Causal effects of government decisions on earthquakes in Groningen

Comi, Giovanni Eugenio; Fitzner, R.J.; Kolumbán, Sandor; Pijpers, F.P.; Pires da Silva Castro, R.M.; Post, R.A.J.; Vromans, A.J.

Publication date
2019

Causal effects of government decisions on earthquakes in Groningen

G.E. Comi[1], R.J. Fitzner[2], S. Kolumbán[2], F.P. Pijpers[3], R.M. Pires da Silva Castro[2], R.A.J. Post[2], and A.J. Vromans[2]

Abstract

The problem of determining causal relations between two or more time series is a fundamental and frequently occurring challenge. It is relevant for a variety of problems ranging from determining chains of events leading to failures in industrial environments to identifying influencing relations between socioeconomic variables. Existing solutions to this problem usually require the involved time series to be available for a prolonged period of time, having a large sample count. This might be an acceptable requirement in industrial environments but not in case of problems from the socioeconomic sector, where time series of monthly aggregates is a typical scenario.

If two time series, $A(t)$ and $B(t)$, are connected in such a way that changes in $A(t)$ can cause changes in $B(t)$, then there is said to be a causal relation between the time series $A \to B$. Quantifying such a relation requires more than merely correlation because of the directionality of the relation $A \to B$ due to the symmetric nature of correlations.

We target causal inference scenarios where the sample count cannot be considered big and we develop a framework that analyzes the time series and determines the existence of causal relations while also identifying their direction.

This paper presents a hypothesis testing framework that, given some assumptions, can be used to reject a null-hypothesis of no causal relation between two time series in a sense that is explained later. The advantage of this new framework is that it is designed to be used when relatively few samples are available in the given time series. This makes the framework especially useful for national statistical agencies, for which time series with more than a few hundred samples are very rare.

[1]Scuola Normale Superiore, Pisa
[2]Eindhoven University of Technology
[3]Statistics Netherlands

This new framework can provide a powerful new tool in assessment of the effects of government policies in a wide range of settings.

1.1 Problem Introduction

The generic problem of causal inference can be formulated as the problem of finding causal relations between a set of quantities. More formally we are given a set of K quantities $(X_k)_{k=1}^K$, these are sampled over some time interval resulting in n samples $(X_k(i))_{i=1}^n$. The goal is to uncover causal relations of the type $X_i \to X_j$ where this relation is interpreted as changing or influencing the values of X_i will result in changes in X_j. A detailed introduction to causal inference can be found in Pearl (2009) and references therein. We focus only on two series, that we denote by X and Y to ease notation and eliminate the subscript.

A typical domain where such a problem arises is that of governmental decision support. Policymakers decide on the goals that they want to achieve. In order to devise policies leading to the desired results it is very important to properly understand the causal relations of the socioeconomic landscape. As governments are trying to gather a broader, more comprehensive, evidence base for their decisions and verifying their results, the importance of causal inference methods capable of dealing with the specific properties of such socioeconomic data is increased.

CBS (Statistics Netherlands) is the national statistical agency of the Netherlands, responsible for all official statistics of and about the country. In recent years, the activity of Statistics Netherlands has expanded to provide more in-depth technical analyses of all data that it collects and archives. The demand for such technical analyses has increased because all government ministries are focusing more on the evidence base for policy decisions, as well as on measurements and evaluations of the effects such decisions have had in terms of societal changes.

A specific case study, for which we determine the causal relationship between data sets, is the effect of government decisions on the prevalence of non-tectonic earthquakes in the Dutch province of Groningen. In 1959 the largest natural gas field in Europe and current tenth-largest natural gas field in the world was discovered in the Dutch province of Groningen (Whaley (2009)). For some decades earthquakes of modest magnitudes have occurred in the Groningen gas field. It is recognized that these events are induced by the extraction of gas from the field. The earthquakes, together with the ground subsidence, have caused damage to housing and other infrastructure over much of the region of north-east Groningen. Following an $M = 3.6$ event near Huizinge in 2012, the 6th earthquake with $M \geq 3.0$ in the province since 2003, and motivated by the public concern raised by these events, an extensive study program was started into the understanding of the hazard and risk due to gas production-induced earthquakes, *cf.* BV (2016).

The Dutch government is supporting financial compensation schemes for inhabitants of the Groningen gas field region who have suffered from the damages caused. It also has reduced the upper limit of the amount of gas that the operator NAM is allowed to extract from the reservoir on several occasions. To enable informed decisions about the level of financial compensation, Statistics Netherlands carries out research on the housing market and reports time series of various relevant indicators Posthumus et al. (2017). However, there is still some uncertainty regarding the mechanisms by which the gas production affects the generation rate of earthquakes in the region. For this reason, Statistics Netherlands works with the inspector Staatstoezicht op de Mijnen, State Supervision of Mines, to assess whether apparent changes/variations in earthquake rates are statistically significant Pijpers (2014a,b) and publishes semiannual updates, the most recent of which are Pijpers (2017); Pijpers and van Straalen (2017). In addition research is pursued to assess the role that the production changes have had in these variations Pijpers (2015, 2016, 2018).

Central in the statistical problem is the question whether or not the production changes imposed by the Dutch government have in fact been causal in reducing earthquake rates. This is a particular example of the more general type of questions posed increasingly often by ministries: to what extent has changing government policy been causal in the registered (un)desired behavior of the policy target. Part of assessments of government policy is the question of causality. It can be argued that, while it is possible to demonstrate that one event or phenomenon cannot be (or have been) the cause of another event or phenomenon, the opposite is impossible. Even the former task is often hampered by the fact that in a complex system a structural change can be quite subtle compared to larger stochastic variations.

In this report we focus on determining relationships between data sets, and in particular on determining causal relationships between data sets. Determining causal relationships between data sets allows one to quantify the effect of policy decisions on real world observable quantities, such as GDP or corporate investment spending. We are aware that causality is a highly sensitive topic due to the many connotations associated with and the different definitions of causality in various different fields in the scientific community. To discriminate between these different notions of causality, we provide a precise definition of causality that can be operationally tested for.

Relevant concepts of interest and of potential use in official statistics are 'information flow' and 'information dissipation', for which we refer to Liang (2014b,a); Daniušis et al. (2010). The data presumed available are time series of measurements of phenomena that may be related. The concept of 'information flow' provides a means to measure the amount of information that can flow from time series X to Y and vice versa. While a positive information flow from X to Y does not automatically imply causality, an information flow equal to 0 implies that there is no transfer mechanism for X to cause Y. The concept of 'information dissipation' may help to measure the extent to which an intervention at a single node in a complex network can affect the entire network or only a small part of it.

The problem with these concepts of 'information flow' and 'information dissipation' is that measuring the necessary quantities requires long time series with thousands of samples and preferably higher orders of magnitude. In official statistics such long time series are rarely available. The highest cadence with which time series are available is normally at 1 month intervals. Long time series are rare in this context as 1000 samples would cover more than 83 years and there are very few questions for which this is the appropriate time horizon. It is already quite rare in official statistics to have a time series exceeding 10 years, i.e. 120 samples.

Most causal inference methods rely on asymptotic results, hence their need for high sample count $n \to \infty$. In this report we devise an elaborate statistical test that checks the correlation, the time delay in the correlation, and a specific functional form between time series X and Y with any finite sample size n. Moreover, the test discriminates between the options 'X causally affects Y', 'Y causally affects X', 'X and Y are correlated without a clear causal direction' and 'X and Y have no correlation'.

The paper is structured as follows. We present the problem formulation and describe the proposed method in Section 1.2 then we illustrate the use of the method on a hypothetical example in Section 1.3. We raise awareness to some important issues that might be encountered when using such a framework in Section 1.4. The last part of the paper describes application of the method to non-blind and blind synthetic data in Section 1.5 and to the actual real world case of the Groningen gas field earthquake data in Section 1.6.

This problem was proposed for the SWI by CBS. The presented work and its conclusions are evaluated by F.P. Pijpers, the CBS representative. He discusses the applicability of the devised statistical test and the results of the application to the Groningen gas field earthquake case-study in Section 1.7.

