Symbolic transfer entropy test for causality in longitudinal data *

Máximo Camacho † Andres Romeu ‡ Manuel Ruiz-Marín §

December 26, 2018

Abstract

In this paper, we use multiple-unit symbolic dynamics and the concept of transfer entropy to develop a non-parametric Granger causality test procedure for longitudinal data. Monte Carlo simulations show that our test displays the correct size and large power in situations where linear panel data causality tests fail such as when the linearity assumption breaks down, when the data generating process is heterogeneous across the cross-section units or presents structural breaks, when there are extreme observations in some of the cross-section units, when the process displays causal dependence in the conditional variance and when the analysis involves qualitative data. We illustrate the usefulness of our proposal with the analysis of the dynamic causal relationships between public expenditure and GDP, between firm productivity and firm size in US manufacturing sectors, and among sovereign credit rating, growth and interest rates.

Keywords: Transfer entropy test, Longitudinal dynamic data, Causality test.
JEL classification: C12, C14, C33, C55

*The authors acknowledge the financial support from MINECO projects ECO2016-76178-P and ECO2015-65637-P, which are co-financed by FEDER funds. This study is part of the collaborative activities carried out under the program Groups of Excellence of the Region of Murcia, the Fundación Seneca, Science and Technology Agency of the Region of Murcia Project 19884/GERM/15. All remaining errors are our responsibility. Data and codes that replicate our results are available from the authors' websites.

†Métodos Cuantitativos para la Economía y la Empresa, Universidad de Murcia, 30100 Murcia, Spain. Email: mcamacho@um.es

‡Fundamentos del Análisis Economico, Universidad de Murcia, 30100 Murcia, Spain. Email: aromeu@um.es

§Métodos Cuantitativos e Informáticos, Universidad Politécnica de Cartagena, 30201 Cartagena, Spain. Email: manuel.ruiz@upct.es

1 Introduction

Testing whether or not an economic variable x causes another y is a central topic in economics.[1] This is particularly important for understanding and interpreting dynamic economic phenomena such as those relating monetary and fiscal policies with output, energy consumption with growth, demand or supply shocks with economic fluctuations, oil prices with speculation, or firms' expenditure in R&D with profits, among many others. Following the lines initiated by Granger (1969), the causality analysis was originally designated in terms of changes in joint distributions. However, in practice, economists focus on assessing causality in time series contexts in terms of incremental predictability through block exogeneity tests. Then, examining causality has been reduced to determine whether the shocks in the time series x_t lead the time series y_t and help to predict it significantly.

Perhaps due to the traditional data scarcity, testing Granger noncausality of x_t and y_t has frequently relied on data of cross-sectional units i (individual, firm, or country), which restricts considerably the analysis. However, thanks to the recent advances in the diffusion of information technology, most data sets tend to have both cross-sectional and time series features, being the number of cross sections frequently larger than the time series dimension. Some examples are the data provided by World Bank Open Data with free and open access to global development data, the economic time-series based on national accounts collected in the Penn World Table, or the data on prices, employment, working conditions, and productivity provided by the US Bureau of Labor Statistics. Thus, performing causality tests on only one cross-sectional unit is unnecessarily partial or local nowadays.

Several causality tests have been proposed in the context of pooling time series data from different units. Being easy to implement, the traditional approach follows the lines initiated by Holtz-Eakin, Newey and Rosen (1988) and assumes homogeneity across cross-section units, in the sense that either causality occurs everywhere or it occurs nowhere in the panel. To overcome this unrealistic assumption, Dumitrescu and Hurlin (2012) proposed a causality test for heterogeneous panel data models whose test statistic is based on averaging the individual Wald statistics of Granger non causality across the cross-section units. In this context, the test handles with different lag orders in the autoregressive processes across the cross-section units and with unbalanced panels.[2]

Common to all these extensions of Granger causality tests to panel data is that a linear, basically autoregressive representation of the time series is required. By means of Monte Carlo simulations, we show that the size and power of these tests are seriously deteriorated when the linearity assumption breaks down, when the data generating process is heterogeneous across the cross-section units, when there are structural breaks or extreme observations in some of the cross-section units, when the causal dependence appears in the conditional variance or when it

[1] Although this paper focuses on economic applications, there has been recently a growing interest in the use of Granger causality to identify causal interactions in several other fields such as neuroscience and neuroimaging (Seth, Barrett, and Barnett, 2015).

[2] Hurlin and Venet (2001) assume homogeneous autoregressive coefficients but they allow for heterogeneity in the regression coefficients slopes. Weinhold (1996) and Nair-Reichert and Weinhold (2001) have considered testing for causality in a panel using random coefficient models.

involves qualitative data. Unfortunately, these scenarios are the norm rather than the exception when the number of cross-section units increases in the panel, as it is the case of most of the data sets provided by institutions such as the World Bank Open Data, the Penn World Table and the US Bureau of Labor Statistics.

To overcome this drawback, we extend the causality test described in Matilla-Garcia, Ruiz-Marin and Dore (2014) to a longitudinal data setup by translating the information set of the time series dynamics into symbols by means of a simple symbolization technique. More concretely, we study all symbols obtained by categorizing/symbolizing $m \geq 1$ consecutive values of the individual time series (namely m-history), which are considered here as ordinal patterns of the m numbers in the case of quantitative data, and as different labels in the case of qualitative data. Under the non-restrictive assumption that the symbolization procedure respects the null hypothesis of non-causality across the cross-unit sections, one can simply stack the symbols across the different cross units on the top of one another.

Then, we use the symbolic transfer entropy associated with the pooled set of symbols to construct a causality test for longitudinal data in which we check whether the stacked symbols of r-lagged series x_{t-r} reduce the entropy (which is a measure of the degree of disorder) of the stacked symbols of y_t conditional to those of y_{t-r}. In addition, we use the concept of net transfer entropy to test directional causality, i.e., whether the stacked symbols of past x_t cause those of current y_t or the other way around, whether the stacked symbols of past y_t cause those of current x_t. Due to the difficult to provide a distribution for the symbolic transfer entropy and the symbolic net transfer entropy measures, we rely on stationary bootstrap methods to compute significance of our causality tests statistics.

By construction, the test is non-parametric and relies on mild assumptions. This feature is reflected in the Monte Carlo analyses provided, suggesting that the test displays correct size and higher power in those cases where causality tests based in linear panel data specifications fail. The reason is that our test avoids the need to rely on a linear parametric representation of the data set which, if the model is misspecified or not consistently estimated, might invalidate the results of those tests based on linear panel data representations. Notably, our causality test for longitudinal data performs well even when the data generating process varies significantly across the cross-sectional units.

In order to illustrate the usefulness of the symbolic transfer entropy test with longitudinal data in practice and how its results can differ from other approaches, we provide three empirical examples. In the first example, that uses annual data collected by the World Bank Open Data from 1961 to 2016 for a set of 100 countries, we detect low evidence of causal relationships between the rate of growth of government expenditures and GDP growth. In the second example, using four-digit data provided by the US Bureau of Labor Statistics Current Employment Statistics survey of 86 industries from 1988 to 2015, we conclude that there is two-way causality between firm size and productivity. In this case, we find that the causality from firm productivity to firm size dominates, although in a non-significant way.

The third example illustrates the additional ability of our transfer entropy test over standard proposals: with a subtle modification, the test can easily be used to examine causality

relationships on panels that include qualitative time series that refer to categorical grades. In particular, we examine the interactions between Fitch's sovereign rating changes, the growth rate of per capita GDP and interest rates from a panel of countries from 1994 to 2012. We fail to detect significant causal relationships across these variables in the short term, which calls into question the role sometimes attributed to rating agencies in the amplification of the business cycles.

The rest of the article is structured as follows. In Section 2, we describe the extensions of Granger causality tests to linear panel data, introduce some notation and present the symbolic transfer entropy test for longitudinal data. In Section 3, we study the size and power of our proposal by means of several Monte Carlo simulations that capture some of the data problems that are common to large panels applications. In Section 4, we illustrate the usefulness of the proposal by using three applications with real data from macro and micro panel data samples. Section 5 concludes.

2 Granger causality for pooled data

2.1 Linear approaches

Let $\{y_{i,t}\}_{t\in I_i}$ and $\{x_{i,t}\}_{t\in I_i}$ be two real valued stationary time series, with t being the time index taking values in the set $I_i = \{1,\ldots,T_i\}$ of cardinality T_i, where i is the cross-section unit, with $i = 1,\ldots,N$. Testing causality usually involves the regression of a fixed coefficient model with fixed individual effects

$$y_{i,t} = \alpha_i + \sum_{k=1}^{K} \gamma_{ik} y_{i,t-k} + \sum_{k=1}^{K} \beta_{ik} x_{i,t-k} + \epsilon_{i,t}, \tag{1}$$

where individual errors $\epsilon_{i,t}$ are independently and normally distributed with mean 0, variance σ_i^2, and are independently distributed across cross-section units.

While not claiming to be exhaustive, there are two significant contributions in the context of Granger causality for panel data. Holtz-Eakin et al. (1988) assume that the unique heterogeneity between cross-section units is due to the fix effect α_i, which implies $\gamma_{ik} = \gamma_k$ and $\beta_{ik} = \beta_k$. Then, the model is estimated by applying instrumental variables to the quasi-differenced autoregressive equations to avoid endogeneity problems. In this context, the procedure devoted to determine the existence of causality is to test the null of absence of causality for all individuals in the panel, ($\beta_k = 0$, for all $k = 1,2,\ldots,K$) against the alternative that N causality relationships exist (there exists $1 \leq k_0 \leq K$ such that $\beta_{k_0} \neq 0$). The test statistic is computed by comparing restricted and unrestricted sum of squared residuals, which has an asymptotic chi-square distribution.

