Inference from Data Partitions

Ricardo Bórquez and Melvin Hinich

May 18, 2010

1 Introduction

Consider a stationary process $X = \{X_t,\ t = 1, 2, ..., T\}$ defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ where $\mathcal{F}$ denotes the Borel sets, $\{\mathcal{F}_t\}$ is a filtration and P is a probability measure which is assumed to be absolutely continuous respect to the Lebesgue measure. For this process we are interested on testing any form of time-dependence (linear or nonlinear). Usually, we would expect a researcher to run some convenient test over the whole sample in order to infer about the kind of dependence that is present in the data. For instance, in the Box and Jenkins modelling strategy for ARMA models it is required some inference over the whole sample using the autocorrelation and partial autocorrelation functions in order to select the apropriate model, and it is also required an inference procedure over the whole sample when selecting a parsimonious model using information criterions such that of Akaike. However, in these examples as well as in several other statistical settings it is implicit that the result of the test does not depend on how we can partitionate the data, otherwise the inference made without this additional information is generally invalid.

In this study, we propose a test of whether a specific form of partitioning the data sample is informative. The test is based on finding evidence of transient dependency (i.e., unstable structure of dependence in the data). This problem can be restated to apply whenever the data is given some order not necessarily in time (e.g., the order of individuals in a cross-section), although our examples are uniquely taken from the time-series context where the problem is more evident. To be precise, let $\left\{X^{(i)},\ i=1,2,...,k\right\}$ be an arbitrary collection of subsamples of X with elements of length N_i and such that $\sum_i N_i = T$. Define a suficient statistic $S \in \mathbb{R}^d$ to be used in the inference procedure for which we know its limit distribution Q and denote $S\left(X^{(i)}\right)$ the statistic evaluated at each $X^{(i)}$, $S(X)$ corresponds to the *same* statistic evaluated at the entire sample.

The problem of transient dependency can be stated as follows. Let $\mathcal{T}: \mathbb{R}^{k+d} \longmapsto \mathbb{R}^d$ be an application over the sets $\left\{S\left(X^{(i)}\right) \in A_i,\ i=1,2,...,k\right\}\ \forall k>2$ and denote the corresponding composition as $\mathcal{T}(X) = \left\{S\left(X^{(i)}\right) \in A_i,\ i=1,2,...,k\right\} \circ \mathcal{T}$. If the partition of the data is informative and if this information is summarized in the parameter ϕ, then we can build a similar test based on $\mathcal{T}(X)$ with similar region α (the size of the test that is based on $S(X)$). That is, we intempt to have (Barklett, 1937): $P(S(X) \in A;\ \phi) = \alpha\ \forall \phi \in \Phi$ where ϕ is a nuisance parameter. But if $\mathcal{T}(X)$ is a sufficient statistic for ϕ then under the null hypothesis the conditional distribution $P\left(S\left(X^{(i)}\right) \in A'_i | \mathcal{T}(X) \in A\right)$ (with $A' = \underset{i}{\cup} A'_i$) will not depend on the parameter ϕ for $i=1,2,...,k$. Thus, evidence supporting transient dependence can be found by rejecting the null hypothesis for *some* subsample based on this conditional test. We show below that this property is satisfied by the family of union-intersection tests.

It is needed first to define formally the relationship between an inference based on $S(X)$ and that of $\mathcal{T}(X)$ which occurs under the null hypothesis.

Proposition 1. *For $S(X)$ to provide the same inference than that of $\mathcal{T}(X)$ it is necessary that $P[S(X) \in A'] \Rightarrow_n Q$ where $A' \in \mathcal{F}$ is a Borel set such that $Q(\partial A') = 0$ ($\partial A'$ is the boundary of A') and Q is proportional to the limit distribution of $\mathcal{T}(X)$.*

