New Exogeneity Tests and Causal Paths: Air Pollution and Monetary Policy Illustrations Using generalCorr Package

Hrishikesh D. Vinod *

January 4, 2018

Abstract

It is shown to be impossible to directly test Engle-Hendry-Richard's 'weak exogeneity,' which relies on 'sequential cuts' of a likelihood function. Hausman-Wu's indirect exogeneity test is akin to medieval-style diagnosis of a disease (endogeneity) by showing that a (dubious) instrumental variables (IV) estimator remedy 'works.' Hence my package 'generalCorr' fills a need for a modern exogeneity test along with new tools for determining causal paths using four orders of stochastic dominance and generalized partial correlation coefficients. We illustrate with air-pollution and variables driving 'excess bond premium,' a known predictor of US recessions.

1 Introduction

Consider a possibly non-linear nonparametric regression:

$$Y = f(X_1, X_2, \ldots X_p) + \epsilon_1, \tag{1}$$

*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, JEL codes C30, C51. I thank Prof. J. Francis for suggesting the 'excess bond premium' application. An earlier version is posted as: `https://ssrn.com/abstract=2982128`

where exogeneity of a regressors X_i was first defined by Koopmans (1950) as requiring that X_i should "approximately cause" Y, the endogenous variable. That is, the causal path $X_i \rightarrow Y$ should hold.

A model obtained by flipping Y and X_i is:

$$X_i = f_2(Y, X_1, X_2, \ldots X_{i-1}, X_{i+1}, \ldots X_p) + \epsilon_2, \tag{2}$$

which specifies the opposite causal path $Y \rightarrow X_i$.

A joint density can be written as a product of a conditional and marginal density (using subscripts 'jt', 'cnd' and 'mar'):

$$f_{jt}(Y, X_i, i = 1, \ldots p) = f_{cnd}(Y, X_j, (j \neq i))|X_i)\ f_{mar}(X_i). \tag{3}$$

Webster's seventh collegiate dictionary defines kernel as "a central or essential part." Accordingly, let us define the essential part of exogeneity as:

Definition (Kernel exogeneity): A variable X_i from the joint density f_{jt} in eq. (3) is exogenous if the data generating process (DGP) for X_i is self-driven, not responding to (dependent on) stimulus from other variables.

Properties of Kernel exogeneity: Kernel exogeneity properties are:

(KE1) It allows one to assess exogeneity by comparing performance of flipped models, since the joint density of eq. (3) can alway be conditioned on the marginal density $f_{mar}(X_i)$ for eq. (1) or $f_{mar}(Y)$ for the flipped eq. (2).

(KE2) Since no parameters are explicitly involved, it is applicable to almost any parametric or nonparametric model of the conditional mean function based on the conditional density f_{cnd}.

Engle et al. (1983) [hereafter, 'EHR'], were the first to use flipped models in the present context. In addition to Koopmans' causality, EHR mention (p. 285) Zellner's predictability-based causality, whereby a causal direction $X_i \rightarrow Y$ requires that X_i in eq. (1) should predict Y better, with a larger R^2 than the flipped eq. (2). EHR focus on flipped ordinary least squares (OLS) estimators for equations (1) and (2), assuming $p = 1$ and f_1, f_2 are linear. Now EHR argue that Koopman-Zellner causality is ambiguous, because the R^2 values of such restricted flipped regressions are identical.

The computing facilities in 1983 when EHR was published (as a lead article in *Econometrica*) did not include easy-to-use nonparametric (kernel) regressions or distribution-free bootstrap tests used here. Hence they could

not have imagined an asymmetric matrix of generalized correlation coefficients proposed in Vinod (2014), whereby the predictive power of flipped models need not be identical.

In the absence of nonparametric tools, EHR rewrite eq. (3) after conditioning on explicit parameters $\lambda = (\lambda_1, \lambda_2)$ as:

$$f_{jt}(Y, X_i, i = 1, \ldots p|\lambda) = f_{cnd}(Y, X_j, (j \neq i)|X_i, \lambda_1)\, f_{mar}(X_i|\lambda_2), \qquad (4)$$

related to a factoring of the likelihood function, needed for maximum likelihood (ML) estimation. Now EHR's widely accepted "weak exogeneity" is complicated because it requires eq. (4) to implement a 'sequential cut' extending Barndorff-Nielson notion of a cut for exponential familiy of distributions.

Definition (EHR weak exogeneity): X_i is weakly exogenous for parameters of interest, ψ, if there exists a re-parameterization $\lambda = (\lambda_1, \lambda_2)$ where

(i) ψ is a function of λ_1, and

(ii) $[f_{cnd}(Y, X_j, (j \neq i))|X_i, \lambda_1)\, f_{mar}(X_i|\lambda_2)]$ operates a 'sequential cut' defined in eq. (4).

Properties of EHR weak exogeneity:

[WE1] *Parameter Distinctions*: A distinction between parameters of interest, ψ, and other (nuisance) parameters λ_2 is a crucial part of the definition.

[WE2] *Granger Causality Irrelevant*: EHR state (p. 290) that "Granger noncausality is neither necessary nor sufficient for weak exogeneity."

[WE3] *Invariance*: EHR assume that ψ are invariant to policy changes to avoid the famous Lucas critique.

[WE4] *Inability to test*: EHR flip a two-equation simultaneous equations model (their equations numbered 27 and 28 versus 30 and 31) to argue on page 288 that "the choice of parameters of interest is the sole determinant of weak exogeneity, which is, therefore not directly testable."

The aim of this paper is to simplify 'weak exogeneity' by reverting back to 'Kernel exogeneity' where no parameters are present, by using nonparametric kernel regressions. We overcome the property [WE4] by developing a new computer intensive bootstrap test for exogeneity.

1.1 Indirect Exogeneity Testing

Lacking a direct exogeneity test, practitioners often use tests for validity of the underlying model as indirect exogeneity tests. For example, the model in eq. (1) is valid if its errors ϵ_1 are statistically independent of regressors, satisfying the orthogonality null:

$$H_0 : E(X_i \epsilon_1) = 0. \tag{5}$$

Wu (1973) had originally provided another indirect exogeneity test, which was later popularized as the Hausman-Wu test (HWT). It defines a vector of contrasts, $d = b_{OLS} - b_{IV}$, between OLS, an efficient but potentially inconsistent (due to endogeneity) estimator and inefficient but consistent (by assumption) IV estimator. The covariance matrix of d can be shown to be $V_d = V(b_{IV}) - V(b_{OLS})$, and a quadratic form, $d'(V_d)^{-1}d$, is asymptotically a $\chi^2(p)$, with p degrees of freedom. The HWT amounts to medieval diagnosing of a disease (endogeneity) by showing that a cure (b_{IV}) works.

Actually, the IV remedy has been found to be seriously flawed as shown by Bound et al. (1995) with a provocative title "the cure can be worse than the disease". This paper illustrates the use of an R package "generalCorr" to develop a new test which does not use any IV estimator. We indicate the very few lines of code needed to assess the preponderance of evidence in support of a causal path using macroeconomic examples which can serve as a template in many areas of research.

Urgency of Replacing HWT (the Hausman-Wu test)

Many authors including Bound et al. (1993) and Kiviet and Niemczyk (2007), have warned that in finite samples IV estimators "have systematic estimation errors too, and may even have no finite moments." Moreover they can be very inefficient (even in large samples) and unnecessarily change the original specification. This paper is motivated by the following disadvantages of HWT:

1. One must replace X_i with *ad hoc*, potentially weak and/or irrelevant instrumental variable $\tilde{Z}_i$ before testing for exogeneity of X_i.

2. The test needs to be repeated for each potential $\tilde{Z}_i$ replacing each X_i.

3. Davidson and MacKinnon (1993) (p. 241) show that degrees of freedom p for the $\chi^2(p)$ test is too large when a subset of X_i are exogenous.

4. The Chi-square sampling distribution is subject to unverified assumptions of linearity and normality, especially unrealistic in finite samples.

1.2 Kernel Estimation of Models 1 and 2

Assuming $p = 1$ in eq. (1) for ease of exposition, without loss of generality (wlog), define Model 1 as a nonlinear non-parametric kernel regression:

$$Y_t = G_1(X_t) + \epsilon_{1t}, \quad t = 1, \ldots, T, \tag{6}$$

where errors are no longer Normal and independent. Our nonparametric estimate $g_1(x)$ of the population conditional mean function $G_1(x)$ is:

$$g_1(x) = \frac{\sum_{t=1}^{T} Y_t K(\frac{X_t - x}{h})}{\sum_{t=1}^{T} K(\frac{X_t - x}{h})}, \tag{7}$$

where $K(.)$ is the well known Gaussian kernel function and h is the bandwidth parameter often chosen by leave-one-out cross validation, Li and Racine (2007) and (Vinod, 2008, Sec. 8.4). It is well known that kernel regression fits are superior to OLS.

