Causality Estimation in Panel Data

H. D. Vinod *

October 7, 2023

Abstract

Evaluation of causal paths from panel data (time series of cross sections or longitudinal data) can use pooled data, ignoring the time and space dimensions. More generally, we want to draw readers' attention to an algorithm `causeSum2Panel(.)`, freely available in the R package 'generalCorr.' It estimates causality directions and strengths, focusing on the time and space dimensions. We describe new tools using the space dimension data to formally test Granger causal directions. We illustrate the uniquely new insights gained from the two dimensions, using three datasets already available in the R package 'plm' for panel linear models, namely Grunfeld, Crime, and Cigar. Among new insights available nowhere else, we identify which regressions suffer from endogeneity issues, causal path directions, and strengths. We indicate fruitful areas for further research in studies of panel data.

1 Introduction

This paper generalizes the following setting for panel data linear regression models with a new focus on the directions and strengths of causal paths. Vinod (2019) has 32-page details, including theorems and propositions. It describes stochastic causality and counterfactual presence when using our kernel regressions. This paper goes beyond ordinary nonlinear nonparametric regressions to accommodate special features of panel data.

Panel data typically has a time dimension as well as a cross-sectional dimension. It is convenient to refer to the cross-sectional dimension as "space," even though it may not be literally appropriate in some contexts. A great deal of panel data literature deals with extensions of the following linear model.

$$y_{it} = \alpha_{it} + \beta'_{it} x_{it} + u_{it}, \tag{1}$$

where the subscript it refers to $i = 1, \ldots n$ spaces at times $t = 1, \ldots, T$, where u_{it} are errors, β_{it} is a long vector of p parameters, and α_{it} intercepts vary with each i and t. We cannot estimate (1), since it has too many $((p+1)nT)$ unknown parameters to be estimated with

only nT observations. If one assumes that $\alpha_{ij} = \alpha$, a constant, and $\beta_{ij} = \beta$, also a constant vector, we can estimate the so-called pooled data ordinary least squares (OLS) with no subscripts for the parameters (α, β).

$$y_{it} = \alpha + \beta' x_{it} + u_{it}, \quad \text{(pooled)} \tag{2}$$

assuming well-behaved white noise errors u_{it}.

Splitting the error term $u_{it} = u_i + \epsilon_{it}$, allowing for distinct individual-related errors, may be realistic. If we treat u_i as parameters to be estimated, we have the so-called fixed-effects (FE) model:

$$y_{it} = \alpha_i + \beta' x_{it} + \epsilon_{it}, \quad \text{(fixed} - \text{effects)} \tag{3}$$

where ϵ_{it} elements are well-behaved, and the individual-specific effects α_i are additional fixed-effects parameters.

If the relations are nonlinear and nonparametric, one uses kernel regressions.

$$y_{it} = f(x_{it}) + u_{it}, \tag{4}$$

where f is a generic nonparametric (no β, α_i parameters) nonlinear function. Actually, one should think of $f(x_{it})$ as a conditional expectation function distinct for each observation. One can then compute averages of partial derivatives $(\partial y / \partial x_{it})$ as approximations to β parameter vector. There are no intercepts, *per se*, in kernel regressions. One can approximate fixed-effect parameters α_i from the difference between summations Σy_{it} of data and summation $\Sigma \hat{y}_{it}$ of kernel regression fitted values. The main appeal of kernel regressions is twofold. First, its fits are usually much better than those of linear models. Second, linearity is often invalid when human agents are involved. We leave the estimation of (β, α_i) outside the scope of this paper.

Instantaneous kernel causality distinguished from Granger's causality is made practical in the R package Vinod (2021) and cited by many, including Allen and Hooper (2018), Allen (2022), Allen and McAleer (2022), Lister and Garcia (2018), Mlynczak and Krysztofiak (2019), Fousekis (2020), Vinod (2020b), Vinod et al. (2023), among others. Since these papers do not deal with panel data, this short paper focuses on describing a newer algorithm in the 'generalCorr' package `causeSum2Panel(.)`. This paper includes R code to facilitate an application to any panel data.

