Abstract

Many scientific disciplines rely on observational data alone. For example, there is no current technology that can control the interaction between the solar wind and the Earth's magnetic field, so space weather studies rely on data collected without performing controlled experiments (i.e. without interventions to the system dynamics). Unfortunately, causal inference with such observational data sets is difficult. It is well known that correlation does not imply causation, which has lead to the development of several different time series causality tools. In this article, we introduce new tools for exploratory causal inference of a pair of time series data sets that are believed to be causally related. The tools, called penchants and leanings, do not depend on any assumed models for the times series, do not rely on any embedding procedures, are computationally straightforward to apply, and follow directly from assumptions of probabilistic causality, which may provide a clearer interpretation of the results than other existing time series causality tools.

Non-parametric casual inference in bivariate time series data

Robert S. Weigel

April 14, 2015

1 Introduction

Casual inference in time series analysis centers on finding "driving" relationships between different time series data sets. Showing the existence, rather than the exact nature, of the "driving" relationship between the sets is often the primary goal. Thus, words like "driving", "causality", and related terms usually do not have straightforward analogs to the same terms used in other fields [**?**, **?**, **?**], e.g. , theoretical quantum or classical mechanics. The development and study of such causal inference techniques is often called *time series causality*. Most of these techniques fall into four broad categories, those related to transfer entropy [**?**], those related to Granger causality [**?**], those related to state space reconstruction (SSR) techniques [**?**], and those related to lagged cross-correlation [**?**, **?**]. These techniques have found application in a wide range of fields including neuroscience (e.g. [**?**]), economics (e.g. [**?**, **?**]), climatology (e.g. [**?**]), and others.

In this article, we introduce another time series causality technique derived directly from the definition of probabilistic causality []. The new technique is then applied to synthetic and empirical time series data sets with known, or intuitive, causal structure. We will discuss the strengths and weaknesses of the technique and try to frame how the technique may be useful for causal inference with empirical data. Formal relationships to the other well known time series causality techniques, such as Granger causality or transfer entropy, will be the subject of future work.

2 Causal Penchant

Define the *causal penchant* as

$$\rho_{EC} := P\left(E|C\right) - P\left(E|\bar{C}\right) \quad . \qquad (1)$$

The motivation for this expression is in te interpretation of ρ_{EC} as a causal indicator; i.e. if C causes (or *drives*) E, then $\rho_{EC} > 0$, and if $\rho_{EC} \leq 0$, then the direction of causal influence in the system is undetermined. If the effect E is assumed to be recorded in one time series and the cause C is assumed to be recorded in a different time series, then the direction of causal influence in the system can be determined by comparing various penchants calculated using both time series.

The penchant definition includes the probability of an assumed effect occurring given a lack of the assumed cause,i.e. $P(E|\bar{C})$. It may be argued that the lack of an assumed cause is unobservable, and thus, the penchant definition is not practically useful. Pearl, and others, argue that causality depends on intervention, which implies the occurrence probability of the assumed effect should be conditioned on performing or not performing an action rather than on the presence or lack of an assumed cause [**?**]. Bunge defines a causal relation as "a relation among events" [**?**], again implying the lack of an assumed cause cannot be used to

2

find causal relations. These issues have been a part of probabilistic definitions of causality at least since the 1960s [?], and we will not attempt to solve them in this article. Our goal is the define a practical tool for causal inference. As such, we will circumvent these philosophical issues in the definition of the penchant by deriving an expression that removes any conditioning on a lack of the assumed cause.

Eqn. 1 can be rewritten using the law of total probability, i.e

$$P(E) = P(E|C)P(C) + P(E|\bar{C})P(\bar{C}) \ , \quad (2)$$

or Bayes theorem, i.e.

$$P(E|\bar{C}) = P(\bar{C}|E)\frac{P(E)}{P(\bar{C})} \ . \quad (3)$$

The definition of probability complements implies $P(\bar{C}) = 1 - P(C)$ and $P(\bar{C}|E) = 1 - P(C|E)$. Applying Eqn. 3 leads to

$$\begin{aligned} P(\bar{C}|E) &= 1 - P(C|E) \\ &= 1 - P(E|C)\frac{P(C)}{P(E)} \end{aligned}$$

Thus,

$$\begin{aligned} P(E|\bar{C}) &= P(\bar{C}|E)\frac{P(E)}{P(\bar{C})} \\ &= \left(1 - P(E|C)\frac{P(C)}{P(E)}\right)\frac{P(E)}{1 - P(C)} \ , \end{aligned}$$

and the second term in Eqn. 1 has been written in terms of the of the first term. This expression implies a penchant calculation containing only the conditional probability that is directly observed from the data, i.e.

$$\rho_{EC} = P(E|C)\left(1 + \frac{P(C)}{1 - P(C)}\right) - \frac{P(E)}{1 - P(C)} \quad (4)$$

This same expression can be derived without Eqn. 3 by using Eqn. 2 to make the appropriate substitution for $P(E|\bar{C})$ into Eqn. 1. This penchant calculation requires only a single conditional probability be estimated from the data.

The use of either Eqn. 2 or Eqn. 3 may eliminate the concern that $P(E|\bar{C})$ is fundamentally unobservable. It may also, however, introduce new philosophical concerns in the definition of the penchant. For example, there has been no discussion of confounding. Assuming, e.g. $P(\bar{C}) = 1 - P(C)$ where $P(C)$ can be estimated from the data alone may be seen as an oversimplification of the dynamics generating the time series data; i.e. it may be seen as an assumption that the assumed effect is only caused by the assumed cause. Thus, it may be argued that the penchant is an indication of predictability rather than causality (similar to arguments made regarding Granger causality [?]). These concerns will not be addressed directly in this article. We emphasize, however, that we use terms like cause, effect, causal inference, and related terms to specifically refer to the expressions presented in this article and not to the general, or intuitive, meanings these terms are given elsewhere.

Assuming $P(E)$ and $P(C)$ can be estimated from the data alone is a key part of the penchant definition. The penchant is non-parametric in the sense that there is no assumed functional relationship between the times series being investigated. It will be shown below that the penchant is, essentially, a structured method for counting the data and using the those counts, along with notions from probabilistic causality, to infer which time series in a given pair might be seen as "driving" the other. Motivation for the penchant is the need for a times series causality tool that relies only on the data. It is often impossible to design experiments, or control interventions into the dynamics as discussed above, for many systems of interest, and it may be difficult to determine functional relationships between the time series data sets.

It follows from Eqn. 1 that

$$\rho_{EC} \in [1, -1] \ , \quad (5)$$

3

but, more importantly for the calculations in the following sections, the penchant is not defined if $P(C)$ or $P(\bar{C})$ are zero (because the conditionals in Eqn. 1 would be undefined). Thus, the penchant is not defined if $P(C) = 0$ or if $P(C) = 1$. The former condition is interpreted intuitively as an inability to determine causal influence between two time series using points that do not appear in one of the series, and the latter condition is interpreted intuitively as an inability to determine causal influence between two time series if one of the data series is constant. The use of Bayes theorem in the derivation of Eqn. 4 implies that same conditions apply to $P(E)$. It will be seen below that there is no *a priori* assignments of "cause" or "effect" to a given time series when using penchants for causal inference. So, operationally, these conditions of $P(C)$ and $P(E)$ only mean that the penchant is undefined between pairs of time series where one series is constant.

The philosophical concerns are perhaps not as important as an answer to the straightforward question of whether or not the penchant is a useful tool for time series causality. The rest of this article will focus on answering that question.

