Impacting Policy By Estimating Causal Links

Hrishikesh D. Vinod *

Abstract

Statistics makes an important impact on society by analyzing quantitative evidence related to public policy issues regarding socioeconomic well-being which must be based only on non-experimental data. Suppes' probabilistic causality theory establishes inequalities among probabilities of events. Instead of events, our "cause" is a self-driven data generating process (DGP). We develop three criteria based on properties residuals of flipped kernel regression conditional expectation functions, $E\,f(X_j|X_i,X_k)$ and $E\,f(X_i|X_j,X_k)$. Our unanimity index aggregates measures of four orders of stochastic dominance and new asymmetric partial correlation coefficients, which yields decision rules for quantifying percent support for competing causal paths $X_i \to X_j, X_j \to X_i, X_i \leftrightarrow X_j$. A simulation supports our decision rules illustrated by many real world examples, including the causes of US recession, policies for encouraging private investment in India and assessment of effective advertising media.

1. Introduction

Estimation and inference regarding causal directions is a fundamental challenge in many sciences, which is linked to a long-standing need in Econometrics for testing exogeneity, without exclusively relying on instrumental variables (IV). Philosophers have debated causality concepts for over a millennium. For example, let the event X_i represent a large change in barometric pressure, and let X_j represent a weather storm. Now intuition suggests that a storm is more likely with barometric pressure change than without. Hence, Suppes' "probabilistic theory of causation," Suppes (1970), claims that if an event X_i causes X_j, the probability of X_j occurring must increase when event X_i has occurred. That is, the causal path $X_i \to X_j$ requires:

$$P(X_j|X_i) > P(X_j) \quad a.e., \tag{1}$$

where $a.e.$ denotes "almost everywhere" in a relevant measure space. Vinod (2019) provides a new proof of a known result that Suppes' condition is logically flawed– neither necessary not sufficient.

Vinod (2019) replaces probabilities of "events" by data generating processes (DGPs) and proves the following theorem implying new empirical approaches for causal paths and extending Koopmans (1950)-type exogeneity testing. Let us state it without proof.

Theorem 1: Stochastic Causality
(a) Assuming data on all confounding and control variable(s) denoted as X_k are available, the causal path $X_i \to X_j$ holds if and only if (iff)

$$\big(P(X_j|X_i,X_k) - P(X_j|X_k)\big) > \big(P(X_i|X_j,X_k) - P(X_i|X_k)\big), \quad a.e. \tag{2}$$

*address: H. D. Vinod, Professor of Economics, Fordham University, Bronx, New York, USA 10458. E-mail: vinod@fordham.edu. Tel. 201-568-5976, Fax 718-817-3518, I thank Prof. J. Francis for suggesting the 'excess bond premium' application. The advertising application is based on joint work with Kurt Jetta and Erick Rengifo. This paper was presented at Joint Statistical Meeting (JSM) in Denver Colorado Convention Center on Sunday, July 28, 2019.

(b) Assuming data on all controls X_k defined in Remark 3 are available, the causal path $X_i \rightarrow X_j$ holds iff

$$(P(X_j|X_i, X_k) - P(X_j)) > (P(X_i|X_j, X_k) - P(X_i)), \quad a.e. \tag{3}$$

Upon replacing above probabilities of events $P(X_i, X_j, X_k)$ by densities $f(X_i, X_j, X_k)$ of DGPs the iff conditions in the Theorem 1 replace $P(.)$ by $f(.)$, and measure numerical differences between resulting $f(.)$ expressions. An R package, Vinod (2016), uses nonlinear, nonparametric kernel regression residuals to measure inequalities between $f(.)$ expressions.

1.1 Pearson correlation coefficient underestimates dependence by 91%

Consider a simple example showing that the traditional Pearson correlation coefficient can be a very poor measure of dependence when the variables have a perfect nonlinear relation, such as $y = sin(x)$.

```
x=1:20; y=sin(x)
reg=lm(y~x)
yhat=fitted(reg)
matplot(cbind(x,y,yhat), main="Plot of x=1:20 and sin(x)", typ="l",
xlab="sequence number", ylab="x, y and OLS fitted y",lwd=2)
nam=c("x", "y", "OLS fitted y")
legend(x=1,y=17,legend=nam, lty=c(1:3), col=1:3,lwd=2)
cor(x,y)# simple correlation between x and y
coef(reg)
```

The graphical output of the above R code in Figure 1 depicts the sequence plot of x as a solid line and $y = sin(x)$ as a dashed line, showing the nonlinear relation between x and y. The figure also contains a dotted line representing ordinary least squares (OLS) fitted straight line with a small positive intercept (0.1735) and barely visible very small negative slope (-0.0118). The graph reveals a well-known fact that when the relationship is nonlinear, OLS fit can be very poor. By contrast, the fitted values $\hat{y}$ based on cross-validated bandwidths of the R package 'np' for nonparametric kernel estimation, Hayfield and Racine (2008), visually coincide with y in the data. Although Figure 1 cannot depict y and np package $\hat{y}$ as distinct lines, we do have nonzero vectors of residuals for further analysis.

The output of the last part of the above code is

```
> cor(x,y)# simple correlation between x and y
[1] -0.0948372
> coef(reg)
(Intercept)           x
0.17352349 -0.01177261
> cor(y,yhat)^2 #squared correlation between y and fitted y
[1] 0.008994095
> summary(reg)$r.sq
[1] 0.008994095
```

The above output shows that the squared Pearson correlation coefficient r^2_{xy} equals the R^2 of the OLS regression of y on x, and also of the flipped regression of x on y.

The R function 'cor.test(x, y)' tests the null hypothesis $\rho_{xy} = 0$, that the population correlation coefficient is zero yields the following output.

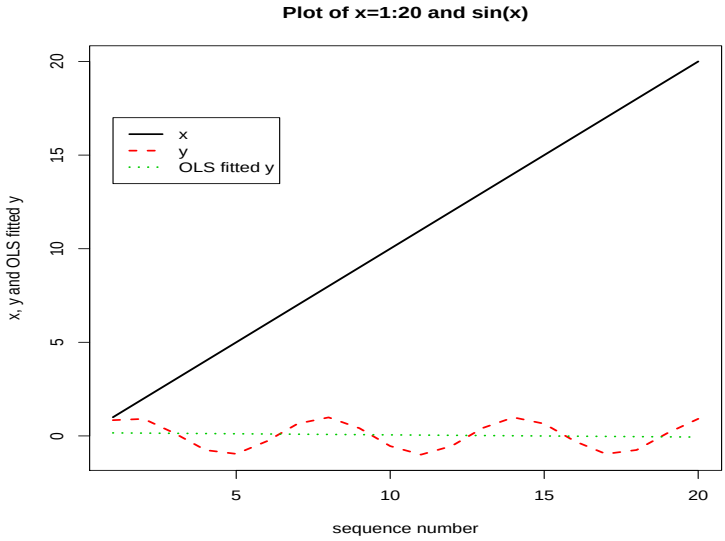

Figure 1

```
Pearson's product-moment correlation
t = -0.40418, df = 18, p-value = 0.6908
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:        -0.5157148  0.3629142
sample estimates: cor -0.0948372
```

Fox (2009) package 'car' helps create a scatterplot between our x and y in Figure 2 depicts the scatter diagram. A slightly downward sloping solid line depicts the linear regression of y on x. A locally best fitting zig-zag fit snaking around the linear fit and its 95% confidence band are also seen in Figure 2.