The R package 'plm' by Croissant and Millo (2008) helps estimate panel linear models, which focus on fixed effect and random effect models using time series of cross sections of data, also known as panel or longitudinal data. The 'plm' package methods do not report causal directions or strengths of causal paths between y and p regressor variables `xj` `= x1, x2, ..., xp` allowing for control variables (if any) `z1, z2, ...`. The R package 'generalCorr' has algorithms for exogeneity testing and estimating causal path directions and strengths for pooled data, that is, without explicitly allowing for special features of panel data.

Regarding the estimation of nonlinear "Granger causality," the R package generalCorr has the function `GcRsqX12()` with details in Vinod (2020a). Its arguments are `y` and potential causal variables `xj = x1, x2, ..., xp`. The algorithm uses the 'np' package

2

by Hayfield and Racine (2008) for Kernel regressions using "local linear" fits and Akaike Information Criterion (AIC) for bandwidth selection options.

$$y = f(Ly, Lx_j) + u, \tag{5}$$

where f is a generic nonlinear nonparametric function, u has errors, Ly denotes the set of all lagged values of y, and Lx_j denotes similar sets with all time lags of x_j variables. The flip side regression replaces y on the left-hand side by x_j of equation (5) in similar notation is,

$$x_j = f(Ly, Lx_j) + u. \tag{6}$$

According to Granger's methods, the causal path $x_j \to y$ is supported by data if the R^2 or the goodness-of-fit for (5) is larger than the R_j^2 (distinguished by the subscript j) or the goodness-of-fit for (6).

Granger causality direction is indicated by the sign of the difference $\Delta R^2 = [R^2 - R_j^2]$. Of course, the difference needs to be significantly positive over and above sampling variation. These differences need to be computed several hundred times using bootstrap resamples to define Granger causality in usual data (not panel). The package generalCorr has functions for computationally demanding bootstrapping.

Panel data offer a way to avoid computationally demanding bootstraps by using Granger calculations along the space dimension. We inject a subscript 'i' to indicate space in $\Delta_i R^2$ sample estimates of differences in goodness of fit for the s-th space dimension. A real-world indication of sampling variability in ΔR^2 values is hidden in the variability among $\Delta_i R^2$ distinct space values. Denote their sample mean by $\bar{\Delta}$. Denote the sample variance by $s^2(\Delta)$. The sample mean is statistically significantly different from zero if the (say 95%) confidence interval for $\bar{\Delta}$ does not contain zero. Kendall and Stuart (1977) section 20.31 describes "Studentization" for this purpose. Denote the Studentized statistic shown to be distributed as a Student's t density with degrees of freedom $df = n - 1$.

$$z = \frac{\bar{\Delta} - \mu}{s(\Delta)} \sim t(0, 1, df), \tag{7}$$

where μ is the true unknown (or null hypothesis value 0) mean of $\bar{\Delta}$, and where the sample standard deviation is in the denominator. Denote the critical value of the t density with appropriate df as $C_r = |t_{\alpha/2}|$. For example, if $\alpha = 0.05$ and degrees of freedom equal 24. the R command `qt(0.025,df=24))` yields -2.063899. Its absolute value is the critical value C_r for this example.

The $(1 - \alpha)\%$ confidence interval is

$$P\{\bar{\Delta} - C_r\, s(\Delta) \le \mu \le \bar{\Delta} + C_r\, s(\Delta)\} = 1 - \alpha. \tag{8}$$

If the absolute value of the statistic on the left-hand side of (7) exceeds C_r, the sign of the statistic implying the causal path is statistically significant. The `causeSum2Panel(.)` algorithm reports a matrix called `outdif`, giving all $i = 1, 2, \ldots, n$ values of $\Delta_i R^2$ for all

3

(y, x_j) pairs. The column headings have a hybrid name y-xj. It also reports a summary matrix grangerAns with two rows. The first row has $\bar{\Delta}$ and second the test statistic from the left-hand side of (7) with $\mu = 0$. The novelty in our Granger causality results is the formal test for Granger causality, without any computer-intensive bootstraps. Hence, these results are worthy of further study.

