalous, whereas $GC\_Anomaly(non\text{-}linear)$ reports VEN as anomalous.

The threshold parameter h_0 is important in the algorithm $GC\_Anomaly(linear)$. Table 3 shows the effect of different values of parameter h_0 (threshold for majority) fed to the Algorithm $GC\_Anomaly(linear)$. The value of h_0 is varied from 0.55 to 0.95; because we are considering the majority condition for finding anomalies, we consider $h_0 > 0.5$. The range of values we tried for h_0, and for which the anomalies are reported, is given in the $Range(h_0)$ column. The number of anomalies remains constant when the anomalies are reported, so we conclude that Granger causality between the time-series is not very sensitive to the value of h_0.

We also conducted experiments with different lags for the regression equations in the Granger causality check of the $GC\_Anomaly(linear)$ algorithm. We used the lag values in the range 1–4 in the Parkinson's Disease (PDSST) and Electric Motor dataset. We limited lag values in the range 1–2 for the World Development Indicators dataset due to lack of sufficient number of observations in the time-series. The $GC\_Anomaly(linear$ algorithm produced different counts of anomalous time-series at different lag values; however, we found no significant relationship pattern between the lag value of the regression and the outputs of the algorithm.

6 Suspicious order behavior in stock market

A stock exchange maintains an *order book*, containing buy–sell orders for all stocks on a particular day. Each order is placed by one *agent*, either for buying or selling shares of a specific company (stock). Depending on the granularity of interest, an agent may be a stock + broker combination, or a stock + broker + dealer combination, or a stock + broker + dealer + investor combination. Thus, an order book of a particular day consists of N time-series, each containing the sequence of orders placed by a particular agent on that day. Each element of these time-series corresponds to one order. An order is a k-vector, consisting of attributes such as, $timestamp, agent\_id, order\_flag, order\_type, limit\_price, quantity, price, order\_status$, etc. Here, $order\_flag$ indicates whether it is a buy order or sell order. Most stock exchanges allow different types of orders to be placed. When $order\_type = market$, it means that the agent wants to match this order to any order (with opposite value for $order\_flag$) which is currently already placed and not fulfilled. When $order\_type = limit$, it means that the agent wants to execute this order only when the $limit\_price$ (specified in the order) is matched; the order remains unfulfilled till such a matching order enters the order book. When $order\_type = market$, the attribute $limit\_price$ is not used. In general, there are other types of orders, but for simplicity we omit them here.

Stock exchanges have complex order matching algorithms, and a *trade* happens when two orders of opposite values of $order\_flag$ are matched; in that case the order book is updated suitably. Example: when a buy order for 300

Table 3 Effect of the parameter h_0.

Dataset	Range(h_0)	Count of anomalies
Elec. motors ($u\_d \Rightarrow i\_d$)	(0.55–0.8)	4
Elec. motors ($u\_d \Rightarrow i\_q$)	(0.55–0.85)	2
Elec. motors ($u\_d \Rightarrow motor\_speed$)	(0.55–0.7)	8
Elec. motors ($u\_q \Rightarrow i\_d$)	(0.55–0.6)	14
Elec. motors ($u\_q \Rightarrow i\_q$)	(0.55–0.65)	11
Elec. motors ($u\_q \Rightarrow motor\_speed$)	(0.55–0.85)	2
Elec. motors ($pm \Rightarrow torque$)	(0.55–0.65)	11
World. dev. indicators ($birth\_skilled\_staff \Rightarrow mortality\_under\_5$)	(0.55–0.75)	13
World. dev. indicators ($contraceptive\_prevalance \Rightarrow population\_growth$)	(0.55–0.8)	2
World. dev. indicators ($prevalence\_underweight \Rightarrow mortality\_under\_5$)	(0.55–0.75)	1
World. dev. indicators ($immunization\_measles \Rightarrow mortality\_under\_5$)	(0.55–0.75)	31
Parkinson's disease (PDSST)	(0.55–0.6)	10

shares of a stock matches a sell order for 1000 shares of the same stock, the buy order has its $order\_status$ is set to the value $completed$, and the order quantity for the sell order is updated to 700; also, a trade record is created for this transaction. Once all shares in a sell (buy) order are sold (bought), the $order\_status$ is marked as $completed$. An agent is allowed to modify any of the orders placed by him any number of times, until $order\_status = completed$, where a modification may involve changes in price or quantity. In this paper, we focus only on equity trading and do not consider any other kind of trading, such as futures or options. Also, we exclude the part of the order book that records the trade information, i.e., we focus only on the orders.