1.2 Mathematical Details

This section describes the technical details of the proposed method. We start by listing all assumptions required by the method in Section 1.2.1, where we also list the user input needed to set up the problem. The exact definition of causality, as used throughout the report, is given in Section 1.2.2. This is followed by the presentation of the algorithm in Section 1.2.3.

1.2.1 Assumptions

In order to decide if there is a causal relation between two time series X and Y we make some assumptions. These are listed and explained in this section for the causal direction $X \to Y$. The fundamental assumption is a presumed relation between X and Y, stated by the model class assumption, Assumption 1.2.1. The model class has to fulfill some invertibility properties, that are formalized in Assumption 1.2.2. The proposed model contains a random noise component E. This noise variable needs to fulfill two conditions that are formalized in Assumptions 1.2.3 and 1.2.4.

Assumption 1.2.1 (Model class). *There is a known model class $\mathcal{M}_{\mathcal{X},\mathcal{E}\to\mathcal{Y}}$ whose elements can be written as a function f_θ parameterized by a vector $\theta \in \mathbb{R}^d$, such that there exists a nominal parameter vector θ_0 for which*

$$Y = f_{\theta_0}(X, E), \tag{1.1}$$

where E is considered as random unmeasured disturbance.

The model class assumption states that we have full knowledge about how the driving and responding time series are related to each other and how the unknown effects influence the responding time series. It is always an option to choose the model class $\mathcal{M}_{\mathcal{X},\mathcal{E}\to\mathcal{Y}}$ as a universal model family (like polynomials or splines), but that comes at a cost. Since the presented framework is statistical in nature and we focus on small sample situations, it is easy to understand that reliable knowledge about the model class would result in better statistical power compared to a universal model classes. This is a point where expert knowledge can come very handy.

It should be highlighted what happens if we assumed a wrong model class $\mathcal{M}_{\mathcal{X},\mathcal{E}\to\mathcal{Y}}$, i.e. it does not contain the actual relation between driver, noise and responding series. Just as with any mathematical results, if the assumptions are violated then the results might also be compromised. These two considerations should be kept in mind when specifying the model class.

Just knowing the model class is not enough to employ the presented framework. This chosen model class needs to fulfill some conditions that are expressed by the invertibility assumption, as follows.

Assumption 1.2.2 (Invertibility). *The model class $\mathcal{M}_{\mathcal{X},\mathcal{E}\to\mathcal{Y}}$ defined by the function $f_\cdot(\cdot,\cdot)$ is called invertible if for every element of it, it can be inverted for both its inputs*

$$\forall \theta, Y, X, E \quad Y = f_\theta(X,E) \Rightarrow X = f_\theta^{-X}(Y,E),\ E = f_\theta^{-E}(X,Y). \tag{1.2}$$

The functions $f_\cdot^{-X}$ and $f_\cdot^{-E}$ define the model classes $\mathcal{M}_{\mathcal{Y},\mathcal{E}\to\mathcal{X}}$ and $\mathcal{M}_{\mathcal{X},\mathcal{Y}\to\mathcal{E}}$ the same way as $f_\cdot$ defines $\mathcal{M}_{\mathcal{X},\mathcal{E}\to\mathcal{Y}}$.

The invertibility assumption means that the model classes $\mathcal{M}_{\mathcal{X},\mathcal{E}\to\mathcal{Y}}$, $\mathcal{M}_{\mathcal{Y},\mathcal{E}\to\mathcal{X}}$ and $\mathcal{M}_{\mathcal{X},\mathcal{Y}\to\mathcal{E}}$ are parameterized by the same parameters and the three models belonging to the same parameter vector θ are coordinatewise inverses of each other.

The noise parameter E of the functional relation is considered as random noise in the system. The remaining two assumptions are related to the statistical properties of this noise term.

Assumption 1.2.3 (Noise causality). *We assume that the noise series E is independent of $\mathcal{M}_{\mathcal{X},\mathcal{E}\to\mathcal{Y}}$ and of θ_0.*

The noise causality assumption ensures that the noise distribution can be defined separately, irrespective of the parameterization of the model and the nominal parameter vector θ_0. In order to make statistical statements we need to make some assumptions about the noise E. Since we focus on a scenario with small sample count we cannot make very stringent assumptions about the distribution of the noise. Group invariance, as given in Assumption 1.2.4 is a mild assumption about the joint distribution of E.

Assumption 1.2.4 (Group invariance). *Let the full noise E be defined over the probability space $(\Omega, \mathcal{F}, \mu)$ as*

$$\forall A \in \mathcal{F} \quad \Pr(E \in A) = \mu(A) \tag{1.3}$$

for any measurable set $A \in \mathcal{F}$. The measure is group invariant under a group $(\mathcal{G}, \cdot)$ if

$$\forall G \in \mathcal{G}, \forall A \in \mathcal{F} \quad \mu(A) = \mu(GA), \tag{1.4}$$

where $\mathcal{G}$ is a set of mappings from Ω into itself.

The definition of group invariance might seem very abstract, let us illustrate it with an example. We can choose $\Omega = \mathbb{R}^n$ and $\mathcal{F}$ be the usual σ-algebra of Borel sets over $\mathbb{R}^n$. Standard examples are symmetric noise, exchangeable noise or sum defined noise.

If the noise distribution is assumed to be symmetric around zero, then one choice of $\mathcal{G}$ can be the set of $n \times n$ diagonal matrices with ± 1 entries on the diagonal. In case of exchangeable noise distributions (like independent and identically distributed noise) $\mathcal{G}$ can be represented as the set of $n \times n$ permutation matrices. When the distribution of E is only a function of $\sum_{t=1}^{n} E(t)$, then $\mathcal{G}$ can be the set of doubly stochastic matrices.

1.2.2 Definition of Causality

Now we present the framework to detect a causal relation between signals X and Y. This section defines precisely what we understand under causality and what needs to be provided by the user of the method in order to obtain an answer. This description is quite abstract, but a clear illustration will be given afterwards in Section 1.3. The user needs to specify as input the following:

- Model classes $\mathcal{M}^{X \to Y}$ and $\mathcal{M}^{Y \to X}$ along with their common parameterization using a vectors $\theta \in \mathbb{R}^d$.
- Estimators for both model classes that can estimate the value of θ.
- Group of transformations $(\mathcal{G}, \cdot)$ under which the joint noise distribution is assumed to be invariant.
- Statistical test for the noise causality given by Assumption 1.2.3 and a noise causality confidence threshold p_{nc}. The test can provide a p-value that shows the plausibility of a triplet (X, Y, θ). Due to the invertibility assumption this triplet uniquely defines the value of E as well, so the test reflects on the statistical properties of E with respect to X and Y given some parameter θ. The user needs to select a confidence level p_{nc} that specifies the confidence level used to decide if the noise causality assumption is fulfilled or not in either of the possible causal directions.
- Subset $\mathcal{C} \subset \mathbb{R}^d$ of the parameter space that is considered as the causality domain. Models corresponding to parameters $\theta \in \mathcal{C}$ are considered to represent a relation between X and Y that can be considered causal for the purposes of the user.
- Model causality confidence threshold $p_{\mathcal{C}}$. The algorithm, outlined later in Section 1.2.3, associates p-values to every parameter vector $\theta \in \mathcal{C}$. The user-specified threshold value $p_{\mathcal{C}}$ is used to decide if there is at least one model in $\mathcal{C}$ that is plausible enough to believe that causality is present.

Definition 1.2.5 (Causality)**.** We say that there is a causal relationship $X \to Y$ between X and Y under the model structures $\mathcal{M}^{X\to Y}$ and $\mathcal{M}^{Y\to X}$ with significance levels p_{nc} and $p_{\mathcal{C}}$ if and only if

1. the noise causality test performed on the noise values $\hat{E}^{X\to Y}$ belonging to the estimate $\hat{\theta}^{X\to Y}$ for the model structure $\mathcal{M}^{X\to Y}_{\mathcal{X},\mathcal{E}\to\mathcal{Y}}$ results in a p-value that is greater than the user given threshold p_{nc} and

2. there is at least one model θ in the causality set $\mathcal{C}$ such that the p-value of the causality hypothesis test with null hypothesis $H_0 : \theta = \theta_0$ and alternative hypothesis $H_1 : \theta \neq \theta_0$ is greater than the user given threshold $p_{\mathcal{C}}$ and

3. at least one of the previous conditions does not hold for the direction $Y \to X$.