3   Causal Leaning

Given a pair of times series $\{\mathbf{X}, \mathbf{Y}\}$, it is difficult to use the penchant directly for causal inference between the pair. Consider the assignment of $\mathbf{X}$ as the cause, C, and $\mathbf{Y}$ as the effect, E, i.e. $\{C, E\} = \{\mathbf{X}, \mathbf{Y}\}$. If $\rho_{EC} > 0$, then the probability that $\mathbf{X}$ drives $\mathbf{Y}$ is higher than the probability that it does not, which is stated more sufficiently as $\mathbf{X}$ has a penchant to drive $\mathbf{Y}$ or $\mathbf{X} \xrightarrow{pen} \mathbf{Y}$. It is possible, however, that the same penchant could be positive with the opposite cause-effect assignment, i.e. $\rho_{EC} > 0 \mid \{C, E\} = \{\mathbf{Y}, \mathbf{X}\} \Rightarrow \mathbf{Y} \xrightarrow{pen} \mathbf{X}$.

Even though it is possible that $\mathbf{X} \xrightarrow{pen} \mathbf{Y}$ and $\mathbf{Y} \xrightarrow{pen} \mathbf{X}$ are both true, such information does not provide information about the causal relationship within the pair $\{\mathbf{X}, \mathbf{Y}\}$.

The *leaning* is meant to address this problem and is defined as

$$\lambda_{EC} := \rho_{EC} - \rho_{CE} \ . \tag{6}$$

A positive leaning implies the cause C drives the effect E more than the effect drives the cause, a negative leaning implies the effect E drives the cause C more than the cause drives the effect, and a null leaning (i.e. $\lambda_{EC} = 0$) yields no causal information for the cause-effect pair $\{C, E\}$.

Consider again the assignment of $\{C, E\} = \{\mathbf{X}, \mathbf{Y}\}$. If $\lambda_{EC} > 0$, then $\mathbf{X}$ has a larger penchant to drive $\mathbf{Y}$ than $\mathbf{Y}$ does to drive $\mathbf{X}$. More verbosely, $\lambda_{EC} > 0$ implies the difference between the probability that $\mathbf{X}$ drives $\mathbf{Y}$ and the probability that it does not is higher than the difference between the probability that $\mathbf{Y}$ drives $\mathbf{X}$ and the probability that it does not. For convenience, this language is boiled down to $\mathbf{X} \xrightarrow{lean} \mathbf{Y}$, as in $\lambda_{EC} > 0 \mid \{C, E\} = \{\mathbf{X}, \mathbf{Y}\} \Rightarrow \mathbf{X} \xrightarrow{lean} \mathbf{Y}$, $\lambda_{EC} < 0 \mid \{C, E\} = \{\mathbf{X}, \mathbf{Y}\} \Rightarrow \mathbf{Y} \xrightarrow{lean} \mathbf{X}$, and $\lambda_{EC} = 0 \mid \{C, E\} = \{\mathbf{X}, \mathbf{Y}\} \Rightarrow$ no conclusion.

It follows from Eqn. 6 and the bound for the penchant that $\lambda_{EC} \in [-2, 2]$. The leaning is a function of four probabilities, $P(C)$, $P(E)$, $P(C|E)$ and $P(E|C)$. The usefulness of the leaning for causal inference will depend on an effective method for estimating these probabilities from times series data and a more careful definition of the cause-effect assignment within the time series pair. These topics will be discussed with a motivating toy model of a dynamical system for which the penchant and leaning calculations are simple enough to perform without any computational aid.

4 Motivating Example

Consider a time series pair $\bar{\mathbf{T}} = \{\mathbf{X}, \mathbf{Y}\}$ with

$$
\begin{aligned}
\mathbf{X} &= \{x_t \mid t \in [0, 9]\} \\
&= \{0, 0, 1, 0, 0, 1, 0, 0, 1, 0\} \\
\mathbf{Y} &= \{y_t \mid t \in [0, 9]\} \\
&= \{0, 0, 0, 1, 0, 0, 1, 0, 0, 1\}.
\end{aligned}
$$

It seems intuitive to say that $\mathbf{X}$ drives $\mathbf{Y}$ because $y_t = x_{t-1}$. However, to show this result using a leaning calculation requires specification of the cause-effect assignment $\{C, E\} = \{\mathbf{X}, \mathbf{Y}\}$. A cause must precede an effect in the cause-effect assignment for consistency with the intuitive definition of causality. It follows that a natural assignment may be $\{C, E\} = \{x_{t-l}, y_t\}$ where $l \in [1, 9]$. This cause-effect assignment will be referred to as the l-standard assignment.

4.1 Defining the penchants

Given $\bar{\mathbf{T}}$, one possible penchant that can be defined using the 1-standard assignment is

$$
\begin{aligned}
\rho_{y_t=1, x_{t-1}=1} &= \kappa \left(1 + \frac{P(x_{t-1} = 1)}{1 - P(x_{t-1} = 1)} \right) \\
&\quad - \frac{P(y_t = 1)}{1 - P(x_{t-1} = 1)},
\end{aligned}
$$

with $\kappa = P(y_t = 1 | x_{t-1} = 1)$. Another penchant defined using this assignment would be the corresponding term with $\kappa = P(y_t = 0 | x_{t-1} = 0)$. These two penchants are called the *observed* penchants because κ can be found directly from the time series data.

Equations for the unobserved penchants corresponding to $\kappa = P(y_t = 0 | x_{t-1} = 1)$ and $\kappa = P(y_t = 1 | x_{t-1} = 0)$ can be written down. These penchants are defined, but in both cases $\kappa = 0 \Rightarrow \rho_{y_t x_{t-1}} < 0$. Thus unobserved penchants imply the effect, $y_t = 0$ or 1 (for this toy model) is most likely not caused by the postulated cause, $x_{t-1} = 1$ or 0, respectively. Using these unobserved penchants to define leanings becomes a comparison of how unlikely postulated causes are to cause given effects. Such comparisons are not as easily interpreted in the intuitive framework of causality, and as such, are not explored as tools for causal inference in this article.

4.2 Finding the penchants from the data

The probabilities in the penchant calculations can be estimated from the time series data with frequency counts, e.g.

$$
P(y_t = 1 | x_{t-1} = 1) = \frac{n_{EC}}{n_C} = \frac{3}{3} = 1 ,
$$

where n_{EC} is the number of times $y_t = 1$ and $x_{t-1} = 1$ appears in $\bar{\mathbf{T}}$, and n_C is related to the number of times the assumed cause, $x_{t-1} = 1$, has appeared in $\bar{\mathbf{T}}$ and is defined in more detail below.

Estimating the other two probabilities in this penchant calculation using frequency counts from $\bar{\mathbf{T}}$ is slightly more subtle. The underlying assumption that the assumed cause must precede the assumed effect must be considered when defining the frequency counts. This concern is addressed by shifting $\mathbf{X}$ and $\mathbf{Y}$ into $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ such that, for any given t, $\tilde{\mathbf{X}}_t$ precedes $\tilde{\mathbf{Y}}_t$, and defining

$$
P(y_t = 1) = \frac{n_E}{L} = \frac{3}{9} \tag{7}
$$

and

$$
P(x_{t-1} = 1) = \frac{n_C}{L} = \frac{3}{9} , \tag{8}
$$

where n_C is the number of times $\tilde{x}_t = 1$, n_E is the number of times $\tilde{y}_t = 1$, and L is the library length of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ (which are assumed to be the same length). For this example, those subsets are

$$
\begin{aligned}
\tilde{\mathbf{X}} &= \{0, 0, 1, 0, 0, 1, 0, 0, 1\} \\
\tilde{\mathbf{Y}} &= \{0, 0, 1, 0, 0, 1, 0, 0, 1\}
\end{aligned}
$$

which are both shorter than there counterparts above by a single value because the penchants are being calculated using the 1-standard cause-effect assignment. It follows that $\tilde{x}_t = x_{t-1}$ and $\tilde{y}_t = y_t$.

