# Generalized divergence criteria for model selection between random walk and AR(1) model

Papa Ngom, Hamza Dhaker, Mendy Pierre, El Hadji Deme

# Generalized divergence criteria for model selection between random walk and AR(1) model

Papa Ngom[a,b,*], Hamza Dhaker[a,b], Pierre Mendy[b], El Hadji Deme[c]

[a]LMA,Laboratoire de Mathematiques Appliquees
[b]LMDAN,Universit Cheikh Anta Diop, Dakar-Fann, BP 5005, Senegal
[c]LERSTAD, Universite Gaston Berger, UFR SAT Saint-Louis, BP 234, Senegal

**Abstract**

We investigate a general class of divergence measures among distributions for model selection. As alternative to the classical test of model choice, we introduce kernel type estimators of $(\psi, \phi)$-divergence for continuous distributions based on model selection criteria in general non parametric case.
We introduce the Divergence Indicator $\mathcal{DI}$ method by proposing a test for choosing between a random walk and a regression one, using a unified divergence measure. Under the assumptions of standard type about model densities, the asymptotic properties estimator of the expected divergence between the true unknown model and the candidate model are established. From the point of the resulting statistics divergence estimator, the performance of the discrepancy criteria is discussed and illustrated in various settings in model selection test.

*Keywords:* divergence measure, Kernel Estimator, Hypothesis testing
*2010 MSC:* 94A17, 62G07, 62G10.

## 1. Introduction

Statistical modeling technique using the functionals of information theory such as divergence measure, is not new. The divergence measures have provided several useful methods in statistical inference. For example, testing statistical hypotheses with type measures of information theory have been elaborated for models with continuous and discrete data. A comprehensive surveys on divergence measure in statistical testing have been proposed. In particular, among others, to Cressie and Read [6] , Nayak [22], Read and Cressie [27], Zografos et al. [33], Salicru et al. [31], Menendez et al. [18, 19, 20], Pardo et al. [24], Morales et al. [21], Zografos [34, 35, 36] and references therein.
Model selection is one of standard tools for time series econometricians for selecting the best model among competitor models. One can consider the model selection criteria as an approximately unbiased estimator of the discrepancy, between the true unknown model and a goodness-of-fit approximating model.

Many others model selection criterion have been introduced so far. One can cite the classical model selection criteria based on least-squares estimation, which makes them sensitive to non normalities in the case of finite samples and outliers.
To solve this drawback, robust versions of classical models criterion, which are not affected by outliers, have been proposed, in first, by Ronchetti [28], Ronchetti and Staudte [30]. Other references on this topic can be found in Maronna et al. [16]. On the other hand, a major problem with these tests (Dickey and Fuller) is that the decision on the level of differencing is then based on the outcome of a test at a significance level. A well known difficulty is that when these tests are applied to the same series, the result is that neither null hypothesis-stationarity or a unit autoregressive root-can be rejected at the usual significance levels.

*Corresponding author
*Email address:* papa.ngom@ucad.edu.sn (Papa Ngom)

More recently, among the proposals for model selection we recall the criteria presented by Karagrigoriou et al. [13], the divergence information criteria (DIC) introduced by Mattheou et al. [17]. The DIC criteria use the density power divergences introduced by Basu et al. [3].

In traditional method, Pearson type chi-square statistics have been used to test whether a specified model is consistent with observed data.

Because divergence statistic provide naturel measures for dissimilarity between the observed data and a specific model, it has been used through an informational criteria for discriminating among competing models. The statistic resulting of divergence estimator is asymptotically distributed as a chi-squared with d degrees of freedom. In this context, the main problem is that each divergence statistic tends to became large without no increase in its degrees of freedom as the sample size increases.

Hence the goodness-of-fit in forming type chi-squared statistics will generally (over) reject the correct specification of evry competitor model.

The most commonly used approach to this issue is through a method for model selection of Akaike (1973) Information Criterion (AIC).

This popular method consists in considering Pearson type chi-square statistics that the lower the value of criterion, the better is the approximated competitor model. In other wods, the model associated to smaller value of chi-square statistic is generally chosen as the best.

It is not at all sur that this approach accurately is entirely satisfactory : these chi-square tests based on the sample are random, in the sense that their actual values are subject (to fluctuation sample). As a consequence in terms of adequation, a model with a smaller value of criteria is not necessarily better than one with the a larger chi-square statistic.