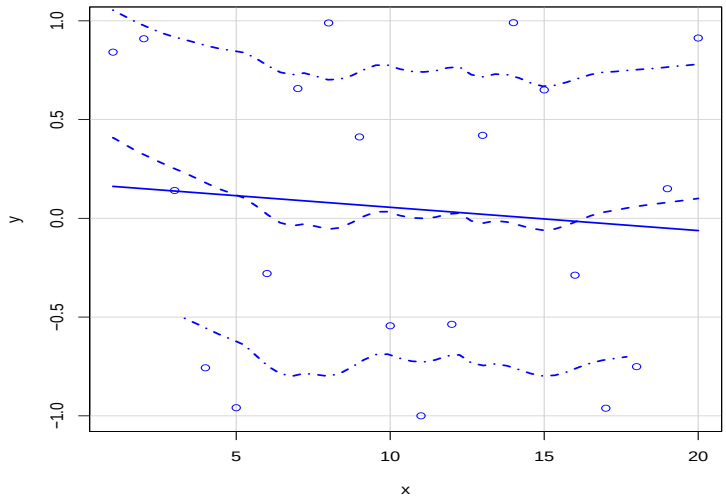

Figure 2

Since Pearson correlation coefficient r_{xy} = −0.09 underestimates dependence by

as much as 91% for our simple example, it highlights a long-standing need for a generalized correlation coefficient between variables related by nonlinear relations.

1.2 Asymmetric Matrix of Generalized Correlations

In addition to providing superior fits, kernel regressions using cross-validation to determine bandwidths can allow us to define a non-symmetric matrix of generalized correlation coefficients. We shall see that generalized correlations can be useful for distinguishing causal paths $X_i \rightarrow X_j$ from $X_j \rightarrow X_i$.

Generalized measures of correlation (GMC) defined by eq. (2) in Zheng et al. (2012) are:

$$GMC(X_j|X_i) = [1 - \frac{E(X_j - E(X_j|X_i))^2}{var(X_j)}], \tag{4}$$

$$GMC[X_i|X_j] = [1 - \frac{E(X_i - E(X_i|X_j))^2}{var(X_i)}], \tag{5}$$

which are computed simply as the R^2 values of flipped kernel regressions. Since R^2 values cannot be negative, any GMC is not a proper generalization of often negative Pearson's correlation coefficient, $r_{ij} \in [-1, 1]$. Vinod (2014) overcomes the range problem by defining a generalized correlation coefficient (distinguished by an asterisk) from the $GMC(X_i|X_j)$ as:

$$r^*_{X_i|X_j} = r^*_{i|j} = \text{sign}(r_{ij})\sqrt{GMC(X_i|X_j)}, \tag{6}$$

where the square root of the GMC is assigned the sign of the Pearson correlation coefficient.

In general, the $R^2 = GMC(X_i|X_j)$ of kernel regression of X_i on X_j is distinct from the $R^2 = GMC(X_j|X_i)$ of kernel regression of X_j on X_i. Note that $r^*_{i|j} \neq r^*_{i|j}$, and that $\{r^*_{i|j}\} \in [-1, 1]$ must hold. Assuming that we have data on $X_i, i = 1, 2, \ldots p$ variables, we can construct a $p \times p$ asymmetric matrix $R^* = \{r^*_{i|j}\}$ of generalized correlation coefficients. The computation of R^* regards X_i, the left-hand-side variable, as the "response" named along the rows of $\{r^*_{i|j}\}$, whereas the right-hand-side regressor X_j is, in some sense, the "cause." How to compute the R^* matrix? An R function 'gmcmtx0(mtx)' of the package "generalCorr" is the answer.

The R code to compute a generalized correlation matrix R^* for our simple example ($y = sin(x)$) is:

```
library(generalCorr);x=1:20; y=sin(x)
options(np.messages=FALSE)
gmcmtx0(cbind(x,y)) #R* matrix
```

The output of the above code is a 2×2 matrix R^* below.

```
   x           y
x  1 -0.04847292
y -1  1.00000000
```

In our simple example, we know that $x = 1 : 20$ is the (driver) "cause" of $y = sin(x)$, implying the causal path $x \rightarrow y$. Rounding to 3 places the output above states that $r^*_{x|y} = -0.048, r^*_{y|x} = -1$. That is the kernel regression y of x has near-perfect fit, $(r^*_{y|x})^2 = 1$, whereas the flipped regression of x on y has a poor fit.

The asymmetry of the matrix of generalized correlation coefficients R^* provides a preliminary indicator of the correct causal path $x \to y$.

Next, let us implement the two flipped kernel regressions called k1 and k2 in the following code, and evaluate the size of the absolute residuals. The following code reports the maximum absolute residual values for both k1 and k2.

```
k1=kern(dep.y=y, reg.x=x,residuals=TRUE)
max(abs(y-fitted(k1)))# this is very small only 1.16*10^{-16}
k2=kern(dep.y=x, reg.x=y,residuals=TRUE)#flipped kernel regression
max(abs(x-fitted(k2)))# this is large 9.5
```

The following output of the above code shows that fitted values $\hat{y}$ of kernel regression model k1 [when the dependent variable (dep.y=y)] is very close to the correct y. The largest error is near zero, $max(|y-\hat{y}|) = 0.0000$, rounded to four digits. By contrast, the alternate model k2 [when the dependent variable (dep.y=x)] the residuals $x-\hat{x}$ are large, with a relatively huge maximum, $max(|x-\hat{x}|) = 9.5$. Clearly, the kernel regression k2 with RHS variable y is a poor choice. The model k1 with x as the 'cause' on the RHS is better.

```
>  max(abs(y-fitted(k1)))
[1] 1.110223e-16
>  max(abs(x-fitted(k2)))
[1] 9.5
```

A comparison of absolute values of generalized correlation coefficients is seen to be a plausible indicator of the causal path $x \to y$. Since $R^*\{ij\}$ matrix, is always between X_i and X_j or only two variables. Two variables do not suffice when we want the study causal paths in the presence of control variables X_k. In general, we need to extend causal identification to remove the effect of a set of control variables before measuring the generalized correlation between variable pairs.

1.3 Generalized Partial Correlations

Since the kernel causality theorem involves X_k variables (confounders or controls) this subsection allows for them by generalizing partial correlation coefficients. Recall the definition of the partial correlation between (X_1, X_2), after removing the effect of (X_3):

$$r_{12;3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)}\sqrt{(1-r_{23}^2)}}. \tag{7}$$

Since $r_{12;3} = r_{21;3}$ holds, traditional partial correlations cannot reveal causal directions. We need asymmetric generalized partial correlation coefficients.

Kendall and Stuart (1977) show that an alternative definition of $r_{12;3}$ is a simple correlation between residuals of the regression: $X_1 = f_1(X_2, X_3) + \epsilon_{1|2,3}$, and similar residuals of the regression: $X_2 = f_2(X_1, X_3) + \epsilon_{2|1,3}$. We use this method in our generalization as follows.