4.3 Mean observed leaning for $\bar{\text{T}}$

The two observed penchants in this example that assume $\mathbf{X}$ causes $\mathbf{Y}$ (i.e. using the 1-standard assignment) are found from the data to be

$$\rho_{y_t=1,x_{t-1}=1} = 1 \tag{9}$$

and

$$\rho_{y_t=0,x_{t-1}=0} = 1 \ . \tag{10}$$

The complements of these observed penchants are found using the complementary 1-standard assignment of $\{C, E\} = \{y_{t-1}, x_t\}$ and are found from the data to be

$$\rho_{x_t=1,y_{t-1}=0} = \frac{3}{7} \tag{11}$$

$$\rho_{x_t=0,y_{t-1}=1} = \frac{3}{7} \ . \tag{12}$$

and

$$\rho_{x_t=0,y_{t-1}=0} = -\frac{3}{7} \ . \tag{13}$$

The *mean observed penchant* is the algebraic mean of the observed penchants, i.e.

$$\langle \rho_{y_t,x_{t-1}} \rangle = \frac{1}{2} \left(\rho_{y_t=1,x_{t-1}=1} + \rho_{y_t=0,x_{t-1}=0} \right)$$
$$= 1$$

and

$$\langle \rho_{x_t,y_{t-1}} \rangle = \frac{1}{3} \left(\rho_{x_t=1,y_{t-1}=0} \right.$$
$$\left. + \rho_{x_t=0,y_{t-1}=1} + \rho_{x_t=0,y_{t-1}=0} \right)$$
$$= \frac{1}{7} \ .$$

The *mean observed leaning* follows from the definition of the mean observed penchants as

$$\langle \lambda_{y_t,x_{t-1}} \rangle = \langle \rho_{y_t,x_{t-1}} \rangle - \langle \rho_{x_t,y_{t-1}} \rangle \tag{14}$$
$$= \frac{6}{7} \ . \tag{15}$$

The positive leaning implies the probability that x_{t-1} drives y_t is higher than the probability that y_{t-1} drives x_t; i.e. $\mathbf{X} \xrightarrow{lean} \mathbf{Y}$ given the 1-standard cause-effect assignment. This result is expected and agrees with the intuitive definition of causality in this example.

4.4 Unobserved penchants

The *unobserved* penchants (using the 1-standard assignment from the beginning of the subsection) for this example are

$$\rho_{y_t=1,x_{t-1}=0} = -1 \tag{16}$$
$$\rho_{y_t=0,x_{t-1}=1} = -1 \tag{17}$$

and their complements are

$$\rho_{x_t=1,y_{t-1}=1} = -\frac{3}{7} \tag{18}$$

These values can incorporated into the averaging calculation to yield a *mean total penchant*; i.e.

$$\langle\langle \rho_{y_t,x_{t-1}} \rangle\rangle = \frac{1}{4} \left(\rho_{y_t=1,x_{t-1}=1} + \rho_{y_t=0,x_{t-1}=0} \right.$$
$$\left. \rho_{y_t=1,x_{t-1}=0} + \rho_{y_t=0,x_{t-1}=1} \right)$$
$$= 0$$

and

$$\langle\langle \rho_{x_t,y_{t-1}} \rangle\rangle = \frac{1}{4} \left(\rho_{x_t=1,y_{t-1}=1} + \rho_{x_t=0,y_{t-1}=0} \right.$$
$$\left. \rho_{x_t=1,y_{t-1}=0} + \rho_{x_t=0,y_{t-1}=1} \right)$$
$$= 0 \ .$$

Thus, the *mean total leaning* (defined analogous to Eqn. 14) would be $\langle\langle \lambda_{y_t,x_{t-1}} \rangle\rangle = \langle\langle \rho_{y_t,x_{t-1}} \rangle\rangle - \langle\langle \rho_{x_t,y_{t-1}} \rangle\rangle = 0$ and would not be useful for casual inference in this example.

4.5 Cause-effect assignment independence

It may be argued that the causal inference above was a little disingenuous in that the assumed cause-effect relationship was known to

be correct. It can be shown, however, that causal inference is independent of the assumed cause-effect relationship. For example, consider the 1-standard cause-effect assignment $\{C, E\} = \{y_{t-l}, x_t\}$. The mean observed leaning would be

$$
\begin{aligned}
\langle \lambda_{x_t, y_{t-1}} \rangle &= \langle \rho_{x_t, y_{t-1}} \rangle - \langle \rho_{y_t, x_{t-1}} \rangle \quad (19) \\
&= -\frac{6}{7} \; , \quad\quad\quad\quad\quad (20)
\end{aligned}
$$

which implies $\mathbf{X} \xrightarrow{lean} \mathbf{Y}$, as expected for this example.

In general, $\lambda_{AB} := \rho_{AB} - \rho_{BA} \Rightarrow -\lambda_{AB} = \rho_{BA} - \rho_{AB} := \lambda_{BA}$. Thus, the causal inference is independent of which times series is initially assumed to be the cause (or effect).

4.6 Weighted Mean Observed Leaning

The *weighted mean observed penchant* is defined similarly to the mean observed penchant but each term is weighted by the number of times that penchant appears in the data; e.g.

$$
\begin{aligned}
\langle \rho_{y_t, x_{t-1}} \rangle_w &= \frac{1}{L-l} \Big(n_{y_t=1, x_{t-1}=1} \rho_{y_t=1, x_{t-1}=1} \\
&\quad + n_{y_t=0, x_{t-1}=0} \rho_{y_t=0, x_{t-1}=0} \Big) \\
&= 1
\end{aligned}
$$

and

$$
\begin{aligned}
\langle \rho_{x_t, y_{t-1}} \rangle_w &= \frac{1}{L-l} \Big(n_{x_t=1, y_{t-1}=0} \rho_{x_t=1, y_{t-1}=0} \\
&\quad + n_{x_t=0, y_{t-1}=1} \rho_{x_t=0, y_{t-1}=1} \\
&\quad + n_{x_t=0, y_{t-1}=0} \rho_{x_t=0, y_{t-1}=0} \Big) \\
&= \frac{3}{63} \; ,
\end{aligned}
$$

where n_{ab} is the number of times the assumed cause a appears with the assumed effect b in the data, L is the library length of the times series data, and l is the lag used in the l-standard cause-effect assignment under which these penchants are being calculated.

The *weighted mean observed leaning* follows naturally as

$$
\begin{aligned}
\langle \lambda_{y_t, x_{t-1}} \rangle_w &= \langle \rho_{y_t, x_{t-1}} \rangle_w - \langle \rho_{x_t, y_{t-1}} \rangle_w \\
&= \frac{60}{63} \; .
\end{aligned}
$$

For this example, $\langle \lambda_{y_t, x_{t-1}} \rangle_w \Rightarrow \mathbf{X} \xrightarrow{lean} \mathbf{Y}$ as expected.