It seems natural to explore new approach to the comparison of stationary models by for taking into account the stochastic nature of these differences. The modest aim of this paper is to address fundamental issues arising from the practical application of that approach. Our concern is considering an inference from the perspective of model selection based on divergence type statistics, by proposing some asymptotically standard normal tests.

Methodology considered here are testing the null hypothesis that the Random Walk is equally close to the data generating process (*DGP*) versus the alternative hypothesis that the Stationary $AR(1)$ model is closer to the *DGP* where closeness of a model is measured according to the discrepancy implicit in the divergence type statistic used.

The plan of the paper is as follows. In Section 2 we present the divergence measures. Then in Section 3 we develop our main results. Section 4 provides the results on nonparametric estimation and specification testing. Finally, in Section 5 we present our conclusion.

## 2. Formal Problem: Definitions and Estimation

One important aspect of statistical modeling is evaluating the fit of the chosen model. Marriott and Newbold [15] discussed the Bayesian goodness of the unit root as follows:

$$\begin{cases} H_0 : \rho = 1, \\ H_1 : \mid \rho \mid < 1 \end{cases}$$

in the model $AR(1)$ with intercept

$$X_t - \mu 1_d = \rho(X_{t-1} - \mu 1_d) + \varepsilon_t,$$

where $d \in \mathbb{N}^*$, the $d$-dimensional vector $1_d = (1, ..., 1)'$, $X_t \in \mathbb{R}^d$, $\forall t$ and $\varepsilon_t$ are *i.i.d* Gaussian vector i.e $N(0_{\mathbb{R}^d}, \sigma^2\Sigma_d)$, $\Sigma_d$ is the identity matrix and $\mu$ is an unknown parameter. Marriott and Newbold [15] proposed to eliminate the parameter $\mu$ considering the sample $(W_1, ..., W_n) \in \mathbb{R}^{d\times n}$ with zero mean vector instead of the sample $(X_1, ..., X_n)$ and

$$W_t = X_t - X_{t-1}, \qquad \forall\ t = 1, ..., n.$$

These authors transforme this problem of test by a comparison one between the two models, following the Bayesian approach:

$$\begin{cases} W_t = \varepsilon_t, & (M1), \\ W_t = \rho W_{t-1} + \varepsilon_t - \varepsilon_{t-1} & (M2). \end{cases}$$

Under the model $(M1)$, the distribution function $W_t$ given by:

$$f_1(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x'\Sigma_d x}{2\sigma^2}\right), \quad x \in R^d.$$

And under the model $(M2)$, the distribution function $W_t$ can be expressed by :

$$f_2(x) = \frac{1}{\sqrt{2\pi\Lambda}} \exp\left(-\frac{x'\Sigma_d x}{2\Lambda}\right), \quad x \in R^d.$$

where

$$W_t = \rho W_{t-1} + \varepsilon_t - \varepsilon_{t-1}$$

and

$$\Lambda = Var(W_t) = \rho^2 Var(W_{t-1}) + Var(\varepsilon_t - \varepsilon_{t-1}) + 2\rho\, cov(W_{t-1}, \varepsilon_t - \varepsilon_{t-1})$$

With a little algebra, we have:

$$\Lambda = \frac{2\sigma^2}{1-\rho}$$

Based on their methods, we propose a new approach based on the $(\psi, \phi)$-divergence in order to find a goodness of fit of the model.

### *2.1. A Brief Review of $(\psi, \phi)$-divergence*

The $\phi$-divergence measure between the probability distributions $p$ and $q$ is defined by

$$\mathcal{D}_\phi(p, q) = \int_{\mathbb{R}^d} \phi\left(\frac{p(x)}{q(x)}\right) q(x)dx, \quad \phi \in \Phi^* \tag{1}$$

where $\Phi^*$ is the class of all convex function $\phi(x)$, $x \geq 0$, such that, $\phi(1) = 0$, $\phi'(1) = 0$ and $\phi''(1) = 1$.
For example: $\phi(x) = x\log(x) - x + 1$, we have Kullback-Leibler divergence

$$\mathcal{D}^{KL}(p, q) = \int_{\mathbb{R}^d} p(x) \log \frac{p(x)}{q(x)} dx.$$

Rényi [28] presented the first parametric generalization of Kullback-Leibler

$$\mathcal{D}^R_\alpha(p, q) = \frac{1}{\alpha - 1} \log \int_{\mathbb{R}^d} \left(\frac{p(x)}{q(x)}\right)^\alpha q(x)dx.$$