A key aspect of real-time surveillance in a stock exchange requires monitoring the order book, and detecting attempts by agents to place orders in a suspicious manner with the aim of influencing or manipulating the trading. Following are some examples of such suspicious order behaviors:

- *Front Running*: An agent having information about large, potentially market moving orders, places orders for personal advantage before these large orders are actually placed.
- *Marking the close*: Placing orders and creating trades in the stock near the end of the day to affect the published closing price.
- *Pegging and capping*: Placing orders for a stock, with prices always above (*pegging*) or always below (*capping*) a pre-specified (undisclosed) threshold price.
- *Pump and dump*: To attract attention in a particular low-trading stock, an agent places many buy orders with higher prices, and when other join and place buy orders at higher prices, then the agent exits by selling large blocks of shares at these inflated prices.
- *Insider trading*: When an agent has access to secret information which is going to affect the share price in near future, he places large buy or sell orders accordingly.
- *Circular Trading/Wash Trades/Pools/Painting the Tape/ Collusion*: Agents collude with other to artificially create an impression of high activity, by placing synchronized orders and trading among themselves. They exit by selling when the price reaches a suitably high level.
- *Pinging and spoofing*: The complete order book of the day is not visible to agents; they typically only know total buy/sell volumes and a summary of orders 5 "ticks" above and below the last traded price. Hence, agents may place *ping* orders, which "explore" the price range at which others have placed orders. Such ping orders are then canceled or modified. Thus, ping orders allow an agent to understand and exploit the market picture, which is not available to other genuine investors. Agents sometimes place *spoofing* orders, outside the bona fide limits, for feigning interest in a stock and hoping to stimulate more activity. Pinging and spoofing are typically used in algorithmic trading.
- Placing "large" orders in the pre-open session with "large" price difference from previous close price.
- Placing orders with "too many" updates.
- Placing and canceling "too many" orders.

Many such patterns of suspicious order behavior can be identified by domain experts and these can be coded into the surveillance system. However, this approach suffers the usual issues plaguing any rule-based system: the patterns are

Table 4 Summary statistics for the order book

Statistic	Max	Average	SD	Q1	Q2	Q3
Buy volume	200,000	1088.65	5713.51	100	200	885
Sell volume	300,000	1467.87	6260.09	100	300	1000
Buy price	58.7	43.011	12.140	43.15	45.15	49.15
Sell price	58.7	45.429	11.229	44.55	48.5	50.15
#orders per agent	1799	2.467	18.463	1	1	2
#buy orders per agent	1741	3.638	26.861	1	1	3
#sell orders per agent	424	1.718	6.587	1	1	1
#modifications per agent	18,606	3.959	147.937	1	1	1
#buy modifications per agent	15,348	13.140	322.272	1	1	2
#sell modifications per agent	3258	2.384	30.659	1	1	1
#modifications per order per agent	979	2.150	12.558	1	1	1
#modifications per buy order per agent	979	5.941	28.926	1	1	1
#modifications per sell order per agent	334	1.345	3.143	1	1	1
Inter-order gap	601.0	0.315	2.951	0	0	0
Inter-order gap per agent (s)	25,623.0	692.303	2046.364	0	9	223

fragile, insensitive to variations, not able to discover newer (emerging) suspicious order behaviors and need to be maintained and updated manually.