Proposition (Kernel Regression is CAN) Assume that
(i) $\{X_t, Y_t\}$ are iid and $g_1(x)$, joint density and error variance functions are twice differentiable.
(ii) K is a bounded second order kernel.
(iii) As $T \to \infty$, $Th^3 \to \infty$ and $Th^7 \to 0$.
Then kernel regression estimate g_1 is consistent and asymptotically Normal (CAN).

Proof: See Theorem 2.7 of Li and Racine (2007) for further details and extensions to multivariate and local polynomial generalizations, including a proof of consistency and asymptotic Normality (CAN).

The flipped kernel regression Model 2 is obtained by interchanging X and Y in eq. (6):

$$X_t = G_2(Y_t) + \epsilon_{2t}, \quad t = 1, \ldots, T. \tag{8}$$

1.3 Intuition behind new exogeneity test

The intuition behind our criteria can be explained with an example used for teaching that correlation is not causation in elementary statistics. The flipped variables are violent crime rate (crim) and police officer deployment (off) rate in 28 European countries in comparable units (Section 5 has data details.). Since most policemen are not murderers, the DGP for 'crim' variable is intuitively likely to be self-driven or exogenous, while the DGP of police officer deployment would respond more to changing crime rates and less likely to be self-driven. Hence we expect 'crim' to be Kernel exogenous according to our definition using eq. (3) above.

If 'crim' is really 'Kernel exogenous' in data, the regression specification: $off = f_1(crim) + \epsilon_1$, should perform superior to: $crim = f_2(off) + \epsilon_2$. Our criteria Cr1 to Cr3 quantify three ways of assessing that superiority.

Since flipped OLS regressions (with one regressor) always yield identical R^2, we must use nonlinear (kernel) regressions for a meaningful comparison which admits distinct R^2 values. Of course, we need appropriate thresholds (τ) and tests admitting bidirectional or indeterminate exogeneity described in the sequel.

EHR's weak exogeneity for this example is clearly unsatisfactory, since it is subject to 'inability to test' property [WE4]. If we let f_1, f_2 be nonparametric, we shall see in Section 5 that without having to specify 'parameters of interest' (ψ) in either of the flipped models, the data clearly support Kernel exogeneity of 'crim' from the causal path: crim→off.

The intuition behind specific criteria is discussed next. Our Cr1 uses the eq. (5) requirement that the covariance between the right hand side variable and regression errors should be 'smaller' for the superior model. Our Cr2 requires absolute values of residuals of the superior specification to be 'smaller.' Our Cr3 requires the superior model to have a larger R^2. Both Cr2 and Cr3 are consistent with Zellner's predictability-based causality mentioned earlier.

1.4 Kernel Regressions and Generalized Correlations

The generalized measures of correlation in Zheng et al. (2012) are:

$$\begin{aligned} GMC(Y|X) &= [1 - \tfrac{E(Y-E(Y|X))^2}{var(Y)}], \\ GMC(X|Y) &= [1 - \tfrac{E(X-E(X|Y))^2}{var(X)}], \end{aligned} \tag{9}$$

which are computed simply as the R^2 values of flipped Models 1 and 2. Since they generally do differ from each other, the ambiguity in Koopmans' method criticized by EHR disappears.

As measures of correlation the non-negative GMC's in the range [0,1] provide no information regarding the up or down overall direction of the relation between Y and X, revealed by the sign of r_{xy}, the Pearson coefficient. Since a true generalization of of r_{xy} should not provide less information, Vinod (2014) and Vinod (2015a) propose the following modification. A general asymmetric correlation coefficient from the $GMC(Y|X)$ is:

$$r^*_{y|x} = \text{sign}(r_{xy})\sqrt{GMC(Y|X)}, \tag{10}$$

where $-1 \leq r^*_{y|x} \leq 1$. A matrix of generalized correlation coefficients denoted by R^* is asymmetric: $r^*_{x|y} \neq r^*_{y|x}$, as desired. A function in the generalCorr package, `gmcmtx0`, provides the R^* matrix from a matrix of data.

Our new test of exogeneity uses the "preponderance of evidence" standard quantified by a comprehensive index, which is a weighted sum of causal direction signs using three criteria Cr1 to Cr3. Our Cr3 which compares R^2 of flipped models is from Vinod (2014). Since elementary statistics teaches us not to rely on R^2 alone, an additional criterion (Cr1) considers evidence from probability distributions of the absolute magnitudes of local orthogonality test statistic by using stochastic dominance (SD). Similarly our second criterion Cr2 compares absolute residuals.

An outline of the remaining paper is as follows. Section 2 further explains kernel causality including our assumptions and a proposition on existence of unanimity index (*ui*) helping to determine from our 'sum' criterion incorporating Cr1 to Cr3 based on some digressions needed for comparing the flipped models. The section ends with our decision rules. Section 3 reports a simulation of our decision rules. Section 4 considers statistical inference using the bootstrap. Section 5 considers examples including the famous Klein I model, air pollution data, and a model that considers what macroeconomic variables drive (cause) 'excess bond premium' known to be a good predictor of US economic recessions. Our examples include bootstrap inference for the new test. Section 6 contains a summary and final remarks.

2 Kernel Causality Explained

Assessing philosophically true causality from non-experimental data is non-trivial, Pearl (2009). Instead, we consider a modified causality called 'kernel causality,' based on an empirical comparison of two flipped models. If the DGP of X_i is more likely to be self-driven (independent) than Y, its the marginal density is appropriate for factoring the joint density in eq. (3). Then, we choose the causal path: $X_i \rightarrow Y$ over the opposite path: $Y \rightarrow X_i$. The name Kernel causality acknowledges our reliance on both kernel regressions and 'Kernel exogeneity.' Recalling property [WE2] of EHR weak exogeneity, Kernel causality has almost nothing to do with Granger causality.

Kernel Causality Assumptions:

Our assumptions are:

(A1) The flipped kernel regressions of Models 1 and 2 from equations (6) and (8) are estimated by leave-one-out local linear kernel estimators: g_1, g_2. The bandwidth selections use local linear cross validation.

(A2) Assume that a data generating process (DGP) consists of three sets of variables, (X, Y, Z), where Z represents non-economic, confounding or control variable(s), if any. Model 1 DGP has X_i independently generated (or 'Kernel exogenous' with autonomous innovations) while the flipped Model 2 has Y independently generated.

(A3) There exists a conditional expectation function $E(Y|X, Z)$ for Model 1 and analogous function $E(X|Y, Z)$ for Model 2 obtained by flipping X_i and Y for the two well-identified models.

Note that we are assuming away functional relations such as Boyle's law (pressure *volume = a constant) which will fail A2, since all component variables (pressure and volume) can be independently generated in a typical laboratory.

Proposition (UI exist): There exists two unanimity index numbers, (UI_1, UI_2), quantifying whether the DGP of X_i is more likely to be self-driven (independent) than Y.

Let τ denote a threshold chosen by the researcher for the application at hand. If $|(UI_1 - UI_2)| < \tau$, the two index numbers are "too close" to

each other. Then the exogeneity of X_i and Y is indeterminate, implying bi-directional kernel causality. It can also mean that both X_i and Y are jointly dependent, perhaps each needing a separate structural equation.

If a majority of Cr1 to Cr3 support the causal path ($X_i \to Y$), assumptions A1 to A3 guarantee that X_i is Kernel exogenous (independently generated) and *kernel* causes Y. We begin with two digressions: (i) stochastic dominance, needed for Cr1 and Cr2, and (ii) partial correlations needed for Cr3.

Digression 1: Stochastic Dominance Notation

Let us describe stochastic dominance (SD) concepts surveyed in Levy (1992) without attempting to summarize the vast and growing published and unpublished literature motivated by financial economists' portfolio choice problem. We say that one density $f(x)$ dominates another density $f(y)$ in the first order (SD1) if their respective empirical cumulative distribution functions (ecdf) satisfy: $F(x) \leq F(y)$. It is well known that SD1 provides a comprehensive picture of the ranking between two probability distributions with a focus on locally defined first moment (mean).

The underlying computation requires bringing the two densities on a common '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) shows how a simple pre-multiplication by a large patterned matrix implements computation of (sd1). Let us use a simple average Av(sd1) whose sign $(+1, 0, -1)$ helps summarize the first order stochastic dominance into only one number.

Second order dominance (SD2) of $f(x)$ over $f(y)$ requires further integrals of ecdf's to satisfy: $\int F(x) \leq \int F(y)$. One computes the numerical integral by using the trapezoidal rule described in terms of a large patterned matrix whose details are given in (Vinod, 2008, ch.4) and in Anderson (1996). The $2T$ estimates of SD2 denoted by (sd2) are locally defined variances. Their simple average is denoted as Av(sd2), whose sign $(+1, 0, -1)$ summarizes the information regarding second order dominance.