We consider the generalized correlations between X_i and X_j after removing the effect of a set of variable(s) in X_k. Define $e_{i|k}$ as the observable residual of kernel regression of X_i on all control variable(s) X_k (but not X_j). Similarly, define $e_{j|k}$ as the residual of kernel regression of X_j on all control variable(s) X_k. Next, we define

a preliminary (symmetric) generalized partial correlation coefficient in the presence of control variable(s) as:

$$e^*_{i|j,k} = \frac{cov(e_{i|k}\ e_{j|k})}{\sigma(e_{i|k})\sigma(e_{j|k})}. \tag{8}$$

This formula computes the usual (symmetric) correlation coefficient between two relevant residuals. Denote the sign of the symmetric correlation in eq. (8) as $\text{sign}(e^*_{i|j,k})$, which also equals $\text{sign}(e^*_{j|i,k})$.

Now, we are ready to define an asymmetric matrix of generalized partial correlation coefficients using the R^2 of nonparametric kernel regression between two residual vectors: $e_{i|k} = f(e_{j|k}) + \epsilon$, as $GMC(e_{i|k}|e_{j|k})$. In general, $GMC(e_{i|k}|e_{j|k}) \neq GMC(e_{j|k}|e_{i|k})$, means that generalized measures of *partial* correlation are asymmetric.

Thus, a generalized partial correlation coefficient for a kernel regression having X_i on the left hand side (LHS) and X_j, X_k on the right hand side (RHS) is defined as:

$$r^*_{i|j,k} = r^*(X_i|X_j; X_k) = \text{sign}(\text{e}^*_{\text{i|j,k}})\sqrt{[\text{GMC}(\text{e}_{\text{i|k}}|\text{e}_{\text{j|k}})]}. \tag{9}$$

Its counterpart on the other side of the diagonal, upon interchanging i and j as a part of the matrix of generalized partial correlations is:

$$r^*_{j|i,k} = r^*(X_j|X_i; X_k) = \text{sign}(\text{e}^*_{\text{i|j;k}})\sqrt{[\text{GMC}(\text{e}_{\text{j|k}}|\text{e}_{\text{i|k}})]}. \tag{10}$$

Recalling our $y = sin(x)$ example, let us create a random control variable z from the uniform density with the code:

```
library(generalCorr);x=1:20; y=sin(x)
set.seed(99); z=runif(20)
options(np.messages=FALSE)
parcor_ijk(xi=x, xj=y, xk=z)
```

The following output of the above code reports 'ouij' denoting generalized partial correlation of X_i with X_j (=cause) after removing the effect of X_k, and 'ouji' denoting generalized partial correlation of X_j with X_i (=cause) after removing the effect of X_k.

```
$ouij
[1] -7.505178e-06
$ouji
[1] -0.7304486
```

Note that absolute values for partial correlation coefficients satisfy $|ouij| < |ouji|$, supporting the correct causal path $X_i \to X_j$ or $x \to y = sin(x)$ in our context.

1.4 Kernel Causality Criteria Cr1 to Cr3

Consistency is a basic requirement in both OLS and kernel regressions. A regression is consistent only when RHS variables are 'exogenous' or uncorrelated with the error terms. For brevity, let us denote the conditional expectation $E\,f(X_j|X_i, X_k)$ by $g_{j|ik}$, where the subscripts after the bar remind the reader that variables (X_i, X_k) are on the RHS. Now, $g_{j|ik}$ is consistently estimated, provided the vector of true unknown

errors $\epsilon_{j|ik}$ is uncorrelated with (X_i, X_k). The well-known Hausman-Wu exogeneity test, Wu (1973), also states the same condition and is called the first criterion, Cr1, implemented in Vinod (2016). Assume we have $t = 1, 2, \cdots, T$ observations on all variables and observable residuals are a good approximation to true unknown errors.

Since the implicit causal direction of a regression model is always (RHS → LHS), the model where X_i is on the LHS while (X_j, X_k) are on the RHS is associated with the causal path $(X_j \rightarrow X_i)$. The consistency of this model requires that each RHS variable be uncorrelated with the error vector. Hence, each (causal) RHS variable should have a "smaller" correlation with the regression residuals.

(Cr1) The first criterion for kernel causality path $X_i \rightarrow X_j$ has X_i on RHS involving conditional expectation function $g_{j|i,k}$. We want the correlation between the RHS variable X_i and residuals $e_{j|i,k}$ to be relatively more close to zero, or "smaller" in absolute value, than the correlation between the flipped path RHS variable X_j and its residuals $e_{i|j,k}$. That is, we want to satisfy the inequalities:

$$|(X_{it})e_{(i|j,k)t}| < |(X_{jt})e_{(j|i,k)t}|, \tag{11}$$

as often as possible.

(Cr2) The second criterion incorporates the idea that path $X_i \rightarrow X_j$ having X_i on the RHS should have "smaller" absolute residuals (superior fit) than those from the flipped path $X_j \rightarrow X_i$ with X_j on the RHS. Hence, upon inserting a subscript for t-th observation, we write:

$$(|e_{(j|i,k)t}|) < (|e_{(i|j,k)t}|) \tag{12}$$

(Cr3) The generalized partial correlations for the path $X_i \rightarrow X_j$ should satisfy:

$$|r^*_{(j|i,k)}| > |r^*_{(i|j,k)}|, \tag{13}$$

where generalized partial correlation coefficients defined in eq. (10) removes the effect of the control variable(s), if any.

The inequalities of equations (11) and (12) are fuzzy in the sense that we do not expect them to be satisfied for each t. One can think of a kernel density as a smoothed histograms. Now kernel densities of all absolute value expressions appearing in (11) and (12) are smoothed histograms. We want the density on the "larger" side of each inequality to be "larger," in some sense. The concept of one density being "larger" than another density, despite overlapping areas, is formalized by requiring that the "larger" density should stochastically dominate the other density.

1.5 Stochastic Dominance and Fuzzy Inequalities

Financial economists need to decide whether to invest in one stock or another, for example, in IBM or GE, based on their forecast returns, (R_{IBM}, R_{GE}). Unfortunately, sometimes $(R_{IBM} > R_{GE})$ while at other times $(R_{IBM} < R_{GE})$. That is, we want to choose even though the return densities overlap. How to choose between the two streams of returns when the inequality is fuzzy? A well-known sophisticated

solution to this old problem suggests choosing the investment which stochastically dominates the other, Levy (1992). The tools solving the investment problem can be tell us the extent to which any fuzzy inequality holds true, despite partly overlapping densities of the two sides.

Let us briefly describe stochastic dominance (SD) tools without attempting to summarize the vast and growing published and unpublished literature motivated by financial economists' portfolio choice problem as explained in (Vinod, 2008, ch.4). A density $f(x)$ dominates $f(y)$ in the first order (SD1) if their respective empirical *cumulative* distribution functions (ecdf) satisfy: $F(x) \leq F(y)$. This is a well-known slightly counter-intuitive result stating that the cumulative density $F(x)$ of the dominating density $f(x)$ must be *smaller* than the cumulative density $F(y)$ of the dominated density $f(y)$. It is also well known that SD1 provides a comprehensive picture of the ranking between two probability distributions focusing on a vector of 'locally-rolling data-based' first moments (means, or moving averages).