It is easy to prove that

$$\lim_{\alpha \longrightarrow 1} \mathcal{D}^R_\alpha(p, q) = \mathcal{D}^{KL}(p, q) = \int_{\mathbb{R}^d} p(x) \log \frac{p(x)}{q(x)} dx,$$
$$\lim_{\alpha \longrightarrow 0} \mathcal{D}^R_\alpha(q, p) = \mathcal{D}^{KL}(q, p) = \int_{\mathbb{R}^d} q(x) \log \frac{q(x)}{p(x)} dx.$$

Rényi are not $\phi$-divergences measures. However, such measures can be written in the following form:

$$\mathcal{D}^\psi_\phi(p, q) = \psi\left(\mathcal{D}_\phi(p, q)\right), \tag{2}$$

where $\psi$ is a differentiable increasing real function mapping from

$$\left[0, \quad \phi(0) + \lim_{t\to\infty} \frac{\phi(t)}{t}\right]$$

onto $[0,\infty)$ ; this condition will be justified in (Proposition 1.1, [23]), with $\psi(0) = 0$, $\psi^{'}(0) > 0$, and $\phi \in \Phi^{*}$. In the following formules we list the functions $\psi$ and $\phi$ that yield to the Rényi divergence measures:

$$\textit{Renyi}: \quad \psi(x) = \frac{1}{\alpha(\alpha-1)}\log(\alpha(\alpha-1)x+1) \qquad \phi(x) = \frac{x^{\alpha}-\alpha(x-1)-1}{\alpha(\alpha-1)} \qquad \alpha \neq 0,1$$

$$\textit{Sharma-Mittal} \quad \psi(x) = \frac{1}{(s-1)}((1+\alpha(\alpha-1)x)^{\frac{s-1}{r-1}}-1) \quad \phi(x) = \frac{x^{\alpha}-\alpha(x-1)-1}{\alpha(\alpha-1)} \qquad \begin{array}{c} s \neq 1 \\ \alpha \neq 0,1 \end{array}$$

$$\textit{Bhattachayya} \qquad \psi(x) = -\log(-x+1) \qquad \phi(x) = -x^{\frac{1}{2}} + \frac{1}{2}(x+1)$$

Now, let $f$ be the unknown true density function (with respect to Lebesgue measure on $\mathbb{R}^d$) of the sample $(W_1,...,W_n)$ with cumulative distribution function $F$. The distance between true density and those of the models can be measured by the $(\psi,\varphi)$-divergence of $f$ and $f_j$, $j = 1,2$ as follows

$$\mathcal{D}_{\phi}^{\psi}(f,f_j) = \psi\left(\mathcal{D}_{\phi}(f,f_j)\right).$$

For a given density of probability $g$ defined on $\mathbb{R}^d$, we start by giving some notation and conditions that are needed for the forthcoming sections. Below, we will work under the following assumptions on $f$ and $g$ to establish our results.

(**F.1**) The functional $\mathcal{D}_{\phi}^{\psi}(f,g)$ as well-defined as (2), in the sense that $\mathcal{D}_{\phi}^{\psi}(f,g)$ is finite.

### 2.2. Nonparametric estimation of $(\psi,\phi)$-divergence

To define our divergence estimator we define, in a first step, a kernel density estimator. Towards this aim, we introduce a measurable function $K(\cdot)$ fulfilling the following conditions.

(**K.1**) $K(\cdot)$ is of bounded variation on $\mathbb{R}^d$

(**K.1**) $K(\cdot)$ is right continuous on $\mathbb{R}^d$,

(**K.2**) $\|K\|_{\infty} = \sup_{x\in\mathbb{R}^d} | K(x) | < \infty$,

(**K.3**) $\int_{\mathbb{R}^d} K(t)dt = 1$.

The well known Akaike-Parzen-Rosenblatt (refer to [1], [25] and [32]) kernel estimator of $f(\cdot)$ is defined, for any $x \in \mathbb{R}^d$, by

$$\widehat{f_{n,h_n}}(x) = \frac{1}{nh_n^d}\sum_{i=1}^{n} K\left(\frac{x-W_i}{h_n}\right),$$

where $0 < h_n \leq 1$ is the smoothing parameter. Assuming that the density $f$ is continuous, one can obtain the normality asymptotic of the estimator $\widehat{f_{n,h_n}}$ under conditions below see [14]. For more details of kernel estimators $\widehat{f_{n,h_n}}$, one can refer to [9], [10], [5], [26], [7], [11], [8] and the references therein, and their limiting behavior.

