Testing Granger Causality in Heterogenous Panel Data Models with Fixed Coefficients

Christophe Hurlin*†

January 2004

Abstract

This paper proposes a simple test of Granger (1969) non causality hypothesis in heterogeneous panel data models with fixed coefficients. It proposes a statistic of test based on averaging standard individual Wald statistics of Granger non causality tests. First, this statistic is shown to converge sequentially to a standard normal distribution with T tends to infinity, followed by N. Second, for a fixed T sample the semi-asymptotic distribution of the average statistic is characterized. In this case, individual Wald statistics do not have a standard distribution. However, under very general setting, we prove that individual Wald statistics are independently distributed with finite second order moments as soon as $T > 5 + 2K$, where K denotes the number of linear restrictions. For a fixed T sample, the Lyapunov central limit theorem is then sufficient to get the semi asymptotic distribution when N tends to infinity. The two first moments of this normal distribution correspond to empirical mean of the corresponding theoretical moments of the individual Wald statistics. The issue is then to propose an evaluation of the two first moments of standard Wald statistics for small T sample. In this paper we propose a general approximation based on the exact moments of the ratio of quadratic forms in normal variables derived from the Magnus (1986) theorem. For a fixed T sample, we propose simple approximations of the mean and the variance of the Wald statistic. Monte Carlo experiments show that these formulas provide an excellent approximation. Given these approximations, we propose an approximated standardized average Wald statistic to test the HNC hypothesis in short T sample. Finally, approximated critical values are proposed for finite N and T sample and compared to simulated critical values in some experiments.

- *Keywords* : Granger Causality, Panel data, Wald Test.
- *J.E.L Classification* : C23, C11

*LEO, University of Orléans. email: christophe.hurlin@univ-orleans.fr. A substantial part of the work for this paper was undertaken in the Department of Economics of the University Paris IX Dauphine, EURIsCO and in CEPREMAP.

†An earlier version of this paper was presented at the EC^2 meeting "Causality and Exogeneity in Econometrics", CORE Louvain La Neuve, December 2001. I am grateful for comments and advices from Anindya Banerjee, Pierre-Yves Hénin, Valentin Patilea, Gilbert Colletaz and the participants of the econometric seminars of the University Paris I and the University of Geneva.

1 Introduction

The aim of this paper is to propose a simple Granger (1969) non causality test in heterogeneous panel data models with fixed coefficients. In the framework of a linear autoregressive data generating process, the extension of standard causality tests for panel data implies to test cross sectional linear restrictions on the coefficients of the model. As usually, the use of cross-sectional information may extend the information set on the causality from a given variable to another. Indeed, in many economic problems it is highly probable that if a causal relationship exists for a country or an individual, it exists also for some other countries or individuals. In this case, the causality can be tested with more efficiency in a panel context with NT observations. However, the use of the cross-sectional information implies to take into account the heterogeneity across individuals in the definition of the causal relationship. As discussed in Granger (2003), the usual causality test in panel asks "*if some variable, say X_t causes another variable, say Y_t, everywhere in the panel [..]. This is rather a strong null hypothesis.*" Then, we propose here a simple Granger non causality test for heterogeneous panel data models. This test allows to take into account both dimensions of the heterogeneity in this context: the heterogeneity of the causal relationships and the heterogeneity of the data generating process (DGP).

Let us consider the standard implication of the Granger causality definition[1]. For each individual, we say that the variable x is causing y if we are better able to predict y using all available information than if the information apart from x had been used (Granger 1969). If x and y are observed on N individuals, the issue consists in determining the optimal information set used to forecast y. Several solutions could be adopted. The most general is to test the causality from the variable x observed on the i^{th} individual to the variable y observed for the j^{th} individual, with $j = i$ or $j \neq i$. The second solution, is more restrictive and is directly derived from the time series analysis. It implies to test causal relationship for a given individual. The cross sectional information is then only used to improve the specification of the model and the power of tests as in Holtz-Eakin, Newey and Rosen (1988). The baseline idea is to assume that there exists a minimal statistical representation, which is common to x and y at least for a subgroup of individuals. In this paper we use such a model. Then, causality tests could be implemented and considered as a natural extension of the standard time series tests in the cross sectional dimension.

However, one of the main specific stakes of panel data models is to specify the heterogeneity between individuals. In this context, the heterogeneity has two main dimensions as discussed in Hurlin and Venet (2001). We propose to distinguish between the heterogeneity of the DGP and the heterogeneity of the causal relationships from x to y. Indeed, the DGP may be different from an individual to another, whereas there exists a causal relationship from x to y for all individuals. More generally, in a p order linear vectorial autoregressive model, we define four kinds of causal relationships. The first, denoted *Homogenous Non Causality* (HNC) hypothesis, implies that there does not exist any individual causality relationships from x to y. The symmetric case is the *Homogenous Causality* (HC) hypothesis, which occurs when there exists N causality relationships, and when the individual predictors of y, obtained conditionally to the past values

[1] The precise Granger causality definition is based on the "*two precepts that the cause preceded the effect and the causal series had information about the effect that was not contained in any other series according to the conditional distributions*" (Granger 2003). The fact that the cause produces a superior forecasts of the effect is just an implication of these statements. However, it does provide suitable post sample tests as discussed in Granger (1980).

of y and x are identical. The dynamics of y is then totally identical for all the individuals of the sample. The two last cases correspond to heterogeneous process. Under the *HEterogenous Causality* (HEC) hypothesis, we assume that there exists N causality relationships as in the HC case, but the dynamics of y is heterogenous. The heterogeneity does not affect the causality result. Finally, under the *HEterogenous Non Causality* ($HENC$) hypothesis, we assume that there exists a subgroup of individuals for which there is a causal relationship from x to y. Symmetrically, there is at least one and at the most $N-1$ non causal relationships in the model. That is why, in this case, the heterogeneity deals with the causality from x to y.

To sum it up, in the HNC hypothesis, there does not exist any individual causality from x to y. On the contrary, in the HC and HEC cases, there is a causality relationships for each individual of the sample. In the HC case, the DGP is homogenous, whereas it is not the case in the HEC hypothesis. Finally in the $HENC$ hypothesis, there is an heterogeneity of the causality relationships since there is a subgroup of N_1 units for which the variable x does not cause y.

In this paper, we propose a simple test of the Homogenous Non Causality (HNC) hypothesis. Under the null hypothesis, there is no causal relationship for all the units of the panel. However, we do not test this hypothesis against the HC hypothesis as Holtz-Eakin, Newey and Rosen (1988). We specify the alternative as the $HENC$ hypothesis. There is two subgroups of units: one with causal relationships from x to y, but not necessarily with the same DGP, and an another subgroup where there is no causal relationships from x to y. For that, our test is lead in an heterogenous panel data model with fixed coefficients. Under the null or the alternative, the unconstrained parameters may be different from individual to another. Then, whatever the result on the existence of causal relationships, we assume that the dynamic of the individual variables may be heterogeneous.

As in the literature on unit root tests in heterogeneous panels, and particularly in Im, Pesaran and Shin (2002), this paper proposes a statistic of test based on averaging standard individual Wald statistics of Granger non causality tests. First, this statistic is shown to converge sequentially in distribution to a standard normal variate when the time dimension T tends to infinity, followed by the individual dimension N. Second, for a fixed T sample the semi-asymptotic distribution of the average statistic is characterized. In this case, individual Wald statistics do not have a standard chi-squared distribution. However, under very general setting, it is shown that individual Wald statistics are independently distributed with finite second order moments as soon as $T > 5 + 2K$, where K denotes the number of linear restrictions. For a fixed T, the Lyapunov central limit theorem is sufficient to get the distribution of the standardized average Wald statistic when N tends to infinity. The two first moments of this normal semi-asymptotic distribution correspond to empirical mean of the corresponding theoretical moments of the individual Wald statistics.

The issue is then to propose an evaluation of the two first moments of standard Wald statistics for small T sample. The first solution consists in using bootstrap simulations. However, in this paper we propose an approximation of these moments based on the exact moments of the ratio of quadratic forms in normal variables derived from the Magnus (1986) theorem for a fixed T sample, with $T > 5 + 2K$. Monte Carlo experiments show that these formulas provide an excellent approximation to the true moments. Given these approximations, we propose an approximated standardized average Wald statistic to test the HNC hypothesis in short T sample. Finally, approximated critical values are proposed for finite T and N sample.

The paper is organized as follows. Section 2 is devoted to the definition of the Granger causality test in heterogenous panel data models. Section 3 sets out the asymptotic distributions of the average Wald statistic. Section 4 derives the semi-asymptotic distribution for fixed T sample and section 5 proposes some results of Monte Carlo experiments. Section 6 extends the results to a fixed N sample and the last section provides some concluding remarks.