The underlying computation requires bringing the two densities on a common x-axis known as 'support,' requiring ecdf's to have up to $2T$ possible jumps or steps. Hence there are $2T$ estimates of $F(x) - F(y)$ denoted by a $2T \times 1$ vector (sd1). Anderson (1996) provides a convenient (trapezoidal) numerical integration tool so that a cumulative sum summarizes SD1.

Second order dominance (SD2) of $f(x)$ over $f(y)$ requires further integrals of ecdf's to satisfy a fortunately analogous inequality: $\int F(x) \leq \int F(y)$ providing a cumulative sum summarizing SD2. Analogous cumulative sums and higher-order integrals summarize SD3 and SD4 for locally moving skewness and kurtosis measures.

We consider decision rules based on a summary of quantified SD1 to SD4 described next.

1.6 Derivation of Causality Decision Rules

By analogy with two streams of investment returns, stochastic dominance measures (SD1 to SD4) of four orders allow us to study fuzzy inequalities of (11) for Cr1 and those of (12) for Cr2. These four orders are roughly associated with the first four locally defined rolling moments (mean, standard deviation, skewness, and kurtosis) yielding four numbers each for Cr1 and Cr2. The signs of these four numbers suggest the causal path direction implied by the four moments of the underlying kernel density. Using appropriate weights for the reliability of sample moments while allowing an option to change weights, Vinod (2016) first obtains an index of the sign implied by Cr1 and another index for Cr2. Our Cr3 based on partial correlation in (13) also yields a comparable sign of. $|r^*_{(i|j,k)}| - |r^*_{(j|i,k)}|$, indicating the direction of the causal path.

We cannot assume that sample estimates of the three distinct criteria will suggest the same causal direction. Hence we need a sample measure of the unanimity among the three criteria, denoted by (ui). Denote the corresponding population unanimity index as (UI) shown to lie in the intuitive closed interval: $[-100, 100]$. It measures both the direction and "unanimity strength" of the indicated causal path.

Individual application determines the threshold τ so that $|ui| < \tau$ means $ui \approx 0$. Choosing a 5% threshold, $\tau = 5$, say, our **Decision Rules** are:

Ru.1: If $(ui < -\tau)$ the causal path is: $X_i \rightarrow X_j$

Ru.2: If $(ui > \tau)$ the causal path is: $X_j \rightarrow X_i$

Ru.3: If ($|ui| \leq \tau$) we obtain bi-directional causality: $X_i \leftrightarrow X_j$, that is, the variables are jointly dependent.

Simulations showing the success of such causality tools are reported in Vinod (2017).

2. Simulation for Checking Decision Rules

We assume that X_1 is exogenous, and generate it independent of any other variable. Next, we generate our X_2 variable which depends on X_1 and also involves adding a noise term, defined from the standard normal deviate, $\epsilon \sim N(0,1)$. Our decision rules are known to perform better in the *absence* of normality and linearity. We hope to create conservative and robust experiments by using ϵ values designed to handicap (not favor) our decision rules.

In the following experiments, X_1 is an independently generated (exogenous) DGP, and hence the causal path is known to be $X_1 \rightarrow X_2$, by construction. We use sample sizes: $T = 50, 100, 300$, to check if our decision rules correctly assess the causal path, despite the handicap of linearity, normality, or both.

Let m denote the count for indeterminate signs when we repeat the experiments $N = 1000$ times. Define the success probability (suPr) for each experiment as:

$$(suPr) = \frac{(count\ of\ correct\ signs)}{N - m}. \tag{14}$$

The simulation considers four sets of artificial data where the causal direction is known to be $X_1 \rightarrow X_2$.

1. Time regressor: $X_1 = \{1, 2, 3, \cdots, T\}$
   $X_2 = 3 + 4X_1 + \epsilon$

2. Unit root Quadratic:

   X_1 has T random walk series from a cumulative sum or standard normals.
   $X_2 = 3 + 4X_1 - 3X_1^2 + \epsilon$

3. Two Uniforms:

   X_1, Z_1 each have T uniform random numbers
   $X_2 = 3 + 4X_1 + 3Z_1 + \epsilon$

4. Three Uniforms:

   X_1, Z_1, Z_2 each have T uniform random numbers
   $X_2 = 3 + 4X_1 + 5Z_1 - 6Z_2 + \epsilon$

The simulation required about 36 hours on a Dell Optiplex Windows 10 desktop running Intel Core i5-7500, CPU at 3.40 GHz, RAM 8 GB, R version 3.4.2.

The large success proportions (suPr) reported in row 7 (for T=50), row 15 (for T=100) and row 23 (for T=300) of Table 1 assume the threshold $\tau = 0$. The results for the four experiments in four columns show that our decision rules using a 'ui' from Cr1 to Cr3 work well. The effect on success probabilities of the choice of the threshold is studied for the $T = 300$ case by using $\tau = 0, 15, 20, 25$, respectively, along rows 21 to 24.

Moreover, since the success probabilities 'suPr' for $\tau = 0$ along rows 7, 14 and 21 increase as $T = 50, 100, 300$ increases, this suggests a desirable asymptotic convergence-type property. Thus, our decision rules are supported by the simulation.

Table 1: Summary statistics for results of using the 'ui' measure for correct identification of causal path indicated by its positive sign using N=1000 repetitions, T=50, 100, 300 sample sizes along three horizontal panels. Success probabilities (suPr) show convergence as T increases in the three panels.

Row	stat.	Expm=1	Expm=2	Expm=3	Expm=4
1	Min.T=50	31.496	-100.000	-100.000	-100.000
2	1st Qu.	63.780	31.496	31.496	-31.496
3	Median	100.000	31.496	31.496	37.008
4	Mean	82.395	33.725	24.386	27.622
5	3rd Qu.	100.000	100.000	37.008	37.008
6	Max.	100.000	100.000	100.000	100.000
7	suPr	1.000	0.793	0.808	0.712
8	Min.T=100	31.496	-100.000	-100.000	-100.000
9	1st Qu.	63.780	31.496	31.496	31.496
10	Median	81.102	31.496	31.496	37.008
11	Mean	74.691	33.106	32.822	35.879
12	3rd Qu.	100.000	100.000	37.008	37.008
13	Max.	100.000	100.000	100.000	100.000
14	suPr	1.000	0.787	0.892	0.803
15	Min.T=300	31.496	-100.000	-31.496	-63.780
16	1st Qu.	81.102	31.496	31.496	37.008
17	Median	81.102	31.496	31.496	37.008
18	Mean	80.357	43.020	42.973	42.117
19	3rd Qu.	100.000	100.000	37.008	37.008
20	Max.	100.000	100.000	100.000	100.000
21	suPr,$\tau = 0$	1.000	0.829	0.987	0.963
22	suPr,$\tau = 15$	1.000	0.833	0.988	0.970
23	suPr,$\tau = 20$	1.000	0.835	0.989	0.971
24	suPr,$\tau = 25$	1.000	0.836	0.989	0.971

3. A Bootstrap for Inference

Statistical inference regarding causal paths and exogeneity uses ui for estimating the population parameter UI.