Similarly, SD of order 3 is estimated by a vector (sd3) of $2T$ locally defined skewness values defined from $\int\int F(x) \leq \int\int F(y)$. The sd3 is further summarized by the sign of Av(sd3). Analogous SD of order 4 for kurtosis requires $\int\int\int F(x) \leq \int\int\int F(y)$ and measures investor 'prudence' according to Vinod (2004). Average of pointwise kurtosis estimates of SD4 are Av(sd4),

whose sign $(+1, 0, -1)$ summarizes the SD4 dominance information.

Remark 1: By analogy with two streams of investment returns, stochastic dominance allows us to study realistic but fuzzy inequalities (may not hold for subsets of points) of the type $(x_t < y_t)$ for $t = 1, \ldots, T$. Stochastic dominance of four orders associated with the four moments yield $2T$ estimates of sd1 to sd4. The signs of their averages, Av(sd1) to Av(sd4), indicate whether the inequality holds true in an overall sense.

Digression 2: Partial Correlations

Note that the partial correlation between (X_1, X_2) after removing the effect of (X_3) is:

$$r_{12;3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)}\sqrt{(1 - r_{23}^2)}}. \tag{11}$$

Kendall and Stuart (1977) show that an alternative definition of $r_{12,3}$ is a simple correlation between residuals of the regression: $X_1 = f(X_2, X_3) + error$ and similar residuals of the regression: $X_2 = f(X_1, X_3) + error$. 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. Let us first define $u_{i,k}$ as the residual of kernel regression of X_i on all control variable(s) X_k. Similarly define $u_{j,k}$ as the residual of kernel regression of X_j on all control variable(s) X_k. Next, we define a symmetric version of generalized partial correlation coefficient in the presence of control variable(s) as:

$$u^*_{ij;k} = \frac{cov(u_{i,k}u_{j,k})}{\sigma(u_{i,k})\sigma(u_{j,k})}, \tag{12}$$

a symmetric correlation coefficient between two relevant residuals.

Now we recall eq. (10) based on GMC's to obtain asymmetric generalized partial correlation coefficients. Denote the sign of the correlation in eq. (12) as $\text{sign}(u^*_{ij;k})$. Finally we are ready to define an asymmetric matrix of generalized partial correlation coefficients using the R^2 of kernel regression: $u_{i,k} = f((u_{j,k}) + err$ as $GMC(u_{i,k}|u_{j,k})$. Note that the generalized partial correlations will be asymmetric since GMC's are asymmetric.

Thus, we can define:

$$r^*(X_i, X_j; X_k) = \text{sign}(\text{u}^*_{\text{ij;k}})\sqrt{[\text{GMC}(\text{u}_{\text{i,k}}|\text{u}_{\text{j,k}})]}. \tag{13}$$

Often, we simplify the notation and write the generalized partial correlations as $r^*_{i,j;k}$. Section 5.5 provides an illustrative example implementing the generalized partial correlation coefficients from data.

2.1 Kernel Causality from Flipped Model Choice

We determine whether X_i drives Y, or vice versa, by considering the evidence from the majority of three criteria. They are:

(Cr1) Our first criterion is based on orthogonality null of eq. (5). If the causal path $X_i \rightarrow Y$ is chosen when the following inequality cumulatively holds for the set of $t = 1, 2, \ldots, T$, using kernel regression residuals of eq. (6) times a right hand side regressor.

$$|(X_{it})\hat{\epsilon_{1t}}| < |(Y_t)\hat{\epsilon_{2t}}|. \tag{14}$$

Package 'generalCorr' older versions ($\leq$ 1.0.9) have absolute values of local gradients of kernel regressions as Cr1.

Some comments on an older version of Cr1 are included here. When one spells out the left hand side (LHS) expression in eq. (14) upon replacing $\hat{\epsilon_{1t}}$ by local linear estimates, we have: LHS= $E(X_{it}Y_t) - E(a_{1t}X_{it} + b_{1t}X^2_{it})$, where the local gradients b_{1t} are multiplied by squares of regressor variables X^2_{it}. The magnitudes of LHS quantities will be most influenced by $b_{1t}X^2_{it}$. Since the flipped regressors are standardized, $E(X^2_{it}) = E(Y^2_t) = 1$, the flip side with a smaller absolute gradient b_{1t} is likely to satisfy eq. (14).

It is therefore not surprising that the older choice of Cr1 based on absolute gradients also works well in simulations. Hence newer versions of 'generalCorr' will provide an option to use the older Cr1. For example, the commands `causeSummary` and `silentPairs` have versions using the older definition of Cr1 (using gradients) as `causeSummary0`, and `silentPairs0`, respectively

(Cr2) The path $X \rightarrow Y$ should have "smaller" absolute residuals (superior local fit) than those of the flipped model, that is, for $t = 1, 2, \ldots, T$:

$$|Y_t - g_1(X_t, Z_t)| = (|\hat{\epsilon_{1t}}|) < |X_t - g_2(Y_t, Z_t)| = (|\hat{\epsilon_{2t}}|). \tag{15}$$

(Cr3) The fit (and forecasts) implied by the path $X \to Y$ should have a larger $R^2 = GMC(Y|X, Z)$ than those of reversed path:

$$|r^*_{(y|x;\, z)}| > |r^*_{(x|y;\, z)}|, \tag{16}$$

where generalized partial correlation coefficients defined in eq. (13) remove the effect of control variable(s), if any.

The inequalities of equations (14) and (15) are fuzzy, requiring stochastic dominance tools summarized in Remark 1 above. Let us begin with some definitions.

Definition 1: According to Legal Information Institute (2017) the preponderance of evidence means a burden to show that greater than 50% of evidence points to something.

Definition 2: Assuming A1 to A3, we say that X is the kernel cause of Y (causal path: $X \to Y$), if at least two of Cr1 to Cr3 criteria satisfying the preponderance of evidence standard support the path.

Definition 3: Bidirectional causality ($X \leftrightarrow Y$) or causality marred by the presence of confounding variable(s) occurs if the evidence does not support either ($X \to Y$) or ($Y \to X$).

Remark 2: If relations are strictly linear and/or the errors are precisely normally distributed, flipped R^2 are almost identical creating an ambiguity of Koopmans' criterion criticized by EHR. Since we are using kernel regressions, not OLS, this problem can disappear. In any case, we are not relying on R^2-based Cr3 alone, but incorporate stochastic dominance for criteria (Cr1 and Cr2) unrelated to normality or linearity.

2.2 Weighted sum index from Cr1, Cr2 and Cr3

Applying Remark 1 to the inequality (14) for Cr1, we compute Av(sdℓ) for $\ell = 1, \ldots 4$, magnitudes from absolute gradients of two flipped models. Define a tolerance constant $\tau = 0.476$, say. If $|Av(sd\ell)| < \tau$, we say that the sign is ambiguous, denoted as zero for the ℓ-th SD. When $|Av(sd\ell)| > \tau$, only the signs of Av(sdℓ) not their magnitudes matter. These signs (sg) from the set $(+1, 0, -1)$, are denoted as $sg_{1\ell}$, where the first subscript 1 refers to Cr1. In practice, the signs sg_{11} to sg_{14} are rarely distinct.

Since it is cumbersome to track four signs, we propose a weighted sum, using the signs, $(+1, 0, -1)$, not magnitudes of Av(sd1) to Av(sd4). Statistical theory suggests that weights on magnitudes should be inversely proportional to the increasing sampling variances of the first four central moments. We choose the following weakly declining weights: (1.2/4, 1.1/4, 1.05/4, 1/4), with an option to change them in the R functions `silentPairs` and `causeSummary` of the 'generalCorr package.

Denote a summary sign index based on Cr1 as sC_1. It is computed as:

$$sC_1 = [1.2 * sg_{11} + 1.1 * sg_{12} + 1.05 * sg_{13} + sg_{14}]/4. \tag{17}$$

When all four (Av(sd1) to Av(sd4)) suggest the same sign, ie, all are (± 1), the largest magnitude of our weighted index of sign by Cr1 is $sC_1 = \pm 1.0875$.

Analogous signs $(+1, 0, -1)$ of Av(sd1) to Av(sd4) representing absolute residuals help define their weighted sum for Cr2 is

$$sC_2 = [1.2 * sg_{21} + 1.1 * sg_{22} + 1.05 * sg_{23} + sg_{24}]/4. \tag{18}$$

As before, if all four dominance measures suggest the same sign, the largest magnitude of sC_2 is 1.0875. Hence, the sign index based on Cr2 lies in the closed interval: $sC_2 \in [-1.0875, 1.0875]$.