In a less restrictive context, Dumitrescu and Hurlin (2012) allow the coefficients to differ across cross-section units, different lag orders, unbalanced panels and cross-sectional dependence. In particular, under the non-causality null hypothesis, there is no causal relationships for any of the units of the panel ($\beta_{ik} = 0$, for all $k = 1,2,\ldots,K$, and $i = 1,2,\ldots,N$). Under

the alternative, there is at least one causal relationship in one cross-section unit ($\beta_{ik} = 0$, for all $k = 1, 2, \ldots, K$, and $i = 1, \ldots, n_1$; and there exists $1 \leq k_0 \leq K$ such that $\beta_{ik_0} \neq 0$, for all $i = N_1 + 1, \ldots, N$; with $N_1 \in [0, N-1]$), which implies that rejecting the null does not exclude that there is no causality for some N_1 individuals.

These authors propose running the N individual regressions implicitly enclosed in (1). Then, for each i, the proposal focuses on performing the F-tests of the K linear hypotheses $\beta_{ik} = 0$ to retrieve the Wald statistics W_i, and computing the average of the N individual Wald statistics

$$W = \frac{1}{N} \sum_{i=1}^{N} W_i. \tag{2}$$

Under the assumption that the Wald statistics W_i are independently and identically distributed across individuals, the average statistic sequentially converges in distribution to a standard normal distribution

$$Z = \sqrt{\frac{N}{2K}} \left(W - K\right) \longrightarrow N\left(0, 1\right), \tag{3}$$

where the asymptotic result requires that the time dimension goes to infinity and then the cross section dimension goes to infinity.[3]

When both the time dimension and the cross section dimension are fixed, the authors propose using the mean Wald statistic stated in (2) and the empirical critical values for the corresponding sizes via stochastic simulations. In addition, under cross-sectional dependence, they mitigate the negative impact on the small-sample properties of the test for alternative forms of dependence by using bootstrapped critical values. In Section 2.3, we describe a block-bootstrap procedure that calculates the bootstrapped critical regions at the desired confidence levels.

2.2 Symbolic analysis: definitions and notation

The basic idea of symbolic dynamics is very simple: divide the phase space into a finite number of regions and label each region by a letter (a symbol) from a certain alphabet (a set of symbols). Instead of following a trajectory point by point, one only keeps recording the alternation of certain appropriate symbols, which, according to the results of Collet and Eckmann (2009), can capture the complete description of the behavior of the dynamical system. In the particular case of this paper, the basic block of analysis is the ordinal pattern because it only demands a totally ordered set to be fully applicable.[4]

We start the notation by defining the symbolization procedure of the single real valued time series $\{y_{i,t}\}$ and $\{x_{i,t}\}$. For a positive integer $m \geq 2$, usually known as the embedding dimension,

[3]Since the time dimension can be small in practice, the authors also provide an approximated standardized statistics for a fixed T.

[4]To facilitate understanding, the Appendix shows several examples that help the reader unfamiliar with symbolic dynamics to understand the basic concepts and definitions described in this section. An example of symbolization based on a non ordinal pattern is analyzed in Section 4.3.

we embed the time series in m-dimensional space as follows:

$$\mathbf{x}_{im}(t) = (x_{i,t}, x_{i,t+1}, \ldots, x_{i,t+m-1}), \tag{4}$$

$$\mathbf{y}_{im}(t) = (y_{i,t}, y_{i,t+1}, \ldots, y_{i,t+m-1}). \tag{5}$$

The m-dimensional histories $\mathbf{y}_{im}(t)$ and $\mathbf{x}_{im}(t)$ summarize the behavior of the time series in the neighborhood of time t, accounting for m subsequent steps. For time series of length T_i only $T_i^* = T_i - m + 1$ m-histories $\mathbf{y}_{im}(t)$ and $\mathbf{x}_{im}(t)$ can be obtained, where the total number of m-histories is $T^* = T_1^* + \ldots + T_N^*$.

Next, we denote by $\mathbf{S}_m$ the symmetric group of cardinality $m!$, which is the group formed by all the permutations of length m. We refer to an element $\mathbf{s} = (s_1, s_2, \ldots, s_m)$ in the group $\mathbf{S}_m$ as a "symbol", where $s_j \in \{0, 1, \ldots, m-1\}$ with $s_j \neq s_h$ for all $j \neq h$. The symbolization procedure of the time series $\{y_{i,t}\}$ consists of mapping each m-history in $\{\mathbf{y}_{im}(t)\}_{t \leq T_i^*}$, for $i = 1, \ldots, N$, into a unique symbol $\mathbf{s} \in \mathbf{S}_m$ satisfying the two following conditions:

$$y_{i,t+s_1} \leq y_{i,t+s_2} \leq \cdots \leq y_{i,t+s_m}, \tag{6}$$

$$s_{j-1} < s_j \text{ if } y_{i,t+s_{j-1}} = y_{i,t+s_j}. \tag{7}$$

In this case, we will say that $\mathbf{y}_{im}(t)$ is of $\mathbf{s}$-type. The first condition imposes the ordinal pattern and the second condition guarantees the uniqueness of the symbol $\mathbf{s}$ for any time series. The latter condition becomes irrelevant when the time series is associated with a continuous distribution, such that equal values have a theoretical probability of occurrence of zero. The symbolization procedure of the time series $\{x_{i,t}\}$ is obtained analogously.

The symbolization procedure of $\{y_{i,t}\}$ and $\{x_{i,t}\}$ implies converting the sequences of m-histories $\{\mathbf{y}_{im}(t)\}_{t \leq T_i^*}$ and $\{\mathbf{x}_{im}(t)\}_{t \leq T_i^*}$ into sequences of ordinal patterns labeled by symbols, $\{\zeta_{im}^y(t)\}_{t \leq T_i^*}$ and $\{\zeta_{im}^x(t)\}_{t \leq T_i^*}$, respectively, where $\mathbf{y}_{im}(t)$ is of $\zeta_{im}^y(t)$-type and $\mathbf{x}_{im}(t)$ is of $\zeta_{im}^x(t)$-type.

Given the time series $\{y_{i,t}\}$ and $\{x_{i,t}\}$ and an embedding dimension m, the probabilities of the symbols in $\mathbf{S}_m$ are computed as their relative frequencies in $\{\zeta_{im}^y(t)\}_{t \leq T_i^*}$ and $\{\zeta_{im}^x(t)\}_{t \leq T_i^*}$, respectively. For example, the estimated probability of any symbol $\mathbf{s} \in \mathbf{S}_m$ for $\{y_{i,t}\}$ is

$$p_{im}^y(\mathbf{s}) = \frac{\sharp \{t = 1, \ldots, T_i^* \mid \mathbf{y}_{im}(t) \text{ is of } \mathbf{s}\text{-type}\}}{T_i^*}, \tag{8}$$

where $\sharp$ denotes the cardinality. Following Matilla-Garcia and Ruiz-Marin (2008), we define the permutation entropy of $\{y_{i,t}\}$ as the Shannon entropy (Shannon, 1948) of the $m!$ distinct symbols as

$$h_m(y_{i,t}) = -\sum_{\mathbf{s} \in \mathbf{S}_m} p_{im}^y(\mathbf{s}) \ln(p_{im}^y(\mathbf{s})), \tag{9}$$

This quantity has a nice interpretation. The amount of information of every m-history forms a random variable whose expected value, on average, is the Shannon entropy, which measures the degree of uncertainty of the time series. It is always non-negative by construction and is bounded by $0 \leq h_m(y_{it}) \leq \ln(m!)$. The lower bound is attained for an increasing or decreasing

sequence of values, such that only one symbol is sufficient to encode the ordinal pattern. The upper bound refers to a completely random time series, where all the $m!$ possible symbols appear with the same probability.

Toward the analysis of causal relationships, we require a multivariate extension of this symbolization procedure. In order to make it simpler, we consider the extension to a 2-dimensional time series, although the case of n-dimensional time series implies a similar reasoning. For a cross-section unit i, consider a 2-dimensional time series $\{y_{i,t}, x_{i,t}\}_{t\in I_i}$. Assuming a fixed embedding dimension m, we embed the vector of bivariate time series into m-dimensional sequences $\{\mathbf{y}_{im}(t), \mathbf{x}_{im}(t)\}_{t\leq T_i^*}$.

In this setting consider the symbol space $\mathbf{S}_m^2 = \mathbf{S}_m \times \mathbf{S}_m$, which is the direct product of two copies of the symmetric group $\mathbf{S}_m$, whose elements are denoted by $\mathbf{s}_{yx} = (\mathbf{s}_y, \mathbf{s}_x)$. Then we say that $\{\mathbf{y}_{im}(t), \mathbf{x}_{im}(t)\}$ is of $\mathbf{s}_{yx}$-type if and only if $\mathbf{y}_{im}(t)$ is of $\mathbf{s}_y$-type and $\mathbf{x}_{im}(t)$ is of $\mathbf{s}_x$-type, forming the bi-dimensional sequence of symbols $\{\zeta_{im}^{yx}(t)\}$. Then, the estimated probabilities of each symbol $\mathbf{s}_{yx} \in \mathbf{S}_m^2$ is estimated as

$$p_{im}^{yx}(\mathbf{s}_{yx}) = \frac{\sharp\,\{t = 1, \ldots, T_i^* \mid (\mathbf{y}_{im}(t), \mathbf{x}_{im}(t)) \text{ is of } \mathbf{s}_{yx}\text{-type}\}}{T_i^*}. \tag{10}$$

The Shannon entropy for the bi-dimensional series $\{y_{it}, x_{it}\}$ is

$$h_m(y_{it}, x_{it}) = -\sum_{\mathbf{s}_{yx}\in\mathbf{S}_m^2} p_{im}^{yx}(\mathbf{s}_{yx}) \ln(p_{im}^{yx}(\mathbf{s}_{yx})), \tag{11}$$

which measures the amount of information common to both time series. One key concept to analyze causality in multivariate time series is the conditional permutation entropy of $\{y_{i,t}\}$ given $\{x_{i,t}\}$. By definition, it is the result of averaging the conditional entropy of $\{y_{i,t}\}$ over all possible values that $x_{i,t}$ may take. It can be shown that the following chain rule applies

$$h_m(y_{i,t}|x_{i,t}) = h_m(y_{i,t}, x_{i,t}) - h_m(x_{i,t}). \tag{12}$$

The conditional entropy is zero when $\{y_{i,t}\}$ is completely determined by $\{x_{i,t}\}$. Conversely, the conditional entropy is equal to the unconditional entropy of $\{y_{i,t}\}$ when they are independent.

