Time Series Leanings

Robert S. Weigel

January 27, 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

$$P(\bar{C}|E) = 1 - P(E|C)\frac{P(C)}{P(E)} \tag{4}$$

Thus,

$$P(E|\bar{C}) = \left(1 - P(E|C)\frac{P(C)}{P(E)}\right)\frac{P(E)}{1 - P(C)} \quad . \tag{5}$$

This expression implies

$$\mathbf{??}\rho_{EC} = P(E|C)\left(1 + \frac{P(C)}{1 - P(C)}\right) - \frac{P(E)}{1 - P(C)} \tag{6}$$

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.

The use of either Eqn. 2 or Eqn. 3 may eliminate the concern that $P(E|\bar{C})$ is fundamental 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

$$\rho_{EC} \in [1, -1] \quad , \tag{7}$$

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. **??** 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 $\mathbf{X}\hookrightarrow\mathbf{Y}$. It is possible, however, that the same penchant could be positive with the opposite cause-effect assignment, i.e. $\rho_{EC}>0\mid\{C,E\}=\{\mathbf{Y},\mathbf{X}\}\Rightarrow\mathbf{Y}\hookrightarrow\mathbf{X}$. Even though it is possible that $\mathbf{X}\hookrightarrow\mathbf{Y}$ and $\mathbf{Y}\hookrightarrow\mathbf{X}$ 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

$$\lambda_{EC}:=\rho_{EC}-\rho_{CE}\quad . \tag{8}$$

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 $\mathbf{X}\rightarrow\mathbf{Y}$, as in $\lambda_{EC}>0\mid\{C,E\}=\{\mathbf{X},\mathbf{Y}\}\Rightarrow\mathbf{X}\rightarrow\mathbf{Y}$, $\lambda_{EC}<0\mid\{C,E\}=\{\mathbf{X},\mathbf{Y}\}\Rightarrow\mathbf{Y}\rightarrow\mathbf{X}$, and $\lambda_{EC}=0\mid\{C,E\}=\{\mathbf{X},\mathbf{Y}\}\Rightarrow$ no conclusion.

It follows from Eqn. 8 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

$$\begin{aligned}\mathbf{X} &= \{x_t\mid t\in[0,9]\}=\{0,0,2,0,0,2,0,0,2,0\}\\ \mathbf{Y} &= \{y_t\mid t\in[0,9]\}=\{0,0,0,2,0,0,2,0,0,2\}\quad .\end{aligned}$$

It seems intuative 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 $l \in [1, 9]$.

and a penchant defined as

$$\rho_{y_t x_{t-1}} = P\left(y_t = 2|x_{t-1} = 2\right) - P\left(y_t = 2|x_{t-1} \neq 2\right)$$

Estimating the probabilities with frequency counts is straightforward for the first term, i.e.

$$P\left(y_t = 2|x_{t-1} = 2\right) = \frac{n_{EC}}{n_C} = \frac{3}{3} = 1 \ ,$$

where n_{EC} is the number of times $y_t = 2$ and $x_{t-1} = 2$ appears in the pair of time series (where, in this case, y_t is assumed to be the effect E caused by the cause C, $x_{t-1} = 2$) and n_C is the number of times the assumed cause, $x_{t-1} = 2$, has appeared. The conditional probability $P(A|B)$ is not defined if $P(B) = 0$, thus $n_C > 0$ by definition, and $n_{EC} <= n_C$ because even though the assumed effect, $y_t = 2$, might occur without the assumed cause, $x_{t-1} = 2$ in the data, the count n_{EC} is defined to be when they occur together. It follows that this definition of $P\left(y_t = 2|x_{t-1} = 2\right)$ can be interpreted as a probability.