Conceptually, the weighted mean observed penchant is preferred to the mean penchant because it accounts for the frequency of observed cause-effect pairs within the data, which is assumed to be a predictor of causal influence. For example, given some pair $\{\mathbf{A}, \mathbf{B}\}$, if it is known that a_{t-1} causes b_t and both $b_t = 0 | a_{t-1} = 0$ and $b_t = 0 | a_{t-1} = 1$ are observed in the data, then comparison of the frequencies with which the pair occur would be used to determine which of the two pairs represents the true cause-effect relationship and which pair represents, e.g., the effects of noise in the system.

For this example, the weighted mean observed leaning provides the same causal inference as the mean observed leaning. The weighted mean calculation will be used in the examples of the following sections.

4.7 Tolerance Domains

If the example time series was the expected time series but the data points were measured in the presence of noise, then the result may be the noisy time series pair $\bar{\mathbf{T}}' = \{\mathbf{X}', \mathbf{Y}'\}$ with

$$
\begin{aligned}
\mathbf{X}' &= \{x'_t \mid t \in [0,9]\} \\
&= \{0, 0, 1.1, 0, 0, 1, -0.1, 0, 0.9, 0\} \\
\mathbf{Y}' &= \{y'_t \mid t \in [0,9]\} \\
&= \{0, -0.2, 0.1, 1.2, 0, 0.1, 0.9, -0.1, 0, 1\} \; .
\end{aligned}
$$

The previous time series pair, $\bar{\mathbf{T}}$ had only five observed penchants, but $\bar{\mathbf{T}}'$ has more due to the noise. It can be seen in the time series

7

definitions that $x'_t = x_t \pm 0.1 := x_t \pm \delta_x$ and $x'_t = x_t \pm 0.2 := x_t \pm \delta_y$. The weighted mean observed leaning for $\bar{\mathbf{T}}'$ is $\langle \lambda_{y'_t, x'_{t-1}} \rangle_w \approx 0.19$.

The effects of noise on the leaning calculations can be addressed by using the tolerances δ_x and δ_y in the probability estimations from the data. For example, the penchant calculation in Eqn. 9 relied on estimating $P(y_t = 1 | x_{t-1} = 1)$ from the data, but if, instead, the data is known to be noisy, then the relevant probability estimate may be $P(y_t \in [1 - \delta_y, 1 + \delta_y] | x_{t-1} \in [1 - \delta_x, 1 + \delta_x])$.

If the tolerances, δ_x and δ_y, are made large enough, then the noisy system (i.e. $\bar{\mathbf{T}}'$) weighted mean observed leaning, $\langle \lambda_{y'_t \pm \delta_y, x'_{t-1} \pm \delta_x} \rangle_w$, can, at least in the simple examples considered here, be made equal to the noiseless system (i.e. $\bar{\mathbf{T}}$) weighted mean observed leaning, i.e. $\langle \lambda_{y'_t \pm \delta_y, x'_{t-1} \pm \delta_x} \rangle_w = \langle \lambda_{y_t, x_{t-1}} \rangle_w$.

The noise in this example did not affect the causal inference, i.e. $\langle \lambda_{y'_t, x'_{t-1}} \rangle_w \Rightarrow \mathbf{X}' \xrightarrow{lean} \mathbf{Y}'$ as expected. However, consider the time series pair $\bar{\mathbf{T}}'' = \{\mathbf{X}, \mathbf{Y}'\}$, which is a clean impulse $\mathbf{X}$ with a noisy response $\mathbf{Y}'$. The 1-standard assignment mean observed penchants for $\bar{\mathbf{T}}''$ are $\langle \rho_{y'_t = a, x_{t-1} = b} \rangle$ with $\{a, b\} = \{0, 0\}$, $\{0, -0.2\}$, $\{0, -0.1\}$, $\{0, -0.1\}$, $\{1, 1.2\}$, $\{1, 0.9\}$, and $\{1, 1\}$, and $\langle \rho_{x_t = c, y'_{t-1} = d} \rangle$ with $\{c, d\} = \{0, 0\}$, $\{-0.2, 1\}$, $\{0.1, 0\}$, $\{1.2, 0\}$, $\{0, 1\}$, $\{0.9, 0\}$, and $\{0, 0\}$, which lead to a weighted mean observed leaning of $\langle \lambda_{y'_t, x_{t-1}} \rangle_w \approx -0.05$. This value implies $\mathbf{Y}' \xrightarrow{lean} \mathbf{X}$, which disagrees with intuition. The signal $x_{t-1} = 1$ is known to cause the response $y_t = 1$, but in the noisy response $x_{t-1} = 1$ is observed to precede three different possible effects (i.e. $y_t = 1.2$, 0.9, or 1). This effect of the noise in $\mathbf{Y}'$ makes it more difficult to correctly infer how x_t might be related to y_t, which is illustrated sharply by the counterintuitive leaning calculation in this example.

It is not always true that a clean impulse and noisy response lead to leanings that disagree with intuition (e.g. the pair $\{\mathbf{X}, \mathbf{Y}'' = \{0, 0, 0, 1.2, 0, 0, 0.9, -0.1, 0, 1\}\}$ has a weighted mean observed leaning of $\langle \lambda_{y''_t, x_{t-1}} \rangle_w \approx 0.43 \Rightarrow \mathbf{X} \xrightarrow{lean} \mathbf{Y}''$ as expected). However, tolerance domains are an important part of using leaning calculations for causal inference.

Tolerance domains, however, can be set too large. If the tolerance domain is large enough to encompass every point in the time series, then the probability of the assumed cause becomes one, which leads to undefined penchants. For example, given the symmetric definition of the tolerance domain used in this section, $\delta_x = 2$ implies $P(x_{t-1} = 1 \pm \delta_x) = 1$, which implies $\langle \lambda_{y'_t, x_{t-1}} \rangle_w$ is undefined.

It follows that the use of leaning calculations for causal inference depends on an understanding on the noise in the data, which can be troublesome if very little is known about the data sources. One strategy is to calculate the leanings with several different tolerances, increasing the size of the tolerance domains to the point where the penchants become undefined, and finding the tolerance domains for which the leaning changes sign. The sizes of these domains can then be compared to suspected noise levels. This strategy, and others, will be discussed in more detail in the example data sections below. If the noise level is known, then the task becomes much simpler and the tolerances should just be set to the known (or estimated) noise levels for the individual time series.

4.8   Stationarity Dependence

Both $\mathbf{X}$ and $\mathbf{Y}$ are stationary in the original example time series pair $\bar{\mathbf{T}}$. Given a time series pair $\bar{\tau}_{10} = \{\mathbf{X}, \mathbf{R} = \{0, 0, 0, 1, 1, 1, 2, 2, 2, 3\}\}$ containing a non-stationary response signal $\mathbf{R}$, the weighted mean observed leaning calculated under the 1-standard assignment with no tolerance domains still leads to a causal inference

that agrees with intuition; i.e. $\langle \lambda_{r_t, x_{t-1}} \rangle_w \approx 0.11 \Rightarrow \mathbf{X} \xrightarrow{lean} \mathbf{R}$ as expected. This result, however, depends on the library length of the data.