The generalization of this procedure to obtain any conditional entropy is straightforward. For example, for any lag time period r, computing

$$h_m(y_{i,t}|y_{i,t-r}, x_{i,t-r}) = h_m(y_{i,t}, y_{i,t-r}, x_{i,t-r}) - h_m(y_{i,t-r}, x_{i,t-r}), \tag{13}$$

involves the following steps:

1. Map each element of $\{y_{i,t}, y_{i,t-r}, x_{i,t-r}\}$ and $\{y_{i,t-r}, x_{i,t-r}\}$ into the symbol spaces $S_m^3 = S_m \times S_m \times S_m$ and $S_m^2 = S_m \times S_m$, respectively. This leads to the sequences of sample symbols $\{\zeta_{i,m}^{y,y_r,x_r}(t)\}$ and $\{\zeta_{i,m}^{y_r,x_r}(t)\}$.
2. Estimate the relative frequencies of each of the different symbols in the symbol spaces S_m^3 and S_m^2 by using the sequences of sample symbols $\{\zeta_{i,m}^{y,y_r,x_r}(t)\}$ and $\{\zeta_{i,m}^{y_r,x_r}(t)\}$, respectively.

3. Use the estimated probabilities to compute the Shannon entropy measures $h_m(y_{i,t}, y_{i,t-r}, x_{i,t-r})$ and $h_m(y_{i,t-r}, x_{i,t-r})$, and the conditional entropy as expression (13) describes.

2.3 A transfer entropy test of longitudinal data

The extension of causality tests to longitudinal data implies pooling the information contained in the cross-sectional units.[5] Using the approach of stacking the raw data of the time series on top of one another involves managing blocks of stacked data. In particular, examining whether values of x delayed r periods help in predicting actual values of y or whether values of y delayed r periods help in predicting actual values of x requires handling with stacked vectors such as

$$\begin{aligned}
\{y_t\} &= \{y_{1,r}, ..., y_{1,T_i}, ..., y_{N,r}, ..., y_{N,T_i}\},\\
\{x_t\} &= \{x_{1,r}, ..., x_{1,T_i}, ..., x_{N,r}, ..., x_{N,T_i}\},\\
\{y_{t-r}\} &= \{y_{1,1}, ..., y_{1,T_i-r}, ..., y_{N,1}, ..., y_{N,T_i-r}\},\\
\{x_{t-r}\} &= \{x_{1,1}, ..., x_{1,T_i-r}, ..., x_{N,1}, ..., x_{N,T_i-r}\}.
\end{aligned} \tag{14}$$

Performing block exogeneity tests in the resulting pooled data would impose the restrictive constraint that the time series relationship of x and y is the same for each cross-section units. Undoubtedly, the constraint is likely to be violated in practice. This motivated Holtz-Eakin et al. (1988) to allow for individual fixed effects and Dumitrescu and Hurlin (2012) to allow for heterogeneous panel data models.

Instead, in this paper we propose pooling the corresponding set of symbols of the cross-section units in order to examine the informational content of the pooled symbols of $\{x_{t-r}\}$ on the pooled conditional entropy $h_m(y_t|y_{t-r}, x_{t-r})$. Computing this entropy would require managing sequences of stacked symbols such as

$$\begin{aligned}
\{\zeta_m^{y,y_r,x_r}(t)\} &= \{(\zeta_{1,m}^{y}(r+1), \zeta_{1,m}^{y_r}(1), \zeta_{1,m}^{x_r}(1)), \ldots, (\zeta_{N,m}^{y}(T_N^*), \zeta_{N,m}^{y_r}(T_N^*-r), \zeta_{N,m}^{x_r}(T_N^*-r))\},\\
\{\zeta_m^{y_r,x_r}(t)\} &= \{(\zeta_{1,m}^{y_r}(1), \zeta_{1,m}^{x_r}(1)), \ldots, (\zeta_{N,m}^{y_r}(T_N^*-r), \zeta_{1,m}^{x_r}(T_N^*-r))\}.
\end{aligned} \tag{15}$$

Needless is to say that the method requires the conversion of the m-histories into the sequences of symbols for each cross-section unit to be stable across units. However, as we will show in the section devoted to the Monte Carlo analysis, this is far from being a restrictive assumption.

Computing the pooled conditional entropy for the pool of symbols is very easy in practice because it can be obtained by adding the conditional entropy of each cross-section unit. In the previous example, the pooled conditional entropy, which is used to measure the potential gains of using lagged values of x to predict actual values of y, is the sum of the conditional entropies of the N cross-section units

$$h_m(y_t|y_{t-r}, x_{t-r}) = \sum_{i=1}^{N} h_m(y_{i,t}|y_{i,t-r}, x_{i,t-r}). \tag{16}$$

[5]To keep the analysis simple, we limit this section to two-dimensional systems. Extending these concepts to larger sets of time series is straightforward.

To develop a causality test for the pooled data, we follow Schreiber (2000) to introduce the Symbolic Transfer Entropy (STE), which measures the information transfer from $\{x_{t-r}\}$ to $\{y_t\}$ given $\{y_{t-r}\}$

$$\mathrm{STE}_{x\to y}(m,r) = h_m(y_t|y_{t-r}) - h_m(y_t|y_{t-r}, x_{t-r}). \tag{17}$$

This quantity can be viewed as the reduction in uncertainty in predicting the current state of y_t (more precisely, the current symbol $\zeta_m^y(t)$), by knowing those of y_{t-r} and x_{t-r}. A large transfer entropy from x to y indicates that the past values of x help to predict current values of y, whereas a small transfer entropy indicates that the current value of y is independent of the past value of x. Thus, the transfer entropy offers a measure of Granger causality of x to y.

Under the null that x does not cause y, it follows that $\mathrm{STE}_{x\to y}(m,r) = 0$, while $\mathrm{STE}_{x\to y}(m,r) > 0$ otherwise. This is the basis to propose the following causality test procedure for longitudinal data. Providing an exact or an asymptotic distribution of the statistic $\mathrm{STE}_{x\to y}(m,r)$ is certainly difficult because the dynamic structure of the processes y_t and x_t may change the symbols' distribution and thus the distribution of the entropy statistic. For this reason, we rely on bootstrap methods, which require that null hypothesis of no causality remains in the re-sampled data to ensure that the test remains asymptotically independent of these data. For this purpose, we re-sample for each cross section unit the time series $\{y_{i,t}\}$ and $\{x_{i,t}\}$ independently.[6]

In particular, we rely on the stationary bootstrap of Politis and White (2004) to preserve the dynamic structure of the time series in the re-sampled processes. Then, the bootstrap test procedure consists on replicating the following steps a large number of times B:

1. Compute the value of $\mathrm{STE}_{x\to y}(m,r)$ for the original pooled time series $\{y_t\}$ and $\{x_t\}$.
2. Sample $\{x_{i,t}^b\}$ and $\{y_{i,t}^b\}$ for $i = 1, 2, \ldots, N$, by using the stationary bootstrap of Politis and White (2004). For each cross-section unit, obtain the sequences of symbols $\{\zeta_{m,b}^{y,y_r,x_r}(t)\}$, $\{\zeta_{m,b}^{y_r,x_r}(t)\}$, $\{\zeta_{m,b}^{y,y_r}(t)\}$, $\{\zeta_{m,b}^{y_r}(t)\}$, compute the relative frequencies of the symbols, and obtain the entropy measures $h_m^b(y_t|y_{t-r})$ and $h_m^b(y_t|y_{t-r}, x_{t-r})$.
3. Use the latter sets of pooled symbols to obtain the bootstrapped realization of the symbolic transfer entropy

$$\mathrm{STE}_{x\to y}^b(m,r) = h_m^b(y_t|y_{t-r}) - h_m^b(y_t|y_{t-r}, x_{t-r}). \tag{18}$$

4. Repeat $(B-1)$ times steps 2 and 3 to obtain B bootstrap realizations of the statistic, $\{\mathrm{STE}_{x\to y}^b(m,r)\}_{b=1}^B$.
5. Compute the bootstrap p_b-value

$$p_b = \frac{1}{B}\sum_{b=1}^{B} I(\mathrm{STE}_{x\to y}^b(m,r) > \mathrm{STE}_{x\to y}(m,r)), \tag{19}$$

   where $I(\cdot)$ is an indicator function taking the value 1 to a true statement and 0 otherwise.
6. Reject the null hypothesis that x does not cause y at a nominal level α if $p_b < \alpha$.

[6]Notice that pairwise re-sampling could preserve the underlying causality in the bootstrapped data.

Testing the null that y does not cause x is performed in the same fashion.

A second problem in causality analysis is determining the direction of causality. Following Caines, Keng and Sethi (1981), there are four possible situations in analysis of the causal directions between x and y. The first case, the one-way causality of x to y, occurs when the series x cause y but the series y does not cause x, which requires $\text{STE}_{y\to x}(m,r) = 0$ and $\text{STE}_{x\to y}(m,r) > 0$. The second case, when there is one-way causality of y to x, y cause x but x does not cause y, and implies $\text{STE}_{y\to x}(m,r) > 0$ and $\text{STE}_{x\to y}(m,r) = 0$. In the third case x and y are independent, which implies $\text{STE}_{y\to x}(m,r) = \text{STE}_{x\to y}(m,r) = 0$. The last case refers to two-way causality, which occurs when $\text{STE}_{y\to x}(m,r) = \text{STE}_{x\to y}$ but $\text{STE}_{y\to x}(m,r) > 0$ and $\text{STE}_{x\to y}(m,r) > 0$.