The second term in the penchant is trickier. Consider the following frequency count,

$$P\left(y_t = 2|x_{t-1} \neq 2\right) = \frac{n_{EnC}}{n_{nC}} = \frac{0}{6} = 0 \ ,$$

where n_{EnC} is the number of times the assumed effect, $y_t = 2$ appears given that the assumed cause $x_{t-1} = 2$ has *not* occurred, i.e. $x_{t-1} \neq 2$, and n_{nC} is the number of times the assumed cause $x_{t-1} = 2$ has *not* occurred, i.e. the number of times $x_{t-1} \neq 2$. There are seven zeros in $\{x_t\}$, so it may be assumed that $n_{nC} = 7$ rather than $n_{nC} = 6$ as used above. But, the condition is on x_{t-1}, not x_t. Thus, the time series actually being compared are

$$\begin{aligned} \tilde{x} &= \{0, 0, 2, 0, 0, 2, 0, 0, 2\} \\ \tilde{y} &= \{0, 0, 2, 0, 0, 2, 0, 0, 2\} \end{aligned}$$

which are both shorter than there counterparts above by a single value. It follows that $\tilde{x}_t = x_{t-1}$ and $\tilde{y}_t = y_t$. This subsetting of the time series will be discussed again later.

Again, this definition can be interpreted as a probability (for very similar reason to those mentioned above). However, this term can be rewritten using Bayes' theorem or the law of total probability as

$$P\left(y_t = 2|x_{t-1} \neq 2\right) = \frac{P\left(y_t = 2\right)}{P\left(x_{t-1} \neq 2\right)} - P\left(y_t = 2|x_{t-1} = 2\right) \frac{P\left(x_{t-1}\right.}{P\left(x_{t-1}\right.}$$

Substituting the frequency counts leads to

$$P\left(y_t = 2|x_{t-1} \neq 2\right) = \frac{n_E}{n_{nC}} - \frac{n_{EC}}{n_C}\frac{n_C}{n_{nC}} = \frac{3}{6} - \frac{3}{3}\frac{3}{6} = 0 \ ,$$

where the only frequency seen here but not used above is n_E the number of times the assumed effect $y_t = 2$ has appeared. Also consider the completeness relation $P\left(x_{t-1} \neq 2\right) = 1 - P\left(x_{t-1} = 2\right)$. This substitution leads to

$$P\left(y_t = 2|x_{t-1} \neq 2\right) = \frac{P\left(y_t = 2\right)}{1 - P\left(x_{t-1} = 2\right)} - P\left(y_t = 2|x_{t-1} = 2\right) \frac{P}{1 -}$$

and

$$P\left(y_t = 2|x_{t-1} \neq 2\right) = \frac{n_E}{1 - n_C} - \frac{n_{EC}}{n_C}\frac{n_C}{1 - n_C} = \frac{3}{3} - \frac{3}{3}\frac{3}{3} = 0 \ .$$

The last expression has removed any reference to $P\left(x_{t-1} \neq 2\right)$ and n_{nC}, which might be considered unobservable. This particular penchant is the same for both calculation methods.

Consider both these calculation methods for all the possible penchants in this pair of time series, i.e.

$\rho_{y_t x_{t-1}}$	$P(E\|C)$	$P(E\|\bar{C})$ method 1	$P(E\|\bar{C})$ method 2	result 1	result 2
$y_t = 2, x_{t-1} = 2$	1	0	0	1	1
$y_t = 2, x_{t-1} = 0$	0	1	1	-1	-1
$y_t = 0, x_{t-1} = 2$	0	1	1	-1	-1
$y_t = 0, x_{t-1} = 0$	1	0	0	1	1