The computation of a Cr3 from the inequality test of (16) states that $X \to Y$ if the sign defined as: $sg_3 = (+1, 0, -1)$ of the absolute difference between flipped partial correlations equals (-1). We denote the sign index based on Cr3 as:

$$sC_3 = sign(|r^*_{(x|y;\,z)}| - |r^*_{(y|x;\,z)}|) \tag{19}$$

where the largest score, $\max(sg_3) = 1$. When $sg_3 < 0$, the causal path by Cr3 is $X \to Y$. Note that index always lies in the closed interval: $sC_3 \in [-1, 1]$.

So far, we have three sign indexes (sC_1, sC_2, sC_3) for the three criteria, summarizing the evidence supporting the causal path: $X \to Y$. Since our definition of kernel causality requires us to consider all three criteria, we compute their 'sum' defined as:

$$sum = sC_1 + sC_2 + sC_3, \tag{20}$$

from the observed sample data. Let us denote the corresponding true unknown population value with upper case letters as 'SUM'. When $(SUM < 0)$ holds, the causal path is $X \to Y$. Based on the preponderance of evidence,

the sign of *sum* suggests the direction of the path, while its magnitude approximates the strength of sample evidence in support of that causal path.

Combining the three largest possible scores verify that: $\max(sum) = 3.175$, and $sum \in [-3.175, 3.175]$, a closed interval. A summary unanimity index is defined as $ui = 100(sum/3.175)$, always in the range [–100,100]. Since the 'sum' and ui measure the extent of agreement among the three criteria, its magnitude is a reasonable indicator of the strength (or unanimity) of evidence for a particular causal path. When the population parameter is smaller than a threshold value, ($UI < \tau$, where $\tau = 15$, say,) we can conclude that the causal path is $X \to Y$.

Single number summarizing Cr1 to Cr3 in 'generalCorr'

The R command `causeSummary(mtx,ctrl=Z, nam=colnames(mtx))` requires a data matrix with p columns called 'mtx' with the first column for the dependent variable and remaining column(s) for regressors. The order of columns is very important. For example, `mtx=cbind(x1,x2,x3)`, where the matrix 'mtx' has three columns, denoted as $p = 3$. Our flipped models fix the first column x1 and pair it with either x2 or x3 for flipping. We do not pair x2 with x3. Thus we always have $p - 1$ possible flipped pairs. The code indicates an error if $p < 2$ or if it is not a matrix. Sometimes one needs to use `as.matrix(mtx)`. Note that control variables are a separate argument (not within mtx), as in: `causeSummary(mtx, ctrl=0)`, where the default value zero means absence of control variable(s).

The output of 'causeSummary' is self-explanatory based on 'preponderance of evidence' from a weighted combination of Cr1 to Cr3. Since we have exactly $(p - 1)$ possible causal path pairs, the summary reports each printed to the screen. For each pair it reports the name of the causal variable, then the name of the response variable, the strength index in terms of unanimity of the sign of the reported causal path. It also reports Pearson correlation coefficient and its p-value for testing the null hypothesis: $\rho = 0$. If the unanimity strength index (ui) is close to zero, in the range $[-15, 15]$, one should conclude that $X \leftrightarrow Y$.

The code `su=causeSummary(mtx);xtable(su)` may be used to create a Latex table of results from the output of the function. It is a matrix of $(p-1)$ rows and 5 columns providing summary of pair-wise causal path results. The first column entitled 'cause' names the causal variable, while the second column entitled 'response' names the response. The third column entitled

'strength' has absolute value of summary strength index, printed above but now in the positive range [0,100], summarizing preponderance of evidence from Cr1 to Cr3 from four orders of stochastic dominance and generalized partial correlations. The fourth column entitled 'corr' has Pearson correlation coefficient while the fifth column entitled 'p-value' is for testing the null of zero Pearson correlation coefficient.

Our notion of causality is not the true philosophical causality, but an approximation where a 'kernel cause' is simply the 'Kernel exogenous' variable defined above using eq. (3) from a flipped pairs of variables.

2.3 Decision Rules

The 'preponderance of evidence' supports one of the three causal paths listed below when the sample unanimity index 'ui' is inside one of the three intervals (two half-open and one closed, using $\tau = 15$). If one uses the sample 'sum' index, the 15% threshold τ for ui translates to the number 0.475.

R1: $X_{1+j} \to X_1$ if $(ui \in (-100, -15])$ or $sum \in (-3.175, -0.476]$.

R2: $X_{1+j} \leftarrow X_1$ if $(ui \in (15, 100])$ *or* $sum \in (0.476, 3.175]$.

R3: $X_{1+j} \leftrightarrow X_1$ if $(ui \in [-15, 15])$ *or* $sum \in (-0.476, 0.476])$.

3 Simulation for checking decision rules

Following our definition of Kernel exogeneity we generate the X_1 variable independently and then define X_2 to depend on X_1 after adding a noise term, $\epsilon \sim N(0, 1)$, a the standard normal deviate. Here X_1 is Kernel exogenous by construction, and hence the causal path is known to be $X_1 \to X_2$, by construction. Our sample sizes are $T = 50, 100, 300$.

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:

$$(suPr) = \frac{(count\ of\ correct\ signs)}{N - m}. \tag{21}$$

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

1. Time regressor:

   $X_1 = \{1, 2, 3, \ldots, T\}$
   $X_2 = 3 + 4X_1 + \epsilon$

2. Unit root Quadratic:

   X_1 has T random walk series from 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 coumns 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 suggest desirable asymptotic convergence-type feature. Thus, our procedure using flipped models to identify independently generated (causal) variables is supported by the simulation.

This simulation uses our newer definition of Cr1 used in the latest versions of 'generalCorr' package (ver. $\geq$ 1.1.0), which is intuitively more directly aligned with orthogonality null of eq. (5). Unfortunately, newer Cr1 is not unequivocally superior to the older Cr1 based on local linear kernel regression gradients based on an earlier limited simulation, assuming $\tau = 0, T = 100$. For example, when $\tau = 0$ the success probabilities using `causeSummary0` command using the older version of Cr1 are: (1.000, 0.905, 0.882, 0.970). These

are quite comparable to (1.000, 0.787, 0.892, 0.803) along Row numbered 14 of Table 1.

Hence, both R functions `causeSummary` and its older version `causeSummary0` may be attempted. Any sign disagreement is clearly suggestive of uncertainty in the estimated causal paths. Then one can perhaps postpone if not avoid the computer intensive bootstraps described in the next section to save computational or time resources.

4 A Bootstrap Exogeneity Test

Statistical inference regarding causal paths and exogeneity uses the 'sum' statistic defined in equation (20) for estimating the parameter 'SUM' mentioned before. Equivalently we can bootstrap the sample unanimity index, $ui = 100(sum/3.175)$, with the parameter UI mentioned above.

What is the sampling distribution of the 'sum' test statistic? We use the maximum entropy bootstrap (meboot) R package described in Vinod and López-de-Lacalle (2009) because it retains the dependence structure (e.g. rankings of countries) in the data recently supported by simulations in Yalta (2016), Vinod (2015b) and elsewhere. Here we use the meboot package to compute a large number ($J = 999$) of resamples of (X, Y, Z) data. These are an approximation to what the data might look like due to random variation in the population, or the ensemble. The observed (X, Y, Z) data represent only one realization from the ensemble. One can, of course, use other bootstrap algorithms.

Recall that sC_1 to sC_3 is a weighted sum of only three numbers (–1, 0, +1), implying an ordered categorical random variable. Since their *sum* defined in equation (20) can have only a finite set of values, the sampling distribution of the *sum* statistic has nonzero mass only at those set of points in the closed interval:

$$sum \in [-3.175, 3.175]. \tag{22}$$

Since computing the *sum* automatically cancels positive numbers with negative numbers, its magnitude measures a weighted vote count, as it were, in favor of the most enduring (empirically supported) sign of the *sum*. If, for example, $sum = -3.175$, reaching the lower limit of the range, Cr1 to Cr3 are unanimity supporting the causal path $X \to Y$.

Let sum_j denote the j-th bootstrap sum where $j = 1, \ldots J$, for each flipped pair. A direct study of the properties of the sampling distribution looks at the summary statistics of the J replicates sum_j, such as: (mean, median, quartiles), etc. The signs of these summary statistics reveal the most preponderant sign in the bootstrap approximation to their population, illustrated later in Table 2 below. The sign of the mode (most frequently observed sum_j) is also of interest.

