Time Series Leanings

Robert S. Weigel

March 25, 2015

1 Introduction

...TBD...

2 Causal Penchant

Define the *causal penchant* as

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

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 details of these comparison is discuss in the following sections, but some potential philosophical issues with this definition will be addressed first.

... **Add discussion of Pearl argument that $P(E|\bar{C})$ is "unobservable".**

Pearl's concerns can be addressed by rewriting Eqn. 1 using the law of total probability, i.e

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

or Bayes theorem, i.e.

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

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,

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

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.

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

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, ... **expand this discussion**

It follows from Eqn. 1 that

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

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

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

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

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

with $\kappa = P\left(y_t = 1 | x_{t-1} = 1\right)$. Another penchant defined using this assignment would be the corresponding term with $\kappa = P\left(y_t = 0 | x_{t-1} = 0\right)$. 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 $\kappa = P\left(y_t = 0 | x_{t-1} = 1\right)$ and $\kappa = P\left(y_t = 1 | x_{t-1} = 0\right)$ 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.

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

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

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

and

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

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 $\bar{\mathbf{T}}$

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

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

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

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

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

and

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

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

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

and

$$\begin{aligned}\langle\rho_{x_t,y_{t-1}}\rangle &= \frac{1}{3}\big(\rho_{x_t=1,y_{t-1}=0}\\ &\quad +\rho_{x_t=0,y_{t-1}=1} + \rho_{x_t=0,y_{t-1}=0}\big)\\ &= \frac{1}{7} \quad .\end{aligned}$$

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

$$\langle\lambda_{y_t,x_{t-1}}\rangle = \langle\rho_{y_t,x_{t-1}}\rangle - \langle\rho_{x_t,y_{t-1}}\rangle \tag{14}$$

$$= \frac{6}{7} \quad . \tag{15}$$

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.

$$\begin{aligned}\langle\langle\rho_{y_t,x_{t-1}}\rangle\rangle &= \frac{1}{4}\big(\rho_{y_t=1,x_{t-1}=1} + \rho_{y_t=0,x_{t-1}=0}\\ &\quad \rho_{y_t=1,x_{t-1}=0} + \rho_{y_t=0,x_{t-1}=1}\big)\\ &= 0\end{aligned}$$

and

$$\begin{aligned}\langle\langle\rho_{x_t,y_{t-1}}\rangle\rangle &= \frac{1}{4}\big(\rho_{x_t=1,y_{t-1}=1} + \rho_{x_t=0,y_{t-1}=0}\\ &\quad \rho_{x_t=1,y_{t-1}=0} + \rho_{x_t=0,y_{t-1}=1}\big)\\ &= 0 \quad .\end{aligned}$$

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

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

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, however, can be made computationally less expensive than the mean calculation, which will be of practical concern for empirical data sets.

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 $\bar{\mathbf{T}}'$ nessicarily has more due to the noise. It can be seen in the time series 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 $L = 50$. Conceptually, 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.

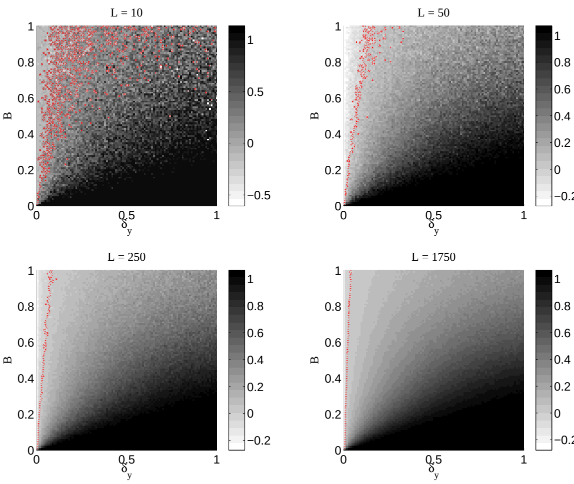

Figure 1: (Color available online.) The unitless leaning is a function of both the noise, the tolerance used for terms from $\mathbf{Y}$, and the library length of the signals. This synthetic data is used to explore the leaning, so $\mathbf{X}$ and $\mathbf{Y}$ do not have explicit units, from which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

Figure **??** 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 $\mathbf{X}\overrightarrow{lean}\mathbf{Y}$ which implies $\tilde{\lambda} > 0$. Figure **??** 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.

Figure **??** shows using the strategy of $\delta_y = B$ always leads to causal inferences that agree with intuition for this example. However, as discussed in section 4.7, knowing B *a priori* may be unrealistic with empirical data sets.