Proof. If $P[S(X) \in A'] \nRightarrow_n Q$ then there exist a collection of disjoint sets A_i for $i=1,2,...,k$ forming a partition of $A = \underset{i}{\cup} A_i$ such that $P[\mathcal{T}(X) \in A] \nRightarrow_n \lambda Q$ where $\lambda = \lambda(X) > 0$ is a constant, but this is not possible because the latter distribution is tight on $\mathbb{R}^d$ for $d \geq 1$ and the finite dimensional distributions form a convergence-determining class on that space (Billingsley, 1999). $\square$

Thus, when the null hypothesis is true the partition of the sample data $\left\{X^{(i)},\ i=1,2,...,k\right\}$ is not informative and its knowledge conduces no further changes to the inference that we can make through the statistic $S\left(X\right)$. We operationalize this proposition as follows. For simplicity assume that $S,\mathcal{T}\in\mathbb{R}^1$. The null hypothesis of the test of interest is described as the intersection of complementary events $\underset{i\leq k}{\cap}\left[S\left(X^{(i)}\right)\leq c\right]$ for some c. A level α test can be formed through the union intersection approach with the maximum order statistic and the rejection region defined as $\left[S\left(X^{(i)}\right)>c\text{ for some }i=1,...,k\right]=\left[\underset{i\leq k}{\max}S\left(X^{(i)}\right)>c\right]$ where $c=c\left(\alpha,n\right)$. To see this, we can note that $\left[S\left(X^{(i)}\right)>c\text{ for some }i=1,...,k\right]\subset\underset{i\leq k}{\cup}\left[S\left(X^{(i)}\right)>c\right]$ and that $P\left(\underset{i\leq k}{\cup}\left[S\left(X^{(i)}\right)>c\right]\right)=\alpha$ (it is only required that S provides a size α test). We show then that the union-intersection test can be used to answer the question of whether a partition of the data sample is informative.

Proposition 2. Under the null hypothesis $\mathcal{T}\left(X\right)=\underset{i\leq k}{\max}S\left(X^{(i)}\right)$ is a suficient statistic for ϕ.

Proof. For union intersection tests we only need to note that $P\left(S\left(X\right)>c|\underset{i\leq k}{\max}S\left(X^{(i)}\right)>c;\ \phi\right)=P\left(S\left(X\right)>c|\underset{i}{\cup}\left\{S\left(X^{(i)}\right)>c\right\};\ \phi\right)=1$ which does not depend on ϕ. $\square$

Then, we identify λQ in Proposition 1 with the limit distribution of $\mathcal{T}\left(X\right)=\underset{i\leq k}{\max}S\left(X^{(i)}\right)$ and $1/\lambda=P\left(S\left(X^{(i)}\right)>c|\underset{i\leq k}{\max}S\left(X^{(i)}\right)>c\right)<1$ for $i=1,2,...,k$.

A well known convergence to types result due to Gnedenko (1943) shows that the limit distribution of the maximum order statistic for a sequence of independent, identically distributed random variables exists and it is one of three types depending on the support of the distribution. Extensions of this result to allow for dependency in the data either in discrete or continuous time are available (e.g. Watson 1954; Welsch, 1971; Durret and Resnick, 1978) and also there are results for stationary processes and some forms of weak dependency (e.g. Berman, 1964; Leadbetter, 1974; Adler, 1978). It is clear that depending upon the particular context a suitable result for a limit distribution of the maxima is often available, and this is enough for our purposes.

2 Testing for Transient Dependence

In this section, we apply the previous results to derive a method for testing transient dependence when $S(X)$ is a centered chi-squared variable. An example is now provided in the context of the Hinich (1996) test for nonlinearity. Consider a zero-mean second-order stationary process for which we are interested on finding significant elements of the third-order cumulants. These are moments of the form $C(r,s) = E(X_t X_{t+r} X_{t+s})$ and their sample counterparts are referred as *bicorrelations*. A stochastic process can show non-zero bicorrelations and still have a white noise representation, which turns out to be a convenient specification for describing time-dependence in many applications.