2 A non causality test in heterogenous panel data models

Let us consider two covariance stationary variables, denoted x and y, observed on T periods and on N individuals. For each individual $i = 1, .., N$, at time $t = 1, .., T$, we consider the following linear model:

$$y_{i,t} = \alpha_i + \sum_{k=1}^{K} \gamma_i^{(k)} y_{i,t-k} + \sum_{k=1}^{K} \beta_i^{(k)} x_{i,t-k} + \varepsilon_{i,t} \tag{1}$$

with $K \in \mathbb{N}^*$ and $\beta_i = \left(\beta_i^{(1)}, ..., \beta_i^{(K)}\right)'$. For simplicity, individual effects α_i are supposed to be fixed. Initial conditions $(y_{i,-K}, ..., y_{i,0})$ and $(x_{i,-K}, ..., x_{i,0})$ of both individual processes $y_{i,t}$ and $x_{i,t}$ are given and observable. We assume that lag orders K are identical for all cross-section units of the panel and the panel is balanced. In a first part, we allow for autoregressive parameters $\gamma_i^{(k)}$ and regression coefficients slopes $\beta_i^{(k)}$ to differ across groups. However, contrary to Weinhold (1996) and Nair-Reichert and Weinhold (2001), parameters $\gamma_i^{(k)}$ and $\beta_i^{(k)}$ are constant. It is important to note that our model is not a random coefficient model as in Swamy (1970): it is a fixed coefficients model with fixed individual effects. We make the following assumptions.

Assumption (A_1) *For each cross section unit $i = 1, .., N$, individual residuals $\varepsilon_{i,t}$, $\forall t = 1, .., T$ are independently and normally distributed with $E(\varepsilon_{i,t}) = 0$ and finite heterogeneous variances $E\left(\varepsilon_{i,t}^2\right) = \sigma_{\varepsilon,i}^2$.*

Assumption (A_2) *Individual residuals $\varepsilon_i = (\varepsilon_{i,1}, .., \varepsilon_{i,T})'$, are independently distributed across groups. Consequently $E(\varepsilon_{i,t}\varepsilon_{j,s}) = 0$, $\forall i \neq j$ and $\forall (t, s)$.*

Assumption (A_3) *Both individual variables $x_i = (x_{i,1}, ..., x_{i,T})'$ and $y_i = (y_{i,1}, ..., y_{i,T})'$, are covariance stationary with $E\left(y_{i,t}^2\right) < \infty$, $E\left(x_{i,t}^2\right) < \infty$, $E(x_{i,t}x_{j,z})$, $E(y_{i,t}y_{j,z})$ and $E(y_{i,t}x_{j,z})$ are only function of the difference $t - z$, whereas $E(x_{i,t})$ and $E(y_{i,t})$ are independent of t.*

This simple two variables model constitutes the basic framework to study the Granger causality in a panel data context. As for time series, the standard causality tests consist in testing linear restrictions on vectors β_i. However with a panel data model, one must be very careful to the issue of heterogeneity between individuals. The first source of heterogeneity is standard and comes from the presence of individual effects α_i. The second source, which is more crucial, is related to the heterogeneity of the parameters β_i. This kind of heterogeneity directly affects the paradigm of the representative agent and so, the conclusions about causality relationships. It is well known that the estimates of autoregressive parameters β_i get under the wrong hypothesis $\beta_i = \beta_j$, $\forall (i, j)$ are biased (see Pesaran and Smith 1995 for an $AR(1)$ process). Then, if we impose the homogeneity of coefficients β_i, the statistics of causality tests can lead to a fallacious inference. Intuitively, the estimate $\widehat{\beta}$ obtained in an homogeneous model will converge to a value close to the average of the true coefficients β_i, and that if this mean is itself close to zero, we risk to accept at wrong the hypothesis of no causality.

Beyond these statistical stakes, it is evident that an homogeneous specification of the relation between the variables x and y does not allow to give some interpretation of the relations of causality as soon as at least one individual of the sample has an economic behavior different from that of the others. For example, let us assume that there exists a relation of causality for a set of N countries, for which vectors β_i are strictly identical. If we introduce into the sample, a set of N_1 countries for which, on the contrary, there is no relation of causality, what are the conclusions? Whatever the value of the ratio N/N_1 is, the test of the causality hypothesis is nonsensical.

Given these observations, we now propose to test the Homogenous Non Causality (HNC) hypothesis. Under the alternative we allow that there exists a subgroup of individuals with no causality relations and a subgroup of individuals for which the variable x Granger causes y. The null hypothesis of HNC is defined as:

$$H_0 : \beta_i = 0 \quad \forall i = 1, ..N \tag{2}$$

with $\beta_i = \left(\beta_i^{(1)}, ..., \beta_i^{(K)}\right)'$. Under the alternative, we allow for β_i to differ across groups. We also allow for some, but not all, of the individual vectors to be equal to 0 (non causality assumption). We assume that under H_1, there are $N_1 < N$ individual processes with no causality from x to y. Then, this test is not a test of the non causality assumption against the causality from x to y for all the individuals, as in Holtz-Eakin, Newey and Rosen (1988). It is more general, since we can observe non causality for some units under the alternative:

$$\begin{aligned} H_1 : \qquad & \beta_i = 0 \quad \forall i = 1, .., N_1 \\ & \beta_i \neq 0 \quad \forall i = N_1 + 1, N_1 + 2, .., N \end{aligned} \tag{3}$$

where N_1 is unknown but satisfies the condition $0 \leq N_1/N < 1$. The fraction N_1/N is necessarily inferior to one, since if $N_1 = N$ there is no causality for all the individual of the panel, and then we get the null hypothesis HNC. In the opposite case $N_1 = 0$, there is causality for all the individual of the sample. The structure of this test is similar to the unit root test in heterogenous panels proposed by Im, Pesaran and Shin (2002). In our context, if the null is accepted the variable x does not Granger cause the variable y for all the units of the panel. On the contrary, let us assume that the HNC is rejected and if $N_1 = 0$, we have seen that x Granger causes y for all the individuals of the panel : in this case we get an homogenous result as far as causality is concerned. The DGP may be not homogenous, but the causality relations are observed for all individuals. On the contrary, if $N_1 > 0$, then the causality relationships is heterogeneous : the DGP and the causality relations are different according the individuals of the sample.

In this context, we propose to use the average of individual Wald statistics associated to the test of the non causality hypothesis for units $i = 1, .., N$.

Definition *The average statistic $W_{N,T}^{Hnc}$ associated to the null Homogenous Non Causality (HNC) hypothesis is defined as:*

$$W_{N,T}^{Hnc} = \frac{1}{N} \sum_{i=1}^{N} W_{i,T} \tag{4}$$

where $W_{i,T}$ denotes the individual Wald statistics for the i^{th} cross section unit associated to the individual test $H_0 : \beta_i = 0$.

In order to express the general form of this statistic, we stack the T periods observations for the i^{th} individual's characteristics into T elements columns as:

$$\underset{(T,1)}{y_i^{(k)}} = \begin{bmatrix} y_{i,1-k} \\ . \\ . \\ y_{i,T-k} \end{bmatrix} \qquad \underset{(T,1)}{x_i^{(k)}} = \begin{bmatrix} x_{i,1-k} \\ . \\ . \\ x_{i,T-k} \end{bmatrix} \qquad \underset{(T,1)}{\varepsilon_i} = \begin{bmatrix} \varepsilon_{i,1} \\ . \\ . \\ \varepsilon_{i,T} \end{bmatrix}$$

and we define two (T, K) matrices :

$$Y_i = \left[y_i^{(1)} : y_i^{(2)} : ... : y_i^{(K)} \right] \qquad X_i = \left[x_i^{(1)} : x_i^{(2)} : ... : x_i^{(K)} \right]$$

Let us denote Z_i the $(T, 2K+1)$ matrix $Z_i = [e : Y_i : X_i]$, where e denotes a $(T, 1)$ unit vector, and $\theta_i = \left(\alpha_i \, \gamma_i' \, \beta_i'\right)'$ the vector of parameters of model. The HNC hypothesis test can be expressed as $R\theta_i = 0$ where R is a $(K, 2K+1)$ matrix with $R = [0 : I_K]$. The Wald statistic $W_{i,T}$ associated to the individual test $H_0 : \beta_i = 0$ is defined for each $i = 1, .., N$ as:

$$W_{i,T} = \widehat{\theta}_i' R' \left[\widehat{\sigma}_i^2 R \left(Z_i' Z_i \right)^{-1} R' \right]^{-1} R \widehat{\theta}_i = \frac{\widehat{\theta}_i' R' \left[R \left(Z_i' Z_i \right)^{-1} R' \right]^{-1} R \widehat{\theta}_i}{\widehat{\varepsilon}_i' \widehat{\varepsilon}_i / \left(T - 2K - 1 \right)}$$

where $\widehat{\theta}_i$ is the estimate of parameter θ_i get under the alternative hypothesis, $\widehat{\sigma}_i^2$ the estimate of the variance of residuals. For a small T sample, the corresponding unbiased estimator[2] may be expressed as $\widehat{\sigma}_i^2 = \widehat{\varepsilon}_i' \widehat{\varepsilon}_i / \left(T - 2K - 1 \right)$. We propose here to express this Wald statistic as a ratio of quadratic forms in normal variables corresponding to the true population of residual (cf. appendix A). This expression is:

$$W_{i,T} = \left(T - 2K - 1 \right) \left(\frac{\widetilde{\varepsilon}_i' \Phi_i \widetilde{\varepsilon}_i}{\widetilde{\varepsilon}_i' M_i \widetilde{\varepsilon}_i} \right) \qquad i = 1, .., N \tag{5}$$

where the $(T, 1)$ vector $\widetilde{\varepsilon}_i = \varepsilon_i / \sigma_{\varepsilon,i}$ is distributed according $N(0, I_T)$ under assumption A_1. The matrix Φ_i and M_i are positive semi definite, symmetric and idempotent (T, T) matrix.

$$\Phi_i = Z_i \left(Z_i' Z_i \right)^{-1} R' \left[R \left(Z_i' Z_i \right)^{-1} R' \right]^{-1} R \left(Z_i' Z_i \right)^{-1} Z_i' \tag{6}$$

$$M_i = I_T - Z_i \left(Z_i' Z_i \right)^{-1} Z_i' \tag{7}$$

where I_T is the identity matrix of size T. The matrix M_i corresponds to the standard projection matrix of the linear regression analysis.