Bootstrap Percentile Confidence Interval: We suggest a large number J of bootstrap resamples of (X, Y, Z) data to obtain $(N_{all})_j$ and $(\text{ui})_j$ using any bootstrap algorithm. These $(j = 1, \cdots J)$ values provide an approximation to the sampling distribution of 'sum' or 'ui.' We can easily sort the J values from the smallest to the largest and obtain the 'order statistics' denoted as $(\text{ui})_{(j)}$, with parenthetical subscripts. Now a $(1 - \alpha)100$ percent confidence interval is obtained from the quantiles at $\alpha/2$ and $1 - \alpha/2$. For example, if $\alpha = 0.05, J = 999$, 95% confidence interval limits are: $(\text{ui})_{(25)}$ and $(\text{ui})_{(975)}$.

Recalling the decision rules Ru.1 to Ru.3 of Section 1.6, if both confidence limits fall inside one of the two half-open intervals, we have a statistically significant conclusion. For example, Ru.1 states that: If $(ui < -\tau)$ the causal path is: $X_i \to X_j$. If instead of a point estimate we have two limits, we want both confidence limits of ui lie in the same half-open interval: $[-100, -5)$. Then, we have a statistically significant conclusion that $X_i \to X_j$, or equivalently that X_i is exogenous.

This paper uses the maximum entropy bootstrap (meboot) R package described in Vinod and López-de-Lacalle (2009) because it is most familiar to me, retains the dependence structure in the data, and is recently supported by simulations in Yalta (2016), Vinod (2015) and elsewhere. An advantage of meboot is that it permits bootstrap inference even if the variables in the model are not stationary. Following Stock (1987) such specification in data levels (without differencing or de-trending) allows the estimators to be super-consistent.

3.1 Summarizing Sampling Distribution of ui

One can use J (=999) resampled estimates of ui to approximate the sampling distribution $f(ui)$. The usual summary statistics of these J values yield preliminary information about its quantiles. We will illustrate this with the help of examples later such as in Table 4.

Another way of studying $f(ui)$ involves computing bootstrap proportion of significantly positive or negative values. Let m denote the bootstrap count of indeterminate signs when (ui)$\in [-\tau, \tau]$. A researcher can change the threshold $\tau = 5$ depending on the problem at hand. Now define a bootstrap approximation to the proportion of significantly positive signs as:

$$P^*(+1) = \frac{(count\ of\ ui_j > \tau)}{J - m}. \tag{15}$$

Similarly, a bootstrap approximation to the proportion of significantly negative signs is:

$$P^*(-1) = \frac{(count\ of\ ui_j < -\tau)}{J - m}. \tag{16}$$

4. Application to US Macroeconomic Policy

This section discusses an application of our tools for predicting a downturn of the US economy. Since the US economy has had one of the longest stretches of steady economic growth, this is a matter of great topical interest. Macroeconomists and Federal Reserve Bank researchers being aware of their failure to forecast the last great recession of 2007-2008 have developed a new data series. For example, Gilchrist and Zakrajek (2012) excess bond premium (EBP) series has been shown to predict recession risk. The term-spread, defined as the difference between long term yield (10-year) and short term yield (1-year) on government securities is shown by Bauer and Mertens (3/5/2018) to be an excellent predictor of recessions.

Instead of directly predicting discrete events like recessions, we are attempting to study what macroeconomic variables drive EBP and term-spread themselves. Our term-spread is denoted as $Dyld$ or difference in yields on 10-year and 6-month government securities and discussed later in subsection 4.1. A negative $Dyld$ is sometimes called the inverted yield curve and described as a predictor of recessions in the popular press. We use the Federal Reserve Bank's fairly long quarterly data set from 1973Q1 to 2017Q1.

The RHS of the following nonparametric regression lists variables which can 'cause' or explain the EBP.

$$\text{EBP} = \text{f}(\text{Yld10}, \text{eFFR}, \text{CrCrea}, \text{CrDstr}, \text{UnemR}, \text{M2}, \text{MbyP}, \text{YbyHrs}, \text{JD}, \text{JC}), \tag{17}$$

where self-explanatory symbols are: yield on 10-year treasury bonds ($Yld10$, not seasonally adjusted), effective federal funds rate ($eFFR$), and credit creation ($CrCrea$, not seasonally adjusted), credit destruction ($CrDstr$, not seasonally adjusted), unemployment rate ($UnemR$), money stock ($M2$, seasonally adjusted billions of dollars), ratio of M to $PGDP$ or real money supply ($MbyP$), ratio of GDP to employee hours, or productivity, ($YbyHrs$), job destruction (JD)), and job creation (JC).

Certain arguments for using separate variables for $CrCrea$ and $CrDstr$ are found in Contessi and Francis (2013), who provide additional references for interested readers. Next, we apply the decision rules of Section 1.6 to check whether the RHS variables in eq. (17) are indeed exogenous with independent self-driving DGP's.

Table 2 explicitly reports for each flipped pair the 'cause' and 'response' such that the LHS variable EBP in eq. (17) is present in all pairs. The column entitled 'strength' reports the absolute value of the unanimity index $|ui|$. The sign of ui determines the direction of the causal path, that is, the variable name in 'cause' column, and also the variable name in the 'response' column. For example, line 6 has $M2$ in the 'cause' column and EBP in the 'response' column, because $ui < 0$ implies that $M2$ is exogenous or causal. The column entitled 'corr' of Table 2 reports Pearson's correlation coefficient with EBP. The column entitled ' p-value' reports the p-value for testing the null of zero correlation. Of course, Kernel causality and exogeneity need not agree with traditional correlation inference, since the latter assumes linearity and normal distributions.

Whenever $ui > 0$, we place EBP in the 'cause' column. Table 2 line 1b reports that ui is positive and smaller than that for $Yld10$ along row 1. We have focused more on EBP than $Dyld$ because EBP has greater independent innovations than $Dyld$ according to line 1b, where $Dyld$ does not 'cause' EBP. The simple correlation between EBP and $Dyld$ is statistically insignificant and lower than the correlation between EBP and $Yld10$ reported along row 1.

Note that only $M2, YbyHrs, JD$ and JC are likely to be self-driven (exogenous) causing the excess bond premium, while all other variables seem to be endogenous, being caused by EBP.

Table 2: Excess Bond Premium and possible causes

	cause	response	strength	corr.	p-value
1	EBP	Yld10	47.244	0.0866	0.25161
1b	EBP	Dyld	31.496	0.0416	0.58258
2	EBP	eFFR	31.496	0.0902	0.23248
3	EBP	CrCrea	31.496	-0.0606	0.42322
4	EBP	CrDstr	31.496	0.2617	0.00043
5	EBP	UnemR	31.496	0.1108	0.14222
6	M2	EBP	31.496	-0.0536	0.47843
7	EBP	MbyP	31.496	0.0195	0.79659
8	YbyHrs	EBP	31.496	-0.0588	0.43693
9	JD	EBP	31.496	0.47	0
10	JC	EBP	31.496	-0.1323	0.07915

4.1 Variables affecting term spread *Dyld*

Before we turn to statistical inference associated with the results we include additional tables created by choosing *Dyld* as the dependent variable. New causal paths and their strengths when the dependent variable in (17) is *Dyld*, or the term spread measured by the difference in yields, not EBP. Our matrix 'mtx' as a part of the R function call 'causeSummary(mtx)' now has *Dyld* as the first column and the remaining ten columns have the variables listed in equation (18). The resulting Table 3 has ten rows when the ten variables are paired with the first variable *Dyld*.