1.2.3 Algorithm

This section describes the algorithm step-by-step that decides on the existence of a causal relationship between X and Y in the sense of Definition 1.2.5. The flowchart representing the algorithm is give in Figure 1.1.

The procedure of determining the causal direction starts with specifying every ingredient given in Section 1.2.2. This entails the model structures and properties of the presumed noise in the data. Beside the theoretical assumptions the input contains the causality domain $\mathcal{C}$ along with the confidence thresholds p_{nc} and $p_{\mathcal{C}}$.

The simplest model structures express relations between time series as a relation between samples at corresponding time instances. For such model structures it can be beneficial to adjust the timing of the signals with respect to each other by an estimated shift $\hat{\tau}$. This is an optional step and time delay can also be incorporated into the model class $\mathcal{M}$ as well.

In most cases when there exists a specific time delay between signals it can be determined very accurately on its own. Thus, we decided to include it as a separate step in the algorithm. In the examples presented later we determine the dominant time delay $\hat{\tau}$ to be the delay that maximizes correlations between $X(t+\tau)$ and $Y(t)$.

In order to decide between the causal directions $X \to Y$ and $Y \to X$ we specify two model structures $\mathcal{M}^{X\to Y}$ and $\mathcal{M}^{Y\to X}$ in such a way that they match each other exactly if there is no noise but they are not coordinatewise inverses of each other with respect to noise, as given in the following definitions.

Definition 1.2.6 (Noiseless match)**.** The two model structures $\mathcal{M}^{X\to Y}$ and $\mathcal{M}^{Y\to X}$, defined by the functions f and g, match each other in the absence of noise if

$$Y = f_\theta(X,0) \Leftrightarrow X = g_\theta(Y,0) \qquad \forall\theta, X, Y. \tag{1.5}$$

Definition 1.2.7 (Coordinatewise inverse)**.** The two model structures $\mathcal{M}^{X\to Y}$ and $\mathcal{M}^{Y\to X}$, defined by the functions f and g, are coordinatewise inverses of each other if

$$f_\theta^{-E}(X,Y) = g_\theta^{-E}(Y,X) \qquad \forall\theta, X, Y. \tag{1.6}$$

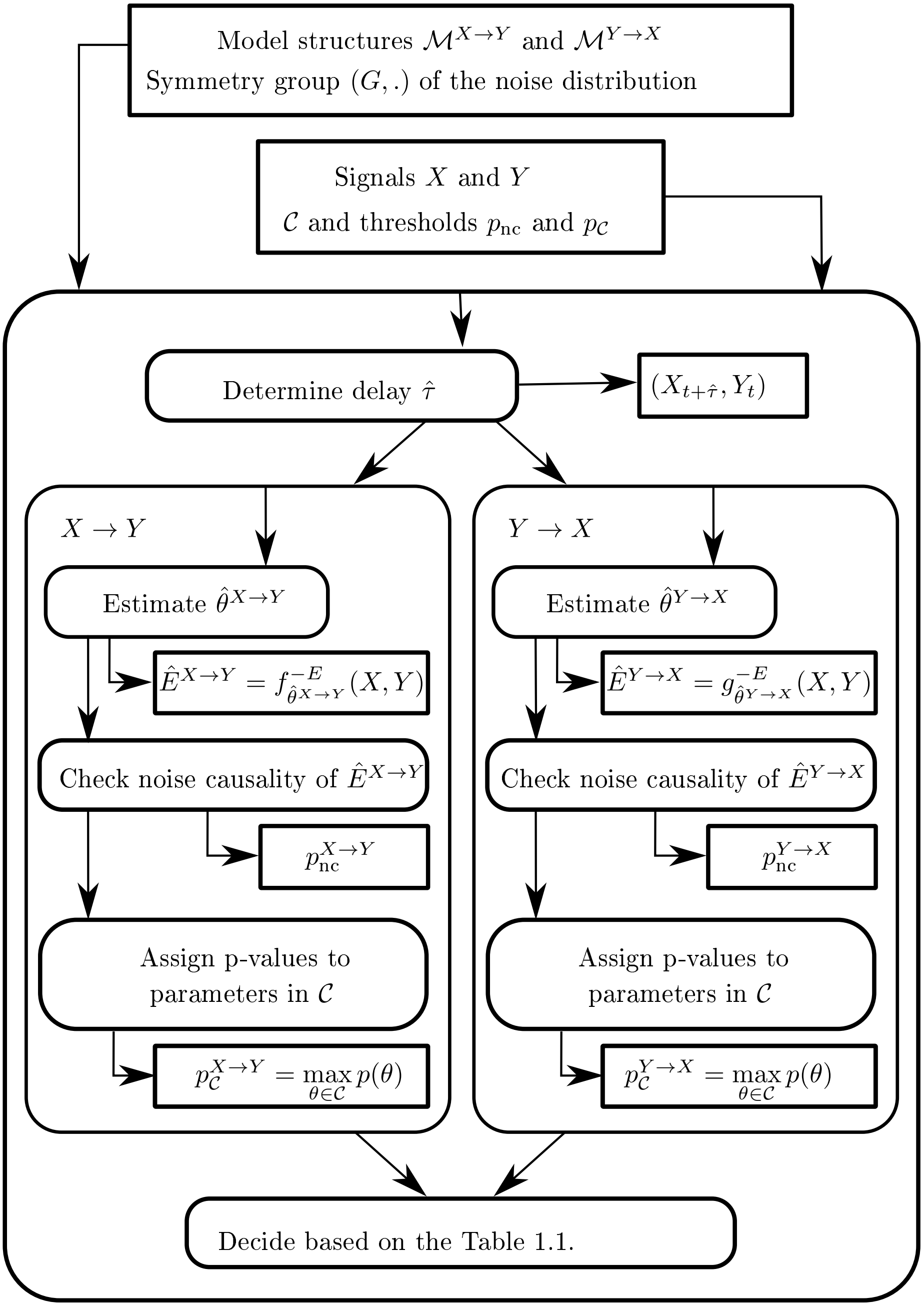


Figure 1.1: Flowchart describing the steps of the algorithm

The rationale behind choosing model classes for the two directions that match each other in the noiseless case is that this choice makes comparison of the two directions meaningful. It also allows a common parametrization of the models, thus the causality domain $\mathcal{C}$ is also common for both directions.

If the model classes would be coordinatewise inverses of each other then the two model classes would describe the same relation between signals, just from a different perspective, one from (X, E) to Y, while the other from (Y, E) to X. The requirement of this difference is what allows discriminating between the causal directions based on the assumed statistical properties of the noise E.

The model structure of both directions comes with an estimation procedure for each direction. This allows estimation of the model parameters for both direction

$$\hat{\theta}_{X\to Y} = \hat{\theta}^{X\to Y}(X, Y) \qquad \hat{\theta}_{Y\to X} = \hat{\theta}^{Y\to X}(Y, X), \tag{1.7}$$

which in turn allows expressing the two noise series for the two causal directions

$$\hat{E}_{X\to Y} = f^{-E}_{\hat{\theta}_{X\to Y}}(X, Y) \qquad \hat{E}_{Y\to X} = g^{-E}_{\hat{\theta}_{X\to Y}}(Y, X). \tag{1.8}$$

At this point we can execute the hypothesis test provided by the user to check the p-value for the noise causality assumption. If we denote the two p-values corresponding to the two directions as $p^{X\to Y}_{\mathrm{nc}}$ and $p^{Y\to X}_{\mathrm{nc}}$, then these are compared to user given threshold p_{nc}. If the p-value corresponding to a direction is smaller than the given threshold, then we conclude that that direction is not feasible. This leaves three options.

