Time Series Leanings

Robert S. Weigel

February 2, 2015

1 Introduction

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 Toy Model

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 cooresponding 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 subsetting $\mathbf{X}$ 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 toy model, those subsets are

$$\begin{aligned}
\tilde{x} &= \{0,0,1,0,0,1,0,0,1\} \\
\tilde{y} &= \{0,0,1,0,0,1,0,0,1\}
\end{aligned}$$

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 toy model 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

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

There are two other observed penchants in this system found using the complentary 1-standard assignment of $\{C,E\} = \{y_{t-1}, x_t\}$, and they are found from the data to be

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

and

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

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

$$\begin{aligned} \langle \rho_{x_t,y_{t-1}} \rangle &= \frac{1}{2}\left(\rho_{x_t=1,y_{t-1}=1} + \rho_{x_t=0,y_{t-1}=0}\right) \\ &= -\frac{3}{7} \ \ . \end{aligned}$$

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

$$\langle \lambda_{y_t,x_{t-1}} \rangle = \langle \rho_{y_t,x_{t-1}} \rangle - \langle \rho_{x_t,y_{t-1}} \rangle \tag{13}$$

$$= \frac{10}{7} \ \ . \tag{14}$$

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 toy model.

The point of this section was simply to motivate the most basic leaning calculations, but there is still much that can be done to use the concepts of penchants and leanings to explore time series causality within this toy model. For example, there are many other l-standard assignments that can be used to find other mean observed leanings and other definitions of "mean" observed penchants can be used to define other mean observed leanings. Averaging techniques for the different possible mean leanings can also be explored. Some of these topics will be discussed as they appear in the examples below.

4.4 Algorithm

Most empirical data sets make observed leaning calculations either tedious, or prohibitively difficult, to do. The calculation itself can be done with a simple algorithm as follows:

1.

5 Simple Example Systems

5.1 Impulse with Noisy Response Linear Example

Consider the linear example dynamical system of

$$X_t = \{0,2,0,0,2,0,0,2,0,0\} \tag{15}$$

$$Y_t = X_{t-1} + B\eta_t, \tag{16}$$

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.

Figure 1: (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.

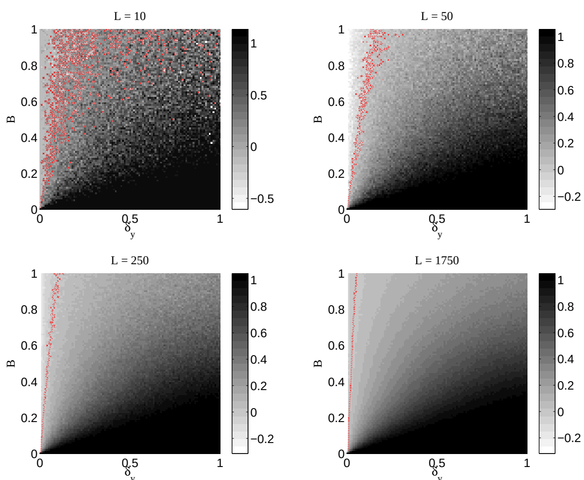

Figure 2: (Color available online.) Leaning as a function of both the noise and the y-tolerance for different library lengths. The red dashed line is the zero contour. See the text for am explanation of the missing data for large δ_y.

5.2 Cyclic Linear Example

Consider the linear example dynamical system of

$$X_t = \sin(t) \tag{17}$$
$$Y_t = X_{t-1} + B\eta_t, \tag{18}$$

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

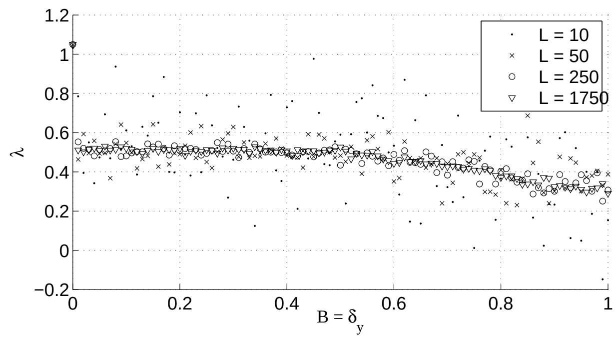

Figure 3: (Color available online.) The leaning agrees with intuition for most noise levels when the y-tolerance is set to the noise level (i.e. $B = \delta_y$).

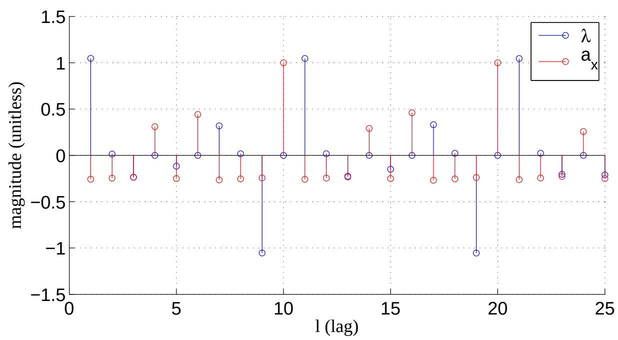

Figure 4: (Color available online.) Different l-standard cause-effect assignments lead to different leanings. The autocorrelation of $\mathbf{X}$, i.e. a_x, is shown in red for reference.

standard Gaussian noise applied at each time step.

5.3 Non-Linear Example

Consider the non-linear dynamical system of

$$X_t = \sin(t) \tag{19}$$
$$Y_t = AX_{t-1}(1 - BX_{t-1}) + C\eta_t, \tag{20}$$

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.

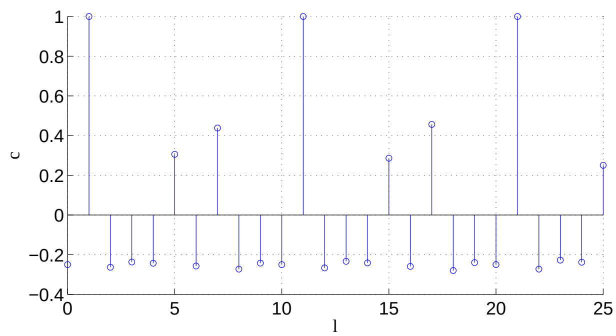

Figure 5: (Color available online.) Lagged cross correlations for reference.

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

$$\frac{dI}{dt} = \frac{V(t)}{L} - \frac{R}{L}I, \tag{21}$$

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. 21 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 $\Delta = C_{VI} - C_{IV} > 0$).

6 Empirical Data

7 Conclusion