$\bar{\tau}_{10}$ is a specific instance of the following time series pair:

$$\bar{\tau}_L = \{\mathbf{X}, \mathbf{R}\} = \{\{x_t\}, \{r_t\}\} \tag{21}$$

where $t \in [0, L]$,

$$x_t = \begin{cases} 0 & \forall\, t \in \{t \mid t \bmod 3 \neq 0\} \\ 1 & \forall\, t \in \{t \mid t \bmod 3 = 0\} \end{cases} \tag{22}$$

and

$$r_t = x_{t-1} + r_{t-1} \tag{23}$$

with $r_0 = 0$. The weighted mean observed leaning, under the 1-standard assignment with no tolerance domains, for $\bar{\tau}_L$ depends on L. As L is increased, the leaning calculation will eventually lead to causal inferences that do not agree with intuition; e.g. $L = 20 \Rightarrow \langle \lambda_{r_t, x_{t-1}} \rangle_w \approx 1.8 \times 10^{-3} \Rightarrow \mathbf{X} \xrightarrow{lean} \mathbf{R}$ and $L = 50 \Rightarrow \langle \lambda_{r_t, x_{t-1}} \rangle_w \approx -2.5 \times 10^{-3} \Rightarrow \mathbf{R} \xrightarrow{lean} \mathbf{X}$.

As the L is increased, the number of possible observed effects for a given observed cause increases. Thus, under the 1-standard assignment $\{C, E\} = \{x_{t-1}, r_t\}$, $x_{t-1} = 1$ precedes three different values, $r_t = 1$, 2, ans 3, if $L = 10$, but it precedes fifteen different values if $L = 50$. The leaning calculations are methods for counting (in a specific way) the number of times (and ways in which) an observed cause-effect pair appears in the data. The causal inference becomes more difficult for non-stationary time series pairs because repeated cause-effect pairs in the data may be more rare than in the cyclic stationary examples. This effect is very similar to the effect seen when the impulse signal was clean but the response was noisy. Unfortunately, it cannot be remedied with tolerance domains for the non-stationary case. For example, for $\bar{\tau}_L$, the cardinality of the set $\{r_t \mid x_{t-1} = 1\} \to \infty$ as $L \to \infty$, and penchants are not defined given a tolerance domain for $\mathbf{R}$ of $\delta_r = \infty$.

These shortcomings of the weighted mean observed leaning when applied to non-stationary data, however, do not imply that causal inference of non-stationary data cannot be done using a different application of the observed penchants. For example, replacing the weighted mean calculation in the weighted mean observed leaning calculation with a median calculation leads to a *median observed leaning*, $[\lambda_{r_t, x_{t-1}}] \approx 5.3 \times 10^{-3} \Rightarrow \mathbf{X} \xrightarrow{lean} \mathbf{R}$ for $L = 50$ as expected, where $[\cdot]$ is used to denote the median. Of course, even though the median leaning calculation agrees with intuition for a library length where the mean leaning calculation did not, there is no reason to believe the median leaning calculation will not also eventually provide counterintuitive causal inferences as L is increased.

A more basic strategy to deal with non-stationary data would be to define the observed penchant using a different cause-effect assignment. For example, the l-standard assignment (with $l = 1$) used above, i.e. $\{C, E\} = \{x_{t-1}, r_t\}$, might be replaced with an l-AR (autoregressive) assignment with $l = 1$ of $\{C, E\} = \{(x_{t-1}, r_{t-1}), r_t\}$. An observed penchant may be calculated with an assumed cause of $(x_{t-1} = 1, r_{t-1} = 0)$ and an assumed effect of $r_t = 1$. The algorithms to compute the observed penchants from the data become more complicated as the cause-effect assignment becomes more complicated, but the basic definition of the penchant provides a very general conceptual framework for causal inference. This work will only use the l-standard cause-effect assignment in the examples that follow.

5   Simple Example Systems

In this section, the leaning, specifically the weighted mean observed leaning using the l-

standard cause-effect assignment for various l, will be applied to dynamical systems and empirical data sets with known causal relationships. The usefulness of the leaning as a tool for causal inference is tested directly with empirical and synthetic time series data sets for which there is an intuitive understanding of the driving relationships within the system.

5.1 Impulse with Noisy Response Linear Example

Consider the linear example dynamical system of

$$\bar{\tau}_L = \{\mathbf{X}, \mathbf{Y}\} = \{\{x_t\}, \{y_t\}\} \tag{24}$$

where $t \in [0, L]$,

$$x_t = \begin{cases} 2 & t = 1 \\ 0 & \forall\, t \in \{t \mid t \neq 1 \text{ and } t \bmod 5 \neq 0\} \\ 2 & \forall\, t \in \{t \mid t \bmod 5 = 0\} \end{cases}$$

and

$$y_t = x_{t-1} + B\eta_t$$

with $y_0 = 0$, $B \in \mathbb{R} \geq 0$ and $\eta_t \sim \mathcal{N}(0, 1)$. Specifically, consider $B \in [0, 1]$. The driving system $\mathbf{X}$ is a periodic impulse with a signal amplitude above the maximum noise level of the response system, and the response system $\mathbf{Y}$ is a lagged version of the driving signal with standard Gaussian noise of amplitude B applied at each time step.

Figure 1 shows how the weighted mean observed leaning using the 1-standard cause-effect assignment, $\tilde{\lambda}$, changes as the noise amplitude B and tolerance δ_y in increments of 0.01. The synthetic data sets $\mathbf{X}$ and $\mathbf{Y}$ are constructed such that intuitively $\mathbf{X}$ drives $\mathbf{Y}$. Thus, it is expected that $\mathbf{X} \xrightarrow{lean} \mathbf{Y}$ which implies $\tilde{\lambda} > 0$. Figure 1 shows that this expectation is met except when $\delta_y < B$ even for a short library length of $L = 10$. Examples of undefined penchants due to large tolerance domains, as discussed in section 4.7, are seen as δ_y is increased in the $L = 10$ example.

(a) $L = 10$ (b) $L = 50$
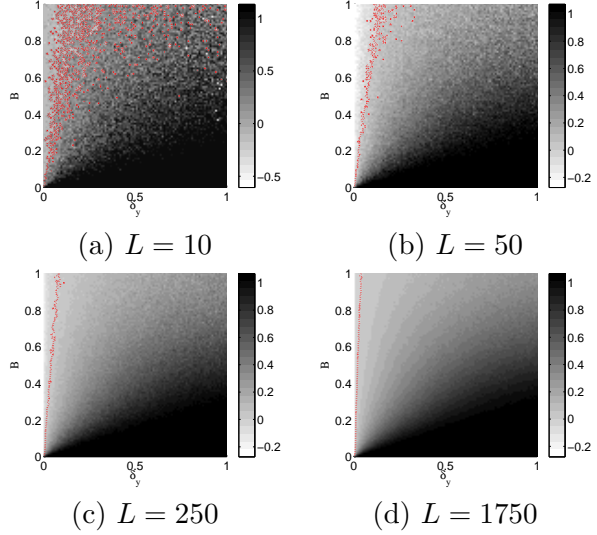
(c) $L = 250$ (d) $L = 1750$

Figure 1: (Color available online.) The unitless leaning is a function of both the noise, the tolerance used for terms from $\mathbf{Y}$, and the library length of the signals. This synthetic data is used to explore the leaning, so $\mathbf{X}$ and $\mathbf{Y}$ do not have explicit units, from which it follows that both δ_y and B are unitless. The red dashed line is the zero contour. See the text for an explanation of the missing data for large δ_y.

Figure 1 shows using the strategy of $\delta_y = B$ always leads to causal inferences that agree with intuition for $L > 10$ in this example. However, as discussed in section 4.7, knowing B *a priori* may be unrealistic with empirical data sets. Consider the following three methods for estimating δ_y from the data:

1. *lagged linear response deviation* - The y-tolerance is set to the mean absolute deviation of y_t from x_{t-1}; i.e. $\delta_y = \langle |y_t - x_{t-1}| \rangle$.