The issue is now to determine the distribution of the average statistic $W_{N,T}^{Hnc}$ under the null hypothesis of Homogenous Non Causality. For that, we first consider the asymptotic case where T and N tends to infinity, and in second part the case where T is fixed.

[2] It is also possible to use the standard formula of the Wald statistic by substituting the term $(T - 2K - 1)$ by T. However, several software (as Eviews) use this normalisation.

3 Asymptotic distribution

We propose to derive the asymptotic distribution of the average statistic $W_{N,T}^{Hnc}$ under the null hypothesis of non causality. For that, we consider the case of a sequential convergence when T tends to infinity and then N tends to infinity. This sequential convergence result can be deduced from the standard convergence result of the individual Wald statistic $W_{i,T}$ in a large T sample. In a non dynamic model, the normality assumption in A_1 would be sufficient to establish the fact for all T, the Wald statistic has a chi-squared distribution with K degrees of freedom. But in our dynamic model, this result can only be achieved asymptotically. Let us consider the expression (5). Given that under A_1 the least squares estimate $\widehat{\theta}_i$ is convergent, we know that $plim\, \varepsilon_i' M_i \varepsilon_i / (T-2K-1) = \sigma_{\varepsilon,i}^2$. It implies that:

$$\underset{T\to\infty}{\text{plim}} \frac{\widetilde{\varepsilon}_i' M_i \widetilde{\varepsilon}_i}{T-2K-1} = \underset{T\to\infty}{\text{plim}} \frac{1}{\sigma_{\varepsilon,i}^2} \left(\frac{\varepsilon_i' M_i \varepsilon_i}{T-2K-1} \right) = 1$$

Then, if the statistic $W_{i,T}$ has a limiting distribution, it is the same distribution of the statistics that results when the denominator is replaced by its limiting value, that is to say 1. Thus, $W_{i,T}$ has the same limiting distribution as $\widetilde{\varepsilon}_i' \Phi_i \widetilde{\varepsilon}_i$. Under assumption A_1, the vector $\widetilde{\varepsilon}_i$ is distributed across a $N(0, I_T)$. Since Φ_i is idempotent, the quadratic form $\widetilde{\varepsilon}_i' \Phi_i \widetilde{\varepsilon}_i$ is distributed as a chi-squared with a number of degrees of freedom equal to the rank of Φ_i. The rank of the symmetric idempotent matrix Φ_i is equal to its trace, that is to say K (cf. appendix A). Then, under the null hypothesis of non causality, each individual Wald statistic converges to a chi-squared distribution with K degrees of freedom:

$$W_{i,T} \underset{T\to\infty}{\overset{d}{\longrightarrow}} \chi^2(K) \qquad \forall i = 1,..,N \tag{8}$$

In other words, when T tends to infinity, individual statistics $\{W_{i,T}\}_{i=1}^N$ are identically distributed. They are also independent since under assumption A_2, residual ε_i and ε_j for $j \neq i$ are independent. To sum it up: if T tends to infinity individual Wald statistics $W_{i,T}$ are *i.i.d.* with $E(W_{i,T}) = K$ and $V(W_{i,T}) = 2K$. Then, the distribution of the average Wald statistic $W_{N,T}^{Hnc}$ when $T \to \infty$ first and then $N \to \infty$, can be deduced from a standard Lindberg-Levy central limit theorem.

Theorem 1 *Under assumption A_2, the individual $W_{i,T}$ statistics for $i = 1,..,N$ are identically and independently distributed with finite second order moments as $T \to \infty$, and therefore by Lindberg-Levy central limit theorem under the HNC null hypothesis, the average statistic W_{HNC}^{Hnc} sequentially converges in distribution.*

$$Z_{N,T}^{Hnc} = \sqrt{\frac{N}{2K}} \left(W_{N,T}^{Hnc} - K \right) \underset{T,N\to\infty}{\overset{d}{\longrightarrow}} N(0,1) \tag{9}$$

with $W_{N,T}^{Hnc} = (1/N) \sum_{i=1}^N W_{i,T}$, where $T, N \to \infty$ denotes the fact that $T \to \infty$ first and then $N \to \infty$.

For a large N and T sample, if the realization of the standardized statistic $Z_{N,T}^{Hnc}$ is superior in absolute mean to the normal corresponding critical value for a given level of risk, the homogeneous non causality (HNC) hypothesis is rejected. This asymptotic result may be useful in some macro panels. However, it should be extended to the case where T and N tend to infinity simultaneously.

4 Fixed T samples and semi-asymptotic distributions

Asymptotically, individual Wald statistics $W_{i,T}$ for each $i = 1, .., N$, converge toward an identical chi-squared distribution. However, this convergence result can not be achieved for any time dimension T, even if we assume the normality of residuals. The issue is then to show that for a fixed T dimension, individual Wald statistics have finite second order moments even they do not have the same distribution and they do not have a standard distribution.

Let us consider the expression (5) of $W_{i,T}$ under assumption A_1: this is a ratio of two quadratic forms in a standard normal vector. Magnus (1986) gives general conditions which insure that the expectations of a quadratic form in normal variables exists. Let us consider the moments $E\left[(x'Ax/x'Bx)^s\right]$, when x is normally distributed vector $N\left(0, \sigma^2 I_T\right)$, A is a symmetric (T,T) matrix and B a positive semi definite (T,T) matrix of rank $r \geq 1$. Let us denote Q a $(T, T-r)$ matrix of full column rank $T-r$ such that $BQ = 0$. If $r \leq T-1$, Magnus (1986)'s theorem identifies three conditions:

(i) If $AQ = 0$, then $E\left[(x'Ax/x'Bx)^s\right]$ exists for all $s \geq 0$.

(ii) If $AQ \neq 0$ and $Q'AQ = 0$, then $E\left[(x'Ax/x'Bx)^s\right]$ exists for $0 \leq s < r$ and does not exist for $s \geq r$.

(iii) If $Q'AQ \neq 0$, then $E\left[(x'Ax/x'Bx)^s\right]$ exists for $0 \leq s < r/2$ and does not exist for $s \geq r/2$.

These general conditions are done in the case where matrices A and B are deterministic. In our case, the corresponding matrices M_i and Φ_i are stochastic, even we assume that exogenous variables X_i are deterministic. However, given a fixed T sample, we propose here to apply these conditions to the corresponding realisation denoted m_i and ϕ_i. First, in our case the rank of the symmetric idempotent matrix m_i is equal to $T - 2K - 1$ (appendix A). Second, since the matrix m_i is the projection matrix associated to the realization z_i of Z_i, we have by construction $m_i z_i = 0$, where z_i of full column rank $2K + 1$, since $T - \text{rank}(m_i) = 2K + 1$ Then, for a given realisation ϕ_i by construction, the product $\phi_i z_i$ is different from zero since

$$\phi_i z_i = z_i \left(z_i' z_i\right)^{-1} R' \left[R \left(z_i' z_i\right)^{-1} R'\right]^{-1} R \neq 0$$

Besides, the product $z_i' \phi_i z_i$ is also different from zero, since

$$z_i' \phi_i z_i = R' \left[R \left(z_i' z_i\right)^{-1} R'\right]^{-1} R \neq 0$$

Then, the Magnus' theorem allows us to establish that $E\left[\left(\widetilde{\varepsilon}_i' \phi_i \widetilde{\varepsilon}_i\right) / \left(\widetilde{\varepsilon}_i' m_i \widetilde{\varepsilon}_i\right)\right]^s$ exists as soon as $0 \leq s < \text{rank}(m_i)/2$. We assume that this condition is also satisfied for $W_{i,T}$:

$$E\left[(W_{i,T})^s\right] = (T - 2K - 1)^s \, E\left[\left(\frac{\widetilde{\varepsilon}_i' \Phi_i \widetilde{\varepsilon}_i}{\widetilde{\varepsilon}_i' M_i \widetilde{\varepsilon}_i}\right)^s\right] \text{ exists if } 0 \leq s < \frac{T - 2K - 1}{2}$$

In particular, given the realizations of Φ_i and M_i, we can identify the condition on T which assures that second order moments ($s = 2$) of $W_{i,T}$ exists.