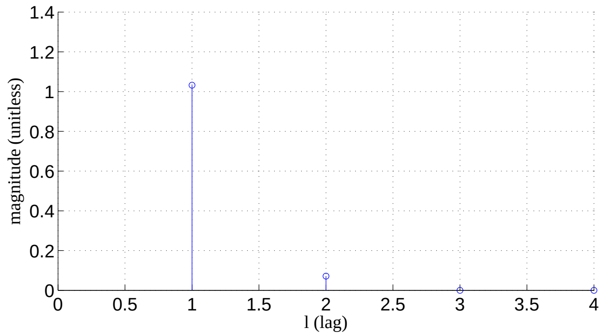

Figure 2: (Color available online.) Different l-standard cause-effect assignments lead to different leanings.

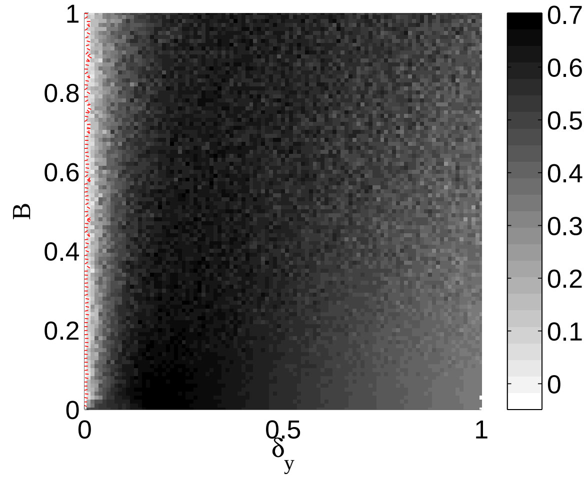

Figure 3: (Color available online.) Leaning as a function of both the noise and the y-tolerance. The red dashed line is the zero contour. See the text for am explanation of the missing data for large δ_y.

5.2 Cyclic Linear Example

Consider the linear example dynamical system of

$$X_t = \sin(t) \tag{25}$$
$$Y_t = X_{t-1} + B\eta_t, \tag{26}$$

with $B \in \mathbb{R} \geq 0$ and $\eta_t \sim \mathcal{N}(0, 1)$. Specifically, consider $B \in [0, 2]$ in increments of 0.02. The response system Y is just a lagged version of the driving signal with varying levels of standard Gaussian noise applied at each time step.

5.3 Non-Linear Example

Consider the non-linear dynamical system of

$$X_t = \sin(t) \tag{27}$$
$$Y_t = AX_{t-1}(1 - BX_{t-1}) + C\eta_t, \tag{28}$$

with $A, B, C \in \mathbb{R} \geq 0$ and $\eta_t \sim \mathcal{N}(0, 1)$. Specifically, consider $A, B, C \in [0, 5]$ in increments of 0.5.

5.4 RL Circuit Example

Both of the previous examples included a noise term, η_t. 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, $V(t) = \sin(\Omega t)$ is the voltage at time t, R is the resistance, and L is the inductance. Eqn. 29 was solved 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.

Consider the situation where $L = 10$ Henries and $R = 5$ Ohms are constant. Physical intuition is that V drives I, and so we expect to find that V CCM causes I (i.e., $C_{VI} > C_{IV}$ or $\Delta = C_{VI} - C_{IV} > 0$).

$$I = \frac{L}{D}e^{-\frac{R}{L}t} + \frac{R}{D}\sin(t) - \frac{L}{D}\cos(t) \tag{30}$$

with $D = L^2 + R^2$.

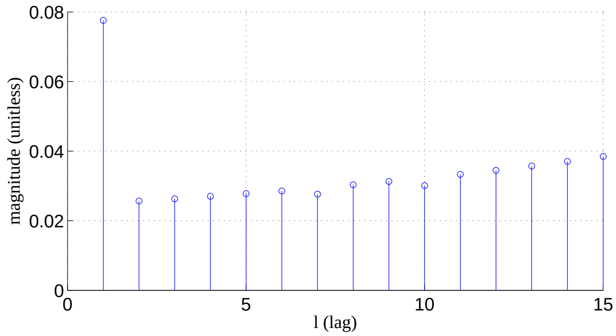

Figure 4: (Color available online.) Different l-standard cause-effect assignments lead to different leanings.

5.5 Impulse with Multiple Noisy Responses Example

$$\bar{\tau}_L = \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\} = \{\{x_t\}, \{y_t\}, \{z_t\}\} \qquad (31)$$

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 $L = 500$.