The Hinich (1996) test is a test for the null hypothesis of a pure white noise process (i.e., a white noise process with independent innovations) against a process having many significant bicorrelations. As usual, the test relies on assumptions about the stability of the dependence structure in the sample. But note that this could be unlikely to occur if the sample covers a relatively long period of time, which is commonly the case in time-series applications. Motivated by this fact, Hinich and Patterson (2005) studied the transient dependence in white noise. Using financial data they found that periods of time dependence do alternate with periods of independence, a result that can have implications regarding the efficiency of financial markets. From a statistical point of view, that result can also have implications on the forecasting ability of linear time-series models. In their setting, Hinich and Patterson (2005) applied the test separately over data grouped in consecutive window frames of fixed but rather short length of time. Thus, a penalty in the size and power of the test is expected for that procedure because of the limited information contained in a single window even when is applied consecutively or overlapped.

Alternatively, we can use a union-intersection approach to control for the size of the test and increase its power. In particular, let $X = \{X_t,\ t = 1, 2, ..., T\}$ be a sequence of linearly filtered data where $EX_t = 0$ and $EX_t^2 = 1$ for all $t \leq T$. The testing procedure employs non-overlapped data windows, thus if N is the window length, then $[X(t_{i+1}), X(t_{i+1} + 1), ..., X(t_{i+1} + N - 1)]$ is the i-th window where $X_t = X(t)$ and $i = 1, 2, ..., k$ and $t = 1, 2, ..., T$. The next non-overlapped window simply considers $t_{i+1} = t_i + N$. Define the statistic $H_i = \sum_r \sum_s G_i^2(r, s)$ where $G_i(r, s) = (N - s)^{-.5} \sum_t X_t X_{t+r} X_{t+s}$ for $0 < r < s$ which is indexed to the window i. The H_i statistic is distributed chi-squared with $(L - 1)(L/2)$ degrees of freedom for a test of size α. L is the number of lags that enters the window and it is determined endogenously as $L = Nb$ with $0 < b < 0.5$ (recommended to maximize the power of the test).

Under the null hypothesis $\{H_i,\ i = 1, 2, ..., k\}$ is a collection of independent and identically distributed random variables, then we can characterize this hypothesis as $\bigcap_i \{H_i \leq c\}$ and its probability as $P\left(\bigcap_i \{H_i \leq c\}\right) = P(H_1 \leq c)^k$ when all the windows have the same length (or in general as $\prod_{i \leq k} P(H_i \leq c_i)$ where $c_i = c_i(\alpha, N_i)$ and $i = 1, 2, ..., k$). The rejection region for the union intersection test is given by $\left\{\max_{i \leq k} H_i > c\right\}$.

Proposition 3. $P(H_1 \leq b_k u)^k \Rightarrow_k \exp(-u^{-\gamma})$, where $\gamma = \gamma(\alpha, k) > 0$ and b_k is a normalizing constant such that $1 - P(H_1 \leq b_k) = \frac{1}{n}$.

Proof. In order to apply Proposition 3 to the H statistic, we have to show that we can write $1 - P(H_1 \leq u) = u^{-\delta} h(u)$ for some $\delta > 0$ and slowly varying function $h(u)$. But it sufficies to assume that $Eu^\delta < \infty$ with $\delta = 1$ (so that the process X is second-order stationary). The rest is a standard result and it can be found in Ferguson (1996), p.95. $\square$

It is immediate that $\max_{i \leq k} H_i$ is distributed reverse weibull with parameters $(\gamma, 1)$. In order to have a similar proposition for considering the case of different lengths on each window, one could apply the results of limit convergence for the maxima on arrays of independent random variables in Serfozo (1982), but the limiting distribution differs from that of Proposition 3.

Figure 1: Size of the test according window length (N) and sample size (T)

2.1 Size and Power of the Test

In this section we provide evidence on the size and power of the union-intersection test of section 2 through a Monte Carlo experiment. We generate pseudo random numbers for the pure noise process from four alternative distributions: Gaussian, t-student (with v degrees of freedom), Uniform and Exponential. Both the size and power of the test vary in a complex manner according to the sample size T and the window length N, which is controlled by adequately choosing the value of the parameter γ. Our Monte Carlo results show that we can use $\gamma = k^{0.2}$ as a valid approximation for most empirical applications.