A further summary of the sampling distribution can be obtained by computing bootstrap proportion of positive or negative values:

$$P^*(+1) = \#(sum_j > 0.5)/J, \quad \text{and} \quad P^*(-1) = \#(sum_j < -0.5)/J, \quad (23)$$

where $\#(sum_j > 0)$ denotes the number of occurrences of positive signs out of J computations while ignoring the magnitudes. Thus $P^*(\pm 1)$ is a bootstrap approximation to the probability of a positive or negative sign in determining the causal path direction.

In the context of simultaneous equation models, consider the null hypothesis that X_j of eq. (1) is exogenous. Then the path implied by eq. (1) should have greater support than (2). We expect the preponderance of evidence supporting a negative 'SUM'.

Define the null and alternative hypotheses for exogeneity as:

$$H_0 : SUM \leq 0, \quad \text{against} \quad H_1 : SUM > 0, \quad (24)$$

Negative values of SUM are desirable, if we want to assure ourselves that the regressor is exogenous. A simple rule for statistical inference is to *reject* the hypothesized exogeneity whenever the bootstrap proportion $P^*(+1)$ sufficiently exceeds $P^*(-1)$ for the problem at hand. The Definition 1 suggests preponderance of evidence or $> 50\%$ standard. In our experience and illustrations below a much larger percentage is often attainable.

5 Application Examples

Let us begin with an example mentioned earlier where the cause is intuitively known to illustrate our statistical inference using the sum and ui statistics. Vinod (2015a) describes a cross section data example where Y denotes the number of police officers per 1000 population, and X denotes the number of crimes per 1000 population in T =29 European countries in 2008.

```
require(generalCorr);require(Hmisc)
attach(EuroCrime)#bring package data into memory
causeSummary(cbind(crim,off))
pcause(crim,off,n999=29)
```

The output of above code given below shows that crime causes officer deployment with strength 100, while bootstrap resampling success proportion is about 0.59.

```
causeSummary(cbind(crim,off))
[1] crim      causes    off       strength= 100
[1] corr=  0.99   p-val= 0
     cause  response strength corr.  p-value
[1,] "crim" "off"    "100"    "0.99" "0"
pcause(crim,off,n999=29) #illustrative bootstrap
[1] 0.5862069
```

A single bootstrap computation for these data when $J = 999$ on a home PC requires about 20 minutes of CPU time. An approximate sampling distribution of 'sum' statistic for these data is depicted in Figure 1. We are using a histogram because the sampling distribution is categorical with nonzero frequency counts only at a finite set of points. The mode is clearly seen at –3.175 in the histogram. suggesting that the path (crime→officer deployment) is not due to random noise, but likely to be present in the population. The descriptive statistics for the set of J values of (sum_j) are: (first quartile=–3.175, median =–1.175, third quartile=1), and proportion of negatives, $P^*(-1) = 0.641$.

5.1 Klein I simultaneous equations model

This section reports the results for our three criteria regarding exogeneity of each of the regressors of the three equations of the famous Klein I model. Let us use the following four-character abbreviations using the upper case trailing L for lagged version of a variable: cons=consumption, coPr=corporate profits, coPL= corporate profits with a lag, wage=wages, inve=investment, capL=capital with a lag, prWg=private sector wages, gnpL=GNP with a lag, and finally, tren=time trend.

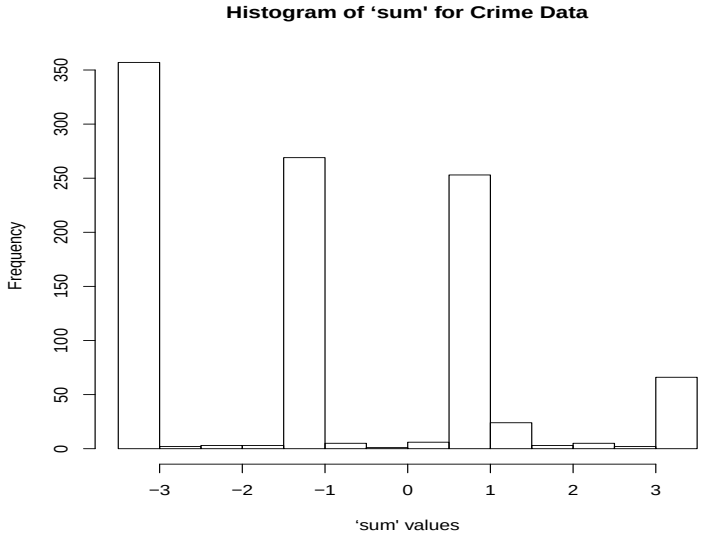

Figure 1: European Crime Data Approximate Sampling Distribution of the *sum* statistic

Klein's specification of the expected consumption equation (stated in terms of fitted coefficients) is:

$$E(\text{cons}) = a_{10} + a_{11}\,\text{coPr} + a_{12}\,\text{coPL} + a_{13}\,\text{wage}. \tag{25}$$

The second (investment) equation of the Klein I model is given by:

$$E(\text{inve}) = a_{20} + a_{21}\,\text{coPr} + a_{22}\,\text{coPL} + a_{23}\,\text{capL}. \tag{26}$$

The third (wage) equation of the Klein I model is given by:

$$E(\text{prWg}) = a_{30} + a_{31}\,\text{gnp} + a_{32}\,\text{gnpL} + a_{33}\,\text{tren}. \tag{27}$$

We report summary statistics for all three criteria combined into the $sum_j, j = 1, \ldots J$ defined in eq. (20) leading to a $J = 999 \times 1$ vector of summary signs, for brevity.

Three columns of Table 2 are for the three equations of the Klein I model. The rows report descriptive statistics: the minimum, maximum, quartiles Q1 and Q3, mean and median based on $J = 999$ bootstrap realizations.

The bottom row of the Table reports the bootstrap probability of a positive result, $P^*(+1)$ defined in eq. (23), which are all close to 0.5. The fact that all equations have the same minimum, maximum, Q1 and Q3 show that the bootstrap variability is considerable in both tails making the causal path subject to sampling variability, implying considerable uncertainty in the estimated 'sum.'

The signs of means and medians are both positive in columns 1 and 3 for consumption and private wage equations, implying that wage appears to be endogenous in the consumption equation (25), while gnp may be endogenous in the private wage equation (27). The $P^*(+1) = 0.481 < 0.5$, along with the negative sign of the mean and the median in the second column entitled 'inve' suggests that coPr appears to be exogenous in the investment equation (26).

5.2 Macro Risk Factors for Excess Bond Premium

US Macroeconomists and Federal Reserve researchers have developed new awareness of their failure to forecast the great recession of 2007-2008. Some have developed new data series. For example, Gilchrist and Zakrajek (2012) have developed excess bond premium (EBP) and shown that it predicts risk of a recession. It is interesting to find what causes the EBP itself, possibly allowing us to understand why EBP predicts recession risk.

Potential causes are: unemployment rate (UnemR), credit creation (CrCrea, not seasonally adjusted), credit destruction (CrDstr, not seasonally adjusted), yield on 10-year treasury bonds (Yld10, not seasonally adjusted), effective federal funds rate (EffFFR), and money stock (M2, seasonally adjusted billions of dollars). Arguments for using separate variables for CrCrea and CrDstr are found in Contessi and Francis (2013) with additional references. We use Federal Reserve's quarterly data from 1973Q1 to 2012Q4, with some data missing. Our software tools can efficiently handle missing data.

We study endogeneity of variables in the following regression model:

$$\text{EBP} = \text{f}(\text{UnemR}, \text{CrCrea}, \text{CrDstr}, \text{Yld10}, \text{EffFFR}, \text{M2}) \tag{28}$$

After getting the data and relevant packages into R memory, we can use the following commands:

```
mtx=cbind(EBP,UnemR,CrCrea, CrDstr,Yld10,EffFFR,M2)
p=NCOL(ntx);print(colnames(mtx)[2:p])
```

```
silentPairs(mtx)#newer version of Cr1
silentPairs0(mtx)#zero suggests older version of Cr1
```

The output of this shows that only CrCrea, CrDstr and M2 are negative implying that they are exogenous.

```
       "UnemR"  "CrCrea" "CrDstr" "Yld10"  "EffFFR" "M2"
NewCr1  3.175   -1.000   -1.000    3.175     3.175 -1.000
OldCr1  1.000   -1.000   -1.000    3.175     1.000 -1.000
```

The above output of 'sum' index is in the range: $[-3.175, 3.175]$. The results in more intuitive translated range: $[-100, 100]$ plus Pearson correlation and its p-values require simple code:

```
su=causeSummary(mtx)
su0=causeSummary0(mtx)#zero suggests older version of Cr1
require(xtable)
xtable(su); xtable(su0)
```

The Latex Table is printed in the following Table 3. Note that only CrCrea, CrDstr and M2 are likely to be independently generated (exogenous) causing the excess bond premium, while the other variables seem to be caused by EBP (endogenous). None of the magnitudes in the column entitled 'strength' is less than the threshold 0.476 for 'sum' according to our decision rule R3, implying that we do not have bidirectional paths.