We begin our R session with the initial code to start R with a clean slate and load the needed packages mentioned above in the current memory of R.

```
rm(list=ls())
options(width=60)
library(generalCorr)
options(np.messages=FALSE)
library(plm)
```

Once the data is available as a *.csv or comma-separated Excel file, it needs to be arranged in a format suitable for the 'plm' package. Our illustrations use the data available in the 'plm' package already in the needed format.

1.1 Grunfeld Data

This subsection explains the usage of the function causeSum2Panel() of the R package 'generalCorr' using the data from the 'plm' package. A balanced panel of 10 observational units (firms) from 1935 to 1954. The time dimension has T=20 years, and the space dimension has n=10 firms.

```
data(Grunfeld)
da=data.frame(Grunfeld);attach(da)
head(Grunfeld,2);tail(Grunfeld,2)
p=2 # set this as number of regressors=length(namXjmtx)=rowfnout
```

The output of the above code provides a look at the data for the first two and the last two observations. We have inserted the header 'seq' for the first column of output representing the sequence number.

```
seq firm year   inv  value capital
1     1 1935 317.6 3078.5     2.8
2     1 1936 391.8 4661.7    52.6
seq  firm year  inv value capital
199   10 1953 6.53 63.51   11.68
200   10 1954 5.12 58.12   14.33
```

The total number of rows in the data is seen to be 200 from the last sequence number. It equals the total number of firms (n=10) times the total number of years (T=20) in the data. The data are arranged with the entire time series for the first firm, followed by the entire time series for the second firm (space unit), and so on. This illustrates how the data needs to be arranged for the 'plm' package.

4

The underlying kernel regression for this data is

$$inv = f(value, capital) + u, \qquad (9)$$

where f denotes a generic nonlinear nonparametric function, and u is our generic notation for errors. The intuitive causal direction from the specification of (9) is from regressors to the dependent variable. The estimation has an "endogeneity problem" if the data analysis suggests a reversed causal direction.

The following code illustrates how to use the `causeSum2Panel()`. The user needs to provide the name of the data (da) and the names of various variables describing the panel study of estimation of (9).

```
cP=causeSum2Panel(da,
rowfnout=2,
colfnout=5,
fnoutNames=c("cause","effect","strength","r","p-val"),
namXs="firm",
namXt="year",
namXy="inv",
namXc=0,
namXjmtx= c("value","capital"),
# chosenTimes=1943:1949,
# chosenSpaces=3:9,
verbo=FALSE)


xtable(cP$strentime,digits=1)
xtable(cP$pearsontime,digits=3)
xtable(cP$strenspace,digits=1)
xtable(cP$pearsonspace,digits=3)
xtable(cP$grandsum)
xtable(cP$grangerAns, digits =3)
```

In the above code, we have included two extra lines (commented out by the preceding # symbol) to show how to select a subset of time and space values. If no subsets are indicated, the default is to include all time and space values in the data. Since this data has $T = 20$, with four (= default) time lags for Granger causality calculations.

The R function 'xtable' allows output printing in the Latex format. The last few lines of the above code produce the Latex tables reported in the six tables here.

Since y is the dependent variable in these regressions based on (9), one would expect that the causal path goes from xj to y. However, the negative averages in the last part of Table 1, entitled 'avg' $(-70.6, -2.8)$, indicate that the causal direction is the reverse of what might be expected. That is, x1='value of the firm' causes y or gross investment. Similarly, x2= 'the firm's capital stock' drives y or gross investment in the firm. These results indicate that the regression suffers from the endogeneity problem. The results are similar in Table 3, except that the regression of y on capital stock x2 has a positive average (=16), suggesting no endogeneity problem for the regressor x2.

Table 1: Summary of causal strengths and directions for various years averaged over 10 firms. Table columns are split into three lines.