In a second step, given $\widehat{f_{n,h_n}}(\cdot)$, we estimate divergences $\mathcal{D}_{\phi}(f,g)$ and $\mathcal{D}_{\phi}^{\psi}(f,g)$ by using the representation (1) and (2) with $f$ and $g$, by setting

$$\widehat{\mathcal{D}}_{\phi}(\widehat{f_{n,h_n}},g) = \int_{\mathbb{R}^d}\phi\left(\frac{\widehat{f_{n,h_n}}(x)}{g(x)}\right)g(x)dx \tag{3}$$

$$\begin{aligned} \widehat{\mathcal{D}}_{\phi}^{\psi}(\widehat{f_{n,h_n}},g) &= \psi\left(\widehat{\mathcal{D}}_{\phi}(\widehat{f_{n,h_n}},g)\right), &(4)\\ &= \psi\left(\int_{\mathbb{R}^d}\phi\left(\frac{\widehat{f_{n,h_n}}(x)}{g(x)}\right)g(x)dx\right) &(5) \end{aligned}$$

The approach use to define the plug-in estimators $\widehat{\mathcal{D}}_{\phi}(\widehat{f}_{n,h_n}, g)$ and $\widehat{\mathcal{D}}^{\psi}_{\phi}(\widehat{f}_{n,h_n}, g)$ are respectively developed in [4] and [12] in order to introduce a kernel-type estimators of Shannon's entropy and divergences.

In the next section, we wish to establish the asymptotic behavior for the estimates $\widehat{\mathcal{D}}_{\phi}(\widehat{f}_{n,h_n}, g)$ $\widehat{\mathcal{D}}^{\psi}_{\phi}(\widehat{f}_{n,h_n}, g)$ and to give in application for testing hypothesis.

## 3. Main Results

First step we study the consistency of the estimator. In a second step we show the asymptotic normality of the term given in the function $\psi$ and to deduce those of the general estimator.

**Theorem 1.** *Suppose that $f$ is uniformly continuous on $]-\infty, +\infty[$, and that the window width $h_n$ satisfies $h_n \longrightarrow 0$ and $nh_n \longrightarrow \infty$ as $n \longrightarrow \infty$*

$$|\mathcal{D}_{\phi}(\widehat{f}_{n,h_n}, g) - \mathcal{D}_{\phi}(f, g)| \longrightarrow 0 \quad \textit{with probability one as} \quad n \longrightarrow \infty \tag{6}$$

Proof.

$$\begin{aligned} |\widehat{\mathcal{D}}_{\phi}(\widehat{f}_{n,h_n}, g) - \mathcal{D}(f, g)| &= |\int_{\mathbb{R}^d} \phi\left(\frac{\widehat{f}_{n,hn}(x)}{g(x)}\right) g(x) - \phi\left(\frac{f(x)}{g(x)}\right) g(x) dx| \\ &\leq \int_{\mathbb{R}^d} |\left(\phi\left(\frac{\widehat{f}_{n,hn}(x)}{g(x)}\right) - \phi\left(\frac{f(x)}{g(x)}\right)\right) g(x)| dx \end{aligned}$$

$\phi$ is a convex function therefore it is locally Lipschitz, so there exists real as $k$ : $\quad |\phi(x) - \phi(x)| \leq k|x - y|$, for $x = \frac{\widehat{f}_{n,hn}(x)}{g(x)}$ and $y = \frac{f(x)}{g(x)}$

$$\begin{aligned} |\left(\phi\left(\frac{\widehat{f}_{n,hn}(x)}{g(x)}\right) - \phi\left(\frac{f(x)}{g(x)}\right)\right) g(x)| &\leq k|\left(\frac{\widehat{f}_{n,hn}(x)}{g(x)} - \frac{f(x)}{g(x)}\right) g(x)| \\ &\leq k|\widehat{f}_{n,hn}(x) - f(x)| \end{aligned}$$

$$|\widehat{\mathcal{D}}_{\phi}(\widehat{f}_{n,h_n}, g) - \mathcal{D}(f, g)| \leq k \int_{\mathbb{R}} |\widehat{f}_{n,hn}(x) - f(x)| dx \tag{7}$$