Causal directions in Table 3 for old Cr1 and new Cr1 are identical. The strengths in rows labeled 1 and '1.old' are distinct with $ui = (100, 31.496)$, respectively. Same discrepancy holds between rows 5 and '5.old.' Thus the difference between two versions of Cr1 are not found to be significant for this example.

What about sampling variability of strength index? The bootstrap inference is computer time intensive. It requires the R function `pcause` as illustrated in the following code.

```
p=NCOL(mtx)
ou2=matrix(NA,nrow=p-1,ncol=2)
for (i in 2:p){
pp=pcause(mtx[,1],mtx[,i],n999=999)
ou2[i-1,1]=colnames(mtx)[i]
ou2[i-1,2]=round(pp,6)  }
```

```
print(ou2)
colnames(ou2)=c("variable", "P(-1,0,1)")
xtable(ou2)
```

The printed output of the above code is suppressed for brevity. Instead, our Table 4 shows that sampling distribution results provide a distinct piece of information not covered by the results about the strength or p-value in Table 3.

Graphics on Pair-wise Relations

Pretty scatterplots with locally best fitting lines for each pair of data have now become possible with a nice R package called 'PerformanceAnalytics' by Carl and Peterson (2010) with the function `chart.Correlation` modified for our purposes in the following code.

```
require(PerformanceAnalytics)
chartCorr2=function(mtx,temp="temp",nam=colnames(mtx)){
p=NCOL(mtx)
#print(c("colnames=",nam))
if (p<2) stop("chartCorr2 has input mtx with <2 columns")
nameoplot=nam[2:p]
print(nameoplot)
for (i in 2:p) {
mypath<-file.path("C:",temp,paste(nameoplot[i-1],".pdf",sep=""))
pdf(file=mypath,width=9,height=7)
chart.Correlation(mtx[,c(1,i)])
dev.off()
}# end i loop
}#end function
chartCorr2(mtx)
```

All figures are analogous. Histograms of the two variables is seen in the diagonal panels. The South West panel has a scatter diagram and locally best fitting free hand curve. The number in the North East panel is the ordinary correlation coefficient whose font size suggests its statistical significance, with stars increasing with 10%, 5% and 1% level. Figures provide visual impressions while the exact correlation coefficients and their p-values are also found in Table 3 with more decimal points.

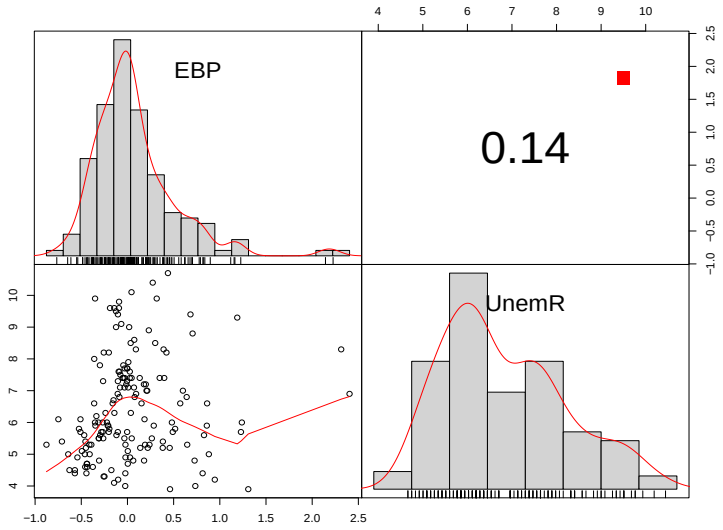

Figure 2: Scatterplot with nonlinear curve: EBP-UnemR

Our evidence including Figure 2 suggests that the variation in UnemR is endogenous, caused by EBP with a scatterplot having a mildly up-down-up pattern.

Our evidence including Figure 3 suggests that the variation in credit creation is exogenous. Its scatterplot is mostly flat and lots of noise.

Our analysis and Figure 4 suggests that the variation in credit destruction is exogenous. This scatterplot is also mostly flat with lots of noise, similar to credit creation.

Our evidence including Figure 5 suggests that the variation in the yield on 10-year notes is endogenous, caused by EBP with a scatterplot having a mildly up-down pattern.

Our analysis and Figure 6 suggests that the variation in the effective federal funds rate is endogenous, caused by EBP with a scatterplot having a mildly up-down pattern. The non-deterministic variation in Effective Federal Funds rate (EffFFR) is less "original or independent" than the corresponding variation in EBP. When EBP is negative and rises toward zero EffFFR increases, but beyond zero it decreases with increase in EBP. It would be interesting to consider the "surprise" component of the effective FFR and its