In the case of two-way causal relation between x and y, we may easily assess the direction of the dominant information transfer between the time series. To this end, we introduce the Net permutation Transfer Entropy (NTE) as follows:

$$\text{NTE}_{xy}(m,r) = \text{STE}_{y\to x}(m,r) - \text{STE}_{x\to y}(m,r). \tag{20}$$

In this context, a positive value of $\text{NTE}_{xy}(m,r)$ means that the net entropy transfer is from y to x while a negative value of this statistic suggests the transfer is the other way around. When $\text{NTE}_{xy}(m,r)$ is zero, it indicates that either y and x are independent, which occurs whenever $\text{STE}_{y\to x}(m,r) = \text{STE}_{x\to y}(m,r) = 0$, or that there is not a predominant direction of causality, which requires $\text{STE}_{y\to x}(m,r) > 0$ and $\text{STE}_{x\to y}(m,r) > 0$.

Using the steps described above to compute the bootstrap symbolic causality test, testing the direction of causality is straightforward. In the first step, $\text{NTE}_{xy}(m,r)$ is computed as the difference between $\text{STE}_{y\to x}(m,r)$ and $\text{STE}_{x\to y}(m,r)$. Repeating steps 2 to 4 provides a number B of bootstrap net transfer entropy statistics $\text{NTE}^b_{xy}(m,r)$. The p-value of the two-tailed test of the null that $\text{NTE}_{xy}(m,r) = 0$ against $\text{NTE}_{xy}(m,r) \neq 0$ can be obtained by computing the averaged number of times that $|\text{NTE}^b_{xy}(m,r)| > |\text{NTE}_{xy}(m,r)|$. In addition, the p-value of the one-tailed test of the null that $\text{NTE}_{xy}(m,r) = 0$ against $\text{NTE}_{xy}(m,r) > 0$ can be obtained by computing the averaged number of times that $\text{NTE}^b_{xy}(m,r) > \text{NTE}_{xy}(m,r)$, whose rejection indicates that the predominant direction of causality is from y to x.[7]

3 Simulation analysis

In this section, we propose several Monte Carlo experiments in order to assess how the Symbolic Transfer Entropy test for causality described above performs in a finite sample of simulated longitudinal data. In all cases, we take the embedding dimension $m = 3$. As a basis of comparison, we also examine the performance of Holtz-Eakin et al. (1988) and Dumitrescu and Hurlin (2012) causality tests in the context of panel data analysis. These two proposals are referred to as HNR and DH henceforth, respectively.

[7]The one-tailed test whose rejection indicates that the predominant direction of causality is from x to y can be obtained in the same fashion.

3.1 Benchmark model

We start the experiment by generating data for $y_{i,t}$ by the homogeneous linear model outlined in (1), which is labeled as HLIN. In each of the $M = 1000$ Monte Carlo simulations, we simulate N vectors of explanatory variables x_i and individual errors ϵ_i of constant size T from independent Gaussian distributions with mean 0 and variances $\sigma^2_{\epsilon i} = \sigma^2_{xi} = 1$, for all $i = 1, \ldots, N$, being the total number of observations TN. With the aim of evaluating the effect of the number of cross-section units N and the amount of time observation points T on the empirical size and power of the causality tests, the experiment is carried out for values of (T, N) that correspond to small $(15, 30)$, medium $(30, 60)$ and large $(60, 120)$ data sets.

Without loss of generality, the individual fixed effects α_i are set to zero and the effect of heterogeneity across section units will be examined through the causality relationships, measured by β_i. One implication of disregarding individual effects in the data generating process of the experiment is that the HNR estimator reduces to a simple OLS/GLS estimate because the procedure does not require first differencing. Nevertheless, we keep the HNR labeling in the presentation of the results in this section to harmonize the notation with Sections 2 and 4.

For each pseudo sample, we generate the data by choosing $y_{i0} = 0$. The lag length of the autoregressive processes, K, and the lag in the causal relationship, r, are set to one. For each sample size, the autoregressive parameters γ_i are constant for all cross-section units, although to assess the effect of the autocorrelation on the tests, its value is allowed to vary from none $\gamma = 0$, to moderate $\gamma = 0.3$ and large $\gamma = 0.9$ autocorrelations.

In this baseline scenario, we keep the lagged dependent variable parameter β fixed across units. In order to visualize its influence on the power of the tests, we rely on the power curves that approximate the probability of rejecting the null of non-causality under the alternative of causality as a function of β. For this purpose, we adopt the response surface strategy developed by Box and Wilson (1951). For each run, we draw β from a uniform distribution on $[0, 1]$, then we compute the test statistics and record the outcome with a binary variable that takes the value of one if the null of non-causality is rejected for a significance level of $\alpha = 0.05$. Finally, we plot the number of times that the tests reject the null over the total number of simulations using a kernel-based estimate of the power curve.[8]

The reason to consider HLIN as the baseline model hangs upon the prospect that HNR should behave optimally because all the assumptions that it requires for consistent and efficient estimation are fulfilled within this specification. Namely, parameters are homogeneous across section units, the functional form is linear, the causal relationships are in the means of the process, and the disturbances are spherical. Then, the aim is to assess the cost of using DH or STE even when the conditions would favor the use of HNR.

Figure 1 plots the power functions of HNR test, DH test and STE test against β for the three different combinations of cross-unit and time-series sample sizes when $\gamma = 0$ (Panel A), $\gamma = 0.3$ (Panel B) and $\gamma = 0.9$ (Panel C) at significance level $\alpha = 0.05$. As expected, the figure shows that HNR clearly outperforms the other two tests regardless of the size of N and T, the

[8]The power function measures the test size at the null $\beta = 0$ and approaches 1 as β meets the upper value. To overcome low performance of the Gaussian kernel-based estimates at these endpoints, we adjusted the window width.

values of the autocorrelation parameter γ, and the magnitude of the causality parameter β. The superior performance is particularly evident when the size of the data set is small. In addition, DH performs better than STE, although the differences in performance tend to decrease as sample size increases.

3.2 Data problems

The rest of our Monte Carlo experiments are designed to analyze the size and power performance of the symbolic transfer entropy causality test for longitudinal data in the presence of large cross-section heterogeneity, non-linearity, structural breaks, outliers, and second-moment causality. In practice, these data problems are the norm rather than the exception, particularly when the number of cross-section units in the panel is high, such as in the case of The World Bank Open Data, the Penn World Table or the US Bureau of Labor Statistics.

The first set of experiments allows for heterogeneity of the causality relationships under the alternative. In particular, we consider a structural break in the impact of x_{t-1} on y_t, in the sense that in each simulation, half of the cross-section units are generated from β while the other half use $-\beta$, with β drawn from a uniform distribution on $[0, 2]$. To facilitate comparisons, the rest of the model parameters are calibrated as in the benchmark experiment. The achieved rejection frequencies for the three different tests as a function of β, for a theoretical level of significance of 0.05, are plotted in Figure 2. Regardless of the sample size and auto correlation parameter values, the plots of the power functions show that HNR and DH tests suffers a dramatic loss in power. Notably, the performance of HNR and DH dramatically deteriorates with the magnitude of the cross-section structural breaks, measured by the magnitude of β, while the performance of STE in terms of power is not much different from the benchmark setup.

Central to the benchmark model is the assumption that the causal relationship between variables is linear in mean. However, there are many evidences of nonlinear causal relationships between variables in economics (for example, Hsieh, 1991). To examine the effect of nonlinearities in the causality tests, in the third experiment we generate data as

$$y_{i,t} = y_{i,t-1}x_{i,t-1} + \epsilon_{i,t}, \tag{21}$$

where $\beta = 1$, $\epsilon_{i,t} \sim N(0, 1)$, and $x_{i,t} \sim N(0, 1)$. By recursive substitution, we can find that $E(y_{i,t} \mid x_{i,t-1} \ldots, x_0) = y_{i,0} \prod_{s=1}^{t} x_{i,t-s}$. Therefore, the partial effect of $x_{i,t-1}$ is given by $y_{i,0} \prod_{s=2}^{t} x_{i,t-s}$, for $i = 1, \ldots, N$. If x has average value of zero across sections, the estimate of β in (1) tends to converge to zero although $x_{i,t-1}$ indeed causes $y_{i,t}$. Then, HNR is expected to miss the causal relationship between x and y, although this drawback does not necessarily affect DH and STE. In line with this reasoning, the performance of HNR in Figure 3 in terms of power is very poor and far below that of DH and STE. Remarkably, although STE and DH have comparative power for large sample sizes, STE clearly outperforms DH for small combinations (N,T).

Due to a growing interest in dynamics of financial data, recent work on causality has also addressed the issue of second order causality or causality in variance (Hafner and Herwartz,

2008). We design the fourth experiment to examine the ability of the causality tests to detect causal relationships in the second moment of the process. For this purpose, we generate data as

$$y_{i,t} = \gamma y_{i,t-1} + \epsilon_{i,t}, \tag{22}$$

where $\epsilon_{i,t} \sim N(0, |x_{i,t-1}|)$, and $x_{i,t} \sim N(0,1)$. Figure 4, which displays the estimated probability of rejection of the tests, shows the low performance of HNR when there is causality in variance but no causality in mean. When the sample size is very low, DH performs slightly better than STE. However, the performance of STE is superior to that of DH as the sample size increases beyond (T,N)=(30,15).