2.1.1 Size

For the estimation of the size of the test we computed the times that the null hypothesis was erroneously rejected, running ten thousand replications in each case. The results are summarized in Figure 1. For a given window length the size of the test increases as the sample size increases which is a standard result. The size also varies with the window length for a given sample size although this is expected. The reason can be associated to the informational content of a single window frame respect to the whole sample, which differs accordingly to the number of windows and the window length.

For a given size of the test and a sample size we can deduce using Figure 1 the window length that is consistent with the asymptotic theory. With $T = 1100$ observations and $\alpha = 0.1$ we should use a window length $N = 90$ observations if Gaussian innovations are assumed and a window length $N = 80$ observations for t-Student innovations. Alternatively, by fixing the window length for a given sample size we can obtain the respective probability of introducing Type I error. For example, consider again $T = 1100$ and $N = 105$. In the case of Gaussian innovations the size of the test is approximately 0.08 but it is near 0.05 for the Uniform innovations.

Note that although the results differ across the four distributions for the innovations such differences are still bounded on values that are commonly used in empirical work. Consequently our results seem to be robust independently of the particular distribution that is assumed. In practice this means that low p-values should be considered as strong evidence in favor of the alternative hypothesis, even if the distribution of the innovations is assumed to have fat tails.

2.1.2 Power

The power of the test is evaluated against two nonlinear models: a nonlinear moving average (NLMA) model and a bilinear (BL) model. The particular specification we use for the NLMA model is $X_t = e_t + \beta e_{t-1} e_{t-2}$ where e_t denotes a zero-mean innovation with variance equal to σ^2. This model permits that the parameter β can take any nonzero value whilst the random variable is clearly not independent yet is white, which has many desirable properties for our study. Note that although there is no correlation between X_t and X_{t+r} for $r \neq 0$ the elements of the third-order cumulants of the process $\{X_t,\ t = 1, 2, ..., T\}$ can be different from zero. In fact, we have that $C(r, s) = \beta\sigma^4$ but there are only six of these elements for this particular process, which makes it very difficult to capture the underlying time-dependence structure based on a nonparametric test. On the other side, the bilinear model can be thought as a reduced form of some higher-order nonlinear moving average process and therefore is characterized by having several non-zero bicorrelations. These models have the property of approximating with arbitrary accuracy any model that reasonably can be represented by Volterra expansions, and consequently they have been proposed as natural nonlinear extensions of ARMA models (Tong, 1990; Granger and Andersen, 1978). For instance, a model of the form $X_t = e_t + \beta X_{t-p} e_{t-q}$ is (second-order) stationary if $|\beta\sigma^2| < 1$ and the series is generally white for $p \neq q$. In our study we use $p = 1$ and $q = 2$.

The results for the test are summarized in Exhibit 1 for each model and two alternative values of the parameters. We report the percentage of correct decisions using a size of 0.01. As is usual the power greatly depends on the values of the model parameters, being more difficult to reject the null hypothesis as their absolute value approaches to zero[1]. The power of the test is higher as the number of windows is higher, which can be achieved by increasing the window length and/or the number of observations. This result differs according to the four distributional alternatives on the innovations, being more sensitive for the case of the Uniform distribution. We also note that the power of the test is higher for the bilinear case than the NLMA, which is expected as the number of possibly nonzero bicorrelations is higher in the former model.

2.2 An Empirical Example

We return to the problem stated in Hinich and Patterson (2005).

References

[1] Adler, 1978

[2] Barklett, 1937

[3] Berman, 1964

[4] Billingsley, 1999

[5] Durret and Resnick, 1978

[6] Ferguson (1996)

[7] Gnedenko (1943)

[8] Granger and Andersen, 1978

[9] Hinich (1996)

[1]If the parameter is exactly zero then the process reduces to a pure white noise

[10] Hinich and Patterson (2005)

[11] Leadbetter, 1974

[12] Serfozo (1982)

[13] Tong, 1990

[14] Watson 1954

[15] Welsch, 1971