For this example, both methods are equal for all the penchants. The main concern here is that the second conditional probability term in the definition of the penchant has been replaced with probabilities of the assumed causes and effects. This substitution addresses the argument that this second conditional probability term is not observable (e.g. Pearl's argument that there's no way to know what an effect given "no cause" actually is supposed to mean). However, it introduces concerns about the underlying assumption that the assumed cause must precede the assumed effect. This concern is partially addressed by subsetting $\{x\}$ and $\{y\}$ into $\{\tilde{x}\}$ and $\{\tilde{y}\}$ because for any given t, $\tilde{x}_t$ precedes $\tilde{y}_t$. But, once is this assumptions still true when frequency counts are used to estimate the necessary probabilities; e.g. when $P(x_{t-1} = 0)$ and $P(y_t = 0)$ are estimated above, is it sufficient to use

$$P(x_{t-1} = 0) = \frac{n_{x0}}{L}$$

and

$$P(y_t = 0) = \frac{n_{y0}}{L} \quad ,$$

where n_{x0} is the number of times $\tilde{x}_t = 0$, n_{y0} is the number of times $\tilde{y}_t = 0$, and L is the library length of both $\{\tilde{x}\}$ and $\{\tilde{y}\}$? The working assumption is yes, but the issue will be addressed primarily by testing the algorithm.

The algorithm used to calculate the penchants calculates only $\rho_{y_t=2,x_{t-1}=2}$ and $\rho_{y_t=0,x_{t-1}=0}$ but not $\rho_{y_t=2,x_{t-1}=0}$ and $\rho_{y_t=0,x_{t-1}=2}$ because the former are the only (y_t, x_{t-1}) pairs that actually appear in the data. The observed mean penchant is then

$$\bar{\rho}_{y_t,x_{t-1}} = \frac{1}{2}\left(\rho_{y_t=2,x_{t-1}=2} + \rho_{y_t=0,x_{t-1}=0}\right) = 1 \quad ,$$

which implies the probability that x_{t-1} drives y_t is higher than the probability that it does not. The observed mean leaning is then

$$\bar{l}_{xy} = \bar{\rho}_{y_t,x_{t-1}} - \bar{\rho}_{x_t,y_{t-1}} \quad ,$$

and the casual inference would be $\bar{l} > 0 \Rightarrow \mathbf{x} \to \mathbf{y}$, $\bar{l} < 0 \Rightarrow \mathbf{y} \to \mathbf{x}$, and $\bar{l} = 0 \Rightarrow ??$; i.e. a positive observed mean leaning implies $\mathbf{x}$ drives $\mathbf{y}$ more than $\mathbf{y}$ drives $\mathbf{x}$ with an analogous conclusion for a negative observed mean leaning.

4.1 Algorithm

5 Simple Example Systems

5.1 Impulse with Noisy Response Linear Example

Consider the linear example dynamical system of

$$X_t = \{0, 2, 0, 0, 2, 0, 0, 2, 0, 0\} \tag{9}$$
$$Y_t = X_{t-1} + B\eta_t, \tag{10}$$

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.

5.2 Cyclic Linear Example

Consider the linear example dynamical system of

$$X_t = \sin(t) \tag{11}$$
$$Y_t = X_{t-1} + B\eta_t, \tag{12}$$

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.

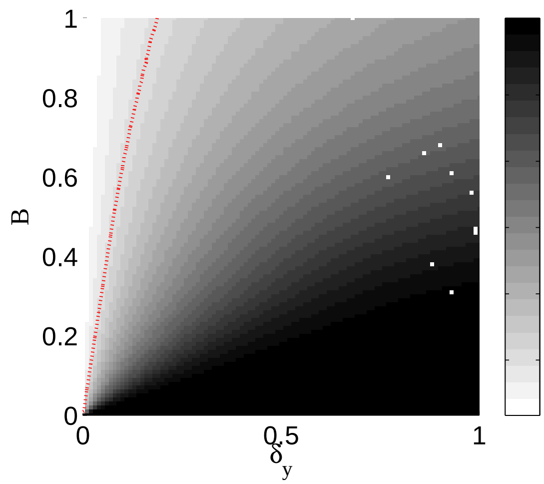

Figure 1: (Color available online.) Leaning as a function of both the noise and the y-tolerance. The red dashed line is the zero contour. See the text for am explanation of the missing data for large δ_y.