	1935	1936	1937	1938	1939	1940	1941	1942
1943	1944	1945	1946	1947	1948	1949	1950	1951
1952	1953	1954	avg					
y-value	-100.0	-77.8	-100.0	11.1	-100.0	-77.8	-100.0	-77.8
-77.8	33.3	-100.0	-100.0	-100.0	-77.8	-100.0	-77.8	-77.8
-77.8	-77.8	44.4	-70.6					
y-capital	-100.0	88.9	33.3	-33.3	-100.0	22.2	44.4	100.0
-100.0	33.3	22.2	22.2	-33.3	33.3	-77.8	-44.4	33.3
-33.3	-11.1	44.4	-2.8					

Table 2: Summary of Pearson correlation coefficients (y,xj) for various years averaged over 10 firms. Positive correlations mean that y and xj move in the same direction. Table columns are split into three lines.

	1935	1936	1937	1938	1939	1940	1941	1942
1943	1944	1945	1946	1947	1948	1949	1950	1951
1952	1953	1954	avg					
y-value	0.930	0.834	0.808	0.797	0.887	0.905	0.919	0.924
0.914	0.934	0.951	0.946	0.944	0.887	0.928	0.926	0.929
0.918	0.941	0.925	0.907					
y-capital	-0.234	-0.142	0.278	0.551	0.463	0.354	0.401	0.419
0.319	0.123	0.155	0.267	0.601	0.584	0.662	0.670	0.636
0.727	0.858	0.919	0.431					

Table 2 correlations are almost all positive, implying the variables xj are positively related to y, with positive averages of 0.907 and 0.431, respectively. Table 4 is similar in that both row averages (0.546, 0.803) are positive.) for each firm (space) where we average over the 20 years.

The grand summary of results in Tables 1 to 4 for Grunfeld data is in Table 5. The summary table reports only the values in columns entitled 'avg' for averages of row value. They are reported for each regressor x1 (value) and x2 (capital) paired with y. The time dimension summaries over space dimension are in the first two columns of Table 5. The column entitled 'strtime' summarizes strengths in Table 1. The column entitled 'cortime' summarizes correlations in Table 2. Similar space dimension summaries are in the last two columns of Table 5, again reporting only the row averages in columns entitled 'strspace' and 'corspace'.

The following interpretation is available from the grand summary Table 5. The negative signs in the column entitled 'strtime' indicate that the causal path goes from [y → value], and [y → capital], where we consider various time values (1935 to 1954) separately. When we use the standard 5% value for significance, the 'strtime' for the row y-capital (-2.78) is not significantly negative, implying that the causal path might be bi-directional [capital ↔ y].

Table 3: summary of causal strengths and directions for various firms (spaces) averaged over 20 years. The columns are split into two rows.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
	9	10	avg					
y-value	-44.4	33.3	-100.0	-66.7	-100.0	55.6	33.3	-88.9
	-100.0	-11.1	-38.9					
y-capital	100.0	-44.4	-44.4	77.8	33.3	33.3	33.3	-77.8
	33.3	NA	16.0					

Table 4: Summary of Pearson correlation coefficients (y,xj) for 10 firms averaged over 20 years. Positive correlations mean that y and xj move in the same direction. Table columns are split into two lines.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
	9	10	avg					
y-value	0.676	0.488	0.319	0.551	0.825	0.975	0.178	0.839
	0.704	-0.096	0.546					
y-capital	0.907	0.546	0.809	0.914	0.727	0.950	0.860	0.758
	0.754	0.802	0.803					

Now, consider the space dimension with ten firms separately, where the column 'strspace' reports a negative value (-38.89) for the row label 'y-value,' implying the causal path [y → value]. The positive 'strspace' of 16.05 suggests the causal path [capital → y].

The interpretation of all positive Pearson correlation coefficients in the second column, 'cortime' and last column, 'corspace' of Table 5, suggest that the variables move in the same direction as y along both time and space dimensions.