Proposition 2 *For a fixed time dimension $T \in \mathbb{N}$, the second order moments of the individual Wald statistic $W_{i,T}$ associated to the test $H_{0,i} : \beta_i = 0$, exist if and only if:*

$$T > 5 + 2K \tag{10}$$

Hence for a small T, individual Wald statistics $W_{i,T}$ are not necessarily identically distributed since matrices Φ_i and M_i are different from an individual to another. Besides, they do not have standard distribution as in previous section. However, the condition which insures the existence of second order moments are the same for all units. The second order moments of $W_{i,T}$ exist as soon as $T > 5 + 2K$ or equivalently $T \geq 6 + 2K$.

For a fixed T sample, the statistic of non causality test $W_{N,T}^{Hnc}$ is the average of non identically distributed variables $W_{i,T}$, but with finite second order moments under the condition of proposition 2. Under assumption A_2, residual ε_i and ε_j for $j \neq i$ are independent. Consequently, individual Wald $W_{i,T}$ for $i = 1, .., N$ are also independent. Then, the distribution of the non causality test statistic $W_{N,T}^{Hnc}$ can be derived according the Lyapunov central limit theorem.

Theorem 3 *Under assumption A_2, if $T > 5 + 2K$ the individual $W_{i,T}$ statistics $\forall i = 1, .., N$ are independently but not identically distributed with finite second order moments, and therefore by Lyapunov central limit theorem under the HNC null hypothesis, the average statistic W_{HNC}^{b} converges. If*

$$\lim_{N \to \infty} \left(\sum_{i=1}^{N} Var\left(W_{i,T}\right) \right)^{-\frac{1}{2}} \left(\sum_{i=1}^{N} E\left[\left| W_{i,T} - E\left(W_{i,T}\right) \right|^{3} \right] \right)^{\frac{1}{3}} = 0$$

the standardized statistic $Z_{N,T}^{Hnc}$ converges in distribution:

$$Z_{N,T}^{Hnc} = \frac{\sqrt{N} \left[W_{N,T}^{Hnc} - N^{-1} \sum_{i=1}^{N} E\left(W_{i,T}\right) \right]}{\sqrt{N^{-1} \sum_{i=1}^{N} Var\left(W_{i,T}\right)}} \underset{N \to \infty}{\xrightarrow{d}} N\left(0, 1\right) \tag{11}$$

with $W_{N,T}^{Hnc} = (1/N) \sum_{i=1}^{N} W_{i,T}$, where $E\left(W_{i,T}\right)$ and $Var\left(W_{i,T}\right)$ respectively denote the mean and the variance of the statistic $W_{i,T}$ defined by equation (5).

The decision of rule is the same as in the asymptotic case: if the realization of the standardized statistic $Z_{N,T}^{Hnc}$ is superior in absolute mean to the normal corresponding critical value for a given level of risk, the homogeneous non causality (HNC) hypothesis is rejected. For large T, the moments used in theorem (3) are expected to converge to $E\left(W_{i,T}\right) = K$ and $Var\left(W_{i,T}\right) = 2K$ since individual statistics $W_{i,T}$ converge in distribution to a chi-squared distribution with K degrees of freedom. Then, we find the conditions of the theorem 1. However, these asymptotic moment values could lead to poor test results, when we have small values of T. The issue is then to evaluate the mean and the variance of the Wald statistic $W_{i,T}$, whereas this statistic does not have a standard distribution for a fixed T.

We have seen in the previous section that the two first moments of $W_{i,T}$ exist, but what are their values? This point is particularly essential, since if we can not propose a general approximation of these moments, it will be necessary to compute these moments via stochastic simulations of the Wald under H_0. However, these simulations necessitate to specify an assumption on the parameters γ_i and α_i, but also on the DGP of exogenous variables $x_{i,t}$. In order to get more general results, we propose here an approximation of $E\left(W_{i,T}\right)$ and $Var\left(W_{i,T}\right)$ based on the results of Magnus (1986) theorem.

Let us consider the expression of the Wald $W_{i,T}$ as a ratio of two quadratic forms in a standard normal vector under assumption A_1:

$$W_{i,T} = (T - 2K - 1) \left(\frac{\widetilde{\varepsilon}_i' \Phi_i \widetilde{\varepsilon}_i}{\widetilde{\varepsilon}_i' M_i \widetilde{\varepsilon}_i} \right) \tag{12}$$

where the $(T, 1)$ vector $\widetilde{\varepsilon}_i = \varepsilon_i / \sigma_{\varepsilon,i}$ is distributed according $N(0, I_T)$ where matrices Φ_i and M_i are idempotent and symmetric (and consequently positive semi-definite). For a given T sample, let us denote respectively ϕ_i and m_i, the realizations of matrices Φ_i and M_i. We propose here to apply the Magnus (1986) theorem to the quadratic forms in a standard normal vector defined as:

$$\widetilde{W}_{i,T} = (T - 2K - 1) \left(\frac{\widetilde{\varepsilon}_i' \phi_i \widetilde{\varepsilon}_i}{\widetilde{\varepsilon}_i' m_i \widetilde{\varepsilon}_i} \right) \tag{13}$$

where matrices ϕ_i and m_i are idempotent and symmetric (and consequently positive semi-definite).

Theorem 4 (Magnus 1986) *Let $\widetilde{\varepsilon}_i$ be a normal distributed vector with $E(\widetilde{\varepsilon}_i) = 0$ and $E\left(\widetilde{\varepsilon}_i \widetilde{\varepsilon}_i'\right) = I_T$. Let P_i be an orthogonal (T, T) matrix and Λ_i a diagonal (T, T) matrix such that*

$$P_i' m_i P_i = \Lambda_i \qquad P_i' P_i = I_T \tag{14}$$

Then, we have, provided the expectation exists for $s = 1, 2, 3..$:

$$E\left[\left(\frac{\widetilde{\varepsilon}_i' \phi_i \widetilde{\varepsilon}_i}{\widetilde{\varepsilon}_i' m_i \widetilde{\varepsilon}_i} \right)^s \right] = \frac{1}{(s-1)!} \sum_v \gamma_s(v) \times \int_0^\infty \left\{ t^{s-1} |\Delta_i| \prod_{j=1}^s [trace\,(R_i)]^{n_j} \right\} dt \tag{15}$$

where the summation is over all $(s, 1)$ vectors $v = (n_1, .., n_s)$ whose elements n_j are nonnegative integers satisfying $\sum_{j=1}^s j n_j = s$

$$\gamma_s(v) = s!\, 2^s \prod_{j=1}^s \left[n_j! \, (2j)^{n_j} \right]^{-1} \tag{16}$$

and Δ_i is a diagonal positive definite (T, T) matrix and R_i a symmetric (T, T) matrix given by:

$$\Delta_i = (I_T + 2\, t\, \Lambda_i)^{-1/2} \qquad R_i = \Delta_i\, P_i' \phi_i P_i\, \Delta_i \tag{17}$$

In our case, we are interested by the two first moments. For the first order moment $(s = 1)$, there is only one scalar $v = n_1$ which is equal to one. Then, the quantity $\gamma_1(v)$ is equal to one. For the second order moment $(s = 2)$, there are two vectors $v = (n_1, n_2)$ which are respectively defined by $v_1 = (0, 1)$ and $v_2 = (2, 0)$. Consequently $\gamma_2(v_1) = 2$ and $\gamma_2(v_2) = 1$. Given these results, we can compute the exact two corresponding moments of the statistic $\widetilde{W}_{i,T}$ as:

$$E\left(\widetilde{W}_{i,T}\right) = (T - 2K - 1) \times \int_0^\infty |\Delta_i| \text{ trace } (R_i)\, dt \tag{18}$$

$$E\left[\left(\widetilde{W}_{i,T} \right)^2 \right] = (T - 2K - 1)^2 \times \left\{ 2 \int_0^\infty t\, |\Delta_i| \text{ trace } (R_i)\, dt + \int_0^\infty t\, |\Delta_i| \left[\text{trace } (R_i) \right]^2 dt \right\} \tag{19}$$

where matrices Δ_i and R_i are defined in theorem (4). Both quantities $|\Delta_i|$ and $\text{trace}\,(R_i)$ can be computed analytically in our model given the properties of these matrices. Since Λ_i is issued from the orthogonal decomposition of the idempotent matrix m_i, with $\text{rank}(m_i) = T - 2K - 1$

(cf. appendix A), this matrix is a zero except the first block which is equal to the $T-2K-1$ identity matrix (corresponding to the characteristic roots of m_i which are non nul). Then, for a scalar $t \in \mathbb{R}^+$, the matrix $\Delta_i = (I_T + 2\,t\,\Lambda_i)^{-1/2}$ can be partitioned as:

$$\underset{(T,T)}{\Delta_i} = \begin{pmatrix} \underset{(T-2K-1,T-2K-1)}{D_i(t)} & \underset{(T-2K-1,2K+1)}{0} \\ \underset{(2K+1,T-2K-1)}{0} & \underset{(2K+1,2K+1)}{I_{2K+1}} \end{pmatrix}$$

where I_p denotes the identity matrix of size p. The diagonal block $D_i(t)$ is defined as $D_i(t) = (1+2t)^{-\frac{1}{2}} I_{T-2K-1}$. Then, the determinant of Δ_i can be expressed as:

$$|\Delta_i| = (1+2t)^{-\left(\frac{T-2K-1}{2}\right)} \tag{20}$$

Besides, the trace of the matrix R_i can be computed as follows. Since for any non singular matrices B and C, the rank of BAC is equal to rank of A, we have here:

$$\text{rank}(R_i) = \text{rank}(\Delta_i\, P_i' \phi_i P_i\, \Delta_i) = \text{rank}(\,P_i' \phi_i P_i)$$

since the matrix Δ_i is non singular. With the same transformation, given the non singularity of P_i, we get:

$$\text{rank}(R_i) = \text{rank}(\,P_i' \phi_i P_i) = \text{rank}(\phi_i)$$

Finally, the rank of the realisation ϕ_i is equal to K, the rank of Φ_i (cf. appendix A).

$$\text{trace}(R_i) = K$$

Given these results, the two first moments (equations 18 and 19) of the statistic $\widetilde{W}_{i,T}$ based for a given T sample on realizations ϕ_i and m_i, can be expressed as:

$$E\left(\widetilde{W}_{i,T}\right) = (T-2K-1) \times K \times \int_0^\infty (1+2t)^{-\left(\frac{T-2K-1}{2}\right)}\; dt$$

$$E\left[\left(\widetilde{W}_{i,T}\right)^2\right] = (T-2K-1)^2 \times \left(2K+K^2\right) \times \int_0^\infty t\;\; (1+2t)^{-\left(\frac{T-2K-1}{2}\right)}\; dt$$

Then, we get the following results.