$$\text{Dyld} = \text{f}(\text{Yld10}, \text{eFFR}, \text{CrCrea}, \text{CrDstr}, \text{UnemR}, \text{M2}, \text{MbyP}, \text{YbyHrs}, \text{JD}, \text{JC}). \quad (18)$$

Table 3: Term Spread between 10-yr to 6-month treasury yields and possible causes

	cause	response	strength	corr.	p-value
1	Yld10	Dyld	31.496	-0.1862	0.0131
2	Dyld	eFFR	100	-0.5463	0
3	Dyld	CrCrea	37.008	-0.1666	0.02668
4	Dyld	CrDstr	100	0.3107	3e-05
5	Dyld	UnemR	31.496	0.5149	0
6	M2	Dyld	31.496	0.2412	0.00122
7	MbyP	Dyld	31.496	-0.0014	0.98522
8	YbyHrs	Dyld	31.496	0.2718	0.00025
9	Dyld	JD	37.008	-0.0403	0.5939
10	Dyld	JC	100	-0.1359	0.07121

The results in Table 3 show that independent variation in *Dyld*, similar to EBP, drives that in variables: $(eFFR, CrCrea, CrDtr, UnemR, JD, JC)$. By contrast, *Dyld* is driven by variables $(Yld10, M2, MbyP, YbyHrs)$. This contrasts with the driver variables $(M2, YbyHrs, JDandJC)$ in Table 2. The common drivers in both tables are money supply $M2$ and productivity $YbyHrs$.

4.2 Bootstrap inference on Estimated Causality Paths for EBP

What about sampling variability of *ui*? We resample the data 999 times using the 'meboot' package to keep the time-series properties of the data unchanged. We report the the summary statistics using the 999 estimates of *ui* for the first five variables in Table 4 and the last five variable in Table 5.

Table 4: Summary statistics of 999 bootstrap estimates of causal directions and strengths, Part 1

	Yld10	eFFR	CrCrea	CrDstr	UnemR
Min.	-31.50	-100.00	-31.50	-31.50	-31.50
1st Qu.	31.50	-31.50	-31.50	31.50	-31.50
Median	47.24	-31.50	31.50	31.50	-31.50
Mean	55.18	-4.88	6.36	29.32	-10.84
3rd Qu.	81.10	31.50	31.50	31.50	31.50
Max.	100.00	100.00	100.00	100.00	100.00

Table 5: Summary statistics of 999 bootstrap estimates of causal directions and strengths, Part 2

	M2	MbyP	YbyHrs	JD	JC
Min.	-31.50	-31.50	-31.50	-31.50	-100.00
1st Qu.	-31.50	31.50	-31.50	-31.50	-31.50
Median	-31.50	31.50	-31.50	-31.50	-31.50
Mean	-31.50	32.77	-31.50	-1.29	-29.55
3rd Qu.	-31.50	31.50	-31.50	31.50	-31.50
Max.	-31.50	100.00	-31.50	100.00	100.00

Table 6 shows that our approximate sampling distribution results provide a distinct piece of information not covered by the results about the strength or the p-value in Table 2. The table contains the proportion of negative or positive (whichever is most prevalent) in each column described as bootstrap success rates, defined in equations (16) and (15).

Table 6: Bootstrap success rates for causal direction using 999 resamples

	variable	P(±1)
1	Yld10	0.9737
2	eFFR	0.6232
3	CrCrea	0.537
4	CrDstr	0.951
5	UnemR	0.7124
6	M2	1
7	MbyP	0.964
8	YbyHrs	1
9	JD	0.5486
10	JC	0.9758

We recommend careful analysis of each causal pair with the help of scatterplots. We include only two plots here for brevity: (i) EBP-UnemR pair where UnemR is found to be endogenous, and (ii) EBP-M2 pair where M2 is found to be exogenous. Histograms of the two variables are seen in the diagonal panels of Figures 3 and 4. The South West panels have a scatter diagram and locally best fitting free hand curve. The number in the North East panels is the ordinary correlation coefficient whose font size suggests its statistical significance.

Figure 3 depicts a scatterplot having a mildly up-down-up pattern. This may explain why the Pearson correlation coefficient of 0.11 is statistically insignificant, since it does not capture nonlinear relations. Endogeneity of unemployment rate suggests that it is likely an effect of recessions and not a cause. Note that Figure 4 suggests that the variation in M2 is weakly exogenous. Its scatterplot is U-shaped and quite noisy, confirming highly insignificant and small Pearson correlation coefficient. Thus a decline in M2 may help predict recessions.

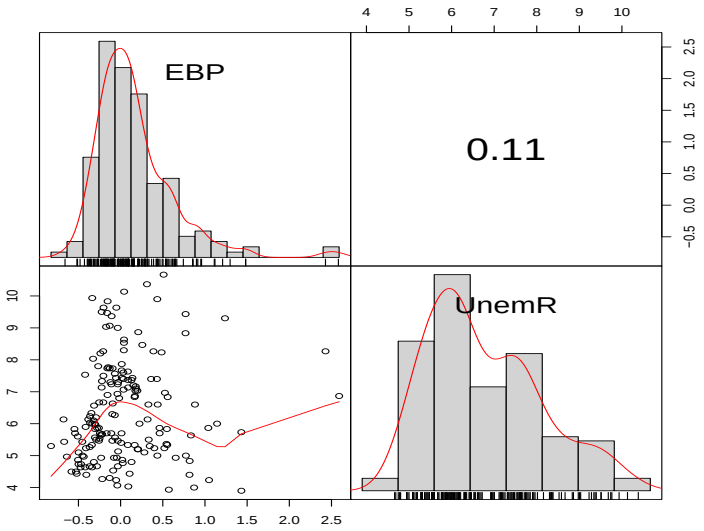

Figure 3: Scatterplot with nonlinear curve: EBP-UnemR

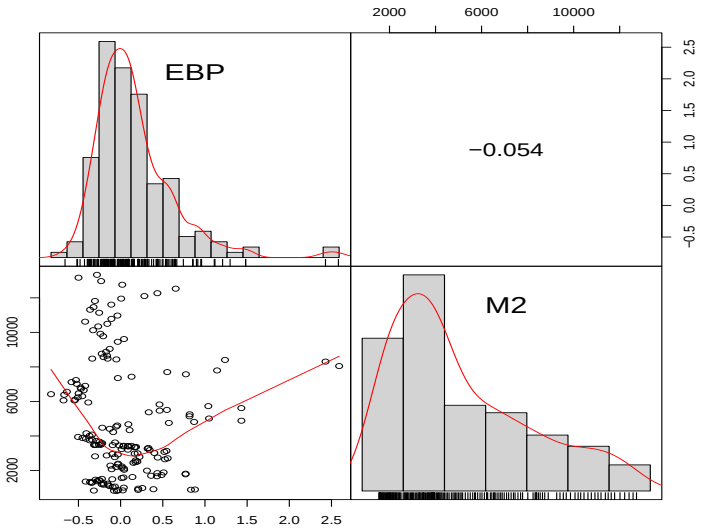

Figure 4: Scatterplot with nonlinear curve: EBP-M2

5. Application to Indian Govt. Infrastructure Investment Policy

In Indian macroeconomic policy, there is considerable debate regarding whether Government expenditures on infrastructure and other investments whether they "crowd out" private sector investments. Vinod et al. (2019) use macro data for fiscal years 2011-2016 and find evidence in support of "crowding in" of private investment through public investment. They find that public infrastructure investment is significant in determining private investment and that a low-interest rate encourages private corporate investment. When the government commits resources for infrastructure, it creates incentives for enhancing private investment.