We have built a surveillance system for a large national exchange, which includes deployment of many different types of anomaly detection techniques for both time-series as well as non-time-series data. As explained, a day's order book consists of N time-series, one for each agent, and each element of each time-series is a order (which is a k-vector). Note that the orders come at irregular time instants. Thus, the problem of detecting agents having suspicious order behavior can be considered (in part) as a problem of detecting causally anomalous time-series among the N time-series in the order book (inter-time-series setting). As emphasized in Sect. 1, agents may have other types of suspicious order behaviors that are *not* detected by our notion of causally anomalous time-series. On the other hand, as already demonstrated in Sect. 5, our techniques are capable of detecting agents with suspicious order behaviors which are *not* detectable by standard techniques. Thus, our notion of causally anomalous time-series, while limited, complements the common notions of anomaly.

For illustration, we use a small subset of the order book in a large stock exchange, for a particular stock on a particular date.[3] In this subset, there were 14,576 agents, with one order time-series for each agent. Total number of unique buy and sell orders are 45,120 and 33,672. The sum of the buy and sell order volumes is 49,283,400 and 50,443,516. Table 4 shows the summary statistics of the order book.

In a stock exchange, agents typically place orders by looking at the previous trades (price and volume) that have happened. This gives rise to the expectation that the orders placed are typically interdependent on orders previously placed (since the trades will happen by matching these orders). Thus, it is appropriate to examine causal relations that may exist between various attributes in the order time-series. We use our two approaches of linear as well as non-linear Granger causality for detecting suspicious order behavior. The idea is to check whether one attribute within an agent's time-series Granger causes another attribute. If this pattern is observed to be true (false) for a "majority" of agents, then any agent for whom this pattern is false (true) is causally anomalous and hence the corresponding order behavior is suspicious. Copeland and Jennings [23,24] indicate a positive causal relation between stock market prices and trading volume in either direction. Hence, we analyze causal relations between these two attributes. Note however, that we work on orders, not trades. Further, as we will show below, we found that a *nonlinear* causal relation between price and volume holds for majority of agents in the order book database of this stock exchange. This piece of discovered domain knowledge should be interesting to financial analysts and econometricians.

Our algorithm $GC\_Anomaly(linear)$ failed to detect the causal relation between price and volume in both directions of the order placement time series for majority of the brokers. It established a non-causal relation for the majority marking the ones with causal relations as anomalous. However, Our algorithm $GC\_Anomaly(non\text{-}linear)$ clearly identified the causal relation for the majority of the time-series and marked the ones having non-causal relations as anomalous. The results of both our algorithms are given in Tables 5 and 6. The results should be interpreted similar to Table 2. As already observed in Sect. 5, here too, the baseline anomaly

[3] This is a proprietary data and we are unable to share it.

Table 5 Results on Order book dataset for linear Granger Causality

Dataset with expected causality	GC_Anomaly (linear)	RRS	LOF	iForest	Mahalanobis	RRS_DTW	GC_Anomaly(non-linear)
Stock Market Data (Buy) ($quantity \Rightarrow price$)	39	5 (0)	5 (0)	5 (1)	5 (0)	5 (4)	7(1)
Stock Market Data (Buy) ($price \Rightarrow quantity$)	40	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	12(1)
Stock Market Data (Sell) ($quantity \Rightarrow price$)	34	5 (0)	5 (0)	5 (0)	5 (0)	5 (3)	8(2)
Stock Market Data (Sell) ($price \Rightarrow quantity$)	31	5 (0)	5 (0)	5 (0)	5 (0)	5 (2)	7(2)

Table 6 Results on Order book dataset for nonlinear Granger Causality

Dataset with expected causality	GC_Anomaly (non-linear)	RRS	LOF	iForest	Mahalanobis	RRS_DTW	GC_Anomaly(linear)
Stock Market Data (Buy) ($quantity \Rightarrow price$)	7	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	39(1)
Stock Market Data (Buy) ($price \Rightarrow quantity$)	12	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	40(1)
Stock Market Data (Sell) ($quantity \Rightarrow price$)	8	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	34(2)
Stock Market Data (Sell) ($price \Rightarrow quantity$)	7	5 (0)	5 (0)	5 (0)	5 (0)	5 (0)	31(2)

detection algorithms do not detect causally anomalous time-series, which again demonstrates that our notion of causally anomalous time-series complements the standard notions of anomaly. In the absence of labeled data, we cannot provide accuracy figure for our results. However, the surveillance officers in the stock exchange have agreed in informal discussions that most of the results indeed *prima facie* indicate suspicious order behaviors.