Devroye and Györfi [9] shows that

$$\int |\widehat{f}_{n,h_n}(x) - f(x)| \longrightarrow 0 \quad \text{with probability one as} \quad n \longrightarrow \infty \tag{8}$$

therefore after Eq 7 and Eq 8 :

$$|\widehat{\mathcal{D}}_{\phi}(\widehat{f}_{n,h_n}, g) - \mathcal{D}(f, g)| \longrightarrow 0 \quad \text{with probability one as} \quad n \longrightarrow \infty$$

**Lemma 1.** *Under the assumptions of Theorem 1*

$$|\mathcal{D}^{\psi}_{\phi}(\widehat{f}_{n,h_n}, g) - \mathcal{D}^{\psi}_{\phi}(f, g)| \longrightarrow 0 \quad \textit{with probability one as} \quad n \longrightarrow \infty \tag{9}$$

Proof. of after Theorem 1, and the effect that $\psi$ is a convex function thus locally Lipschitz.

**Lemma 2.** *Let $K(\cdot)$ satisfy **(K.1-2-3-4)** and let $f(\cdot)$ be a bounded density fulfill **(F.1)**. Suppose that $\phi \in C^1([0, \infty))$ and there exist a measurable and Lebesgue-integrable function $F(x)$ such that $|\ \phi'(\frac{f(x)}{g(x)})\ | < F(x)$, Then*

*(i) if $f \neq g$ we have*

$$\sqrt{nh_n^d}\left(\mathcal{D}_{\phi}(\widehat{f}_{n,h_n}, g) - \mathcal{D}_{\phi}(f, g)\right) \to \mathcal{N}\left(0, \left(\int_{\mathbb{R}^d} \sigma(x)\phi'\left(\frac{f(x)}{g(x)}\right) dx)\right)^2\right),$$

*where* $\sigma^2(x) := f(x)\int K^2(z)dz$

*(ii) if* $f = g$ *we have*

$$\frac{2nh_n^d}{\phi''(1)\int K^2(z)dz}\widehat{\mathcal{D}}_\phi(\widehat{f}_{n,h_n}, g) \longrightarrow \chi^2(d)$$

Proof. • if $f \neq g$

The first order Taylor expansion of $\phi\left(\frac{\widehat{f}_{n,h_n}(x)}{g(x)}\right)$ around $\frac{f(x)}{g(x)}$ gives

$$\int_{\mathbb{R}^d}\phi\left(\frac{\widehat{f}_{n,h_n}(x)}{g(x)}\right)g(x)dx = \int_{\mathbb{R}^d}\phi\left(\frac{f(x)}{g(x)}\right)g(x)dx + \int_{\mathbb{R}^d}\left(\frac{\widehat{f}_{n,h_n}(x)}{g(x)} - \frac{f(x)}{g(x)}\right)\phi'\left(\frac{f(x)}{g(x)}\right)g(x)dx + \int_{\mathbb{R}^d} o\left(\|\frac{\widehat{f}_{n,h_n}}{g} - \frac{f}{g}\|\right)g(x)dx$$

$$\widehat{\mathcal{D}}_\phi(\widehat{f}_{n,h_n}, g) = \mathcal{D}_\phi(f,g) + \int_{\mathbb{R}^d}\left(\frac{\widehat{f}_{n,h_n}(x)}{g(x)} - \frac{f(x)}{g(x)}\right)\phi'\left(\frac{f(x)}{g(x)}\right)g(x)dx + \int_{\mathbb{R}^d} o\left(\|\frac{\widehat{f}_{n,h_n}}{g} - \frac{f}{g}\|\right)g(x)dx$$

$$\widehat{\mathcal{D}}_\phi(\widehat{f}_{n,h_n}, g) - \mathcal{D}_\phi(f,g) = \int_{\mathbb{R}^d}\left(\widehat{f}_{nh_n}(x) - f(x)\right)\phi'\left(\frac{f(x)}{g(x)}\right)dx + \int_{\mathbb{R}^d} o\left(\|\frac{\widehat{f}_{n,h_n}}{g} - \frac{f}{g}\|\right)g(x)dx \tag{10}$$

note that we have from Theorem 2.2. p. 339 of Bulinski. A and Shashkin. A [2]

$$\sqrt{nh_n^d}(\widehat{f}_{nh_n}(x) - f(x)) \to \mathcal{N}(0, \sigma^2(x)). \tag{11}$$

Then $\sqrt{nh_n^d}\int_{\mathbb{R}^d} o\left(\|\frac{\widehat{f}_{n,h_n}}{g} - \frac{f}{g}\|\right)g(x)dx = \sqrt{nh_n^d}\,o(O_p((nh_n^d)^{\frac{-1}{2}}))\int_{\mathbb{R}^d} g(x)dx = o_p(1)$