Finally, panel data estimators can be strongly biased in the presence of a small percentage of outliers (Bramati and Croux, 2007), which can have a large impact on causality analysis. In the fifth experiment, we investigate the robustness of the three causality tests in the presence of outliers in the longitudinal data. In this experiment, we pay attention to the case of block-concentrated outliers as the situation in which the outlying observations are concentrated in few time-series, which is very frequent in panels of economic data for which outliers are present for few cross-section units while others are not contaminated at all. For this purpose, we generate data as in the benchmark case but entering as outlying points the first observation of the first cross-section unit and the last observation of the last cross section unit. In particular, we impose $y_{12} = x_{11} = -10$ and $y_{NT} = x_{NT-1} = 10$ and we plot in Figure 5 the empirical size of the three causality tests. It is obvious that the distortion introduced by these two outliers in the sample greatly bias the results of the HNR and DH tests leading to an oversize problem that worsens as the sample size decreases. By contrast, the figure shows that STE is a robust test in the sense that its empirical size is barely affected by the presence of outliers in the data set.

4 Empirical applications

As a practical illustration of our approach, we now analyze the causal relationships between public expenditure and GDP, firm productivity and firm size, and sovereign credit rating and economic fundamentals. Unless otherwise stated below, we perform the symbolic transfer entropy tests for the embedding parameter $m = 3$.

4.1 Public Expenditure and GDP

The role of government spending in economic growth, and vice versa, has been a long-lasting source of debate in economics, and has recently generated intense literature since the Reinhart and Rogoff (2010) controversial contribution. Some economists argue that government spending cause economic growth, either in a negative way, because large government expenditure leads to public debt slowing growth, or in a positive way, because expansionary fiscal policies accelerate economic activity. By contrast, some others support the alternative that it is economic growth what causes public sector expenditures to expand as the budget constraints become less rigid or to reduce as a policy of countercyclical response. This section attempts to provide a new

contribution to this debate by examining the worldwide short-term causal relationship between government spending and economic growth.

To this purpose, we have collected data from the World Bank Open Data, a free-access service which processes a huge amount of macroeconomic data for a large set of countries around the world. In spite of the interest of examining the causal relationships between expenditure and growth in a world context, as the World Bank acknowledges, under-investment in national statistical systems in some developing countries results in data of poor quality for these countries, which are often affected by some of the data problems described below, resulting in low performance of standard causality tests for longitudinal data.[9] In this context, we believe that symbolic transfer entropy tests are the appropriate approaches to evaluate the causal relationships between government spending and growth across a large set of countries.

The data cover the period from 1961 to 2016 at annual frequency for a set of 100 countries. Regional areas and countries with less than one-third of observations were omitted from the analysis. The measure of aggregate economic activity is GDP per capita (Gross Domestic Product divided by midyear population).[10] The measure of government expenses is general government final consumption expenditure. Both aggregates are measured in constant 2010 US dollars and are used in growth rates to deal with unit root issues.

Table 1 presents the basic results of the non-causality tests. Each panel corresponds to a different lag length specification (from one to three years). In the first row, we examine causality from the growth rates of government expenses (Exp) to GDP growth, then GDP to Exp, and finally the net transfer entropy, i.e., the predominant direction of information transfer between these two macroeconomic aggregates. According to expression (20), a negative value of this net effect implies prevalence of Exp causing GDP growth while a positive value of the net effect implies prevalence of GDP growth over Exp. To evaluate the statistical significance of NTE, one and two tailed p-values are reported. Starting with HNR, the p-values reported in the table show a rejection of the null hypothesis of non-causality of Exp to GDP, except when we consider two lags and no causality from GDP to Exp. However, regardless of the lag length that we consider, the p-values for DH show strong causality in both directions, from GDP to Exp and from Exp to GDP.

Notably, our symbolic transfer entropy test suggests that causality between government expenditures and growth is much more limited. For lags of one and three years, the p-values of the null of noncausality are far above 0.05 regardless of whether we test causality from the growth rate of government expenses to GDP growth or the other way around, which supports the view that there is no causal effect between these two variables. If any, the entropy test only detects one-way causality from GDP growth to the rate of growth of government expenditures achieved with a lag of two years. This result complements those of De Vita, Trachanas and Luo (2018), who also find no robust evidence of a long-run causal effect between debt and growth.

[9]To (partially) overcome this drawback, the World Bank has been leading initiatives such as Open Data for Development (https://www.idrc.ca/en/initiative/open-data-development), a partnership of several institutions to promote the deployment of open data in developing countries.

[10]GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources.

4.2 Firm productivity and size

Gauging the firm-specific determinants of firm productivity is an important research issue in productivity and growth literature. Focusing on firm size, Diaz and Sanchez (2008) argue that firm size could tend to reduce firm productivity due to the increase in organizational and managerial complexity. However, larger firm size could also foster productivity growth because large size firms are more accessible to market and technology (Biesebroeck, 2005) and take advantage of all the increasing returns associated with R&D (Pagano and Schivardi, 2003). By contrast, it might be the case of reverse causality if, as in most of the empirical literature on cross-country growth, large productivity creates larger firms.

In this section, we use our symbolic transfer entropy test to examine the causal relationships between firm size and productivity growth in a longitudinal data set. For this purpose, we look at the industry-level annual data provided by the US Bureau of Labor Statistics Current Employment Statistics survey. To classify the data, we follow the four-digit industry format of the North American Industry Classification System (NAICS). The effective sample period is 1988 to 2015 and the data set comprises a total of 86 industries.

Productivity data refers to multifactor productivity, which is a measure that represents the amount of goods and services that can be produced relative to the amount of various measured inputs such as labor, capital, and intermediate purchases that are consumed or used to produce those goods and services. As a measure of firm size, we use the number of employees, which represents the total number of wage and salary workers, self-employed workers, and unpaid family workers working at various occupations (jobs) within business establishments. An individual who works multiple jobs at separate establishments would have each job included in the number of employees.

Both series appear to be non-stationary in levels but stationary in the first differences for logarithmic form. Since we are interested in examining the short-term dynamics, the causality tests, whose results are displayed in Table 2, are conducted by using the time series in growth rates. Our findings can be summarized in the following way. Both HNR and DH detect some short-term evidence of bi-directional causality between firm size and productivity, which appears only at a one-period lag. However, STE detects bi-directional causality regardless of the lag length that we consider. The negative differences between the pairwise instances of transfer entropy measures mean that the symbolic transfer entropy is from firm size to firm productivity, although the p-values of the null of symmetric information flow indicate that the net transfer is not statistically significant at any lag length.

4.3 Credit rating and economic fundamentals

The global credit rating agencies have come under intense scrutiny regarding their sovereign debt ratings in the wake of the Great Recession. The controversy on the role of rating agencies raised when public debt of crisis-hit countries were relegated to the lowest status and the agencies also downgraded the credit worthiness of many solvent countries, mainly major Eurozone economies,

which could have accelerated the Eurozone sovereign debt crisis.[11]

In this context, a line of research focuses on the influence of rating changes into economic aggregates and financial markets. While not claiming to be exhaustive, Kaminsky and Schmukler (2002) find that changes in rating have a significant impact on yield spreads and stock returns. According to Brooks et al. (2004), it is rating downgrades rather than rating upgrades that is driving a negative wealth impact on both the domestic stock market and the dollar value of the country's currency.

Besides, some studies focus on examining whether sovereign ratings respond to changes in economic and financial fundamentals. Among others, Broto and Molina (2016) show that ratings react to changes in GDP growth and US short-term rate, although their relevance as leading indicators for sovereign ratings changes depends on the underline specification used to obtain these results.

The analysis developed in this section complements these studies by examining the causal relationships between economic fundamentals and sovereign debt ratings. Moreover, this section illustrates the additional ability of our transfer entropy test over standard proposals to examine causality relationships on panels that can include qualitative time series that refer, for example, to categorical grades. With a slight modification, the techniques developed in this paper for ordinal patterns are still valid in this setup.

To achieve our objective, we collect the data on sovereign debt ratings of Fitch's foreign currency rating, which reflects the sovereign's ability and willingness to service its debt in foreign currency. Ratings grade countries' ability to meet its financial obligations according to a letter grade. The highest and safest grade is AAA, with lower grades moving to double and then single letters (AA or A) and down the alphabet from there to D, which implies entering into bankruptcy. The modifiers arithmetic symbols "+" and "-" denote relative status within major rating categories (for example, the rating category AA has three rating levels, AA+, AA, and AA-).

For sovereign debt ratings, we propose the following symbolization technique. First, we construct a country specific time series $z_{i,t}$ that takes the value of -1 if Fitch downgrades country i rating in year t, the value of 0 if its rating is unchanged, and the value of 1 if its rating is upgraded. Then, symbols are used in the analysis as categorical instead of as ordered variables. Now, we take the embedding dimension $m = 1$ to build a 1-dimensional history $\zeta^z_{im}(t)$ that reflect the annual changes in rating grades. With this modification, the symbolic transfer entropy test for longitudinal data described in the paper can easily be applied in this context.[12]

To examine the interactions between sovereign ratings and economic activity, we use the growth rate of per capita GDP, provided by the panel World Bank Open Data as in Section 4.1. The impact of sovereign ratings on the financial sector is examined by using real interest rate also provided by the panel World Bank Open Data, defined as the lending interest rate

[11]Ferri, Lui and Stiglitz (1999) find a procyclical behavior of rating agencies, which could magnify the effects of booms and busts.

[12]As a robustness check, we also performed the analysis using $m = 2$ and $m = 3$ with and without ordering pattern. The results and main conclusions did not imply a qualitative change with respect to those presented in the paper and are available from the authors upon request.

adjusted for inflation as measured by the GDP deflate. The GDP panel comprehends annual data of 99 countries from 1994 to 2012, while the real interest rates panel contains observations of 47 countries for the same period.