2. *normalized standard deviation* - The y-tolerance is set to the standard deviation of $\{|\mathbf{Y} - \langle \mathbf{Y} \rangle|\}$ where $\langle \mathbf{Y} \rangle$ is the mean of $\mathbf{Y}$; i.e. $\delta_y = \sigma_{|y_t - \langle y_t \rangle|}$.

3. *n-bin mean standard deviation* - The y-tolerance is set to the mean standard devi-

B	method 1	method 2	method 3
0.0	1.0	1.0	1.0
0.1	0.40	1.0	0.48
0.5	0.39	0.79	0.26
0.8	0.30	0.44	0.10

Table 1: $\tilde{\lambda}$ using three different estimation methods for δ_y.

ation of n bins of $\mathbf{Y}$; i.e. $\delta_y = \langle \sigma_{B_i} \rangle$ where B_i is the ith bin of an n-bin histogram of $\mathbf{Y}$.

Table 1 shows $\tilde{\lambda}$ for instances of Eqn. 24 with $B = 0$, 0.1, 0.5, and 0.8 given $L = 100$, and $n = 5$ in method 3 listed above.

The three different methods yield different values for the leaning, but all the methods lead to the same causal inference, $\mathbf{X} \xrightarrow{lean} \mathbf{Y}$, which agrees with intuition for this example. These methods are meant to be examples of using the data to set δ_y if B is not known. These methods are not expected to be reasonable estimates for δ_x and δ_y in general. For example, method 1 assumes a linear relationship between $\mathbf{X}$ and $\mathbf{Y}$ that may be unreasonable to assume in general. However, Table 1 shows different methods for setting δ_y can lead to the same causal inference. Setting the tolerances requires an understanding of the noise in the times series data. The leaning is meant to be part of an exploratory causal analysis of the time series data and cannot exist independently of other exploratory analysis of the data, including analysis of the noise levels.

This linear impulse response example has only included calculations of leaning using the 1-standard assignment. The 1-standard assignment is expected to be useful for causal inference given Eqn. 24. However, deciding which l-standard assignment to use given empirical, rather than synthetic, data sets may be more difficult. It is expected that a several different l-standard assignments would be used as part of any exploratory causal analysis using leaning. The next section introduces an example that plots the leanings for a set of different l-assignments and shows the maximum leaning in the set is near the expected value, i.e. near the lag value that appears explicitly in the dynamical system used to create the synthetic data sets.

5.2 Cyclic Linear Example

Consider the linear example dynamical system of

$$\bar{\tau}_L = \{\mathbf{X}, \mathbf{Y}\} = \{\{x_t\}, \{y_t\}\} \tag{25}$$

where $t \in [0, L]$,

$$x_t = \sin(t)$$

and

$$y_t = x_{t-1} + B\eta_t$$

with $y_0 = 0$, $B \in \mathbb{R} \geq 0$ and $\eta_t \sim \mathcal{N}(0, 1)$. Again, we will only consider $B \in [0, 1]$. This example is very similar to the previous one, except that the driving system $\mathbf{X}$ is sinusoidal.

Figures 2 and 3 were calculated for an instance of Eqn. 25 with $L = 41$ generated by sampling one period of $\mathbf{X}$; i.e. $t \in \{0, f\pi, 2f\pi, 3f\pi, \ldots, 2\pi\}$ with $f = 1/20$ implies $L = 41$. Figure 2 shows the weighted mean observed leaning using the 1-standard assignment, λ, is always positive given $\delta_y = B$. So, as was seen in the previous example, the leaning implies $\mathbf{X} \xrightarrow{lean} \mathbf{Y}$, which agrees with intuition for this example.

It has been argued that lagged cross-correlation techniques are the preferred casual inference tool in most situations because of their simplicity [?]. The lagged cross-correlation is defined as

$$\chi_{xy}^l = \frac{\mathrm{E}\left[(x_t - \mu_x)(y_{t-l} - \mu_y)\right]}{\sigma_x \sigma_y} \quad , \tag{26}$$
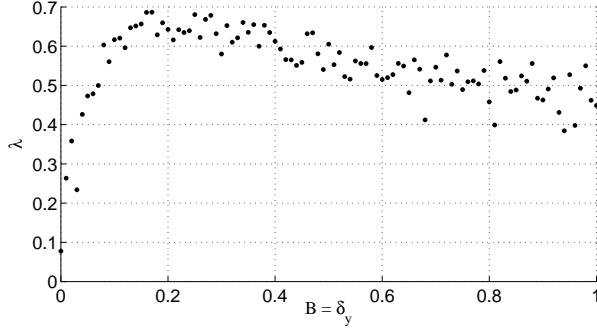
Figure 2: Given $\delta_y = B$, the leaning λ is always positive.

where $E[z_t]$ is the expectation value of $\{z_t\}$, $\mu_{x(y)}$ is the mean of $\mathbf{X}$ ($\mathbf{Y}$), and $\sigma_{x(y)}$ is the standard deviation of $\mathbf{X}$ ($\mathbf{Y}$). The cross-correlation is often used for causal inference by introducing a difference quantity [?]

$$\delta\chi_{xy}^l = \chi_{xy}^l - \chi_{yx}^l \ . \tag{27}$$

The sign of $\delta\chi_{xy}^l$ is used, similar to the leaning approach, to determine the causal inference; i.e. $\delta\chi_{xy}^l > 0$ implies $\mathbf{X}$ "causes" $\mathbf{Y}$ and $\delta\chi_{xy}^l < 0$ implies $\mathbf{Y}$ "causes" $\mathbf{X}$ [?].
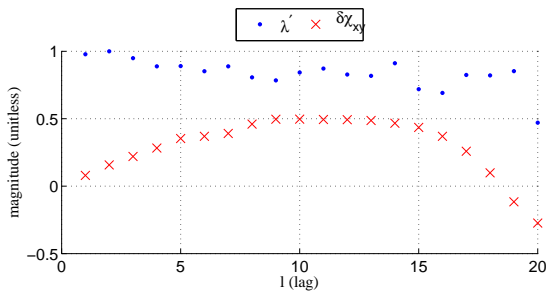

Figure 3: (Color available online.) The unitless, normalized leaning, λ', can be plotted for different l-standard cause-effect assignments along with the cross correlation, χ for the same lags, l, to show how the two values compare for this simple cyclic example.

Figure 3 shows how $\delta\chi_{xy}^l$ compares to the leaning given $l \in [1, 21]$ for instances of Eqn.

25 with $B = 0.5$. In Figure 3, the leaning has been normalized for presentation clarity as

$$\lambda' = \frac{\lambda_l}{\max_{l \in [1,15]} \lambda_l} \ , \tag{28}$$

where λ_l is the weighted mean observed leaning using the l-standard assignment (λ_1 is plotted in Figure 2). Both λ' and $\delta\chi_{xy}^l$ lead to the same causal inference, i.e. $\mathbf{X}$ "drives" $\mathbf{Y}$, for $l \in [1, 19]$, although only the leaning agrees with intuition for $l = 20$ and $l = 21$ in this example. Thus, both tools agree with intuition for small lags in this simple cyclic example. The leaning, however, has its maximum values near the smallest lags, which is expected given Eqn. 25, while the cross-correlation difference has its maximum values at lags that do not explicitly appear in Eqn. 25.

The cross-correlation difference technique is also known to be unreliable given nonlinear dynamics [?]. Leanings of data sets generated from nonlinear dynamics will be discussed in Section 5.4. Neither of the previous examples has been physically motivated, so the next section discusses exploratory causal inference of synthetic data sets generated from the well-known dynamics of a physical system.