Therefore, the random variables

$$\sqrt{nh_n^d}\left(\widehat{\mathcal{D}}_\phi(\widehat{f}_{n,h_n}, g) - \mathcal{D}_\phi(f,g)\right)$$

and

$$\sqrt{nh_n^d}\int_{R^d}\phi'\left(\frac{f(x)}{g(x)}\right)\left(\widehat{f}_{nh_n}(x) - f(x)\right)dx$$

have the same asymptotic distribution. By 11 we have

$$\sqrt{nh_n^d}\int_{R^d}\phi'\left(\frac{f(x)}{g(x)}\right)\left(\widehat{f}_{nh_n}(x) - f(x)\right)dx \to \mathcal{N}\left(0, \left(\int_{\mathbb{R}^d}\sigma(x)\phi'\left(\frac{f(x)}{g(x)}\right)dx\right)^2\right)$$

• if $f = g$

The second order Taylor expansion of $\phi\left(\frac{\widehat{f}_{n,h_n}(x)}{g(x)}\right)$ au around $\frac{f(x)}{g(x)}$ gives

$$\begin{aligned}\int_{\mathbb{R}^d}\phi\left(\frac{\widehat{f}_{n,h_n}(x)}{g(x)}\right)g(x)dx \quad &= \int_{\mathbb{R}^d}\phi\left(\frac{f(x)}{g(x)}\right)g(x)dx + \int_{\mathbb{R}^d}\left(\frac{\widehat{f}_{n,h_n}(x)}{g(x)} - \frac{f(x)}{g(x)}\right)\phi'\left(\frac{f(x)}{g(x)}\right)g(x)dx \\ &+ \frac{1}{2}\int_{\mathbb{R}^d}\left(\frac{\widehat{f}_{n,h_n}(x)}{g(x)} - \frac{f(x)}{g(x)}\right)^2\phi''\left(\frac{f(x)}{g(x)}\right)g(x)dx + \int_{\mathbb{R}^d} o\left(\|\frac{\widehat{f}_{n,h_n}}{g} - \frac{f}{g}\|^2\right)g(x)dx \\ &= \frac{1}{2}\int_{\mathbb{R}^d}\left(\widehat{f}_{n,h_n}(x) - f(x)\right)^2\frac{\phi''(1)}{g(x)}dx + \int_{\mathbb{R}^d} o\left(\|\frac{\widehat{f}_{n,h_n}}{g} - \frac{f}{g}\|^2\right)g(x)dx\end{aligned}$$

$$\begin{aligned}\widehat{\mathcal{D}}_\phi(\widehat{f}_{n,h_n}, g) \quad &= \frac{1}{2}\int_{\mathbb{R}^d}\left(\widehat{f}_{n,h_n}(x) - f(x)\right)^2\frac{\phi''(1)}{g(x)}dx + \int_{\mathbb{R}^d} o\left(\|\frac{\widehat{f}_{n,h_n}}{g} - \frac{f}{g}\|^2\right)g(x)dx \\ &= \frac{1}{2}\int_{\mathbb{R}^d}\left(\widehat{f}_{n,h_n}(x) - f(x)\right)^2\frac{\phi''(1)}{f(x)}dx + \int_{\mathbb{R}^d} o\left(\|\frac{\widehat{f}_{n,h_n}}{g} - \frac{f}{g}\|^2\right)g(x)dx \\ &= \frac{1}{2}\int_{\mathbb{R}^d}\left(\frac{\widehat{f}_{n,h_n}(x) - f(x)}{(f(x))^{1/2}}\right)^2\phi''(1)dx + \int_{\mathbb{R}^d} o\left(\|\frac{\widehat{f}_{n,h_n}}{g} - \frac{f}{g}\|^2\right)g(x)dx\end{aligned}$$

$$\frac{2nh_n^d}{\phi''(1)\int K^2(z)dz}\widehat{\mathcal{D}}_\phi(\widehat{f}_{n,h_n},g)=\int_{\mathbb{R}^d}\left(\frac{\sqrt{nh_n^d}(\widehat{f}_{n,h_n}(x)-f(x))}{\sigma(x)}\right)^2dx+\int_{\mathbb{R}^d}o\left(\|\frac{\widehat{f}_{n,h_n}}{g}-\frac{f}{g}\|^2\right)g(x)dx$$

from Eq.11

$$\frac{2nh_n^d}{\phi''(1)\int K^2(z)dz}\widehat{\mathcal{D}}_\phi(\widehat{f}_{n,h_n},g)\longrightarrow\chi^2(d)$$