Table 3 shows the results for the one-way transfer entropy tests and the net transfer entropy tests between the sovereign rating changes and the economic activity, as measured by GDP growth (left panel), and the financial market, as measured by interest rates (right panel). According to the p-values reported in this table, which are above 0.05 in all cases, we failed to detect significant causal relationships either between rating changes and growth or between rating changes and interest rates. This result is consistent with the view that rating decisions add value in providing new information to economic and financial markets which is not already reflected in other economic aggregates and market prices. In addition, this result does not support the potential role in amplifying the negative effect of recessions sometimes attributed to credit rating.[13]

5 Conclusion

Recent advances in Granger causality have extended this analysis to the context of longitudinal data by imposing coefficient restrictions that can be easily tested using relatively standard techniques in the context of linear panel data. Two significant contributions that follow this approach are Holtz-Eakin et al. (1988) and Dumitrescu and Hurlin (2012).

In spite of their simplicity, the size and power of these tests can be seriously deteriorated when the linearity assumption breaks down, when the data generating process is heterogeneous across the cross-section units, or when there are structural breaks or extreme observations in some of the cross-section units. In addition, these techniques are unable to offer a simple solution to handling qualitative time series data that refer, for example, to categorical grades.

In this paper, we translate the problem into symbolic dynamics and propose a non-parametric causality test based on the concept of transfer entropy. We check the robustness of our test with several Monte Carlo experiments under different scenarios. Our results indicate that the test displays correct size in situations where linear tests fail and increased power to detect causality.

Besides its robustness, the test is simple to implement and we advocate for its generalized use as a complementary tool in causality analysis. We illustrate the usefulness of our proposal by analyzing the dynamic relationships between public expenditure and GDP growth, between firm size and firm productivity, and between changes in Fitch's sovereign ratings and GDP growth and between these rating changes and interest rates. In the first case, we find very limited, if any, causality of GDP to public expenditure and not the other way around. In the case of firm size and productivity, we find bidirectional causality with no predominant causation. Finally,

[13]Some authors such as Kiff at al, (2010) find that the market impact of credit rating may be affected by the *pre-rating* warnings rather than the actual rating changes. To test this extreme, we recoded the series of Fitch's credit rating warnings for the same period using the symbols 0 to 5 to denote "No warning", "Negative, "Rating watch negative", "Stable", "Rating watch positive", and "Positive" announcements respectively, and we conducted the symbolic transfer entropy test. Although we omit the results to save space (they are available upon request), they show again no clear casual relationship in any direction.

in the case of credit rating and GDP growth or interest rates, we find no relevant causality relationship between credit rating and the economic and financial magnitudes in any direction.

References

[1] Biesebroeck, J. 2005. Firm size matters: Growth and productivity growth in African manufacturing. *Economic Development and Cultural Change* 53: 545-83.

[2] Box, G., and Wilson, K. 1951. On the experimental attainment of optimum conditions. *Journal of Royal Statistical Society, Series. B* 13: 1-45.

[3] Bramati, M., and Croux, C. 2007. Robust estimators for the fixed effects panel data model. *Economics Journal* 10: 521–540.

[4] Brooks, R., Faff, R., Hillier, D., and Hillier, J. 2004. The national market impact of sovereign rating changes. *Journal of Banking and Finance* 28: 233-250.

[5] Broto, C., and Molina, L. 2016. Sovereign ratings and their asymmetric response to fundamentals. *Journal of Economic Behavior and Organization* 130: 206-224.

[6] Caines, P., Keng, C., and Sethi, S. 1981. Causality analysis and multivariate autoregressive modelling with an application to supermarket sales analysis. *Journal of Economic Dynamics and Control* 3: 267-298.

[7] Collet, P., and Eckmann, J. 2009. Iterated maps on the interval as dynamical systems. Birkhauser, Basel.

[8] De Vita, G., Trachanas, E., and Luo, Y. 2018. Revisiting the bi-directional causality between debt and growth: Evidence from linear and nonlinear tests. *Journal of International Money and Finance* 82: 55-74.

[9] Diaz, M. and Sanchez, R. 2008. Firm size and productivity in Spain: A stochastic frontier analysis. *Small Business Economics* 30: 315-23.

[10] Dumitrescu, E., and Hurlin, C. 2012. Testing for Granger non-causality in heterogeneous panels. *Economic Modelling* 29: 1450-1460.

[11] Ferri, G., Lui, G., and Stiglitz, J. 1999. The procyclical role of rating agencies: Evidence from East Asian crisis. *Economic Notes* 28: 335-355.

[12] Granger, C. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424-438.

[13] Hafner, Ch., and Herwartz, H. 2008. Testing for Causality in Variance Using Multivariate GARCH Models. *Annales d'Économie et de Statistique* 89: 215-241.

[14] Holtz-Eakin, D., Newey, W., and Rosen, H. 1988. Estimating vector autoregressions with panel data. *Econometrica* 56: 137-195.

[15] Hsieh, D. 1991. Chaos and nonlinearity dynamics: application to financial market. *Journal of Finance* 46: 1839-1877.

[16] Hurlin, C., and Venet, B. 2001. Granger causality tests in panel data models with fixed coefficients. *Working Paper*. EURIsCO, Université Paris IX Dauphin.

[17] Kaminsky, G., and Schmukler, S. 2002 Emerging markets instability : Do sovereign ratings affect country Risk and stock returns? *The World Bank Economic Review* 16: 171-195.

[18] Kiff, J. and Kisser, M. 2010. The Uses and Abuses of Sovereign Credit Ratings, in book *Global Financial Stability Report*, International Monetary Fund.

[19] Matilla-Garcia M., and Ruiz-Marin, M. 2008. A non-parametric independence test using permutation entropy. *Journal of Econometrics* 144: 139-155.

[20] Matilla-Garcia M., Ruiz-Marin, M., and Dore, M. 2014. A permutation entropy based test for causality: The volume-stock price relation. *Physica A* 398: 280-288.

[21] Nair-Reichert, U., and Weinhold, D., 2001. Causality tests for cross-country panels: a look at FDI and economic growth in less developed countries. *Oxford Bulletin of Economics and Statistics* 63: 153-171.

[22] Pagano, P., and Schivardi, F. 2003. Firm size distribution and growth. *Scandinavian Journal of Economics* 105: 255-274.

[23] Politis, D., and White, H. 2004. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* 23: 53–70.

[24] Reinhart, C., and Rogoff, K. 2010. Growth in a time of debt. *American Economic Review* 100: 573-78.

[25] Seth, A., Barrett, A., and Barnett, L. 2015. Granger vausality analysis in neuroscience and neuroimaging. *The Journal of Neuroscience* 35: 3293–3297.

[26] Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters* 85: 461–464.

[27] Shannon, C. 1948 A mathematical theory of communication. *Bell System Technical Journal* 27: 623–656.

[28] Weinhold, D. 1996. Tests de causalité sur données de panel: une application à l'étude de la causalité entre l'investissement et la croissance. *Economie et Prévision* 126: 163-175.

Appendix. Ordinal Pattern and Entropy, a simple example

Consider the finite time series $\{y_t\}$ of seven values

$$\{y_1 = 3, y_2 = 9, y_3 = 7, y_4 = 6, y_5 = 5, y_6 = 10, y_7 = 4\}. \tag{A.1}$$

Using an embedding dimension $m = 3$, the 3-dimensional histories $\mathbf{y}_3(t)$, for $t = 1, \ldots, 5$ are

$$\{\mathbf{y}_3(t)\} = \{(3, 9, 7), (9, 7, 6), (7, 6, 5), (6, 5, 10), (5, 10, 4)\}. \tag{A.2}$$

For $m = 3$, the symmetric group is

$$S_3^1 = \{(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)\}. \tag{A.3}$$

Each one of the five 3-histories $\mathbf{y}_3(t)$ can be uniquely mapped into a symbol in S_3^1. For example, for $t = 1$ we have that $\mathbf{y}_3(1) = (3, 9, 7)$ and $y_1 = 3 < y_3 = 7 < y_2 = 9$, which implies that $\mathbf{y}_3(t)$ is of $(1, 3, 2)$-type. Following in the same fashion, the sequence of symbols in this example is

$$\{\zeta_{i3}^{y}(t)\} = \{(1, 3, 2), (3, 2, 1), (3, 2, 1), (2, 1, 3), (3, 1, 2)\}. \tag{A.4}$$

The probabilities of the symbols in S_3^1, are computed as their relative frequencies in $\{\zeta_3^y(t)\}$. Then, the estimated probabilities are $p(1, 2, 3) = 0$, $p(1, 3, 2) = \frac{1}{5}$, $p(2, 1, 3) = \frac{1}{5}$, $p(2, 3, 1) = 0$, $p(3, 1, 2) = \frac{1}{5}$, and $p(3, 2, 1) = \frac{2}{5}$. Using these probabilities, the permutation entropy is $h_3(y_{i,t}) = -3\frac{1}{5}\ln(\frac{1}{5}) - \frac{2}{5}\ln(\frac{2}{5}) \approx 1.332$.

Now, consider the time series $\{x_{i,t}\}$ that also takes seven values

$$\{x_1 = 2, x_2 = 8, x_3 = 5, x_4 = 4, x_5 = 1, x_6 = 2, x_7 = 3\}. \tag{A.5}$$

In the same way, the permutation entropy is $h_3(x_t) \approx 1.332$.