Table 7: Symbols using up to 7 characters for variables in alphabetic order used in reporting causality paths in Table 8.

Code	Description
FornInv	Log Foreign Investment
LongCPI	Long-term 10 year yield Rate adjusted by CPI
LongWPI	Long-term 10 year yield Rate adjusted by WPI
LongYld	10 Year Long Term Yield
Ogap	Output Gap
PbNnInf	Log Public Non-Infrastructure Investment
PubInfr	Log Public Infrastructure Investment
PubInv	Log Public Investment
PvtInv	Log Private Investment
ReLngY	Real 10 years Long Term Yield Rate
ReTbill	Real 91 Treasury Bills Rate
ShrtCPI	real short-term T-bill Rate adjusted by CPI
ShrtRat	91-day Treasury Bills rate
ShrtWPI	real short-term T-bill Rate adjusted by WPI

Causal paths between thirteen variables paired with private investment using symbols (up to seven characters in length) listed in Table 7 are reported in Table 8. The R function 'causeSummary' of the 'generalCorr' package mentioned earlier yields these results.

Table 8: Causal paths between selected variables using symbols in Table 7

	cause	response	strength	corr.	p-value
1	PvtInv	ShrtRat	100	-0.672	0.00032
2	LongYld	PvtInv	31.496	-0.6862	0.00021
3	Ogap	PvtInv	100	0.0118	0.95626
4	PvtInv	PubInv	100	0.9868	0
5	PvtInv	PubInfr	100	0.9595	0
6	PvtInv	PbNnInf	100	0.9733	0
7	PvtInv	ReTbill	31.496	0.5854	0.00265
8	PvtInv	ReLngY	31.496	0.6745	3e-04
9	PvtInv	FornInv	31.496	0.4027	0.05103
10	PvtInv	ShrtCPI	31.496	0.5854	0.00265
11	PvtInv	LongCPI	31.496	0.6745	3e-04
12	PvtInv	ShrtWPI	100	0.7766	1e-05
13	PvtInv	LongWPI	100	0.8393	0

The numbers in the column entitled 'corr.' of Table 8 are Pearson correlation coefficients. These are the usual symmetric correlation coefficients, measuring the nature of 'linear' dependence between the variables named in the first two columns. Since all p-values are near zero except for the output gap along line 3 of Table 8, all relations in the table have statistically significant Pearson correlation coefficients. However, the symmetry of the matrix of Pearson correlation coefficients means that they cannot suggest anything about the underlying causal directions.

When the value in the 'strength' column of Table 8 exceeds 15, the causal direction determination is strong enough to be believed as a preliminary indicator of the true causal direction. The DGP representing private investment is the most important exogenous driver of the Indian economy. It stands to reason that all variables except $LongYld$ and $Ogap$ along lines 2 and 3 of Table 8 show that long-term yield and output gap influence the private investment, $PvtInv$, but all other variables are sensitive to independent variation in $PvtInv$. It is interesting that 'real' short term or long term interest rates adjusted by the consumer price inflation (CPI) or wholesale price inflation (WPI) give same causal path from $PvtInv$ to various interest rates along rows 10 to 13.

Of course, the true causal direction is unknown in the absence of double-blind controlled experiments. This is the best we can assess using certain nonparametric kernel regressions and stochastic dominance of four orders from passively observed data. The estimated path ($PvtInv \to PubInfr$ with the strength of 100) along line 5 of Table 8 suggests that the growth in $PvtInv$ drives public infrastructure spending. Thus, government infrastructure spending on politically glamorous projects will not help grow the $PvInv$. On the other hand, a focus on infrastructure items directly used by the private sector is expected to achieve long-run growth of the Indian economy.

6. Application to Microeconomic Advertising Policy

Advertising research has long relied on linear regressions between sales (Sal) and advertising expenditures on cold remedies. Our scattergrams of data on sales and 32 advertising media outlets, Adx, for seven retailers show nonlinear ups and downs. We allow for two control variables, all other ad media (Oad) and total ad expenditures (TD). Next, we compute the ui to assess the strength of the causal path $Adx \to Sal$ for each advertising media identifying the effective ones. If, on the other hand, the data support the path $Sal \to Adx$, there may be potential savings in those advertising media expenditures.

The results in Table 9 have Adx identify the rows. The columns refer to a particular retailer, whose names are slightly abridged. If the available number of observations is inadequate for reliable kernel regression estimation, the unanimity index ui values in the body of the Table (along some rows and columns) are left blank. Recall that the control variable Oad is a generic notation for all advertising *other* than Adx, and the second control variable TD is short for 'total distribution,' which is separately computed for TV and Print media.

As can be seen in Table 9 there is no advertising media having a consistent causal impact among all the retailers under consideration. It appears that TVCBEFR (TV cable early fringe), TVCBMOR (TV cable morning), TVNTMOR (TV network morning) and TVSPMOR (TV sports during morning hours) are the ones with the majority of ui have the desirable negative sign. For these cases, we can argue that the preponderance of evidence supports the causal path $Adx \to Sal$. An interesting

Table 9: Causal Path Sign Unanimity Index or ui.

	CVS	KMAR	MEIJ	RAID	TARG	WALG	WALM	ALL
TVCBDAY	-31	-31	37	-100	-31	37	100	31
TVCBEFR	-66	-100	16	-100	-100	-31	100	-100
TVCBLFR	31	100	-31	-31	-31	37	100	66
TVCBLNW	-37	31	37	31	100	-31	31	37
TVCBMOR	-31	-37	31	-37	-100	50	31	-31
TVCBONG	-31	31	100	-31	37	100	-31	66
TVCBPRM	100	31	37	31	100	31	37	66
TVCBPRA	66	100	3	31	31	31	66	-66
TVNTDAY	37	-3	37	39		100	100	100
TVNTEFR								
TVNTLFR								
TVNTLNW								
TVNTMOR	-100	-31	-37	-3	-100	-100	92	-66
TVNTPRM	-31	31	37	13	-37	-37	-31	66
TVNTPRA								
TVSPDAY	47	-31	100	-66		31	-37	-3
TVSPEFR	-3	37	100	-66		37	-31	-3
TVSPLFR	37	31	100	3		-100	31	-31
TVSPLNW	31	100	100	-37		-31	37	100
TVSPMOR	-37	-31	-50	31	-3	-37	31	-37
TVSPONG	-3	-92	37	-13		100	31	100
TVSPPRM	100	-31	100	64		100	-66	-31
TVSPPRA	31	-31	37	37		3	-31	37
TVSYTDY	-31	-31	100	37		29	-31	48
PRGPBHG	3	37	3	100	37	66	37	75
PRGPGOO								
PRGPOPR								
PRGPPEO								
PRGPFAM	66	3	37	100	3	3	100	100
PRSIFIT	3	100	100	66	66	3	66	3
PRSIPAR	29		66		37	3	3	66
FSI								

This table presents the Causal Path unanimity index Results based on decision rules from Section 1.6.

result is that only early fringe and morning advertising appear to have a causal impact on units sold. All entries in Table 9, which are positive are where potential savings are possible by spending less on those media.