5.3   RL Circuit Example

Consider a series circuit containing a resistor, inductor, and time varying voltage source related by

$$\frac{dI}{dt} = \frac{V(t)}{L} - \frac{R}{L}I, \tag{29}$$

where I is the current at time t, $V(t) = \sin(t)$ is the voltage at time t, R is the resistance, and L is the inductance. The time series pair for this example is then

$$\bar{\tau}_L = \{\mathbf{V}, \mathbf{I}\} = \{\{V_t\}, \{I_t\}\} \tag{30}$$

where $\mathbf{V}$ is the set of discrete values of $V(t)$ evaluated using $t \in \{0, f\pi, 2f\pi, 3f\pi, \dots, 8\pi\}$

12

with $f = 1/10$ and $\mathbf{I}$ is the set of discrete values found either by solving Eqn. 29 numerically or by evaluating the analytical solution, i.e.

$$I(t) = \frac{L}{D}e^{-\frac{t}{\tau}} + \frac{R}{D}\sin(t) - \frac{L}{D}\cos(t) \quad (31)$$

with $D = L^2 + R^2$ and $\tau = L/R$, for the same time set used for $\mathbf{V}$.

Physical intuition is that V drives I, and so we expect to find that $V \xrightarrow{lean} I$. The weighted mean observed leaning using the 1-standard assignment, λ_1, can be used to test this expectation. Unlikely, the previous examples, however, there is no noise term in the dynamics (such as B in Eqn. 24 and 25), so setting δ_I will not be as straightforward.

Table 2 shows λ_1 for both the analytical solution and a numerical solution to Eqn. 29 was using the *ode45* integration function in MATLAB. The time series $V(t)$ is created by defining values at fixed points and using linear interpolation to find the time steps required by the ODE solver. Two different physical scenarios are considered in which L and R are constant, $L = 10$ Henries and $R = 5$ Ohms and $L = 5$ Henries and $R = 20$ Ohms.

The previously discussed strategy of increasing δ_I until the leaning becomes undefined and then reporting the leaning calculated using the largest δ_I for which it is defined would lead to the a causal inference that agrees with intuition for this example. Specifically, from Table 2(a) $\delta_I = 10^{-2} \Rightarrow \lambda_1 \approx 0.7 \Rightarrow V \xrightarrow{lean} I$, as expected.

Discussion on setting the tolerance domains has centered on understanding the noise in the system. This example illustrates that the "noise" being considered does not need to be a physical noise source in the system (there are no explicit noise terms in Eqn. 30). For example, the numerical tolerance of the ODE solver was set to 10^{-3} in both Table 2(a) and (b), and the table shows setting $\delta_y = 10^{-3}$ would lead

to causal inferences that agree with intuition in both cases. However, $\delta_y = 0$ leads to causal inferences that do not agree with intuition in both cases even for the analytical solution.

Consider, for example, the peak values of $\mathbf{V}$. The time steps of these peaks are $\mathbf{T}_{peak} = \{t|V_t = 1\} = \{6, 26, 46, 66\}$. The values of $\mathbf{I}$ given $\tau = 0.25$ that immediately follow these peaks are $\mathbf{I}_{0.25}^{peak} = \{I_t|t \in \{7, 27, 47, 67\}\}$. The same values given $\tau = 2$ will be labeled $\mathbf{I}_2^{peak}$. The standard deviation of the first set is $\sigma_{0.25}^{peak} \approx 10^{-6}$ and the standard deviation of the second set is $\sigma_2^{peak} \approx 10^{-2}$. Table 2(a) (for $\sigma_{0.25}^{peak}$) and (b) (for σ_2^{peak}) shows setting δ_I to the appropriate standard deviation of the peaks would lead to causal inferences that agree with intuition. Rather than physical noise levels, the noise levels used to set the tolerance domains for the leaning calculations is better thought of as the spread in the possible values of an assumed effect that may reasonably be considered due to the same assumed cause.

This example can also illustrate the importance of sample frequency and periods. The leaning calculation requires an assumed cause and effect pair to appear in the data enough times to provide a good estimate of the probability. Thus, data that is sampled for too few periods or too sparsely can lead to counter-intuitive leanings. For example, if there is only a single peak in the assumed driving time series because of poor sampling, then there can only be a single response value, which would be insufficient to reliably provide the conditional probabilities in the leaning calculation for that assumed cause-effect pair. For Eqn. 30 with the analytical solution for $\mathbf{I}$, if $\delta_I = 10^{-3}$ and $\tau = 0.25$, then $t \in \{0, f\pi, 2f\pi, 3f\pi, \ldots, 2\pi\}$ with $f = 1/10$ leads to $\lambda_1 = -0.045$ and $t \in \{0, f\pi, 2f\pi, 3f\pi, \ldots, 3\pi\}$ with $f = 2/3$ leads to $\lambda_1 = -0.167$, both of which disagree with intuition.

δ_I	λ_1 (*ode45*)	λ_1 (analytical)	δ_y	λ_1 (*ode45*)	λ_1 (analytical)
0	-0.132	-0.089	0	-0.132	-0.132
10^{-6}	-0.132	0.493	10^{-6}	-0.132	-0.132
10^{-5}	-0.108	0.548	10^{-5}	-0.120	-0.096
10^{-4}	0.188	0.564	10^{-4}	0.011	0.098
10^{-3}	0.582	0.581	10^{-3}	0.398	0.386
10^{-2}	0.730	0.727	10^{-2}	0.676	0.675
10^{-1}	*undefined*	*undefined*	10^{-1}	0.314	0.315
(a) $R = 20\ \Omega,\ L = 5$ H			(b) $R = 5\ \Omega,\ L = 10$ H		

Table 2: The leaning λ_1 depends on both δ_I and the method for evaluating $\mathbf{I}$ in this example. These two tables show that the values of δ_I for which the leaning starts to agree with intuition can also depend on the physical system parameters (e.g., τ).

The examples so far have all had a linear relationship between the driving signal and the response signal. Of the four broad categories of time series causality tools, transfer entropy [] and SSR methods [] are the two categories that can be applied to nonlinear datasets without modification. The conceptual framework of granger causality is not restricted by the linearity of the data set [], but the original formulation by Granger must be modified to do so []. Lagged cross-correlation techniques are known to be unreliable if the data sets are generated by nonlinear dynamics [?, ?]. The next example explores the use of leaning calculations in nonlinear systems.

5.4 Nonlinear Example

Consider the nonlinear dynamical system of

$$\bar{\tau}_L = \{\mathbf{X}, \mathbf{Y}\} = \{\{x_t\}, \{y_t\}\} \qquad (32)$$

where $t \in [0, L]$,

$$x_t = \sin(t)$$

and

$$y_t = Ax_{t-1}\left(1 - Bx_{t-1}\right) + C\eta_t,$$

with $y_0 = 0$, with $A, B, C \in \mathbb{R} \geq 0$ and $\eta_t \sim \mathcal{N}(0, 1)$. Specifically, consider $A, B, C \in$ $[0, 1]$ and $t \in \{0, f\pi, 2f\pi, 3f\pi, \ldots, 6\pi\}$ with $f = 1/30$, which implies $L = 181$.