6 Empirical Data

7 Conclusion

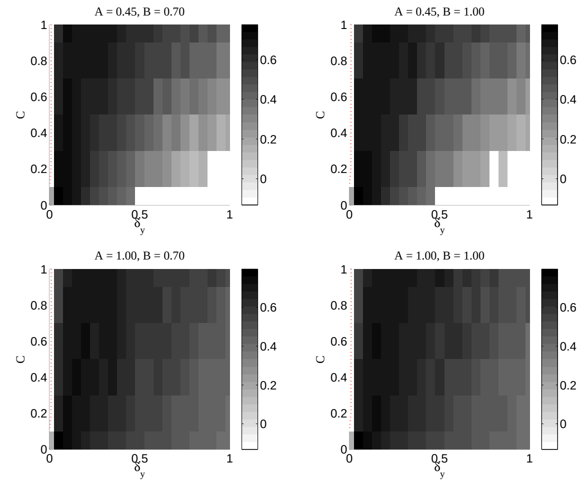

Figure 5: (Color available online.) Leaning as a function of both the noise and the y-tolerance for values of A and B. The red dashed line is the zero contour. See the text for am explanation of the missing data for large δ_y.

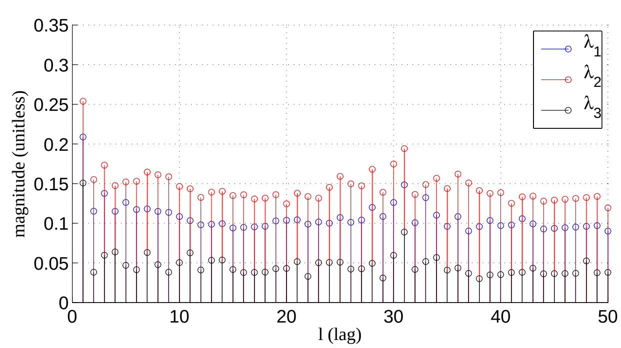

Figure 6: (Color available online.) Different l-standard cause-effect assignments lead to different leanings.

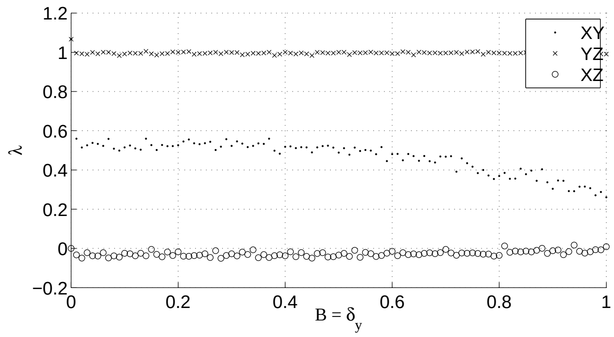

Figure 7: $\{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$

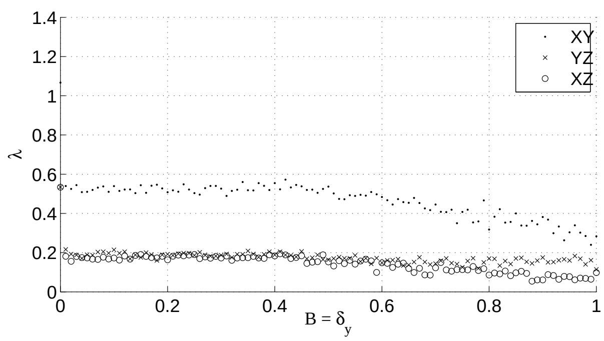

Figure 8: $\{\mathbf{X}, \mathbf{Y}, \mathbf{Z}'\}$

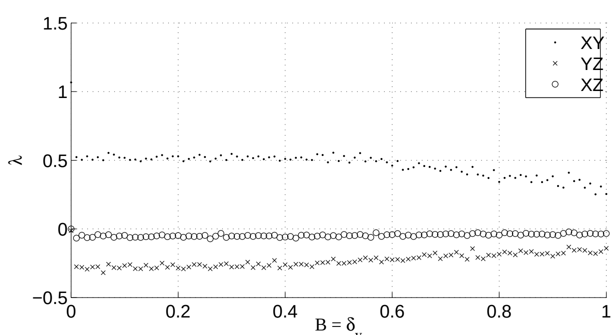

Figure 9: $\{\mathbf{X}, \mathbf{Y}, \mathbf{Z}''\}$