- If both hypothesis tests reject the noise causality assumption, then we conclude that we chose a wrong model and noise structure, the whole process should be restarted.
- If the noise causality assumption is rejected for one direction, but not in the other, then the algorithm can continue on the prevailing branch.
- If both directions accept the noise causality assumption, then the algorithm continues on both branches.

Based on the accepted noise causality and the group invariance properties, a p-value can be associated to every parameter θ in the causality region $\mathcal{C}$. This can be done using data perturbation methods, as described in Kolumbán et al. (2015); Kolumbán (2016). This is a particular choice for certifying the relevance of different models, but other methods can also be used.

- If there is at least one parameter in the causal region that is not rejected in only one causal direction, then we conclude that causality is detected in that direction.
- If there is at least one parameter in the causal region that is not rejected in both causal directions, then we conclude that it seems there is a strong connection in

both direction but we could not discriminate between them. This we translate into a conclusion that there is a clear connection between the time series, but no clear causality direction can be detected.

- If there are no accepted parameters in the causal domain in either of the directions, then we conclude that there is no causal relation between the two time series.

The conclusion of the algorithm based on the different parameters and calculated values is given in Table 1.1.

$p_{\mathcal{C}}^{Y\to X} > p_{\mathcal{C}}$ (columns 2–3)

$p_{\mathrm{nc}}^{Y\to X} > p_{\mathrm{nc}}$ (columns 1–2)

$p_{\mathrm{nc}}^{X\to Y} > p_{\mathrm{nc}}$ (rows 1–2)

$p_{\mathcal{C}}^{X\to Y} > p_{\mathcal{C}}$ (rows 2–3)

<table>
<tr><td>no causal link</td><td>$Y \to X$</td><td>$\mathcal{M}^{Y\to X}$ rejected</td><td>no causal link</td></tr>
<tr><td>$X \to Y$</td><td>correlation</td><td>$X \to Y$</td><td>$X \to Y$</td></tr>
<tr><td>$\mathcal{M}^{X\to Y}$ rejected</td><td>$Y \to X$</td><td colspan="2" rowspan="2">noise assumptions rejected</td></tr>
<tr><td>no causal link</td><td>$Y \to X$</td></tr>
</table>

Table 1.1: Decisions based on the values $p_{\mathrm{nc}}^{X\to Y}$, $p_{\mathrm{nc}}^{Y\to X}$, $p_{\mathcal{C}}^{X\to Y}$ and $p_{\mathcal{C}}^{Y\to X}$

1.2.4 Accurateness of functional relation

The framework outlined in the previous section relies on an appropriate choice for the functional relationship of interest. One could think of several commonly used statistical tests to verify this assumption. For illustration purposes, we present an example under the assumption of additive noise. More specifically, we assume

$$Y(t) = f_{\theta_0}(X(t)) + E(t), \tag{1.9}$$

where E is a random noise term and $f_{\theta_0}(\cdot)$ denotes the real relation between time series $X(t)$ and $Y(t)$.

We can choose to estimate θ_0 using standard methods, like non-linear least squares. Let $\hat{\theta}$ denote this estimate. This results in the noise estimate

$$\hat{E}_{X\to Y} = Y - f_{\hat{\theta}}^{-1}(Y). \tag{1.10}$$

To test the noise causality assumption, Assumption 1.2.3, the residuals $\hat{E}(t)$ are binarized by comparison to the median of all $\hat{E}(t)$, in order to obtain $\hat{E}_B(t)$. Next, we proceed by ordering $\hat{E}_B(t)$ with respect to increasing order of $X(t)$. To test for independence between X and $\hat{E}_B$ and within the new sequence $\hat{E}_B$ we apply the Wald-Wolfowitz runs test, for which we refer to Sheskin (2004)[Chapter 10]. If the hypothesis of independence is rejected at a level of p_{nc}, we will reject Assumption 1.2.3 and stop our procedure. If independence is not rejected, we proceed with the next step of our outline.

If binarized values ordered according to the magnitude of X at the corresponding time fail the Wald-Wolfowitz test it means that independence from X can be rejected, rejecting Assumption 1.2.3.

1.2.5 I.I.D. and Exchangeable Noise

Once the noise causality assumption is not rejected, we proceed to checking the statistical hypothesis that serves as foundation for certifying models in the causality region. Our suggestion for using data-perturbation methods relies on the group invariance assumption, Assumption 1.2.4. This section presents the hypothesis test corresponding to the assumption of exchangeable noise, such as independent and identically distributed (i.i.d.) noise, where the group of permutations is the invariance group. One possible choice for testing if a series of variables is an independent sequence is the Wald-Wolfowitz test. In this context the test is used on the binarized version of $\hat{E}_{X\to Y}$ directly to check if that is an independent sequence.

It is of great importance to note that, if $X \to Y$ is the true causal direction, then assuming the causal relation in the wrong direction $Y \to X$ the hypothesis for both Assumption 1.2.3 and Assumption 1.2.4 should be rejected in theory. The model class for the direction $Y \to X$ is defined as $X(t) = f_{\theta_0}^{-1}(Y(t)) + \tilde{E}$, with an independent noise $\tilde{E}$. Using the relation of the direction $X \to Y$, we can write

$$X(t) \quad = \quad f_{\theta_0}^{-1}(Y(t)) + (X(t) - f_{\theta_0}^{-1}(Y(t))) = f_{\theta_0}^{-1}(Y(t)) + \tilde{E},$$

so that $\tilde{E} = X(t) - f_{\theta_0}^{-1}(Y(t))$ which is obviously not independent of neither Y, nor θ_0.

1.3 Illustrative Example

Since the definition of the framework was given in a rather abstract manner, we give an hypothetical example in this section. This illustrates the general methodology of finding causal relations between two time series that are relatively short. We prove an example for every building block of the method. This is done using an applied example that is not used later on in the report.

To illustrate how to use our method to find causal relations in the sense described above, we employ a situation that can arise in official statistics: *does there exist, and if so describe, a causal relation between the yearly time series of cigarette prices and the percentage of smokers in the population?*

Functional relation: A-priori it is unclear what the causal relation should be. Did the price increase to keep revenue from dropping due to lower percentage of smokers? Did the percentage of smokers drop due to the too high price of cigarettes for a part of the population? Moreover, it is a-priori not clear what the functional relation should be. Or was there a mutual feedback effect?

The problem imposes several natural constraints: the percentage must be between 0 and 100, the price must be positive, the addictive nature of smoking and the income inequality of the population shows that percentage does not drop to 0 if prices tend towards infinity, a price of 0 for the cigarettes does not relate to a 100% share of smokers in the population, and higher prices should give lower percentage of smokers due to socioeconomic theories.

Let S denote the percentage of smokers and let P denote the price of a pack of cigarettes. Then, an educated guess of the functional relationship is

$$S(t) = f_{\theta_0}(P(t)) + E(t) = a_0 + \frac{b_0}{P(t-\tau) + c_0} + E(t),$$

with $a_0, b_0, c_0, \tau > 0$ and $t = 1, 2, \ldots, 12$ denotes the monthly aggregate time resolution. τ denotes the dominant time delay and the parameter vector is composed of $\theta_0 = [a_0,\ b_0,\ c_0]^T$.

Noise Causality Assumption: We assume that the random noise E is independent of P and is an i.i.d. sequence. We set the p-value for all hypothesis tests at 0.95 and start our method.

Estimating of parameters: We use the correlation function to estimate τ with $\hat{\tau}$ and we estimate the values of a, b and c with $\hat{a}$, $\hat{b}$ and $\hat{c}$ using a least-squares method.