7 Conclusions and future work

In this paper, we described a novel inter-time-series setting for detection of entire anomalous time-series in a database of multi-variate time-series. We formalized a new notion of causally anomalous time-series in this setting, and proposed techniques for identifying causally anomalous time-series in a database of time-series. The idea is to look for "frequently observed" causal dependence between univariate attribute time-series in the given database, and identify those time-series as causally anomalous that violate this dependence. Granger causality uses vector autoregression to find the effect of one attribute for prediction of values of another attribute in the same time-series (cause and effect). Granger causality is a linear causal relation. However, sometimes the causal relationships might be nonlinear. Hence, we proposed algorithms that used both linear and nonlinear Granger causality to identify these causal relations. We used public-domain time-series databases from different domains (economics, engineering and medicine) to demonstrate the effectiveness of our techniques, as compared to a set of strong baseline anomaly detection techniques, to identify causally anomalous time-series. Here, we used available domain knowledge to *a priori* identify causal relations that are expected to hold between various attributes. This shows that the approach can accommodate (and validate) domain knowledge about causal relations among time-series attributes. We observed that the results of our algorithms that use linear and nonlinear Granger causality often complement each other. Finally, we used our techniques to identify agents with suspicious order behaviors in a real-life dataset from a large stock exchange, again demonstrating the effectiveness of our techniques, as compared to a set of strong baseline anomaly detection techniques. We found that a *nonlinear* causal relation between price and volume holds for majority of agents in the order book database of our stock exchange. This piece of discovered domain knowledge can be validated for order books other stock exchanges, and thus should be interesting to financial analysts and econometricians.

As mentioned in Sect. 1, our notion of causally anomalous time-series is quite limited—it captures only one particular type of anomaly. It would fail to identify any time-series in the database as anomalous, if no causal relations are found to hold among the attribute time-series. It only applies to multi-variate time-series in the inter-time-series setting. It identifies an entire time-series as causally anomalous or not; it does not identify which regions in an anomalous time-series make it anomalous. In some applications this is sufficient, as we have shown in Sect. 5 as well as in the real-life case-study in Sect. 6. In other applications identifying anomalous segments within a given time-series is important.

Another limitation of work is that it uses only pairwise Granger causality among variables in a time-series. Section 1) several other real-life applications where our approach would be useful. We plan to explore some of them. At present, we have used causal relations between only pairs of attributes in a time-series. We are exploring how multivariate time-series causality techniques can be applied to generalize our approach. However, a similar limitation will apply even when we use multivariate Granger causality: how does one choose the appropriate subset of variables to test for Granger causality? Also, in our experience, users find it easier to understand pairwise Granger causality when it exists. Hence, we have taken a novel approach of appealing to the existing domain-knowledge to identify likely pairs of variables to test for Granger causality. Such knowledge exists in many domains, and we have given many examples of such domain knowledge in the paper. Moreover, since we identify time-series which violate expected causal relations as anomalous, this violation of domain knowledge automatically makes the time-series interesting for end-users. To the best of our knowledge, this is the first time domain knowledge is used to generate hypotheses for causality testing. Still, using only pair-wise Granger causality to identify anomalous time-series may lead to false positives (when ground-truth is already available) due to underlying noisiness. Hence, as future work, we will consider aggregating the results of various pair-wise Granger causality tests to come up with a more robust identification of anomalous time-series.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Granger, C.: Investigating causal relations by econometric models and cross-spectral methods. Econometrica **37**(3), 424–438 (1969)

2. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: a survey. ACM Comput. Surv. **41**(3), 1–58 (2009)
3. Ester, M., Kriegel, H., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Second International Conference on Knowledge Discovery and Data Mining (KDD), pp. 226–231 (1996)
4. Breunig, M.M., Kriegel, H.-P., Ng, R.T., Sander, J.: Lof: identifying density-based local outliers. In: ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 93–104 (2000)
5. Chandola, V., Cheboli, D., Kumar, V.: Detecting anomalies in a time series database. In: UMN TR09-004 (2009)
6. Keogh, E., Lin, J., Fu, A.: Hot sax: efficiently finding the most unusual time series subsequence. In: International Conference on Data Mining (ICDM), pp. 226–233 (2005)
7. Keogh, E., Lin, J., Lee, S., Herle, H.V.: Finding the most unusual time series subsequence: algorithms and applications. Knowl. Inf. Syst. **11**(1), 1–27 (2006)
8. Qiu, H., Liu, Y., Subrahmanya, N., Li. W.: Granger causality for time-series anomaly detection. In: 12th IEEE International Conference on Data Mining (ICDM), pp. 1074–1079 (2012)
9. Tatusch, M., Klassen, G., Bravidor, M., Conrad, S.: Show me your friends and i'll tell you who you are. Finding anomalous time series by conspicuous cluster transitions. In: Australasian Conference on Data Mining (AusDM 2019): Data Mining, pp. 91–103 (2019)
10. Tatusch, M., Klassen, G., Conrad, S.: Behave or be detected! identifying outlier sequences by their group cohesion. In: International Conference on Big Data Analytics and Knowledge Discovery (DaWaK 2020), pp. 333–347 (2020)
11. Chakrabarti, D., Kumar, R., Tomkins, A.S.: Evolutionary clustering. In: 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06), pp. 554–560 (2006)
12. Brock, W.: Causality, chaos, explanation and prediction in economics and finance. In: Beyond Belief: Randomness, Prediction and Explanation in Science, pp. 230–279 (1991)
13. Baek, E., Brock, W.: A nonparametric test for independence of a multivariate time series. In: Statistica Sinica 2, pp. 137–156 (1992)
14. Denker, M., Keller, G.: On u-statistics and von-mises statistics for weakly dependent processes. Zeitschrift fur Wahrscheinlichkeitstheorie und Vervandte Gebiete **64**, 505–522 (1983)
15. Hiemstra, C., Jones, J.: Testing for linear and nonlinear granger causality in the stock price-volume relation. J. Finance **49**, 1639–1664 (1994)
16. Milton, J.S., Arnold, J.C.: Introduction to Probability and Statistics: Principles and Applications for Engineering and the Computing Sciences, 4th edn. McGraw-Hill Education, London (2002)
17. Bilgin, O., Kazan, F.A.: The effect of magnet temperature on speed, current and torque in PMSMs. In: *XXII International Conference on Electrical Machines (ICEM)*, pp. 2080–2085 (2016)
18. Isenkul, M., Sakar, B., Kursun, O.: Improved spiral test using digitized graphics tablet for monitoring parkinson's disease. In: 2nd International Conference on e-Health and Telemedicine (ICEHTM), pp. 171–175 (2014)
19. Ramaswamy, S., Rastogi, R., Shim, K.: Efficient algorithms for mining outliers from large data sets. In: ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 427–438 (2000)
20. Liu, F.T., Ting, K.M., Zhou, Z.-H.: Isolation forest. In: Eighth IEEE International Conference on Data Mining (ICDM), pp. 413–422 (2008)
21. Berndt, D., Clifford, J.: Using dynamic time warping to find patterns in time series. In: Workshop on Knowledge Discovery in Databases (KDD) (1994)
22. Keogh, E.J., Pazzani, M.J.: Derivative dynamic time warping. In: SIAM International Conference on Data Mining, pp. 1–11 (2001)
23. Copeland, T.E.: A model of asset trading under the assumption of sequential information arrival. J. Finance **31**(4), 1149–1168 (1976)
24. Jennings, R.H., Starks, L.T., Fellingham, J.C.: An equilibrium model of asset trading with sequential information arrival. J. Finance **36**(1), 143–161 (1981)