7. Final Remarks

This paper suggests extending Suppes' probabilistic causality theory by replacing inequalities among probabilities of events by inequalities among densities of data generating processes. We first define kernel regression residuals $e_{j|i,k} = X_j - E\,f(X_j|X_i, X_k)$ and $e_{i|j,k} = X_i - E\,f(X_i|X_j, X_k)$, where we flip X_i and X_j, with details available in Vinod (2019). The causal path $X_i \to X_j$ should satisfy by our first criterion Cr1 the inequality:

$$|(X_{it})e_{(i|j,k)t}| < |(X_{jt})e_{(j|i,k)t}|$$

based on a formal consistency requirement for regression models. Criterion Cr2 compares the goodness of fit via absolute values of residuals, $(|e_{(j|i,k)t}|) < (|e_{(i|j,k)t}|)$. Citerion Cr3 requires generalized partial correlations satisfy: $|r^*_{(j|i,k)}| > |r^*_{(i|j,k)}|$. Since Cr1 to Cr3 are co-equal criteria, we aggregate their empirical values (Cr1 and Cr2 use four numbers for four orders of stochastic dominance) into a unanimity index, ui. The index yields our decision rules to help choose between causal paths $X_i \to X_j$, $X_j \to X_j$, and $X_i \leftrightarrow X_j$. Bootstrap inference tools are also available in the R package 'generalCorr.'

We claim easy computation of *ui* and decision rules for assessing causal path directions and strengths by using very few lines of R code. We illustrate the claim by using examples from macroeconomic and micro-economic data. The ability to incorporate control variables in our analysis is new and particularly valuable for causality estimation and testing using observational data. There are several potential applications in all scientific areas, including exploratory hypothesis formulation, Big Data, and artificial intelligence.

One recent paper, Lister and Garcia (2018), uses our decision rules to conclude that global warming causes anthropod deaths. Another paper, Allen and Hooper (2018), uses them to explore the causes of volatility in stock prices.

References

Allen, D. E., Hooper, V., 2018. Generalized correlation measures of causality and forecasts of the VIX using non-linear models. Sustainability 10 ((8): 2695), 1–15.
URL https://www.mdpi.com/2071-1050/10/8/2695

Anderson, G., 1996. Nonparametric tests of stochastic dominance in income distributions. Econometrica 64(5), 1183–1193.

Bauer, M. D., Mertens, T. M., 3/5/2018. Economic forecasts with the yield curve, fRBSF Economic Letter, Federal Reserve Bank of San Francisco, California.
URL https://www.frbsf.org/economic-research/publications/economic-letter/2018/march/economic-forecasts-with-yield-curve

Contessi, S., Francis, J., 2013. U.S. commercial bank lending through 2008:q4: new evidence from gross credit flows. Economic Inquiry 51(1), 428–444.

Fox, J., 2009. car: Companion to Applied Regression. R package version 1.2-14.
URL http://CRAN.R-project.org/package=car

Gilchrist, S., Zakrajek, E., 2012. Credit spreads and business cycle fluctuations. American Economic Review 102(4), 1692–1720.

Hayfield, T., Racine, J. S., 2008. Nonparametric econometrics: The np package. Journal of Statistical Software 27 (5), 1–32.
URL http://www.jstatsoft.org/v27/i05/

Kendall, M., Stuart, A., 1977. The Advanced Theory of Statistics, 4th Edition. Vol. 1. New York: Macmillan Publishing Co.

Koopmans, T. C., 1950. When is an equation system complete for statistical purposes. Tech. rep., Yale University.
URL http://cowles.econ.yale.edu/P/cm/m10/m10-17.pdf

Levy, H., 1992. Stochastic dominance and expected utility: Survey and analysis. Management Science 38(4), 555–593.

Lister, B. C., Garcia, A., 2018. Climate-driven declines in arthropod abundance restructure a rainforest food web. Proceedings of the National Academy of Sciences Oct. 15, 1–10.
URL http://www.pnas.org/content/early/2018/10/09/1722477115.full.pdf

Stock, J. H., 1987. asymptotic properties of least squares estimators of cointegating vectors. Econometrica 55(5), 1035–1056.

Suppes, P., 1970. A probabilistic theory of causality. Amsterdam: North-Holland.

Vinod, H. D., 2008. Hands-on Intermediate Econometrics Using R: Templates for Extending Dozens of Practical Examples. World Scientific, Hackensack, NJ, ISBN 10-981-281-885-5.
URL http://www.worldscibooks.com/economics/6895.html

Vinod, H. D., 2014. Matrix algebra topics in statistics and economics using R. In: Rao, M. B., Rao, C. R. (Eds.), Handbook of Statistics: Computational Statistics with R. Vol. 34. North Holland, Elsevier Science, New York, Ch. 4, pp. 143–176.

Vinod, H. D., 2015. New bootstrap inference for spurious regression problems. Journal of Applied Statistics.
URL http://www.tandfonline.com/doi/full/10.1080/02664763.2015.1049939

Vinod, H. D., 2016. generalCorr: Generalized Correlations and Initial Causal Path. Fordham University, New York, R package version 1.1.2, 2018, has 3 vignettes.
URL https://CRAN.R-project.org/package=generalCorr

Vinod, H. D., 2017. Generalized correlation and kernel causality with applications in development economics. Communications in Statistics - Simulation and Computation 46 (6), 4513–4534, posted online: 29 Dec 2015.
URL https://doi.org/10.1080/03610918.2015.1122048

Vinod, H. D., 2019. New exogeneity tests and causal paths. In: Vinod, H. D., Rao, C. R. (Eds.), Handbook of Statistics: Econometrics Using R. Vol. 41. North Holland, Elsevier, New York, Ch. 2, pp. 1–38.
URL https://doi.org/10.1016/bs.host.2018.11.011

Vinod, H. D., Karun, H., Chakraborty, L. S., 2019. Encouraging private corporate investment in india. In: Vinod, H. D., Rao, C. R. (Eds.), Handbook of Statistics: Econometrics Using R. Vol. 41. North Holland, Elsevier, New York, Ch. 2, pp. 1–31.
URL https://doi.org/10.1016/bs.host.2019.01.003

Vinod, H. D., López-de-Lacalle, J., 2009. Maximum entropy bootstrap for time series: The meboot R package. Journal of Statistical Software 29 (5), 1–19.
URL http://www.jstatsoft.org/v29/i05/

Wu, D.-M., 1973. Alternative tests of independence between stochastic regressors and disturbances. Econometrica 77(5), 733–750.

Yalta, A. T., 2016. Bootstrap inference of level relationships in the presence of serially correlated errors: A large scale simulation study and an application in energy demand. Computational Economics 48, 339–366.

Zheng, S., Shi, N.-Z., Zhang, Z., September 2012. Generalized measures of correlation for asymmetry, nonlinearity, and beyond. Journal of the American Statistical Association 107 (499), 1239–1252.