Residual Noise: Then, we determine $\hat{E}$ with

$$\hat{E}(t) = S(t) - \hat{a} - \frac{\hat{b}}{P(t-\hat{\tau}) + \hat{c}}.$$

Plausibility of the chosen function relation: Testing whether or not $f_{\hat{\theta}}$ accurately represents f_{θ_0} can be done based on the residuals $\hat{E}$. They should be independent of P if $\hat{\theta}$ is an accurate representation of θ_0. We can check this condition by looking at the inner product between the centered versions of $\hat{E}$ and P. This check effectively looks at how often the two sequences have deviation from their average with the same or opposite sign. For an independent random process the sum of these product values should be near 0, while correlated sequences will differ significantly from 0. Correlation would mean that the noise is dependent on P and, hence, $f_{\hat{\theta}}$ is not an accurate representation of the true relation.

If the hypothesis based on Assumption 1.2.3 is rejected, then a different model class should be chosen. If the hypothesis is accepted, then $f_{\hat{\theta}}$ does accurately represent f_{θ_0} and, thus, one can proceed with testing the hypothesis corresponding to Assumption 1.2.4.

Testing the residual noise: To test Assumption 1.2.4, we have to check two conditions: $\hat{E}(t)$ is random, and $\hat{E}$ is identically distributed. Tests for this were suggested in previous sections based on the Wald-Wolfowitz test.

If the hypothesis of Assumption 1.2.4 is rejected, then either $\hat{E}$ is not random or not identically distributed. This would indicate that a hidden behavior is not taken into account in f_θ. Therefore, a different model class should be chosen. If the hypothesis is accepted, then we can proceed to checking models in the causality domain. A possible choice of the causality domain could be

$$\mathcal{C} = \left\{\theta \in \mathbb{R}^3 : |\theta(2)| > 0.1\right\}. \tag{1.11}$$

In this case, the model structure $\mathcal{M}^{X \to Y}_{\mathcal{X},\mathcal{E} \to \mathcal{Y}}$ is defined as

$$S(t) = a + \frac{b}{P(t-\tau) + c} + E^{X \to Y}(t), \tag{1.12}$$

while $\mathcal{M}^{X \to Y}_{\mathcal{Y},\mathcal{E} \to \mathcal{X}}$ is defined as

$$P(t) = \frac{b}{S(t+\tau) - E^{X \to Y}(t+\tau) - a} + c. \tag{1.13}$$

This should be compared to $\mathcal{M}^{Y \to X}_{\mathcal{Y},\mathcal{E} \to \mathcal{X}}$, defined as

$$P(t) = \frac{b}{S(t+\tau) - a} + c + E^{Y \to X}(t). \tag{1.14}$$

Let us assume that the noise estimates belonging to the two directions both passed the statistical tests. Furthermore, there are parameters in $\mathcal{C}$ that cannot be rejected with the specified confidence. In this case, we would conclude that there is a clear relation between the two quantities but we cannot determine any directionality.

1.4 Incorrect Use or Interpretation

Our notion of causality refers to noise properties (the Noise Causality Assumption 1.2.3), and to an educated guess for the functional relationship (Assumption 1.2.1). The use of our method is, therefore, highly dependent on the user knowledge about the problem.

In this section we treat problems that are specific to our method of detecting a causal relation and are not necessarily part of the familiar statistical paradoxes. All problems shown in this section come down to the following advice to any user of our method: without knowledge of your problem set-up, procedure of data gathering, amount of noise in the data, interpretation of the results of our method could lead to incorrect conclusions on which harmful policies can be based.

There are a few remarks that we would like to make about possible pitfalls. One of these is that there are multiple hypothesis tests performed on the same dataset before reaching a conclusion, even on one directional branch. These are by no means independent of each other, so claiming an overall confidence level would be a mistake.

It can easily happen that we are given two time series X and Y but in fact both of these are influenced by a third unmeasured signal with different delays. In such cases it can happen that the algorithm indicates a causal direction that would go against temporal order of events. This still fits our definition of causality.

Another potential situation where the method would be prone to detecting causal relations in the wrong direction is where the signal to noise ratio in X is much larger than for Y. In such cases the algorithm is prone to not rejecting an invalid functional relation in the direction $X \to Y$ because the modeling error with respect to the true functional relation is not big enough to be statistically significant. In such cases on the $X \to Y$ direction prevails from the two possibilities. We note that, if the functional relation is correct, then this issue is not present.

1.5 Synthetic Data Sets

When developing an algorithm, it is helpful to examine its efficacy using synthetic data. Since the desired results are known, a comparison of the actual and desired results is possible and the efficacy of the algorithm can be judged in a setting similar to future use cases.

The synthetic data, in this case a few time series, can be prepared with particular properties or complications that mimic real data. We used synthetic data to explore particular issues that may stretch the capabilities of the algorithm, or even violate some of its assumptions. If, in the latter case, the algorithm does indeed fail to function as desired, this provides a powerful argument for the need to ensure that the data or the problems, to which this algorithm is applied to, do satisfy such constraints.

To ensure that the algorithm is applied as objectively as possible to each problem, three tests are carried out ‘blind’. The CBS representative (the fourth author) prepared several synthetic datasets to be employed in the algorithm tests, keeping in mind the typical problems and issues that data of a national statistical agency (NSI) will suffer from. The intended direct application, at least in the first instance, is by NSIs. The other members of the research team (all authors) were given very little information concerning the process by which the time series were constructed. The results of these data set test-runs are explained in Section 1.5.2.

$(X \to Y)$	$\hat{\alpha}_{X\to Y} = 3.93792$	$\hat{\beta}_{X\to Y} = 2.01604$
$(Y \to X)$	$\hat{\alpha}_{Y\to X} = 4.00647$	$\hat{\beta}_{Y\to X} = 2.09265$
true value	$\alpha = 4$	$\beta = 2$

Table 1.2: Estimated parameters for the non-blind test

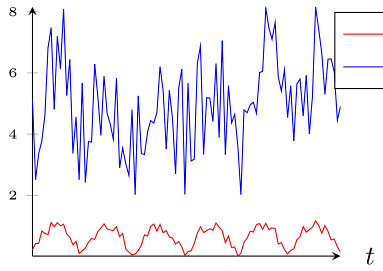


Figure 1.2: Data of the non-blind test

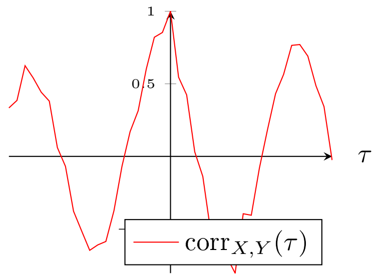


Figure 1.3: Correlation of X and Y

1.5.1 Non-blind synthetic data set

To get an idea about the power of our method we apply it to about a dozen synthetic data sets ranging from combining multiple independent time series to more functional relations between the time series. We explain one of these non-blind tests to show the algorithm once more in detail.

Consider two time series $X(t), Y(t)$ with 200 data points, e.g. $t = 1, 2, \dots, 200$, as shown in Figure 1.2. The first step is to identify the possible time shift between the times series, which we conjecture to maximize the shifted correlation:

$$\mathrm{corr}_{X,Y}(\tau) = \sum_{t=\max\{0,\tau\}}^{\min\{T,T+\tau\}} X(t) \cdot Y(t-\tau). \tag{1.15}$$

In our case this is the this case for 0, see Figure 1.3. As second step, we guess the function form (correctly) to be

$$Y(t) = \alpha + \beta X(t)^2 \qquad X(t) = \frac{1}{\sqrt{\beta}} \sqrt{|Y(t) - \alpha|}.$$

The third step is to fit the model to the data. The fitted values are given in Table 1.2.

Using these fitted models we compute the implied/residual noise

$$\hat{E}_{X\to Y}(t) = \hat{Y}(t) - \hat{\alpha}_Y + \hat{\beta}_Y X(t)^2, \quad \hat{E}_{Y\to X}(t) = X(t) - \frac{1}{\sqrt{\hat{\beta}_X}} \sqrt{|Y(t) - \hat{\alpha}_X|}.$$

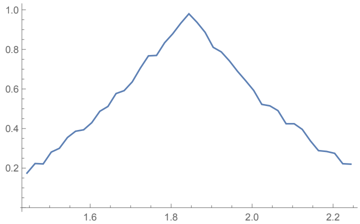


Figure 1.4: Region of trust, $X \to Y$ for parameter β.