5.3 Non-Linear Example

Consider the non-linear dynamical system of

$$X_t = \sin(t) \tag{13}$$
$$Y_t = AX_{t-1}\left(1 - BX_{t-1}\right) + C\eta_t, \tag{14}$$

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

$$\frac{dI}{dt} = \frac{V(t)}{L} - \frac{R}{L}I, \tag{15}$$

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. 15 was solved

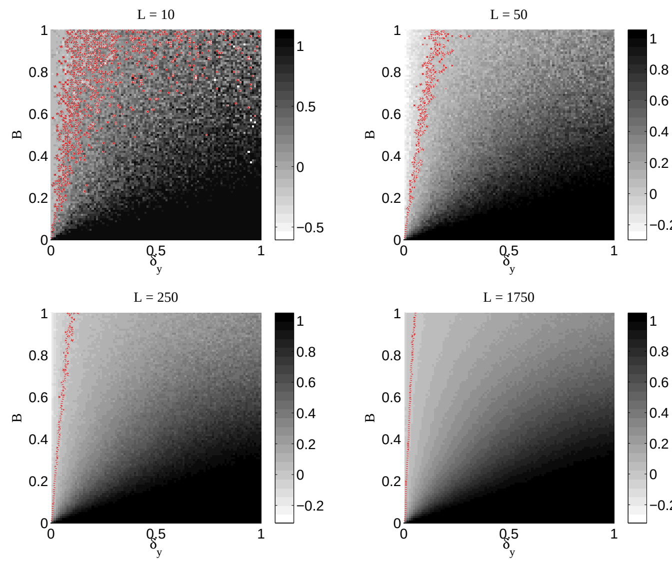

Figure 2: (Color available online.) Leaning as a function of both the noise and the y-tolerance for different library lengths. The red dashed line is the zero contour. See the text for am explanation of the missing data for large δ_y.

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

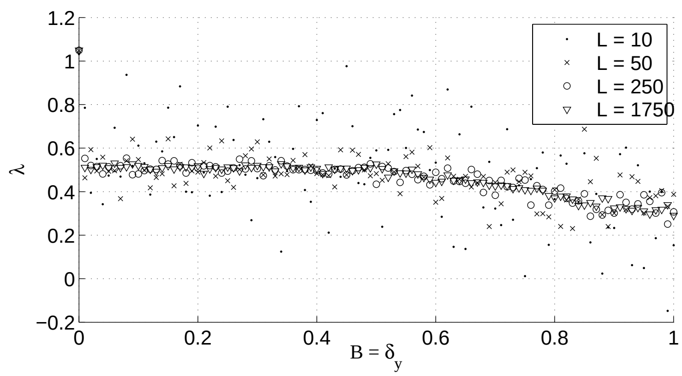

Figure 3: (Color available online.) The leaning agrees with intuition for most noise levels when the y-tolerance is set to the noise level (i.e. $B = \delta_y$).

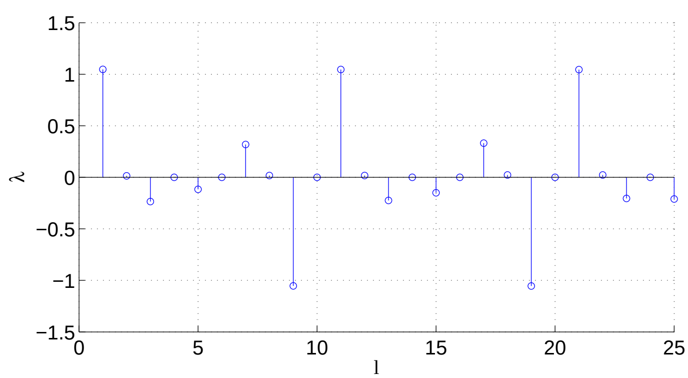

Figure 4: (Color available online.) Different l-standard cause-effect assignments lead to different leanings.