Proposition 5 *For a fixed T sample, where T satisfies the condition of proposition (2), given realizations ϕ_i and m_i of matrices Φ_i and M_i (equations 6 and 7), the exact two first moments of the individual statistics $\widetilde{W}_{i,T}$, for $i = 1, ..., N$, defined by equation (13) are respectively:*

$$E\left(\widetilde{W}_{i,T}\right) = K \times \frac{(T-2K-1)}{(T-2K-3)} \tag{21}$$

$$Var\left(\widetilde{W}_{i,T}\right) = 2K \times \frac{(T-2K-1)^2 \times (T-K-3)}{(T-2K-3)^2 \times (T-2K-5)} \tag{22}$$

as soon as the time dimension T satisfies $T \geq 6 + 2K$.

The proof of this proposition is done in appendix B. It is important to verify that for large T sample, the moments of the individual statistic $\widetilde{W}_{i,T}$ tend to the corresponding moments of the asymptotic distribution of $W_{i,T}$ since $\forall i = 1, ..., N$:

$$\lim_{T\to\infty} E\left(\widetilde{W}_{i,T}\right) = K \qquad \lim_{T\to\infty} Var\left(\widetilde{W}_{i,T}\right) = 2K$$

Both moments correspond to the moments of a $F(K, T-2K-1)$. Indeed, in this dynamic model the F distribution can be used as an approximation of the true distribution of the statistic $W_{i,T}/K$ for a small T sample. Then, the use of the Magnus theorem given the realizations ϕ_i and m_i to approximate the true moments of the Wald statistic is equivalent to assert that the true distribution of $W_{i,T}$ can be approximated by the F distribution.

For a given $T \geq 6+2K$ sample, we propose in this paper to approximate the two first moments of the individual Wald statistic $W_{i,T}$ by the two first moments (equation 21 and 22) of the statistics $\widetilde{W}_{i,T}$ based on the realizations ϕ_i and m_i of the stochastic matrices Φ_i and M_i. More precisely, we propose to approximate the two quantities used in theorem 4 for the finite T distribution of the average statistic $W_{N,T}^{Hnc}$. If $T \geq 6+2K$, the following approximations are supposed to be valid:

$$N^{-1}\sum_{i=1}^{N} E\left(W_{i,T}\right) \simeq N^{-1}\sum_{i=1}^{N} E\left(\widetilde{W}_{i,T}\right) = K \times \frac{(T-2K-1)}{(T-2K-3)}$$

$$N^{-1}\sum_{i=1}^{N} Var\left(W_{i,T}\right) \simeq N^{-1}\sum_{i=1}^{N} Var\left(\widetilde{W}_{i,T}\right) = 2K \times \frac{(T-2K-1)^2 \times (T-K-3)}{(T-2K-3)^2 \times (T-2K-5)}$$

Given these approximations, we propose to compute an approximated standardized statistic $\widetilde{Z}_{N,T}^{Hnc}$ for the average Wald average statistic $W_{N,T}^{Hnc}$ of the HNC hypothesis.

$$\widetilde{Z}_{N,T}^{Hnc} = \frac{\sqrt{N}\left[W_{N,T}^{Hnc} - E\left(\widetilde{W}_{i,T}\right)\right]}{\sqrt{Var\left(\widetilde{W}_{i,T}\right)}} \tag{23}$$

After simplifications, the standardized statistic $\widetilde{Z}_{N,T}^{Hnc}$ is defined as:

$$\widetilde{Z}_{N,T}^{Hnc} = \sqrt{\frac{N}{2 \times K} \times \frac{(T-2K-5)}{(T-K-3)}} \times \left[\frac{(T-2K-3)}{(T-2K-1)} W_{N,T}^{Hnc} - K\right] \tag{24}$$

For a large N sample, under the Homogenous Non Causality (HNC) hypothesis, we assume that the statistic $\widetilde{Z}_{N,T}^{Hnc}$ follow approximately the same distribution as the standardized average Wald statistic $Z_{N,T}^{Hnc}$.

Proposition 6 *Under assumptions A_1 and A_2, for a fixed T dimension with $T > 5+2K$, the approximated standardized statistic $\widetilde{Z}_{N,T}^{Hnc}$ converges in distribution:*

$$\widetilde{Z}_{N,T}^{Hnc} = \sqrt{\frac{N}{2 \times K} \times \frac{(T-2K-5)}{(T-K-3)}} \times \left[\frac{(T-2K-3)}{(T-2K-1)} W_{N,T}^{Hnc} - K\right] \underset{N \to \infty}{\xrightarrow{d}} N(0,1) \tag{25}$$

with $W_{N,T}^{Hnc} = (1/N)\sum_{i=1}^{N} W_{i,T}$.

The test of the HNC hypothesis is built[3] as follows. For each individual of the panel, we compute the standard Wald statistics $W_{i,T}$ associated to the individual hypothesis $H_{0,i} : \beta_i = 0$ with $\beta_i \in \mathbb{R}^K$ Given these N realizations, we get a realization of the average Wald statistic

[3] If one uses the standard definition of the Wald statistic with the T normalization, it is necessary to adapt the formula (25) by substituting the quantity $T-2K-1$ by T. More precisely, if the Wald individual statistic

$W_{N,T}^{Hnc}$. Given the formula (25) we compute the realization of the approximated standardized statistic $\widetilde{Z}_{N,T}^{Hnc}$ for the T and K values. For a large N sample, if the value of $\widetilde{Z}_{N,T}^{Hnc}$ is superior in absolute mean to the normal corresponding critical value for a given level of risk, the homogeneous non causality (HNC) hypothesis is rejected.

5 Monte Carlo simulation results

In this section, we propose Monte Carlo experiments to examine the accuracy of the approximated standardized statistic $\widetilde{Z}_{N,T}^{Hnc}$. The model used is:

$$y_{i,t} = \alpha_i + \sum_{k=1}^{K} \gamma_i^{(k)} y_{i,t-k} + \sum_{k=1}^{K} \beta_i^{(k)} x_{i,t-k} + \varepsilon_{i,t} \tag{26}$$

considered under H_0, that is under the assumption of Homogenous Non Causality hypothesis $\beta_i = 0$. The others parameters of the model are calibrated as follows. The auto-regressive parameters $\gamma_i^{(k)}$ are drawn according to a uniform distribution on $[-10, 10]$ under the constraint that the roots of $\Gamma(z) = \sum_{k=1}^{K} \gamma_i^{(k)} z^k$ lie outside the unit circle in order to satisfy the assumption A_3. The fixed individual effects α_i, $i = 1, .., N$ are simulated according a $N(0, 1)$. Individual residuals are drawn in normal distribution (assumption A_1) with zero means and heterogeneous variances $\sigma_{\varepsilon,i}^2$. The variance $\sigma_{\varepsilon,i}^2$ are generated according a uniform distribution on $[0.5, 1.5]$.

In the first set of experiments, we compute by Monte Carlo simulations two estimates of the mean and the variance of the Wald statistic $W_{i,T}$ under the non causality hypothesis for a given unit i with a residual variance $\sigma_{\varepsilon,i}^2$. The explanatory variable $x_{i,t}$ is simulated according a $N(0, 1)$ distribution for each set of data. The sequence $\{y_{i,t}\}_{t=1}^{T}$ is simulated under the HNC hypothesis $\beta_i = 0$ for a particular realization of normal residuals. Given these pseudo samples, we compute the Wald statistic $W_{i,T}$ associated to the individual non causality test. For different T sample size and lag order K, we use 50 000 replications of the Wald statistics to compute the empirical mean and variance of $W_{i,T}$. The results of the simulated moments are reported in table 1. We can verify that for large T sample, the simulated mean and variance converge to their asymptotic respective values, that is K and $2K$.