Figure 4 shows the weighted mean observed leaning using the 1-standard assignment, i.e. λ_1, agrees with intuition over the considered domains of A, B, and C if the tolerance domain for $\mathbf{Y}$ is set to the noise level, i.e. $\delta_y = C$. The result of $\mathbf{X} \xrightarrow{lean} \mathbf{Y}$ shows that causal inference using leanings on data sets generated from nonlinear dynamics can be performed similarly, and can lead to similarly intuitive results, as the data sets generated from linear dynamics.

Proponents of SSR time series causality tools have pointed out the limitations of tools like lagged cross-correlation and Granger causality when the dynamics exhibit chaotic behavior [?]. Chaotic dynamics are explore in the next section.

5.5 Coupled Logistic Map Example

Consider the nonlinear dynamical system of

$$\bar{\tau}_L = \{\mathbf{X}, \mathbf{Y}\} = \{\{x_t\}, \{y_t\}\} \qquad (33)$$

where $t \in [0, L]$,

$$x_t = x_{t-1}\left(r_x - r_x x_{t-1} - \beta_{xy} y_{t-1}\right)$$

and

$$y_t = y_{t-1}\left(r_y - r_y y_{t-1} - \beta_{yx} x_{t-1}\right)$$

14

(a) $C = 0.2$

(b) $C = 0.4$

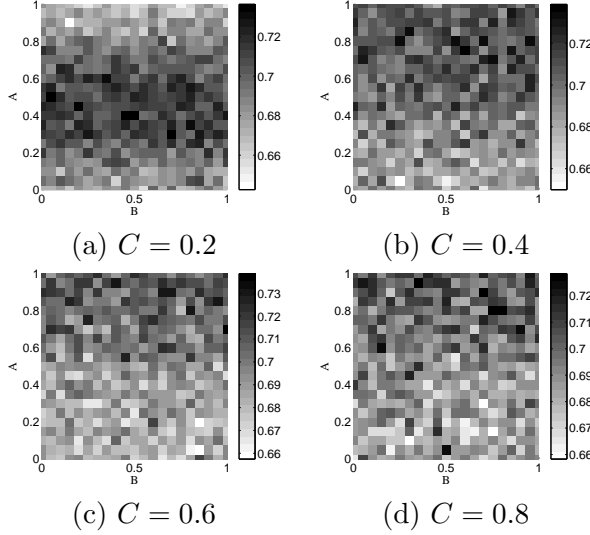
(c) $C = 0.6$

(d) $C = 0.8$

Figure 4: Leaning, λ_1, is a function of all three unitless parameters in Eqn. 32, A, B, and C. The leaning agrees with intuition in this example for all the tested values of A, B, and C given $\delta_y = C$.

tuitive causal inference for this system. The conjectures presented in this paragraph, however, are supported by the leaning calculations (using the 1-standard assignment).
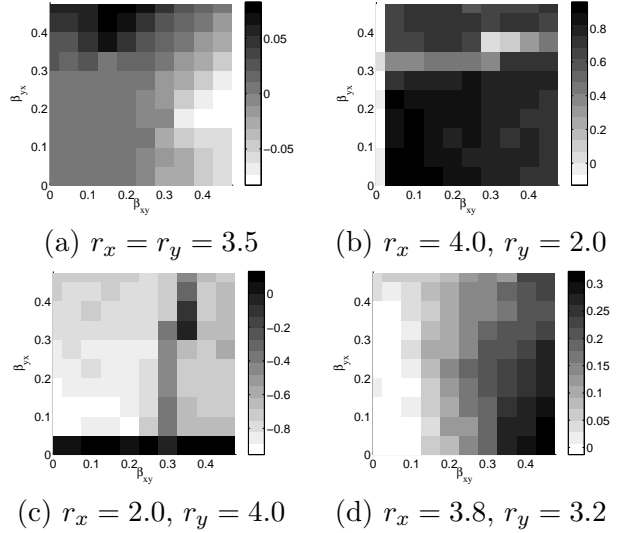

(a) $r_x = r_y = 3.5$

(b) $r_x = 4.0$, $r_y = 2.0$

(c) $r_x = 2.0$, $r_y = 4.0$

(d) $r_x = 3.8$, $r_y = 3.2$

Figure 5: Leaning, λ_1, is a function of four unitless parameters in Eqn. 33, r_x, r_y, β_{xy}, and β_{yx} (along with the initial conditions x_0 and y_0, which are fixed in this example). The tolerance domains are set as $\delta_x = \sigma_{x_t - \langle x_t \rangle}$ and $\delta_y = \sigma_{y_t - \langle y_t \rangle}$. The leaning is defined using the 1-standard assignment, so $\lambda_1 > 0 \Rightarrow \mathbf{X} \xrightarrow{lean} \mathbf{Y}$.

where the parameters $r_x, r_y, \beta_{xy}, \beta_{yx} \in \mathbb{R} \geq 0$. This pair of equations is a specific form of the two-dimensional coupled logistic map system often used to model population dynamics [?] and it was the example system used by Sugihara *et al.* in their introduction of cross convergent mapping (CCM; an SSR time series causality tool) [?].

The idea is that $\beta_{xy} > \beta_{yx}$ intuitively implies $\mathbf{Y}$ "drives" $\mathbf{X}$ more than $\mathbf{X}$ "drives" $\mathbf{Y}$, and vice versa. Such intuition, however, can be difficult to justify for all instances of Eqn. 33. The x_{t-1} term that appears in y_t can be seen as a fuction of x_{t-2} with coeffcents of $\beta_{yx} r_x$. These product coefficients suggest that if $r_x > r_y$, then $\mathbf{X}$ may be seen as the stronger driver in the system even if $\beta_{yx} < \beta_{xy}$. The same argument can be made, with the appropriate substituions, to show that $\mathbf{Y}$ may be seen as the stronger driver in the system even if $\beta_{xy} < \beta_{yx}$. As such, there is no clear in-

Figure 5 shows four instances of Eqn. 33 with different values for r_x and r_y. Each instance has a library length of $L = 500$ and initial conditions of $x_0 = 0.4$ and $y_0 = 0.4$. There is no clear, intuitive driver in this example, so both tolerance domains must be set in the leaning calculation. The leaning is calculated using the 1-standard cause-effect assignment and estimated tolerance domains of $\delta_x = \sigma_{x_t - \langle x_t \rangle}$ and $\delta_y = \sigma_{y_t - \langle y_t \rangle}$.

5.6 Impulse with Multiple Noisy Responses Example

$$\bar{\tau}_L = \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\} = \{\{x_t\}, \{y_t\}, \{z_t\}\} \quad (34)$$

where $t \in [0, L]$,

$$x_t = \begin{cases} 2 & t = 1 \\ 0 & \forall\, t \in \{t \mid t \neq 1 \text{ and } t \bmod 5 \neq 0\} \\ 2 & \forall\, t \in \{t \mid t \bmod 5 = 0\} \end{cases}$$

$$
\begin{aligned}
y_t &= x_{t-1} + B\eta_t \\
z_t &= y_{t-1} \\
z'_t &= y_{t-1} + y_t = y_{t-1} + x_{t-1} + B\eta_t \\
z''_t &= y_{t-1} + x_{t-1} + z_{t-1}
\end{aligned}
$$

with $y_0 = 0$, $B \in \mathbb{R} \geq 0$, $\eta_t \sim \mathcal{N}(0, 1)$, and $L = 500$.

Figure 6: $\{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$

Figure 7: $\{\mathbf{X}, \mathbf{Y}, \mathbf{Z}'\}$

Figure 8: $\{\mathbf{X}, \mathbf{Y}, \mathbf{Z}''\}$

6 Empirical Data

7 Conclusion