**Theorem 2.** *We consider the $\mathcal{D}_\phi^\psi(f,g)$ defined in Eq.2, then we have*

$$\sqrt{nh_n^d}\left(\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{n,h_n},g)-\mathcal{D}_\phi^\psi(f,g)\right)\to\mathcal{N}\left(0,\left\{\psi'\left(\int_{\mathbb{R}^d}\phi(\frac{f(x)}{g(x)})g(x)dx\right)\int_{\mathbb{R}^d}\sigma(x)\phi'(\frac{f(x)}{g(x)})dx\right\}^2\right)$$

Proof. A direct application of the Delta Method.

## 4. Applications for Testing Hypothesis

In this section, we use the estimators $\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{n,h_n},f_j)\quad j=1,2$ to find the perform statistical tests on the model defined in Section 2.

### 4.1. Goodness-of-Fit test

For completeness, we look at $\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{n,h_n},f_j)$ in the usual way, i.e as a goodness-of-fit statistic. From the uniform-in-bandwidth consistency of $\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{n,h_n},f_j)$ for $\mathcal{D}_\phi^\psi(f,f_j)$, the null hypothesis when using the statistic $\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{n,h_n},f_j)$ can be given as $H_0:\mathcal{D}_\phi^\psi(f,f_j))=0$. Under the alternative hypothesis $H_1:\mathcal{D}_\phi^\psi(f,f_j)\neq0$.

### 4.2. Test for Model Selection

Introduce the divergence Indicator $\mathcal{DI}=\mathcal{D}_\phi^\psi(f,f_1)-\mathcal{D}_\phi^\psi(f,f_2)$. An estimator of the divergence indicator is defined as:

$$\widehat{\mathcal{DI}}_n:=\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{nh_n},f_1)-\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{nh_n},f_2).$$

Using the divergence indicator, we develop the following test hypothesis on the model under study

- $H_0^{eq}:\mathcal{DI}=0$ means that the two models are equivalent.
- $H_1^{M_1}:\mathcal{DI}<0$ means that model $M_1$ is better than model $M_2$.
- $H_1^{M_2}:\mathcal{DI}>0$ means that model $M_2$ is better than model $M_1$.

$\widehat{\mathcal{DI}}_n$ converges to zero under the null hypothesis $H_0^{eq}$, but it converges to a strictly negative or positive constant when $H_1^{M_1}$ or $H_1^{M_2}$ hold. These properties actually justify the use of $\widehat{\mathcal{DI}}_n$ as a model selection indicator and common procedure of selecting the model with highest goodness-of-fit.

**Theorem 3.** *Under the assumptions of Lemma 2.*
*1) Under the null hypothesis $H_0^{eq}$, $\sqrt{nh_n^d}\widehat{\mathcal{DI}}_n\longrightarrow\mathcal{N}(0,\Gamma^2)$*
*2) Under the $H_1^{M_1}$ hypothesis $\sqrt{nh_n^d}\widehat{\mathcal{DI}}_n\longrightarrow-\infty$*
*3) Under the $H_1^{M_2}$ hypothesis $\sqrt{nh_n^d}\widehat{\mathcal{DI}}_n\longrightarrow+\infty$*
*with*

$$\Gamma^2=\left\{\int_{\mathbb{R}^d}\left[\psi'\left(\int_{\mathbb{R}^d}\phi(\frac{f(x)}{f_1(x)})f_1(x)dx\right)\phi'(\frac{f(x)}{f_1(x)})-\psi'\left(\int_{\mathbb{R}^d}\phi(\frac{f(x)}{f_2(x)})f_2(x)dx\right)\phi'(\frac{f(x)}{f_2(x)})\right]\sigma(x)dx\right\}^2$$

Proof.

$$
\begin{aligned}
\sqrt{nh_n^d}\widehat{\mathcal{DI}}_n &:= \sqrt{nh_n^d}\left(\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{n,h_n},f_1)-\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{n,h_n},f_2)\right)\\
&= \sqrt{nh_n^d}\left\{\left[\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{n,h_n},f_1)-\mathcal{D}_\phi^\psi(f,f_1)\right]-\left[\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{n,h_n},f_2)-\mathcal{D}_\phi^\psi(f,f_2)\right]\right\}+\sqrt{nh_n^d}\left[\mathcal{D}_\phi^\psi(f,f_1)-\mathcal{D}_\phi^\psi(f,f_2)\right]\\
&= \sqrt{nh_n^d}\left\{\left[\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{n,h_n},f_1)-\mathcal{D}_\phi^\psi(f,f_1)\right]-\left[\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{n,h_n},f_2)-\mathcal{D}_\phi^\psi(f,f_2)\right]\right\}+\sqrt{nh_n^d}\mathcal{DI}
\end{aligned}
$$