For different values of T and N, we compare the simulated moments of $W_{i,T}$ to the exact moments of $\widetilde{W}_{i,T}$. These exact moments computed from (21) and (22) are reported on table 2. As we can observe there is few differences between simulated moments of $W_{i,T}$ (table 1) and the exact moments of $\widetilde{W}_{i,T}$ (table 2). For instance, for $T = 15$ and $K = 1$, the simulated mean is 1.21 whereas the exact mean is 1.20. The simulated variance is 3.97 and the exact one is 3.96. Even with a higher lag-order, the approximations are quite well. With $K = 3$ and $T = 25$ the simulated and exact mean (variance) are respectively 3.37 and 3.38 (10.17 and 10.31). For larger lag-order and very short T sample (for instance $T = 6 + 2K$ the minimum size for the existence of second order moment), the approximation is less accurate.

$W_{i,T}$ is defined as:

$$W_{i,T} = \left\{ \widehat{\theta}_i' R' \left[R \left(Z_i' Z_i \right)^{-1} R' \right]^{-1} R \widehat{\theta}_i \right\} \;/\; \left[\widehat{\varepsilon}_i' \widehat{\varepsilon}_i / T \right]$$

then the standardize average Wald statistic $\widetilde{Z}_{N,T}^{Hnc}$ is defined as:

$$\widetilde{Z}_{N,T}^{Hnc} = \sqrt{\frac{N}{2 \times K} \times \frac{(T-4)}{(T+K-2)}} \times \left[\left(\frac{T-2}{T} \right) W_{N,T}^{Hnc} - K \right]$$

6 Fixed T and fixed N distributions

If N and T are fixed, the standardized statistic $\widetilde{Z}_{N,T}^{Hnc}$ and the average statistic $W_{N,T}^{Hnc}$ do not converge to standard distributions under the HNC hypothesis. Two solutions are then possible: the first consists in using the mean Wald statistic $W_{N,T}^{Hnc}$ and to compute the exact sample critical values, denoted $c_{N,T}(\alpha)$, for the corresponding sizes N and T via stochastic simulations. We propose the results of an example of such a simulation in table 3. As in Im, Pesaran and Shin (2002), the second solution consists in using the approximated standardized statistic $\widetilde{Z}_{N,T}^{Hnc}$ and to compute an approximation of the corresponding critical value for a fixed N. Indeed, we can show that:

$$Prob\left(\widetilde{Z}_{N,T}^{Hnc} < \widetilde{z}_{N,T}(\alpha)\right) = Prob\left(W_{N,T}^{Hnc} < c_{N,T}(\alpha)\right)$$

where $\widetilde{z}_{N,T}(\alpha)$ is the α percent critical value of the distribution of the standardized statistic under the HNC hypothesis. The critical value $c_{N,T}(\alpha)$ of $W_{N,T}^{Hnc}$ is defined as:

$$c_{N,T}(\alpha) = \widetilde{z}_{N,T}(\alpha)\sqrt{N^{-1}var\left(\widetilde{W}_{i,T}\right)} + E\left(\widetilde{W}_{i,T}\right)$$

where $E\left(\widetilde{W}_{i,T}\right)$ and $Var\left(\widetilde{W}_{i,T}\right)$ respectively denote the mean and the variance of the individual Wald statistic defined by equations (21) and (22). Given the result of proposition (6), we know that the critical value $\widetilde{z}_{N,T}(\alpha)$ corresponds to the α percent critical value of the standard normal distribution, denoted z_α if N tends to infinity whatever the size T. For a fixed N, the use of the normal critical value z_α to built the corresponding critical value $c_{N,T}(\alpha)$ is not founded, but however we can propose an approximation $\widetilde{c}_{N,T}(\alpha)$ based on this value.

$$\widetilde{c}_{N,T}(\alpha) = z_\alpha\sqrt{N^{-1}var\left(\widetilde{W}_{i,T}\right)} + E\left(\widetilde{W}_{i,T}\right) \tag{27}$$

or equivalently:

$$\widetilde{c}_{N,T}(\alpha) = z_\alpha \times \frac{(T-2K-1)}{(T-2K-3)} \times \sqrt{\frac{2K}{N} \times \frac{(T-K-3)}{(T-2K-5)}} + \frac{K \times (T-2K-1)}{(T-2K-3)} \tag{28}$$

On the table (4), the simulated 5% critical values $c_{N,T}(0.05)$ get from 50 000 replications of the model under H_0 with $K = 1$ are reproduced (cf. table 3). The approximated 5% critical values $\widetilde{c}_{N,T}(0.05)$ are also reported where the corresponding values are get from the equation (28). As we can observe, both critical values are very similar: the same result can be obtained for larger lag-order K.

7 Conclusion

In this paper, we propose a simple Granger (1969) causality test in heterogenous panel data models with fixed coefficients. Under the null hypothesis, there is no causal relationship for all the units of the panel. We specify the alternative as the $HENC$ hypothesis. There is two subgroups of units: one with causal relationships from x to y, but not necessarily with the same DGP, and an another subgroup where there is no causal relationships from x to y.

As in the unit root literature, our statistic of test is the average of individual Wald statistics associated to the standard Granger causality tests based on time series. We derive the asymptotic distribution of this statistic when T and N tend sequentially to infinity. For fixed T sample, the semi-asymptotic distribution is characterized. In this case, individual Wald statistics do not have a standard chi-squared distribution. However, under very general setting, it is shown that individual Wald statistics are independently distributed with finite second order moments as soon as $T > 5 + 2K$, where K denotes the number of linear restrictions. For a fixed T, the Lyapunov central limit theorem is sufficient to get the distribution of the standardized average Wald statistic when N tends to infinity. The two first moments of this normal semi-asymptotic distribution correspond to empirical mean of the corresponding theoretical moments of the individual Wald statistics.

The issue is then to propose an evaluation of the two first moments of standard Wald statistics for small T sample. In this paper we propose a general approximation of these moments based on the exact moments of the ratio of quadratic forms in normal variables derived in the Magnus (1986) theorem. For a fixed T sample, we propose two approximations of the mean and the variance of the Wald statistic. Monte Carlo experiments show that these formulas provide an excellent approximation to the true moments. Given these approximations, we propose an approximated standardized average Wald statistic to test the HNC hypothesis in short T sample. Finally, approximated critical values are proposed for finite N sample.

Our aim is now to study the size and the power of our panel Granger causality test in several configurations. When there is at least one parameter in the dynamics of the endogenous variable which is common to all individual, it is evident that our panel test would more powerful than individual tests lead on individual times series. However, in more general cases the results are no so obvious.

Appendix

A Individual Wald statistics

The individual Wald statistic $W_{i,T}$ associated to the test $H_0 : \beta_i = 0$, which can be expressed as $R\theta_i = 0$, is defined as, $\forall i = 1, .., N$:

$$W_{i,T} = \frac{\widehat{\theta}_i' R' \left[R \left(Z_i' Z_i \right)^{-1} R' \right]^{-1} R \widehat{\theta}_i}{\widehat{\varepsilon}_i' \widehat{\varepsilon}_i / \left(T - 2K - 1 \right)}$$

where $\widehat{\theta}_i$ and $\widehat{\varepsilon}_i$ respectively denote a convergent estimate of θ_i and the estimated residuals for the cross section unit i get under $H_{1,i} : \beta_i \neq 0$. First, we can express the residual sum of squares $\widehat{\varepsilon}_i' \widehat{\varepsilon}_i$ as a quadratic form defined in the true population of residual ε_i. For that, we introduce the standard projection matrix M_i.

$$\widehat{\varepsilon}_i' \widehat{\varepsilon}_i = \varepsilon_i' \left[I_T - Z_i \left(Z_i' Z_i \right)^{-1} Z_i' \right] \varepsilon_i = \varepsilon_i' M_i \varepsilon_i$$

where ε_i is the true population of residual for the unit i of the model (1). The numerator of the Wald statistic is also defined as a quadratic form in the same normal vector ε_i, if we consider the expression of $\widehat{\theta}_i = \theta_i + \left(Z_i' Z_i \right)^{-1} Z_i' \varepsilon_i$. Since under $H_{0,i}$ we have $R\theta_i = 0$, we get $R\widehat{\theta}_i = R \left(Z_i' Z_i \right)^{-1} Z_i' \varepsilon_i$ and consequently:

$$\widehat{\theta}_i' R' \left[R \left(Z_i' Z_i \right)^{-1} R' \right]^{-1} R \widehat{\theta}_i = \varepsilon_i' Z_i' \left(Z_i' Z_i \right)^{-1} R' \left[R \left(Z_i' Z_i \right)^{-1} R' \right]^{-1} R \left(Z_{ci}' Z_{ci} \right)^{-1} Z_{ci}' \varepsilon_i$$

The Wald statistic is then defined as a ratio of quadratic form defined in a normal $N \left(0, I_T \right)$ vector $\widetilde{\varepsilon}_i = \varepsilon_i / \sigma_{\varepsilon,i}$ under assumption A_1.

$$\frac{W_{i,T}}{T - 2K - 1} = \frac{\varepsilon_i' \Phi_i \varepsilon_i}{\varepsilon_i' M_i \varepsilon_i} = \frac{\widetilde{\varepsilon}_i' \Phi_i \widetilde{\varepsilon}_i}{\widetilde{\varepsilon}_i' M_i \widetilde{\varepsilon}_i} \tag{29}$$

where the matrices M_i and Φ_i are idempotent, symmetric and consequently semi positive definite.