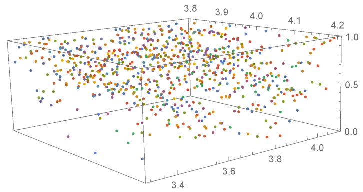


Figure 1.5: Region of trust, $Y \to X$ for parameters α and β.

Using the Wald-Wolfowitz test for both residual noises, we see that is very plausible that both are random. To finally make a statement about the causal relation, and even the direction of this relations we check whether the noise $\hat{E}_Y$ is independent of X and whether $\hat{E}_X$ is independent of Y. Using Spearman's rank test and obtain:

$$(\hat{E}_{X\to Y}, X) \qquad p - \text{value} = 0.965 \text{ independence very plausible}$$
$$(\hat{E}_{X\to Y}, Y) \qquad p - \text{value} = 0.0292 \text{ independence unlikely}$$

Thus, we have concluded that most likely X is causing/driving the signal Y, which is the case as the data was generated using

$$Y(t) = 4 + 2X(t)^2 + 2U_{y,t} \qquad X(t) = \left|\sin\left(6\pi\frac{t}{200}\right) + 0.2U_{x,t}\right|,$$

where $U_{y,t}, U_{x,t}$ are i.i.d real-valued random variable chosen uniformly form the interval $[-1, 1]$.

The last step is to assign p-values to parameters in the causality domain. These values can be calculated by relying on the assumption that the noise is an i.i.d. sequence. Figure 1.4 shows these p-values for the β parameter in the $X \to Y$ direction. If, for example, the causality domain $\mathcal{C}$ is defined as $\{|\beta| > 0.1\}$ then the algorithm would conclude that indeed there is a causal relation $X \to Y$. We note that the α dimension of this figure is not shown because the additive constant can be considered as part of the i.i.d. noise as a shift in the expectation.

For completeness we also test the conjectured function form in the $Y \to X$ direction. These p-values are presented in Figure 1.5. We see no clear structure here, but this is irrelevant, as these p-values are calculated based on the i.i.d. noise assumption which was already rejected for this direction.

For this setting our method worked beautifully. After further testing we found that this method remains useful until the signal consist of around 50% noise (we used around 20%), then the perturbation became so large that the result of the Spearman's rank test were not conclusive anymore.

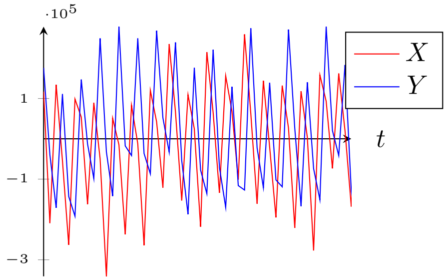


Figure 1.6: Part of the first data set

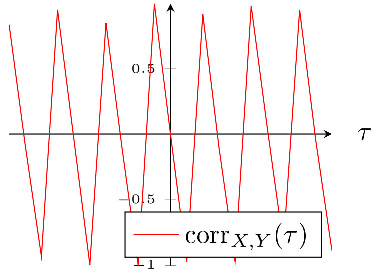


Figure 1.7: Correlation of Y and Y

1.5.2 Blind synthetic data set

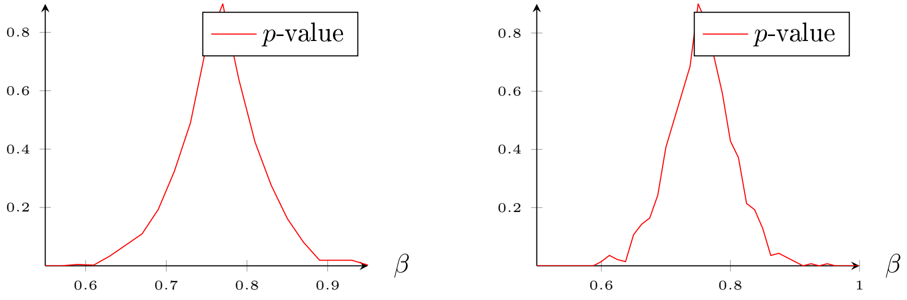
The team was given three sets of time series by the CBS representative (the fourth author) to carry out 'blind' test and demonstrate the power of the proposed method. The aim is as asses whether a particular expectation for the direction of causality can be verified/falsified from their data. For NSIs this is a realistic setting, as it is not unusual that a third party wishes to assess whether a particular expectation for the direction of causality can be verified/falsified from the data.

First data set The first data set consisted of two alternating signals, as displayed in Figure 1.6, we compute the shifted correlation (1.15) to find the most plausible shifts, see Figure 1.7. The periodicity of the correlation suggests some periodicity in the signals as well. We chose the shift that was minimal and with the highest absolute value $\hat{\tau} = -1$.

Examining the shifted time series, a linear fit seamed reasonable. We fitted the models accordingly and computed the residual noise. The Spearman rank test yields that both directions are plausible, with p-values 0.6735 and 0.38656 also indicating a possible dominant direction, namely $Y = \alpha + \beta X + E$.

We go forward with assigning p-values to different parameters values. As already mentioned earlier, an additive parameter α is irrelevant in case of an additive i.i.d. noise model. The p-values for parameter β are given on Figure 1.8. Based on these figures we cannot distinguish between the two directions so the algorithm would conclude that there is a strong relation between the signals but no directionality can be determined. This is due to the simultaneous presence of two characteristics of the model for the direction $Y \to X$, that is $X = \frac{1}{\beta}Y - \frac{1}{\beta}(\alpha + E)$. On the one hand the functional form matches the tested one. On the other hand the noise $-\frac{1}{\beta}(\alpha + E)$ is also an i.i.d. sequence. This illustrates the importance of nonlinear relations for detecting causal relations.

Second data set In practice one might want to analyze relations between multiple properties. Our method can be extended to more than two time series, to which we



Direction $X \to Y$. Direction $Y \to X$.

Figure 1.8: Reliability of the linear description as function of β.

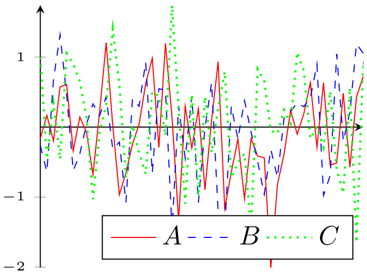


Figure 1.9: Part of the second data set.

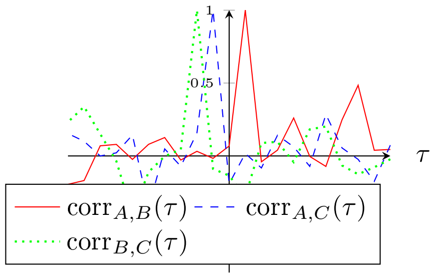


Figure 1.10: Correlation functions.

only restricted for representational purposes. The second data consists of three time series, denote by A, B and C, see Figure 1.9.

First, we computed the pairwise correlation to identify the time shift, see Figure 1.10. Comparing the shifted time series we conjecture a functional form to be:

$$C(t) \quad = \quad \alpha_C A(t-1) + \beta_C B(t-2) + E_{A,B\to C}, \tag{1.16}$$

or other combinations of this:

$$\begin{aligned} A(t) &= \beta_A B(t-1) + \gamma_A C(t+1) + E_{B,C\to A}, \\ B(t) &= \alpha_B A(t+1) + \gamma_B C(t+2) + E_{A,C\to B}. \end{aligned}$$

For each combination, we fitted the parameters and compute the residual noises. Then, we use the Spearman rank test to conclude whether the residual noise is independent of the drivers. As can be see in Table 1.3 the function form stated in (1.16) is the only plausible from (assuming a linear dependence). Figure 1.11 shows the p-values assigned to different parameter values using a data perturbation method given in Kolumbán (2016). If the region with elevated p-values is considered to be part of the causality region, then we can conclude a found causal relation.