We apply kernel regression instead of usual linear regressions in Granger causality assessments based on F tests. We compute differences between two R^2 values of flipped regressions to determine which flip has more significant support in the data. Panel data provide as many time series as there are spaces, each yielding its own differences between two R^2 values. We compute the average of these differences to be -0.1113 in Table 6 for the (y-value) pair. It being negative, the causal path is [y → value]. However, it is not significantly negative. Similarly, the average difference is positive (=0.0723), implying the causal path [capital → y], which is statistically significant. These computations assume a time lag of four years in Granger causality.

1.2   Crime Data

This subsection uses 'Crime' data from 'plm' package for illustration. It has a panel of n=90 space units (counties) from 1981 to 1987 (T=7). We select as our dependent variable (y) by the command `myDep="crmrte"` representing the crime rate. The dataset has a great many variables, of which we select $p = 4$ by the command `myRHS=c( "prbconv", "pctymle" , "density", "wser")`. They are prbconv = probability of conviction, pctymle = percentage

Table 5: Grand summary of signed strengths and Pearson correlations for Grunfeld data.

	strtime	cortime	strspace	corspace
y-value	-70.56	0.91	-38.89	0.55
y-capital	-2.78	0.43	16.05	0.80

Table 6: Grunfeld data Granger causality test based on difference of two R^2 values. The degrees of freedom are n-1 (=9) and time lags (=4) are used.

	y-value	y-capital
mean of differences in Rsq	-0.111	0.072
test statistic	-1.875	3.516

of young males (between ages 15 to 24), density =people per square mile, and wser = weekly wage in the service industry. The underlying kernel regression for this data is

$$crmrte = f(prbconv, pctymle, density, wser) + u, \tag{10}$$

where $f(.)$ denotes a generic nonlinear nonparametric function, and u is our generic notation for errors.

The following code implements the causal path code for panel data estimation of (10).

```
library(generalCorr)
options(np.messages=FALSE)
library(plm)
data(Crime)
da=data.frame(Crime)
head(Crime,2);tail(Crime,2)
crm=causeSum2Panel(da, f
colfnout=5, # always 5
fnoutNames=c("cause","effect","strength","r","p-val"),
namXs="county",
namXt="year",
namXy="crmrte",
namXc=0,
namXjmtx= c( "prbconv", "pctymle" ,   "density", "wser"),
rowfnout=4, # number of regressors
chosenTimes=NULL,
chosenSpaces=NULL,
ylag=1, #for Granger causality using time series data
verbo=FALSE)
```

The above code creates an R object called 'crm'. Further results are accessed as follows. Note that since T=7 is small, we have chosen ylag=1, not the default 4.

```
xtable(crm$strentime,digits=1)
```

```
xtable(crm$pearsontime,digits=3)
xtable(crm$strenspace,digits=1)
xtable(crm$pearsonspace,digits=3)
xtable(crm$grandsum)
xtable(crm$grangerAns, digits=3)
```

The grand summary of results for crime data is in Table 7. Let us discuss the interpretation of results for the four columns of that table. The signed causal strengths estimated in the first column entitled 'strtime' are all negative and larger than 5, suggesting statistical significance at the conventional level. Therefore, averaging over spaces suggests a causal path from the dependent variable crime rate to the four regressors. The same conclusion holds for three of four regressors in the third column entitled 'strspace' averaging over 1981 to 1987. The positive value 33.96 along the line 'y-pctymle' suggests the path [pctymle $\rightarrow$ y] is seen when averaged over the years.

The Pearson correlations in column 2 of Table 7 give a mixed picture of the crime data. The negative (-0.24) means that when the probability of conviction goes up, the crime rate decreases in column 'cortime' where one averages over space units (counties). However, when we see the positive 'cortime' values (0.23, 0.71, 0.26), it means that when the percentage of young males, population density, or weekly wage in the service sector increases, the crime rate also increases. Thus larger male proportion, population density, and wages are unlikely to reduce crime.