$$\Phi_i = Z_i \left(Z_i' Z_i \right)^{-1} R' \left[R \left(Z_i' Z_i \right)^{-1} R' \right]^{-1} R \left(Z_i' Z_i \right)^{-1} Z_i' \tag{30}$$

$$M_i = I_T - Z_i \left(Z_i' Z_i \right)^{-1} Z_i' \tag{31}$$

where $Z_i = [e : Y_i : X_i]$ and R are respectively $(T, 2K+1)$ and $(K, 2K+1)$. The rank of the symmetric idempotent matrix Φ_i is equal to its trace, which is equal to K since:

$$\begin{aligned}
\text{trace}(\Phi_i) &= \text{trace} \left\{ Z_i \left(Z_i' Z_i \right)^{-1} R' \left[R \left(Z_i' Z_i \right)^{-1} R' \right]^{-1} R \left(Z_i' Z_i \right)^{-1} Z_i' \right\} \\
&= \text{trace} \left\{ \left[R \left(Z_i' Z_i \right)^{-1} R' \right]^{-1} R \left(Z_i' Z_i \right)^{-1} Z_i' Z_i \left(Z_i' Z_i \right)^{-1} R' \right\} \\
&= \text{trace} \left\{ \left[R \left(Z_i' Z_i \right)^{-1} R' \right]^{-1} R \left(Z_i' Z_i \right)^{-1} R' \right\} \\
&= \text{trace} \left(I_K \right)
\end{aligned}$$

The rank of the projection matrix M_i is also equal to its trace :

$$\text{trace}(M_i) = \text{trace} \left[I_T - Z_i \left(Z_i' Z_i \right)^{-1} Z_i' \right]$$

$$
\begin{aligned}
&= \operatorname{trace}(I_T) - \operatorname{trace}\left[Z_i \left(Z_i' Z_i\right)^{-1} Z_i'\right] \\
&= \operatorname{trace}(I_T) - \operatorname{trace}\left[Z_i' Z_i \left(Z_i' Z_i\right)^{-1}\right] \\
&= \operatorname{trace}(I_T) - \operatorname{trace}(I_{2K+1}) \\
&= T - 2K - 1
\end{aligned}
$$

B Exact moments of individual Wald $\widetilde{W}_{i,T}$

The two noncentered moments of $\widetilde{W}_{i,T}$ are respectively defined as:

$$E\left(\widetilde{W}_{i,T}\right) = (T-2K-1) \times K \times \int_0^\infty (1+2t)^{-\left(\frac{T-2K-1}{2}\right)} \, dt$$

$$E\left[\left(\widetilde{W}_{i,T}\right)^2\right] = (T-2K-1)^2 \times \left(2K+K^2\right) \times \int_0^\infty t \; (1+2t)^{-\left(\frac{T-2K-1}{2}\right)} \, dt$$

Let us denote for simplicity $\widetilde{T} = (T-2K-1)/2$. For the first order moment, we get:

$$
\begin{aligned}
E\left(\widetilde{W}_{i,T}\right) &= 2\widetilde{T} \times K \times \int_0^\infty (1+2t)^{-\widetilde{T}} \, dt \\
&= 2\widetilde{T} \times K \times \left[\frac{(1+2t)^{-\widetilde{T}+1}}{2\left(-\widetilde{T}+1\right)}\right]_0^\infty \\
&= \frac{2\widetilde{T} \times K}{2\left(\widetilde{T}-1\right)}
\end{aligned}
$$

Since the quantity $2\left(\widetilde{T}-1\right) = T-2K-3$ is strictly different from zero under the condition of proposition (2), we get

$$E\left(\widetilde{W}_{i,T}\right) = K \times \frac{(T-2K-1)}{(T-2K-3)} \tag{32}$$

For the second order moment, the definition is:

$$E\left[\left(\widetilde{W}_{i,T}\right)^2\right] = 4\,\widetilde{T}^2 \times \left(2K+K^2\right) \times \int_0^\infty t \; (1+2t)^{-\widetilde{T}} \, dt$$

By integrating by parts, this expression can be transformed as:

$$E\left[\left(\widetilde{W}_{i,T}\right)^2\right] = 4\,\widetilde{T}^2 \times \left(2K+K^2\right) \times \left\{\left[\frac{t \times (1+2t)^{-\widetilde{T}+1}}{2\left(-\widetilde{T}+1\right)}\right]_0^\infty - \frac{1}{2\left(-\widetilde{T}+1\right)} \times \int_0^\infty (1+2t)^{-\widetilde{T}} \, dt\right\}$$

Under under the condition of proposition (2) we have $\widetilde{T} > 1$, then:

$$
\begin{aligned}
E\left[\left(\widetilde{W}_{i,T}\right)^2\right] &= \frac{4\widetilde{T}^2 \times \left(2K+K^2\right)}{2\left(\widetilde{T}-1\right)} \times \int_0^\infty (1+2t)^{-\widetilde{T}} \, dt \\
&= \frac{4\widetilde{T}^2 \times \left(2K+K^2\right)}{2\left(\widetilde{T}-1\right)} \times \left[\frac{(1+2t)^{-\widetilde{T}+2}}{2\left(-\widetilde{T}+2\right)}\right]_0^\infty \\
&= \frac{4\widetilde{T}^2 \times \left(2K+K^2\right)}{2\left(\widetilde{T}-1\right)} \times \frac{1}{2\left(\widetilde{T}-2\right)}
\end{aligned}
$$

After simplifications, we have :

$$E\left[\left(\widetilde{W}_{i,T}\right)^{2}\right]=\frac{\widetilde{T}^{2}\times\left(2K+K^{2}\right)}{\left(\widetilde{T}-1\right)\left(\widetilde{T}-2\right)}=\frac{\left(T-2K-1\right)^{2}\times\left(2K+K^{2}\right)}{\left(T-2K-3\right)\left(T-2K-5\right)} \tag{33}$$

Under the condition $T>5+2K$, this second order moment exists as it was previously established in proposition (2).

Finally, we can compute the second order centered moment, $Var\left(\widetilde{W}_{i,T}\right)$ as:

$$\begin{aligned} Var\left(\widetilde{W}_{i,T}\right) &= E\left[\left(\widetilde{W}_{i,T}\right)^{2}\right]-E\left(\widetilde{W}_{i,T}\right)^{2} \\ &= \frac{\left(T-2K-1\right)^{2}\times\left(2K+K^{2}\right)}{\left(T-2K-3\right)\left(T-2K-5\right)}-\left[\frac{K\times\left(T-2K-1\right)}{\left(T-2K-3\right)}\right]^{2} \end{aligned}$$

After simplifications, we have:

$$Var\left(\widetilde{W}_{i,T}\right)=2\,K\times\frac{\left(T-2K-1\right)^{2}\times\left(T-K-3\right)}{\left(T-2K-3\right)^{2}\left(T-2K-5\right)} \tag{34}$$

References

[1] Amemiya, T. (1985), "Advanced Econometrics", *Harvard University Press*

[2] Granger, C.W.J (1969), "Investigating causal relations by econometric models and cross-spectral methods", *Econometrica*, 37(3),424-438.

[3] Granger, C.W.J (1980), "Testing for causality", *Journal of Economic Dynamics and Control*,2, 329-352

[4] Granger C.W.J (2003), "Some aspects of causal relationships", *Journal of Econometrics*, 112, 69-71

[5] Holtz-Eakin D., Newey W. and Rosen H.S. (1988), "Estimating vector autoregressions with panel data", *Econometrica*, 56, 1371-1396.

[6] Hsiao, C. (1986), "Analysis of panel data", *Econometric society Monographs N0 11.* Cambridge University Press

[7] Hurlin C. (2001), "A note on causality tests in panel data models with random coefficients", *Mimeo*

[8] Hurlin C., and Venet B., (2001), "Granger causality tests in panel data models with fixed coefficients", *Working Paper Eurisco* 2001-09, Université Paris IX Dauphine

[9] Im, K.S., Pesaran, M.H., and Shin, Y. (2002), "Testing for Unit Roots in Heterogenous Panels", revised version of the *DAE, Working Paper* 9526, University of Cambridge.

[10] Judson, R.A., and Owen, A.L. (1999), "Estimating dynamic panel data models: a guide for macroeconomists", *Economic Letters*, 65, 9-15.

[11] Magnus, J.R. (1986), "The exact moments of a ratio of quadratic forms in normal variables", *Annales d'Economie et de Statistiques*, 4, 96-109.

[12] 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

[13] Pesaran, H.M., and. Smith, R (1995), "Estimating long-run relationships from dynamic heterogenous panels", *Journal of Econometrics*, 68, 79-113

[14] Swamy, P.A. (1970), "Efficient inference in a random coefficient regression model", *Econometrica*, 38, 311-323

[15] 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.