	A	B	C
$E_{A,B\to C}$	0.760615	0.907167	-
$E_{B,C\to A}$	-	$1.494 * 10^{-17}$	0.703413
$E_{A,C\to B}$	$1.367 * 10^{-12}$	-	0.989447

Table 1.3: p-value of the test for independence of the two signals.

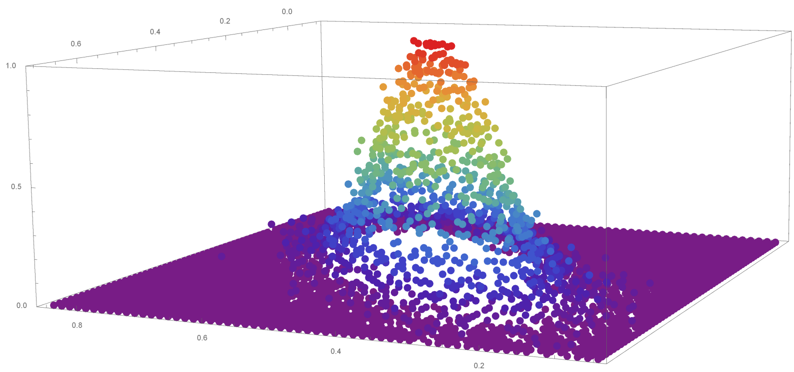


Figure 1.11: Likely region in which the parameter of the functional form (1.16).

Third data set This data set consists of two time series, shown in Figure 1.12. We compute the possible shift between the time series. Looking at the quite involved shape of the data it was hard to come up with a suitable functional form. In practice we would have asked a domain expert for a reasonable guess of the functional. As this was not an option we used a small family of standard functions (low order polynomials, exponential, logarithm) and have chosen the most suitable form that used only two parameters

$$Y(t) = \alpha + \beta X(t-1) + \log(1 + X(t-1)) + E_t. \tag{1.17}$$

Assuming zero noise, $E_t = 0$, and remembering that the Lambert W-function is defined by the relation $z = W(z)e^{W(z)}$ we can express $X(t)$ as

$$X(t) = -1 + \frac{1}{\beta} W(\beta e^{Y(t+1)-\alpha+\beta}). \tag{1.18}$$

Then, we employed our method: We fit the parameters and computed the residual noise and tested for independence of the noise to the corresponding into, which resulted in the p-values 0.079 and 0.055, respectively. Using our favorite threshold of 5%, we could accept both direction, even if we see that it might be unlikely. In the hope of gaining more insight and maybe even a clear indication as in Figures 1.4,1.5, we compute the region of trust, but could not extract any additional information.

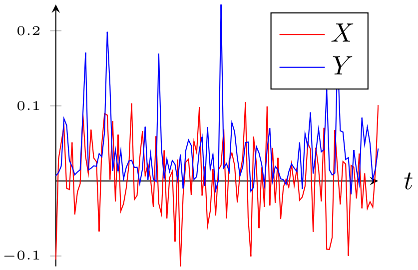


Figure 1.12: The third data set.

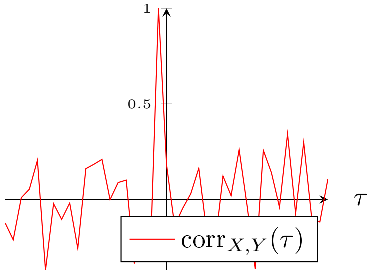


Figure 1.13: Correlation of X and Y

Comment of the fourth author on this data set: A violation of one of the assumptions of the model is introduced in setting up the synthetic dataset: correlated errors between driving and responding time series. In addition, the functional relation between the two time series was not obvious. The rest of the team was not informed of this issue, and therefore the expectation was that either no inference can be made concerning information flow, or even that an erroneous conclusion is reached. While in normal business practice every effort is made to avoid such issues concerning error correlations, there may be cases where such an issue nevertheless occurs. It is important for CBS to have an explicit example of the risks and consequences of such problems, to prevent the inexpert use of the algorithm as a black box.

1.6 Groningen Earthquake Case-study

Time series data on the gas production [4] and the induced tremor catalog[5] in the North of the Netherlands can be found online. Nowadays, there is no doubt whether the number of induced earthquakes and the amount of gas produced are correlated, see e.g. Sijacic et al. (2017). In addition, causal relations between the cumulative production and cumulative number of earthquakes have been proposed. Bourne and Oates made use of some geomechanical expert knowledge to propose a nonlinear exponential relation between the total compaction and total number of events Bourne and Oates (2015). A more empirical approach, directly linking the cumulative production since 1991 to the cumulative number of events since 1991, has been proposed by Hagoort Hagoort (2017). Since compaction data is not publicly available, we will subject Hagoort's proposed relation to our new developed machinery. In this model, $X(t)$ represents the cumulative production in Nm^3, scaled by the factor 10^{-15}, t the months after 1st of January, 1991, and $Y(t)$ the cumulative number of induced earthquakes (with magnitude ≥ 1.5) up until that time. Hagoort proposed the functional relation

$$Y(t) = a \cdot X(t)^2 + b \cdot X(t). \tag{1.19}$$

[4] `https://www.nam.nl/feiten-en-cijfers/gaswinning.html`

[5] `https://www.knmi.nl/kennis-en-datacentrum/dataset/aardbevingscatalogus`

We investigate whether it seems reasonable to assume the causal relation

$$Y(t) = f(X(t), E) = a \cdot X(t)^2 + b \cdot X(t) + E(t).$$

Firstly, the most appropriate time-delay is derived according to the maximizer of the time series correlation. Indeed, a time-delay of 0 months seems most appropriate according to the data, see Figure 1.14.

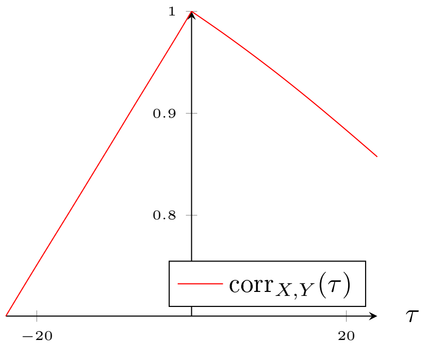


Figure 1.14: Correlation of X and Y

The NLS fit $\hat{f}(X(t))$ is attained at $\hat{a} = 0.027$ and $\hat{b} = -0.275$. A comparison between $\hat{f}(X(t))$ and $Y(t)$ is presented in Figure 1.15.

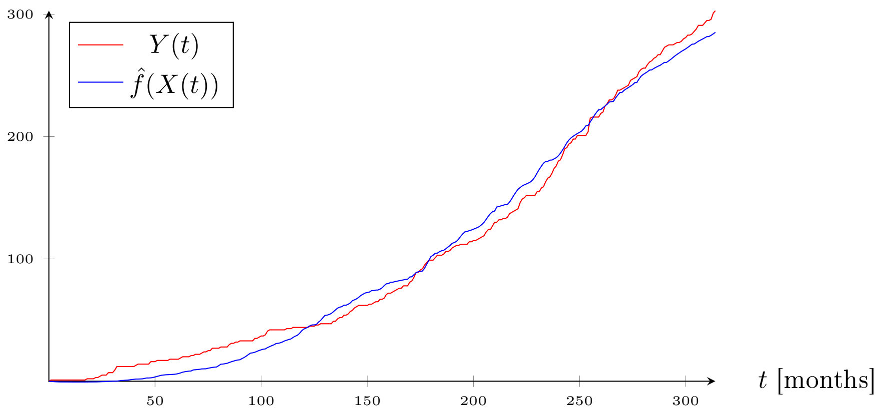


Figure 1.15: Accumulated quake count $Y(t)$ and its hypothesized model $\hat{f}(X(t))$

Subsequently, the residuals $\hat{E}(t)$ are sorted in increasing order of $X(t)$. The behavior of $\hat{E}(t)$ versus $X(t)$ can be found in Figure 1.16.