From these two Universities time series, the bivariate time series $\{y_t, x_t\}$ is

$$\{(3, 2), (9, 8), (7, 5), (6, 4), (5, 1), (10, 2), (4, 3)\}. \tag{A.6}$$

The sequence of symbols in this example is

$$\{\zeta_{i3}^{yx}(t)\} = \{[(1, 3, 2), (1, 3, 2)], [(3, 2, 1), (3, 2, 1)], [(3, 2, 1), (3, 2, 1), [(2, 1, 3), (2, 3, 1)], [(3, 1, 2), (1, 2, 3)]\}. \tag{A.7}$$

Using this sequence of symbols, the estimated probabilities for all the symbols in S_3^2 are zero, with the exception of $p([(1, 3, 2), (1, 3, 2)]) = \frac{1}{5}$, $p([(1, 3, 2), (1, 3, 2)]) = \frac{2}{5}$, $p([(2, 1, 3), (2, 3, 1)]) = \frac{1}{5}$, and $p([(3, 1, 2), (1, 2, 3)]) = \frac{1}{5}$. The permutation entropy is $h_3(y_t, x_t) = -3\frac{1}{5}\ln(\frac{1}{5}) - \frac{2}{5}\ln(\frac{2}{5}) \approx 1.332$. In this example, the conditional entropy is $h_3(y_t|x_t) = h_3(y_t, x_t) - h_3(x_t) = 0$.

Table 1: Causality tests: Governement expenditure and growth

	HNR		DH		STE		
Direction	Stat	p-value	Stat	p-value	Stat	1-tail pval	2-tail pval
Panel A: $r = 1$							
Exp→GDP	-3.242	0.001	5.703	0.000	0.008	0.240	-
GDP→Exp	1.441	0.150	17.815	0.000	0.008	0.165	-
Net (Exp - GDP)	-	-	-	-	-0.001	0.455	0.820
Panel B: $r = 2$							
Exp→GDP	1.376	0.169	7.223	0.000	0.018	0.290	-
GDP→Exp	0.166	0.868	22.069	0.000	0.024	0.005	-
Net (Exp - GDP)	-	-	-	-	-0.005	0.050	0.110
Panel C: $r = 3$							
Exp→GDP	-2.397	0.017	6.065	0.000	0.017	0.520	-
GDP→Exp	-0.750	0.453	10.386	0.000	0.019	0.320	-
Net (Exp-GDP)	-	-	-	-	-0.001	0.345	0.700

Notes. Columns labeled as HNR and DH show the results of Holtz-Eakin, Newey, and Rosen (1988) and Dumitrescu and Hurlin (2012) causality tests for longitudinal data. Last column show the results of the Symbolic Transfer Entropy (STE) causality test and the one-tailed and two-tailed net transfer entropy test.

Table 2: Causality tests: Firm size and productivity

	HNR		DH		STE		
Direction	Stat	p-value	Stat	p-value	Stat	1-tail pval	2-tail pval
Panel A: $r = 1$							
Size→TFP	-6.850	0.000	6.189	0.000	0.025	0.005	-
TFP→Size	3.150	0.002	6.471	0.000	0.026	0.000	-
Net (Size - TFP)	-	-	-	-	-0.001	0.510	0.925
Panel B: $r = 2$							
Size→TFP	1.126	0.260	1.469	0.315	0.057	0.000	-
TFP→Size	-1.430	0.153	-0.237	0.855	0.053	0.000	-
Net (Size - TFP)	-	-	-	-	-0.004	0.290	0.570
Panel C: $r = 3$							
Size→TFP	0.358	0.720	0.704	0.580	0.049	0.010	-
TFP→Size	-0.122	0.903	0.073	0.945	0.049	0.010	-
Net (Size - TFP)	-	-	-	-	-0.001	0.450	0.905

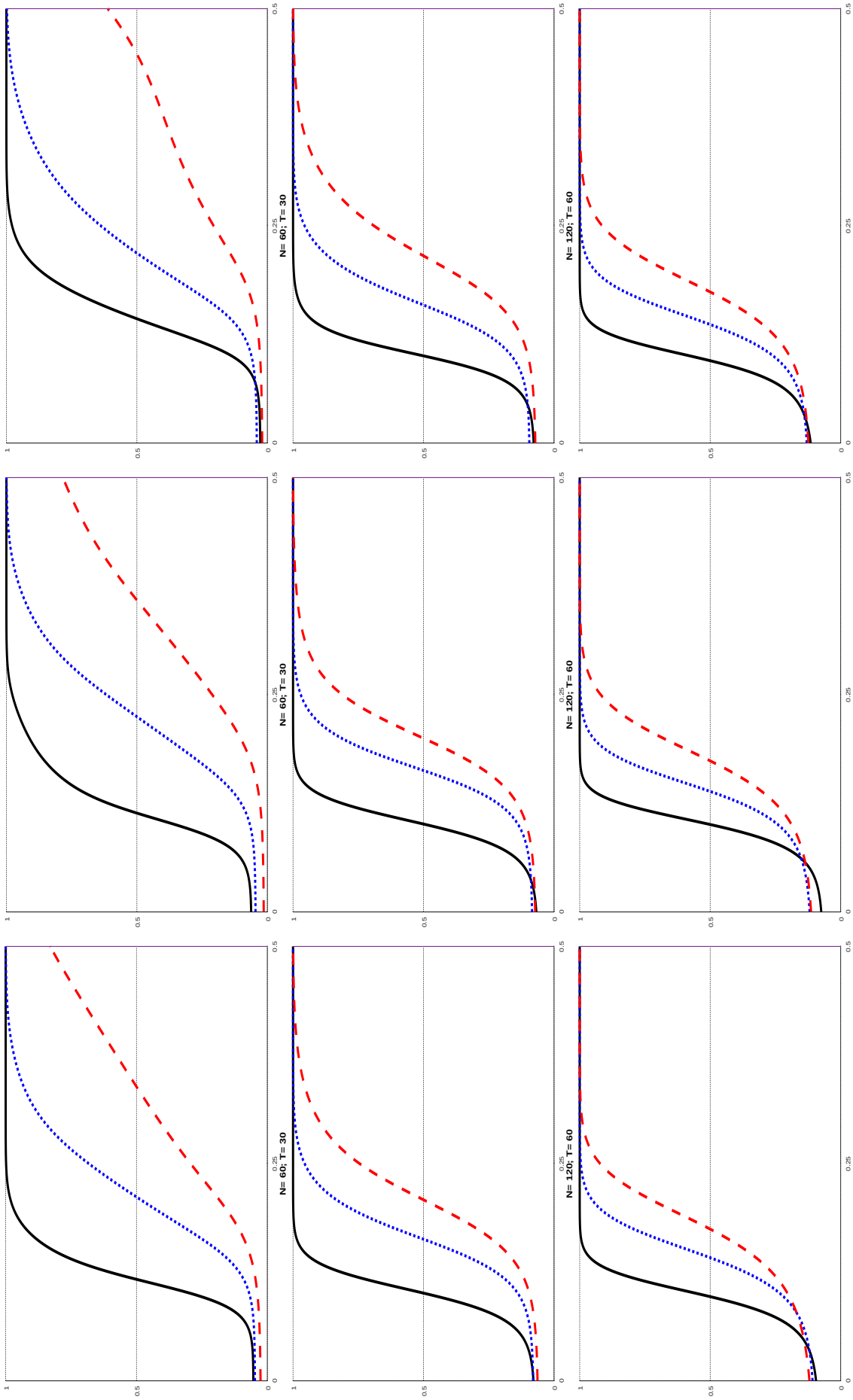
Notes. See notes of Table 1.

Table 3: Causality tests: Fitch sovereign rating, GDP growth and interest rate

GDP growth				Interest rate			
Direction	Stat	1 p-value	2 p-value	Direction	Stat	1 p-value	2 p-value
Panel A: $r = 1$							
Rating→GDP	0.017	0.060	-	Rating→I. Rate	0.026	0.650	-
GDP→Rating	0.025	0.115	-	I. Rate→Rating	0.033	0.895	-
Net effect	-0.008	0.610	0.610	Net effect	-0.007	0.735	0.740
Panel B: $r = 2$							
Rating→GDP	0.028	0.890	-	Rating→I. Rate	0.075	0.260	-
GDP→Rating	0.023	0.500	-	I. Rate→Rating	0.035	0.930	-
Net effect	0.005	0.825	0.830	Net effect	0.040	0.095	0.095
Panel C: $r = 3$							
Rating→GDP	0.035	0.675	-	Rating→I. Rate	0.064	0.925	-
GDP→Rating	0.024	0.465	-	I. Rate→Rating	0.033	0.995	-
Net effect	0.011	0.580	0.580	Net effect	0.031	0.435	0.435