Table 1: Simulated Moments of Individual Wald $W_{i,T}$ Statistics

	$K=1$		$K=2$		$K=3$		$K=4$	
T	$E(W_{i,T})$	$V(W_{i,T})$	$E(W_{i,T})$	$V(W_{i,T})$	$E(W_{i,T})$	$V(W_{i,T})$	$E(W_{i,T})$	$V(W_{i,T})$
8	1.68	18.09	—	—	—	—	—	—
9	1.52	11.02	—	—	—	—	—	—
10	1.40	8.31	3.25	44.30	—	—	—	—
11	1.32	5.82	2.93	22.86	—	—	—	—
12	1.29	5.26	2.75	16.57	4.76	63.00	—	—
13	1.25	4.64	2.63	14.55	4.41	40.92	—	—
14	1.23	4.18	2.56	12.59	4.09	28.49	6.56	360.65
15	1.21	3.97	2.50	11.11	3.94	23.34	5.89	65.47
16	1.18	3.79	2.45	9.41	3.81	19.52	5.51	43.62
17	1.17	3.64	2.40	8.66	3.71	15.68	5.19	30.59
18	1.15	3.30	2.37	8.16	3.63	14.56	5.10	27.29
19	1.14	3.24	2.32	7.60	3.57	14.29	4.99	25.21
20	1.12	3.02	2.29	7.40	3.52	13.30	4.88	21.69
25	1.09	2.81	2.23	6.19	3.37	10.17	4.58	15.59
30	1.08	2.69	2.16	5.48	3.33	9.43	4.43	12.82
40	1.06	2.39	2.12	5.11	3.23	7.71	4.27	11.16
45	1.05	2.42	2.12	4.95	3.16	7.73	4.22	10.37
50	1.03	2.31	2.08	4.71	3.17	7.58	4.22	10.28
100	1.03	2.17	2.02	4.27	3.07	6.93	4.08	9.12
500	1.00	2.05	1.99	4.42	3.00	5.99	4.02	8.16
1000	1.00	2.00	1.99	3.95	3.00	6.16	4.01	8.53
∞	1.00	2.00	2.00	4.00	3.00	6.00	4.00	8.00

Notes: The auto-regressive parameters $\gamma_i^{(k)}$ are drawn according to a uniform distribution on $[-10, 10]$ under the constraint that the roots of $\Gamma(z) = \sum_{k=1}^{K} \gamma_i^{(k)} z^k$ lie outside the unit circle. The fixed individual effects α_i, $i = 1, .., N$ are simulated according a $N(0,1)$ and the exogenous variables $x_{i,t}$ used to estimate the model under H_1 are simulated in a $N(0,1)$ distribution. The empirical moments of the Wald statistic $W_{i,T}$ are evaluated under the null HNC hypothesis.

Table 2: Exact Moments of the Statistic $\widetilde{W}_{i,T}$

	$K=1$		$K=2$		$K=3$		$K=4$	
T	$\mathrm{E}(\mathrm{W}_{i,T})$	$\mathrm{V}(\mathrm{W}_{i,T})$	$\mathrm{E}(\mathrm{W}_{i,T})$	$\mathrm{V}(\mathrm{W}_{i,T})$	$\mathrm{E}(\mathrm{W}_{i,T})$	$\mathrm{V}(\mathrm{W}_{i,T})$	$\mathrm{E}(\mathrm{W}_{i,T})$	$\mathrm{V}(\mathrm{W}_{i,T})$
8	1.67	22.22	—	—	—	—	—	—
9	1.50	11.25	—	—	—	—	—	—
10	1.40	7.84	3.33	55.56	—	—	—	—
11	1.33	6.22	3.00	27.00	—	—	—	—
12	1.29	5.29	2.80	18.29	5.00	100.00	—	—
13	1.25	4.69	2.67	14.22	4.50	47.25	—	—
14	1.22	4.27	2.57	11.90	4.20	31.36	6.67	155.56
15	1.20	3.96	2.50	10.42	4.00	24.00	6.00	72.00
16	1.18	3.72	2.44	9.39	3.86	19.84	5.60	47.04
17	1.17	3.54	2.40	8.64	3.75	17.19	5.33	35.56
18	1.15	3.39	2.36	8.07	3.67	15.37	5.14	29.09
19	1.14	3.27	2.33	7.62	3.60	14.04	5.00	25.00
20	1.13	3.16	2.31	7.26	3.55	13.04	4.89	22.19
25	1.10	2.82	2.22	6.17	3.38	10.31	4.57	15.67
30	1.08	2.64	2.17	5.63	3.29	9.09	4.42	13.22
40	1.06	2.44	2.12	5.08	3.19	7.97	4.28	11.17
45	1.05	2.38	2.11	4.92	3.17	7.67	4.24	10.65
50	1.04	2.33	2.09	4.81	3.15	7.45	4.21	10.28
100	1.02	2.15	2.04	4.36	3.07	6.62	4.09	8.94
500	1.00	2.03	2.01	4.07	3.01	6.11	4.02	8.16
1000	1.00	2.01	2.00	4.03	3.01	6.05	4.01	8.08
∞	1.00	2.00	2.00	4.00	3.00	6.00	4.00	8.00

Notes: The exact moments of statistic $\widetilde{W}_{i,T}$ for different values of T and K are computed according equations (21) and (22) of proposition 5.

Table 3: Exact Critical Values for $W_{N,T}^{Hnc}$ with a lag order $K = 1$.

$N \backslash T$	10	15	20	25	30	40	50	70	100
					1% Critical Values				
5	5.98	4.30	3.86	3.62	3.54	3.37	3.28	3.24	3.18
10	4.54	3.27	2.88	2.77	2.63	2.55	2.51	2.44	2.40
15	3.80	2.78	2.52	2.41	2.31	2.25	2.21	2.14	2.10
20	3.37	2.56	2.31	2.19	2.12	2.04	2.03	1.98	1.94
25	3.21	2.37	2.15	2.06	1.99	1.93	1.90	1.86	1.84
50	2.62	1.99	1.82	1.73	1.71	1.65	1.62	1.59	1.57
100	2.21	1.72	1.59	1.53	1.50	1.46	1.44	1.41	1.39
					5% Critical Values				
5	3.54	2.87	2.66	2.53	2.47	2.39	2.36	2.33	2.28
10	2.97	2.38	2.19	2.10	2.04	1.98	1.95	1.91	1.88
15	2.68	2.15	1.99	1.91	1.85	1.80	1.77	1.74	1.71
20	2.49	2.01	1.86	1.79	1.75	1.69	1.67	1.64	1.62
25	2.40	1.92	1.78	1.71	1.66	1.62	1.60	1.57	1.55
50	2.10	1.70	1.57	1.52	1.49	1.45	1.42	1.40	1.39
100	1.89	1.54	1.44	1.39	1.36	1.33	1.31	1.29	1.27
					10% Critical Values				
5	2.74	2.30	2.16	2.08	2.02	1.98	1.95	1.92	1.90
10	2.43	2.01	1.87	1.80	1.76	1.71	1.69	1.66	1.64
15	2.26	1.87	1.74	1.68	1.64	1.60	1.57	1.55	1.52
20	2.15	1.78	1.66	1.60	1.57	1.52	1.50	1.48	1.46
25	2.08	1.72	1.60	1.55	1.51	1.47	1.45	1.42	1.41
50	1.89	1.57	1.46	1.41	1.38	1.35	1.33	1.31	1.30
100	1.75	1.46	1.37	1.32	1.29	1.26	1.25	1.22	1.21

Notes: The exact critical values for the average statistic $W_{N,T}^{Hnc}$ are computed via stochastic simulations with 50 000 replications. The individual Wald statistics $W_{i,T}$ are built under the HNC hypothesis, where the auto-regressive parameters $\gamma_i^{(k)}$ are drawn according to a uniform distribution on $[-10, 10]$ under the constraint that the roots of $\Gamma(z) = \sum_{k=1}^{K} \gamma_i^{(k)} z^k$ lie outside the unit circle. The fixed individual effects α_i, $i = 1, .., N$ are simulated according a $N(0,1)$ and the exogenous variables $x_{i,t}$ used to estimate the model under H_1 are simulated in a $N(0,1)$ distribution.

Table 4: Comparison of exact and approximated critical values with a lag-order $K = 1$.

Simulated 5% Critical Values $c_{N,T}\,(0.05)$								
$N \backslash T$	10	15	20	25	30	40	50	100
5	3.54	2.87	2.66	2.53	2.47	2.39	2.36	2.28
10	2.97	2.38	2.19	2.10	2.04	1.98	1.95	1.88
15	2.68	2.15	1.99	1.91	1.85	1.80	1.77	1.71
20	2.49	2.01	1.86	1.79	1.75	1.69	1.67	1.62
25	2.40	1.92	1.78	1.71	1.66	1.62	1.60	1.55
Approximated 5% Critical Values $\widetilde{c}_{N,T}\,(0.05)$								
$N \backslash T$	10	15	20	25	30	40	50	100
5	3.46	2.66	2.44	2.34	2.27	2.21	2.17	2.10
10	2.86	2.24	2.06	1.97	1.92	1.87	1.84	1.78
15	2.59	2.05	1.89	1.81	1.77	1.72	1.69	1.64
20	2.43	1.93	1.79	1.72	1.68	1.63	1.61	1.56
25	2.32	1.85	1.72	1.65	1.61	1.57	1.55	1.50

Notes: The exact critical values for the average statistic $W_{N,T}^{Hnc}$ are computed from equation (28). The simulated critical values are taken from table (3).