The Pearson correlation reporting column 4 is entitled 'corspace' in Table 7. The values are (-0.20, 0.01, -0.07, -0.04), having three negatives and one positive value, also giving a mixed picture. Detailed output by time dimension and space dimension similar to Tables 1 to 4 are printed by our code but omitted here for brevity.

Table 7: Grand summary of signed strengths and Pearson correlations for crime data.

	strtime	cortime	strspace	corspace
y-prbconv	-52.38	-0.24	-24.20	-0.20
y-pctymle	-69.84	0.23	33.96	0.01
y-density	-77.78	0.71	-11.98	-0.07
y-wser	-14.29	0.26	-17.10	-0.04

The R output of `xtable(cP$grangerAns)` in the above code ends with all positive signs of average causal paths, implying the regression path from each regressor to y remains so according to nonlinear Granger causality. However, the first column (y-prbconv) of Table 8 test statistic (=0.966) is statistically insignificant, implying that the causal path [prbconv $\rightarrow$ y] is tenuous.

1.3   Cigar Data

Cigar data in the R package 'plm' has a panel of n=46 observations from 1963 to 1992 (T=30). The space dimension is US states. In total, we have 1380 rows of data. The

Table 8: Summary of Granger causality study with average $\bar{\Delta}$ and its test statistic for crime data. The degrees of freedom are n-1 (=89), and one-year time lags are used.

	y-prbconv	y-pctymle	y-density	y-wser
mean of differences in Rsq	0.020	0.183	0.137	0.109
test statistic	0.966	7.495	6.135	4.845

kernel regression model is

$$sales = f(pop16, price, pimin, ndi) + u, \tag{11}$$

where f denotes a generic nonlinear nonparametric function, and u is our generic notation for errors.

The dependent variable is cigarette sales in packs per capita (sales). Our four regressors are population above the age of 16 (pop16), price per pack of cigarettes (price), minimum price in adjoining states per pack of cigarettes (pimin), and per capita disposable income (ndi). After cleaning the memory, the following code brings the cigar data into the current memory of R.

```
library(plm);data(Cigar)
da=data.frame(Cigar)
head(Cigar,2);tail(Cigar,2)
```

The following code processes the data to estimate causal path strengths and directions in our estimation of (11). Since T=30 is large enough, we use the default time lag of four years in assessing Granger causality by leaving the argument 'ylag' unspecified. The code produces various tables below, though we include only the last two summary tables.

```
cP=causeSum2Panel(da,
rowfnout=4,
colfnout=5,
fnoutNames=c("cause","effect","strength","r","p-val"),
namXs="state",
namXt="year",
namXy="sales",
namXc=0,
namXjmtx= c("pop16","price","pimin","ndi"),
verbo=FALSE, verboLatex=TRUE)

xtable(cP$strentime,digits=1)
xtable(cP$pearsontime,digits=3)
xtable(cP$strenspace,digits=1)
xtable(cP$pearsonspace,digits=3)
xtable(cP$grandsum)
xtable(cP$grangerAns)
```

Similar to the section on crime data, we report only the grand summary for both time and space dimensions in Table 9.

Recall that negative values of strengths in the first column of Table 9 indicate a reverse causation from [y → xj]. When we average over all the states, the variables pop16 and ndi have the endogeneity problem in regressions for various years. All positive strengths in the third column, 'strspace' suggest the intuitive causal directions [xj→y]. The magnitude of strengths exceeds 5, implying they are statistically significant. In Table 9, all strengths are significant. We find that [y → pop16] and [y → ndi] along the time dimension after averaging over the spaces.

Pearson correlations are in the second column, 'cortime,' and the fourth column, 'corspace' of Table 9. They are negative in both columns 2 and 4 for the first three rows of the table. They suggest that more populous states have larger cigarette sales. If we report correlations for each year separately and average over all states, then higher income leads to larger sales. However, if we average over the years, state-wise sales are lower in richer states.