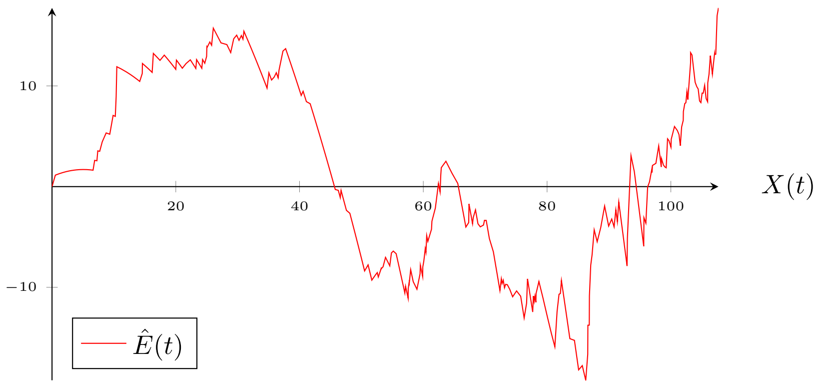


Figure 1.16: $\hat{E}(t)$ versus $X(t)$

These residuals are binarized by comparing them to the median of

$$\{\hat{E}(t), 1 \leq t \leq 325\}.$$

The $p-$value based on the Wald-Wolfowitz test equals $3.51806 \cdot 10^{-59}$, from which we clearly reject randomness of this sequence. As a result we reject hypothesis 1 and reject that the relation proposed by Hagoort is reasonable.

We conclude that geomechanical experts should think about an appropriate model class $\mathcal{M}$ relating the gas production, or derived quantities such as compaction, before our method can be of use in practice. If one would like to apply our machinery, more specifically, if one would like to use the model with i.i.d. noise, then it seems more appropriate to model the monthly number of earthquakes instead of the cumulative number of earthquakes.

1.7 Conclusion

In this paper a new method is described to assess in which direction, if any, information is flowing between two or more time series. Previous work in this area assumed that time series would be available with large numbers of samples. For the purposes of National Statistical Institutes (NSIs), this would be a nearly insuperable obstacle, since obtaining long uniform time series for official statistics is very difficult indeed: a cadence of measurements of more than once a month is not feasible, and in most cases such time series are available for 10 to 20 years at best. The statistical test presented here does not quantify an amount of information transferred, but instead tests whether a hypothesis of zero or non-zero information transfer can be rejected. This allows meaningful conclusions to be drawn with far shorter time series than the previous methods.

A convenient shorthand term for the technique is that an assessment is made of *causality*. In writing such shorthand it is important to emphasize that this term here is intended to be used with a specific technical meaning. If this method accepts ('does not reject') the hypothesis that 'variations in time series A' cause 'variations in time series B' this does not imply exclusivity: **no** statement is made as to whether an (unobserved) variable C caused variations in both A and B, rather than A directly affecting B. Also, even if variations in A cause some of the variations in B, there could nevertheless still be variables C, D etc. that **also** affect B in some way. In other words 'A causes B' is **not** the same statement as 'only A causes B'.

Most algorithms and proofs in mathematics and statistics make certain assumptions about the data or settings in which they can be applied, and this algorithm is no exception. The details are set out in Section 1.2, but it is important to emphasize that there is a crucial role for a plausible relationship between the time series, which in addition must have a non-linear term or terms. This means that an expert in the research field where this method is to be applied needs to be involved in order to provide such a context-relevant, plausible model. Another important aspect of the data is measurement noise. Perhaps paradoxically, methods that assess transfer of information require the presence of noise, i.e. intrinsic stochasticity or measurement noise. Without it, these tests cannot provide an answer. The noise must itself have some known symmetry properties (for example be i.i.d.) in order for the test to be able to provide a conclusive answer.

Using several practical examples on synthetic data, the importance of the assumptions is illustrated, by showing that a violation of the assumptions used to build the method produces either inconclusive results, which is relatively benign, or conclusive but erroneous results. Such examples provide valuable insight which can also serve as guidance when this method is to be applied to a new problem at NSIs such as Statistics Netherlands. In addition, this method is applied to one particular indicator of seismic activity in Groningen. For a long time, doubts were cast on the causal relationship between the production of natural gas from the Groningen gas reservoir and seismic events in the region. While this point of view has now changed, it is nevertheless useful to assess the value of an indicator such as the time series of cumulative numbers seismic events above a threshold magnitude. The test was applied to see whether a quadratic relationship with cumulative production appears a valid causal model. The tests demonstrate that such a model is rejected. The interpretation of this result is that the path from gas extraction to earthquakes is not purely one of compaction driving the generation of earthquakes. Furthermore, at a more technical/operational level, it appears that using non-cumulative production and earthquake numbers, i.e. rates per month or some other appropriate time interval, may work better as primary data in the framework presented here.

Bibliography

SJ Bourne and SJ Oates. An activity rate model of induced seismicity within the groningen field (part 1). nam report, 2015.

Nederlandse Aardolie Maatschappij BV. A technical addendum to the winningsplan

Groningen 2016 production, subsidence, induced earthquakes and seismic hazard and risk assessment in the Groningen field. Technical report, Nederlandse Aardolie Maatschappij (NAM), 2016.

Povilas Daniušis, Dominik Janzing, Joris Mooij, Jakob Zscheischler, Bastian Steudel, Kun Zhang, and Bernhard Schölkopf. Inferring deterministic causal relations. In *UAI'10 Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, pages 143–150, 2010.

Jacques Hagoort. Empirical model for induced earthquakes in the groningen gas field. *Unpublished manuscript. Amsterdam, Newzealand*, 2017.

Sándor Kolumbán. *System Identification in Highly Non-Informative Environment*. University Press, 2016. ISBN 978-9-4619739-8-6.

Sándor Kolumbán, István Vajk, and Johan Schoukens. Perturbed datasets methods for hypothesis testing and structure of corresponding confidence sets. *Automatica*, 51(0):326 – 331, 2015. ISSN 0005-1098. doi: http://dx.doi.org/10.1016/j.automatica.2014.10.083. URL `http://www.sciencedirect.com/science/article/pii/S0005109814004646`.

X San Liang. Causality between time series. *arXiv preprint arXiv:1403.6496*, 2014a.

X San Liang. Unraveling the cause-effect relation between time series. *Physical Review E*, 90(5):052150, 2014b.

Judea Pearl. Causal inference in statistics: An overview. *Statist. Surv.*, 3:96–146, 2009. doi: 10.1214/09-SS057. URL `https://doi.org/10.1214/09-SS057`.

F. P. Pijpers. Phase 0 report 1 : significance of trend changes in ground subsidence in Groningen. Technical report, Statistics Netherlands, 2014a.

F. P. Pijpers. Phase 0 report 2 : significance of trend changes in tremor rates in Groningen. Technical report, Statistics Netherlands, 2014b.

F. P. Pijpers. A phenomenological relationship between gas production variations and tremor rates in Groningen. Technical report, Statistics Netherlands, 2015.

F. P. Pijpers. A phenomenological relationship between reservoir pressure and tremor rates in Groningen. Technical report, Statistics Netherlands, 2016.

F. P. Pijpers. Supplementary material to CBS reports on earthquake frequencies. Technical report, Statistics Netherlands, 2017.

F. P. Pijpers. Improved time resolution relationship between pressure and earthquake rates in Groningen. Technical report, Statistics Netherlands, 2018.

F. P. Pijpers and V. van Straalen. Trend changes in tremor rates in Groningen : update October 2017. Technical report, Statistics Netherlands, 2017.

H. Posthumus, M. Joosten, and B. Peeters. Woningmarktontwikkelingen rondom het groningenveld, 1e kwartaal 1995 t/m 2e kwartaal 2017. Technical report, Centraal Bureau voor de Statistiek, 2017.

David Sheskin. *Handbook of parametric and nonparametric statistical procedures.* Boca Raton: Chapman & Hall Crc., 2004.

Danijela Sijacic, Frank Pijpers, Manuel Nepveu, and Karin van Thienen-Visser. Statistical evidence on the effect of production changes on induced seismicity. *Netherlands Journal of Geosciences*, 96(5):s27–s38, 2017.

Jane Whaley. The groningen gas field, 2009.