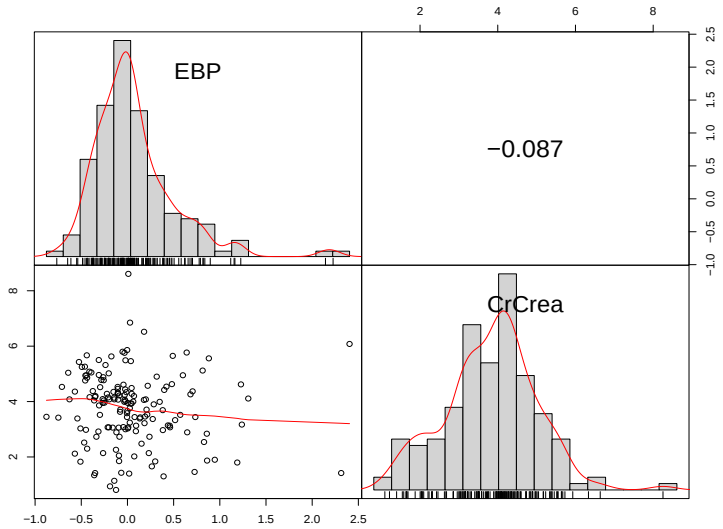

Figure 3: Scatterplot with nonlinear curve: EBP-CrCrea

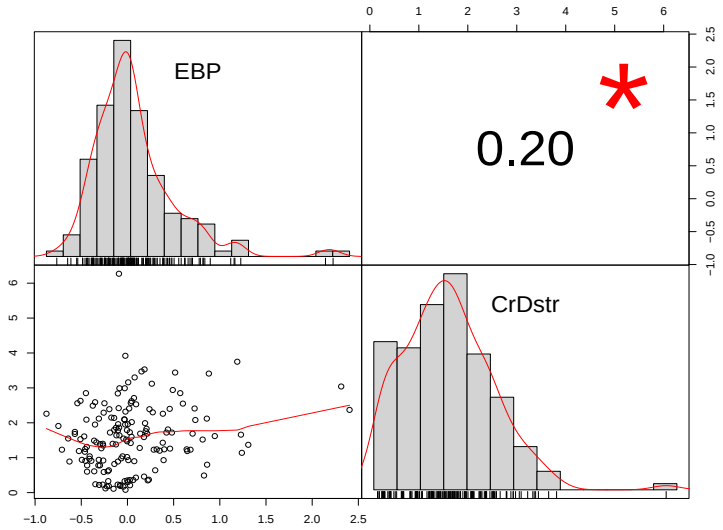

Figure 4: Scatterplot with nonlinear curve: EBP-CrDstr

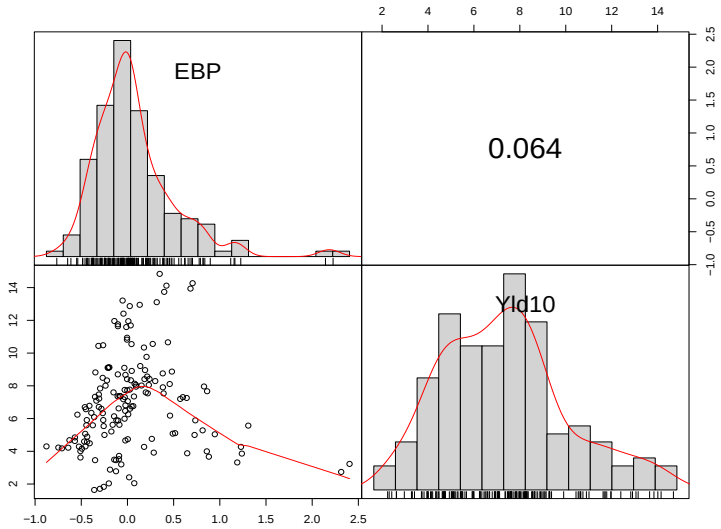

Figure 5: Scatterplot with nonlinear curve: EBP-Yld10

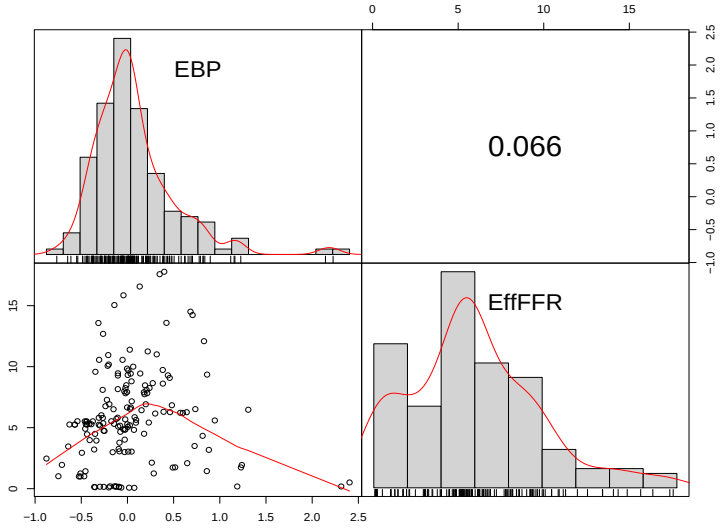

Figure 6: Scatterplot with nonlinear curve: EBP-EffFFR

relationship with the EBP

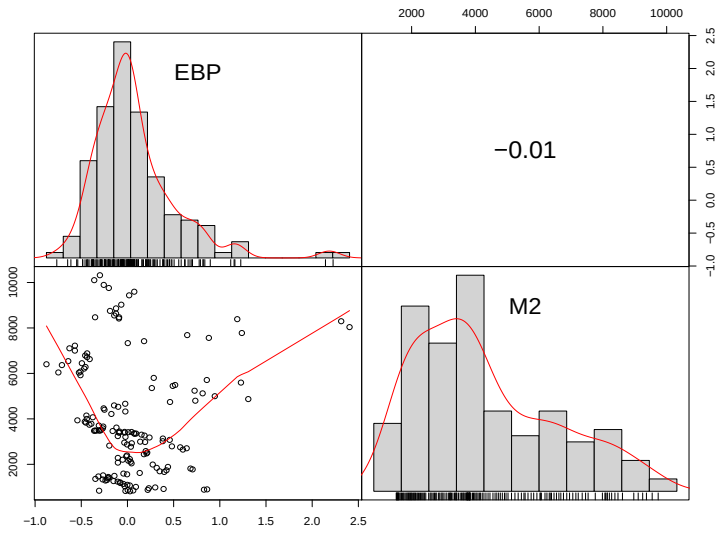

Figure 7: Scatterplot with nonlinear curve: EBP-M2

Our evidence including Figure 7 suggests that the variation in money stock M2 is exogenous with a scatterplot having a mildly down-up pattern. The non-deterministic variation in EBP is less"original or independent" than the corresponding variation in money stock M2.). The graphics reveals that when EBP is negative and rises toward zero as M2 decreases, but beyond the zero EBP M2 increases with increase in EBP.

5.3 Airquality data

Our next example shows how the `causeSummary` function of the package provides reasonable results showing that all meteorological variables are exogenous for Ozone (ppb) air pollution in New York in 1973, using some famous data always available in R.

```
library(generalCorr)
c1=causeSummary(as.matrix(airquality))
library(xtable)
xtable(c1)
```

The results in Table 5 show that solar radiation (Langleys) and temperature (degrees F) have strongly independent variation, influencing Ozone pollution levels with high strength of 100 for both, suggesting unanimity of Cr1 and Cr2 criteria at all four stochastic dominance levels and further confirmed by Cr3. The results in Table 6 using older Cr1 are almost identical.

Other variables: Wind (mph), month number (1:12) and Day number (1:31) also affect Ozone, but the causal direction is not unanimous. Hence the strength index is only 31.496 for them. Not surprisingly, high wind reduces Ozone pollution is indicated by the significantly negative (–0.6015) Pearson correlation coefficient with a near zero p-value. Additional comments about Tables 5 and 6 are omitted for brevity.

We use following code to generate a table of bootstrap results.

```
options(np.messages=FALSE)
bb=bootPairs(airquality, n999=999)
ap=apply(bb$out,2,summary)
ap2=rbind(ap,bb$probSign)#P* at the bottom of summary table
xtable(ap2,digits=3)
```

The results are summarized in Table 7, where the 'sum' index is in the range $[-3.175, 3.175]$. We can focus of the means to obtain the overall effect. The bottom row of Table 7 reports the relative frequency of negative values according to the definition (21) implying a success probability in obtaining a negative sign after removing from the denominator all bootstrap estimates m lying in the bidirectional range $[-0.05, 0.05]$. For our example, $m = 0$ for all columns. The bottom line shows that the negative signs in all columns are very reliably estimated. It may be convenient to simply set $m = 0$ in the denominator $(N - m)$, leading to conservative estimates of success rates.

5.4 'silentMtx' illustrated with 'mtcars' automobile data

In some engineering applications the causal direction is up to the engineer in the sense that she can change engineered settings for one variable to study its effect on some other variable. We use well known 'mtcars' data always available in R to describe the function 'silentMtx' which prints a signed matrix of unanimity indexes in the range [–100, 100] for each pair of variables allowing for some variables to be treated as control. Let us use the sixth variable 'wt' or weight of the car as the control.

```
require(np);require(generalCorr);options(np.messages=FALSE)
silentMtx(mtcars[,1:4],ctrl=mtcars[,6])
silentMtx0(mtcars[,1:4],ctrl=mtcars[,6])
```

The interpretation of signed unanimity indexes is self-explanatory in the following R outpout.

```
[1] "Negative index means the column named variable
 kernel-causes row named"
[1] "Positive index means the row named variable
kernel-causes column named"
[1] "abs(index)=sign unanimity by weighted sum of
3 signs from Cr1 to Cr3"
#using silentMtx command for newer Cr1 version
       mpg     cyl    disp       hp
mpg  100.000  37.008 -31.496 -100.000
cyl  -37.008 100.000  37.008   18.110
disp  31.496 -37.008 100.000   37.008
hp   100.000 -18.110 -37.008  100.000

#using silentMtx0 command for older Cr1
mpg  100.000 -31.496 -31.496 -100.000
cyl   31.496 100.000 -31.496  -31.496
disp  31.496  31.496 100.000  -31.496
hp   100.000  31.496  31.496  100.000
```

For example, the negative elements [1,4] =(–100, –100) in the upper and lower parts of the above output matrix suggest that the column 4 'horse power variable' kernel causes the 'miles per gallon' or the row 1 variable, or: 'hp'→'mpg'. The absolute values of the unanimity index (=100) suggests that the path direction is unanimously supported by Cr1 to Cr3 under both definitions of Cr1.

The elements at the diagonally opposite locations [4,1] in the output matrix have the opposite positive sign, meaning reverse causal path with the same meaning: Column 1 variable is kernel caused by the row 4 variable or 'mpg'←'hp'. Both paths are exactly the same even though the signs are opposite, as they should be. Of course, the signs and magnitudes of all pairs do not match for the two distinct definitions of Cr1.

If the argument matrix 'mtx' has p rows, 'silentPairs' provides a useful summary vector with $(p-1)$ elements, focused on the first column paired with all other columns in the range [–3.175, 3.175]. By contrast, 'silentMtx' provides a useful summary matrix of all causal path pairs converted to the intuitive range [–100, 100].

5.5 'parcorMany' illustrated with 'mtcars' data

The R function `parcorMany` creates a matrix of generalized partial correlation coefficients between all pairs of variables after removing the effect of remaining variables and also after removing the effect of control variables if any, when the dependencies are computed from kernel regressions.

```
parcorMany(mtcars[,1:4],ctrl=mtcars[,6])
```

Since we have four basic variables and one control variable, we have `choose(4,2)` or six pairs or three flipped pairs. In the following output column entitled nami and namj provide names of X_i and X_j while partij and partji provide the partial correlations. The column entitled 'rijMrji' reports the difference between their absolute values useful for our third criterion Cr3: `(abs(partij) - abs(partji))`.

```
> parcorMany(mtcars[,1:4],ctrl=mtcars[,6])
      nami    namj   partij    partji    rijMrji
[1,] "mpg"   "cyl"  "-0.0033" "-0.3428" "-0.3395"
[2,] "mpg"   "disp" "0.0634"  "0.0421"  "0.0213"
[3,] "mpg"   "hp"   "-0.0845" "-0.0883" "-0.0037"
```

This function is included at the request of a package user.