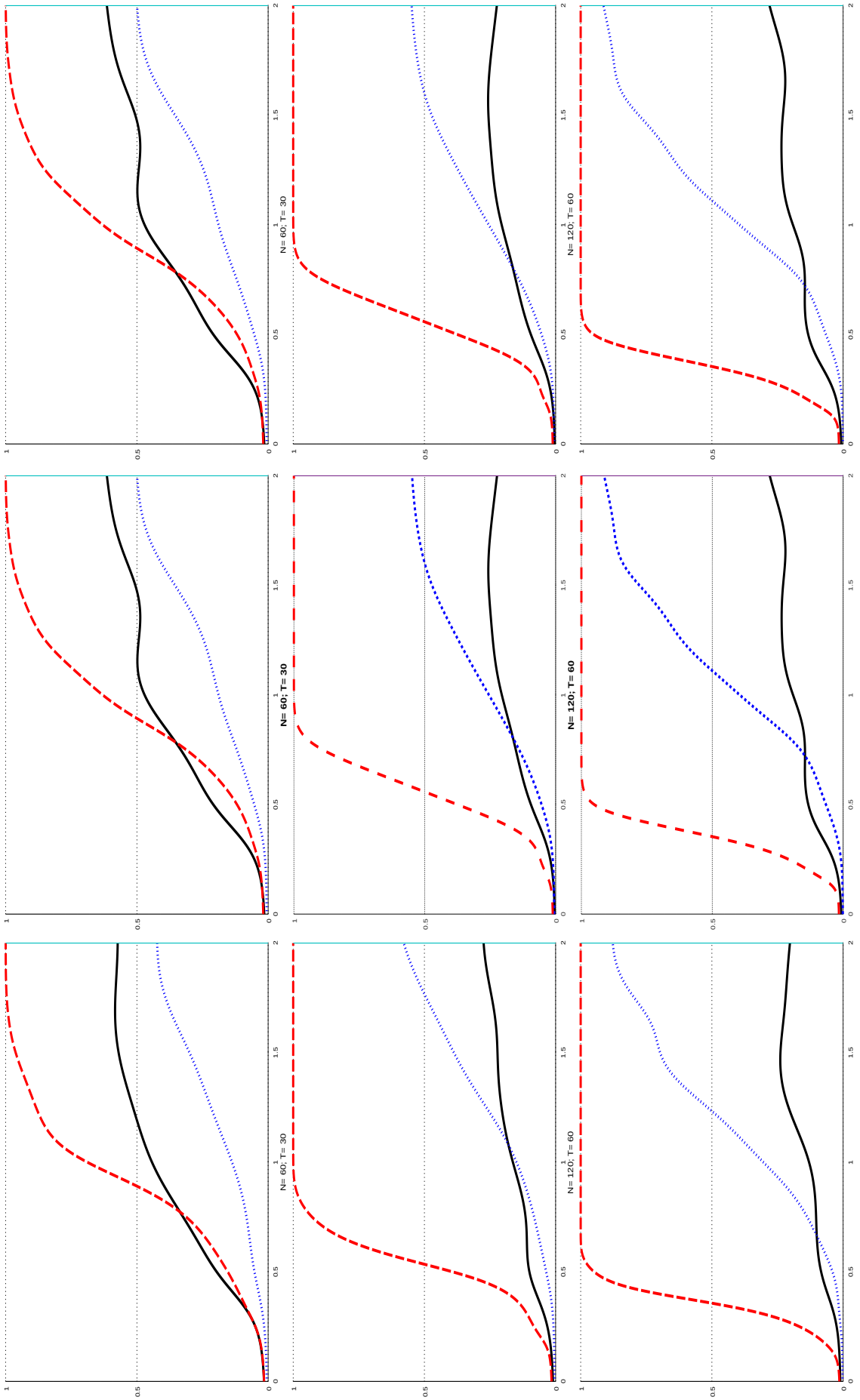
Notes. See notes of Table 1.

Figure 1: Power functions: benchmark model



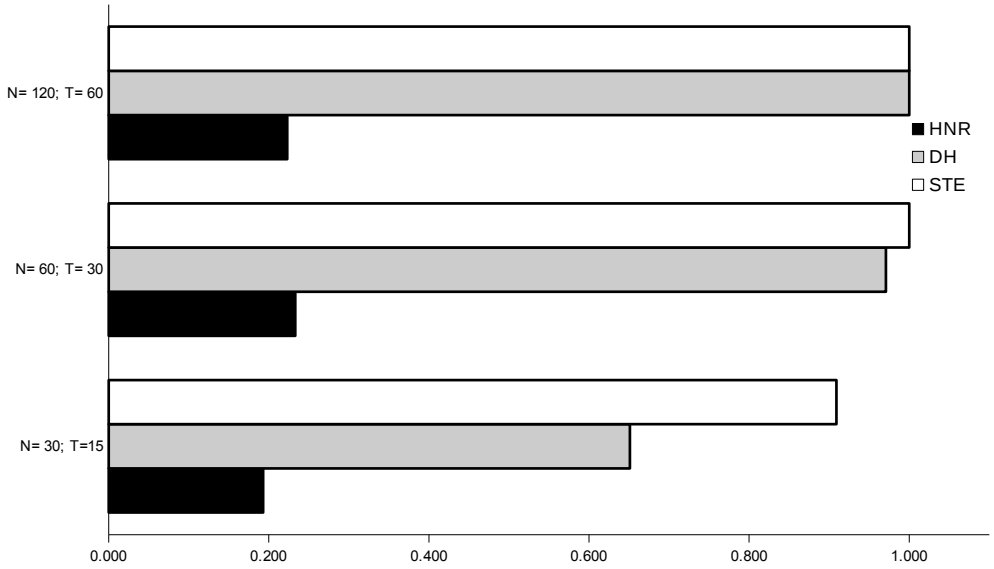
Notes. The figures display the rejection rates across the Monte Carlo simulations of the null of non-causality for a level of significance $\alpha = 0.05$ in a homogeneous linear model, for three different values of the autocorrelation parameter. Acronyms HNR, DH, and STE refer to the non-causality tests for longitudinal data of Holtz-Eakin, Newey, and Rosen (1998, solid lines), Dumitrescu, and Hurlin (2012, dotted lines) and the Symbolic Transfer Entropy test (dashed lines), respectively.

Figure 2: Power functions: structural break



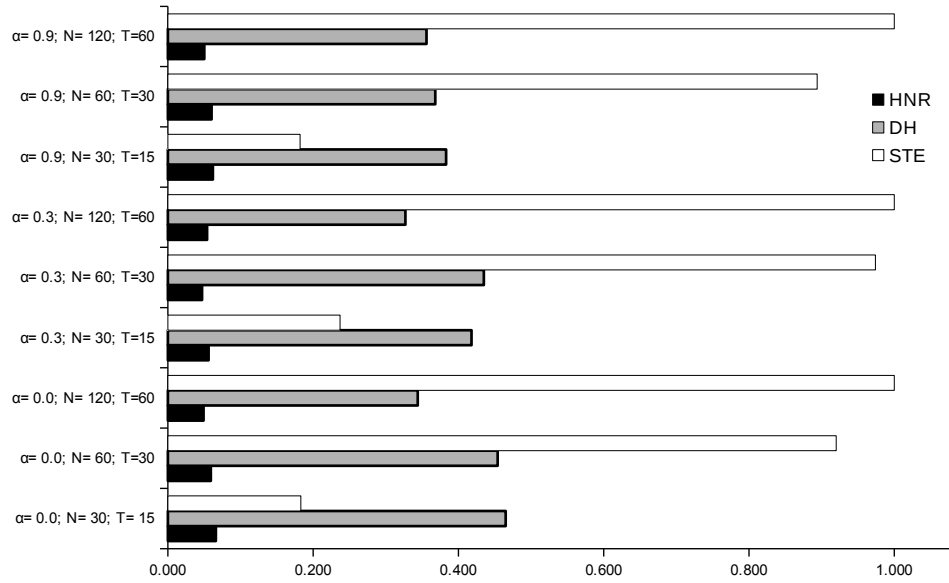
Notes. The figures display the rejection rates across the Monte Carlo simulations of the null of non-causality for a level of significance $\alpha = 0.05$ for three different values of the autocorrelation parameter. Acronyms HNR solid lines, DH dotted lines, and STE dashed lines. In this case, half of the simulations use β and the other half use $-\beta$, with $\beta \in [0, 2]$.

Figure 3: Power functions: nonlinear causality



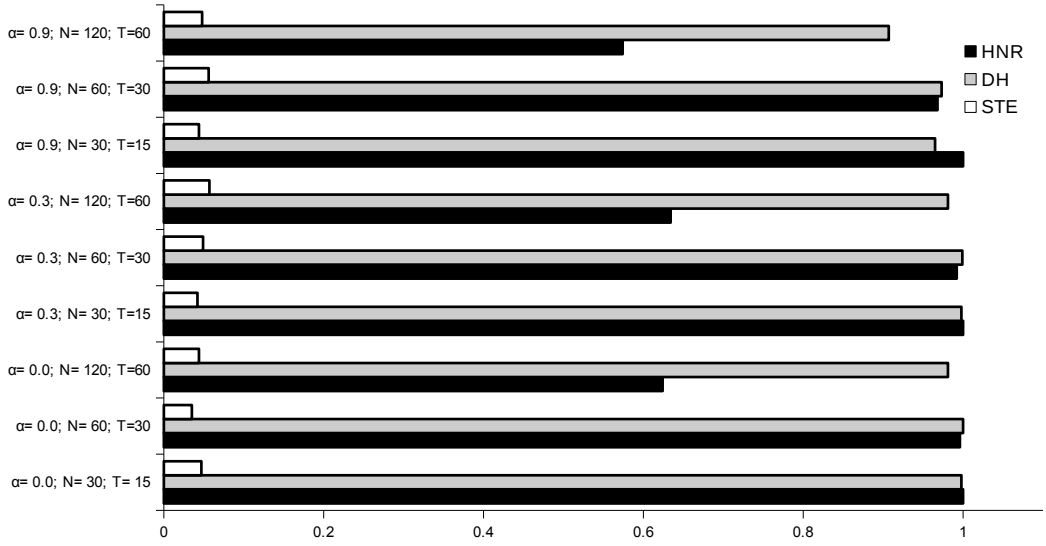
Notes. See the notes of Figure 1. In this case, we generate data as $y_{i,t} = y_{i,t-1}x_{i,t-1} + \epsilon_{i,t}$, where $\epsilon_{i,t} \sim N(0,1)$ and $x_{i,t} \sim N(0,1)$.

Figure 4: Power functions: causality in variance



Notes. See the notes of Figure 1. In this case, we generate data as $y_{i,t} = \gamma y_{i,t-1} + \epsilon_{i,t}$, where $\epsilon_{i,t} \sim N(0, |x_{i,t-1}|)$ and $x_{i,t} \sim N(0,1)$.

Figure 5: Size functions: the effect of outliers



Notes. See the notes of Figure 1. In this case, we generate data as $y_{i,t} = \gamma y_{i,t-1} + \epsilon_{i,t}$, where $\epsilon_{i,t} \sim N(0,1)$, and we impose $y_{12} = x_{11} = -10$ and $y_{N\bar{T}} = x_{N\bar{T}-1} = -10$.