Table 9: Grand summary of signed strengths and Pearson correlations for Cigar data.

	strtime	cortime	strspace	corspace
y-pop16	-45.2	-0.103	26.3	-0.201
y-price	53.6	-0.426	91.8	-0.545
y-pimin	55.6	-0.125	90.6	-0.533
y-ndi	-20.0	0.242	37.4	-0.422

The R output of `xtable(cP$grangerAns)` in the above code ends with all four positive signs of average causal paths implying the regression path from each regressor to y, remains so according to nonlinear Granger-causality. The test statistics in the second row of Table 10 suggest that all path directions are statistically significant.

Table 10: Summary of Granger causality study with average $\bar{\Delta}$ and its test statistic for cigar data. The degrees of freedom are n-1 (=45), and four time lags are used.

	y-pop16	y-price	y-pimin	y-ndi
mean of differences in Rsq	0.132	0.117	0.154	0.144
test statistic	7.137	8.824	8.331	8.795

2 Further Extensions

Panel data regressions have mostly been linear, even though linear models have poorer fits, leading to much larger errors. Their better-fitting nonlinearization creates problems for fixed-effect models since kernel regressions do not have direct estimates of the intercepts. Estimation of average partial derivatives $(\partial y/\partial x_{it})$ as approximations to β parameters is straightforward using the R packages 'np' and 'generalCorr.' This area deserves further attention. Another neglected area of research is applying Granger-causality to the time dimension of Panel data using the bootstrap.

This paper illustrates newer tools in 'generalCorr' that exploit the special features of the time and space dimensions in Panel data. We show how to use `causeSum2 Panel()` to identify which regressors suffer from the endogeneity issues. We also estimate causal path directions and strengths available nowhere else.

We claim novelty in the development of equation (7) for estimating the statistical significance of Granger-causality path directions, without using computer-intensive bootstraps. Finally, this paper provides R code templates for practitioners to implement the function using three commonly used Panel data sets.

References

Allen, D. E., 2022. Cryptocurrencies, diversification and the covid-19 pandemic. Journal of Risk and Financial Management 15 (3).
URL https://www.mdpi.com/1911-8074/15/3/103

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

Allen, D. E., McAleer, M., 2022. "generalized measures of correlation for asymmetry, non-linearity, and beyond": Some antecedents on causality. Journal of the American Statistical Association 117 (537), 214–224.

Croissant, Y., Millo, G., 2008. Panel data econometrics in R: The plm package. Journal of Statistical Software 27 (2), 1–43.
URL http://www.jstatsoft.org/v27/i02/

Fousekis, P., 2020. Returns and volume: Kernel causality in the US futures markets for agricultural, energy and metal commodities. Studies in Economics and Finance.
URL https://doi.org/10.1108/SEF-10-2019-0416

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. 2. New York: Macmillan Publishing Co.

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

Mlynczak, M., Krysztofiak, H., 2019. Cardiorespiratory temporal causal links and the differences by sport or lack thereof. Frontiers in Physiology 10 (45), 1–14.

Vinod, H., Jetta, K., Rengifo, M., 2023. An econometric identification of ineffective common-cold advertising. International Journal of Pharmaceutical and Healthcare Marketing.
URL https://doi.org/10.1108/IJPHM-02-2021-0022z

Vinod, H. D., 2019. New exogeneity tests and causal paths. In: Vinod, H. D., Rao, C. R. (Eds.), Handbook of Statistics: Conceptual Econometrics Using R. Vol. 41. North-Holland, Elsevier, New York, Ch. 2, pp. 33–64.
URL https://doi.org/10.1016/bs.host.2018.11.011

Vinod, H. D., 2020a. Pandemic-proofing Out-of-sample Portfolio Evaluations. SSRN eLibrary.
URL http://ssrn.com/paper=3737870

Vinod, H. D., 2020b. Updating statistical measures of causal strength. Science and Philosophy 8 (1), 3–20.

Vinod, H. D., 2021. generalCorr: Generalized Correlations and Initial Causal Path. Fordham University, New York, R package version 1.2.0, 2021, has 6 vignettes.
URL https://CRAN.R-project.org/package=generalCorr