6 Summary and Final Remarks

We show that Engle et al. (1983) or EHR's "weak exogeneity" is not directly testable as it involves arbitrarily defined parameters of interest (ψ) and nuisance parameters (λ_2). Hausman-Wu indirect exogeneity tests use IV estimators which can "do more harm than good" (Bound et al., 1995, p. 449), and are criticized as being "very inefficient" by Kiviet and Niemczyk (2007), Dufour, and others. Medicine has long rejected medieval-style diagnoses of

diseases by simply showing that a cure works. Hence there is a long-standing need for a practical exogeneity test.

We define Kernel exogeneity and suggest that exogenous variables X_i should have independently generated DGP (self-driven innovation) and according to Koopmans should "approximately cause" the dependent variables Y. EHR correctly show that Koopmans' methods cannot unambiguously identify the causal direction, since two flipped linear regressions, (Y on X_i) and (X_i on Y), have the same R^2. We show that modern computing tools and concepts including Zheng et al. (2012) allow us to remove the linearity assumption and use newly defined asymmetric generalized partial correlations.

We assess whether a variable has independently generated DGP by using three distinct quantifiable criteria: Cr1 to Cr3. Our unanimity index (ui) considers preponderance of evidence using two out of three criteria Cr1 to Cr3. The Cr3 uses 'goodness of fit' when it compares generalized (partial) correlation coefficients, first suggested in Vinod (2014). Since it is not safe to rely on goodness of fit alone, the other two criteria (Cr1, Cr2) here yield two fuzzy inequalities. The Cr1 involves absolute values of an orthogonality statistic from the cross product of regressor and local residual of kernel regression.

The R package 'generalCorr' has an option to use an older version of Cr1 involving absolute values of the gradients of local linear kernel regressions. The Cr2 involves absolute values of local kernel regression residuals.

Financial economics has long ago developed tools for a comprehensive study of fuzzy inequalities between stock market returns of two competing investment opportunities (e.g., mutual funds) called stochastic dominance of orders 1 to 4 (or SD1 to SD4). See a survey in Levy (1992) and discussion of SD4 in Vinod (2004). Our sample statistics measuring SD1 to SD4 are called Av(sd1) to Av(sd4) which are further aggregated by using weights inversely related to their sampling variances. Weighted sums quantify the Cr1 and Cr2.

Vinod (2015a) reports favorable simulations using Cr3 alone. Our decision rules based on the 'sum' or unanimity index ui statistics incorporating all three criteria are simulated in section 3 with high success rate.

Our new bootstrap test for exogeneity in section 4 can do statistical inference for the 'sum' statistic, using about a thousand estimates. Descriptive statistics of these estimates, illustrated in Table 2, provide a view of their sampling distribution to assess the preponderant sign and hence the causal

direction, as well as, unanimity strength.

If the causal path is from the left hand side variable to a right hand side variable, or is bidirectional, endogeneity problem exists and researchers may well add extra equations leading to a simultaneous equations model. Koopmans' "departmental principle" gives practitioners some flexibility in designating certain non-economic variables as exogenous, without having to add extra equations and eliminating a need for exogeneity testing.

We illustrate the new bootstrap exogeneity test using the famous Klein I simultaneous equations model. Our Section 5.2 considers a novel model explaining the 'excess bond premium' (EBP) known to be a good predictor of US recessions. We study detailed relation between EBP and six variables including various criteria and graphics, providing software tools for implementation based on the R package 'generalCorr.' Our evidence suggests that the variation in three variables: credit creation (CrCrea), credit destruction (CrDstr) and money stock (M2), is exogenous (independenly generated) causing changes in EBP.

Clearly, practitioners can use our summary functions implemented with very few lines of code. The ability to treat potentially confounding variables as control may be particularly valuable. It is straightforward to extend and modify our tools, if indicated by future research, since they are open source.

References

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

Bound, J., Jaeger, D. A., Baker, R., 1993. The Cure Can Be Worse than the Disease: A Cautionary Tale Regarding Instrumental Variables. NBER Working Paper No. 137.
URL `http://ssrn.com/paper=240089`

Bound, J., Jaeger, D. A., Baker, R., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak. Journal of the American Statistical Association 90, 443–450.

Carl, P., Peterson, B. G., 2010. PerformanceAnalytics: Econometric tools for

performance and risk analysis.
URL http://CRAN.R-project.org/package=PerformanceAnalytics

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.

Davidson, R., MacKinnon, J. G., 1993. Estimation and Inference in Econometrics. New York: Oxford Univ. Press.

Engle, R. F., Hendry, D. F., Richard, J.-F., 1983. Exogeneity. Econometrica 51, 277–304.

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

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

Kiviet, J. F., Niemczyk, J., 2007. The asymptotic and finite-sample distributions of OLS and simple IV in simultaneous equations. Computational Statistics & Data Analysis 51, 3296–3318.

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

Legal Information Institute, 2017. Wex Legal Dictionary. Cornell Law School, Ithaka, NY.
URL https://www.law.cornell.edu/wex/preponderance_of_the_evidence

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

Li, Q., Racine, J. S., 2007. Nonparametric Econometrics. Princeton University Press.

Pearl, J., 2009. Causality: Models, Reasoning and Inference. New York: Wiley.

Vinod, H. D., 2004. Ranking mutual funds using unconventional utility theory and stochastic dominance. Journal of Empirical Finance 11(3), 353–377.

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., 2015a. Generalized correlation and kernel causality with applications in development economics. Communications in Statistics - Simulation and ComputationAccepted Nov. 10, 2015.
URL http://dx.doi.org/10.1080/03610918.2015.1122048

Vinod, H. D., 2015b. 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., 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., 2012. Generalized measures of correlation for asymmetry, nonlinearity, and beyond. Journal of the American Statistical Association 107, 1239–1252.

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

Table 2: Klein I model: Bootstrap summary statistics for 'sum' of eq. (20) using 999 resamples to represent the population. A positive mean and median with a large $P^*(+1)$ imply the relevant regressor might not be exogenous.

	cons	inve	prWg
Minimum	-3.1750	-3.1750	-3.1750
1st Quartile, Q1	-1.1750	-1.1750	-1.1750
Median	1.0000	-0.9250	0.0875
Mean	0.4443	-0.1892	0.1874
3rd Quartile, Q3	1.1750	1.1750	1.1750
Maximum	3.1750	3.1750	3.1750
$P^*(+1)$	0.597	0.481	0.504

Table 3: Excess Bond Premium and possible causes using new Cr1 and old Cr1 indicated by row names

	cause	response	strength	corr.	p-value
1	EBP	UnemR	100	0.1443	0.06875
1.old	EBP	UnemR	31.496	0.1443	0.0688
2	CrCrea	EBP	31.496	-0.087	0.27387
2.old	CrCrea	EBP	31.496	-0.087	0.2739
3	CrDstr	EBP	31.496	0.1998	0.01132
3.old	CrDstr	EBP	31.496	0.1998	0.0113
4	EBP	Yld10	100	0.064	0.42165
4.old	EBP	Yld10	100	0.064	0.4216
5	EBP	EffFFR	100	0.0657	0.40915
5.old	EBP	EffFFR	31.496	0.0657	0.4091
6	M2	EBP	31.496	-0.0103	0.8976
6.old	M2	EBP	31.496	-0.0103	0.8976

Table 4: Bootstrap success rates for causal direction using 999 resamples

	variable	P($\pm$1)
1	UnemR	0.801802
2	CrCrea	0.927928
3	CrDstr	0.626627
4	Yld10	0.947948
5	EffFFR	0.600601
6	M2	1

Table 5: Ozone pollution and its various known causes using newer Cr1

	cause	response	strength	corr.	p-value
1	Solar.R	Ozone	31.496	0.3483	0.00018
2	Wind	Ozone	100	-0.6015	0
3	Temp	Ozone	31.496	0.6984	0
4	Month	Ozone	100	0.1645	0.0776
5	Day	Ozone	100	-0.0132	0.88794

Table 6: Ozone pollution and its various known causes using older Cr1

	cause	response	strength	corr.	p-value
1	Solar.R	Ozone	100	0.3483	2e-04
2	Wind	Ozone	31.496	-0.6015	0
3	Temp	Ozone	100	0.6984	0
4	Month	Ozone	31.496	0.1645	0.0776
5	Day	Ozone	31.496	-0.0132	0.8879

Table 7: Variability of 'sum' over 999 bootstrap resamples using airquality data

	Solar.R	Wind	Temp	Month	Day
Min.	-3.175	-3.175	-3.175	-3.175	-3.175
1st Qu.	-3.175	-2.575	-1.500	-1.600	-1.000
Median	-3.175	-1.000	-1.175	-1.000	-1.000
Mean	-2.347	-1.539	-1.520	-1.531	-0.957
3rd Qu.	-1.175	-1.000	-1.175	-1.000	-1.000
Max.	1.975	1.175	1.000	-0.500	2.025
$P^*(-1)$	0.9459	0.9299	0.9710	1.0000	0.9760