127 ◦ Under the null hypothesis $H_0^{eq}$, we have: $\mathcal{DI}=0$

$$
\sqrt{nh_n^d}\widehat{\mathcal{DI}}_n = \sqrt{nh_n^d}\left[\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{n,h_n},f_1)-\mathcal{D}_\phi^\psi(f,f_1)\right]-\sqrt{nh_n^d}\left[\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{n,h_n},f_2)-\mathcal{D}_\phi^\psi(f,f_2)\right] \tag{12}
$$

A first order Taylor expansion of $\psi(y)$ around $y=y_0$ at $y=\widehat{y}$ gives

$$
\psi(\widehat{y})=\psi(y_0)+\psi'(y_0)(\widehat{y}-y_0)+o(\|\widehat{y}-y_0\|).
$$

128 Now for $y_0=\int_{\mathbb{R}^d}\phi(\frac{f(x)}{f_j(x)})f_j(x)dx$ and $\widehat{y}=\int_{\mathbb{R}^d}\phi(\frac{\widehat{f}_{n,h_n}(x)}{f_j(x)})f_j(x)dx$, with $j=1,2$ we get

$$
\begin{aligned}
\psi\left(\int_{\mathbb{R}^d}\phi(\frac{\widehat{f}_{n,h_n}(x)}{f_j(x)})f_j(x)dx\right) &= \psi\left(\int_{\mathbb{R}^d}\phi(\frac{f(x)}{f_j(x)})f_j(x)dx\right)+\psi'\left(\int_{\mathbb{R}^d}\phi(\frac{f(x)}{f_j(x)})f_j(x)dx\right)\left[\int_{\mathbb{R}^d}\phi(\frac{\widehat{f}_{n,h_n}(x)}{f_j(x)})f_j(x)dx-\int_{\mathbb{R}^d}\phi(\frac{f(x)}{f_j(x)})f_j(x)dx\right]\\
&\quad +o\left(\|\int_{\mathbb{R}^d}\phi(\frac{\widehat{f}_{n,h_n}(x)}{f_j(x)})f_j(x)dx-\int_{\mathbb{R}^d}\phi(\frac{f(x)}{f_j(x)})f_j(x)dx\|\right)
\end{aligned}
$$

129

$$
\begin{aligned}
\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{n,h_n},f_j)-\mathcal{D}_\phi^\psi(f,f_j) &= \psi'\left(\int_{\mathbb{R}^d}\phi(\frac{f(x)}{f_j(x)})f_j(x)dx\right)\left[\int_{\mathbb{R}^d}\phi(\frac{\widehat{f}_{n,h_n}(x)}{f_j(x)})f_j(x)dx-\int_{\mathbb{R}^d}\phi(\frac{f(x)}{f_j(x)})f_j(x)dx\right]\\
&\quad +o\left(\|\int_{\mathbb{R}^d}\phi(\frac{\widehat{f}_{n,h_n}(x)}{f_j(x)})_j(x)dx-\int_{\mathbb{R}^d}\phi(\frac{f(x)}{f_j(x)})f_j(x)dx\|\right)\\
&= \psi'\left(\mathcal{D}_\phi(f,f_j)\right)\left[\widehat{\mathcal{D}}_\phi(\widehat{f}_{nh},f_j)-\mathcal{D}_\phi(f,f_j)\right]+o\left(\|\widehat{\mathcal{D}}_\phi(\widehat{f}_{nh},f_j)-\mathcal{D}_\phi(f,f_j)\|\right)
\end{aligned}
$$

130 from Eq.10, replacing $g$ by $f_j$, we have

$$
\begin{aligned}
\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{n,h_n},f_j)-\mathcal{D}_\phi^\psi(f,f_j) &= \psi'\left(\mathcal{D}_\phi(f,f_j)\right)\left\{\int_{\mathbb{R}^d}(\widehat{f}_{nh_n}(x)-f(x))\phi'(\frac{f(x)}{f_j(x)})dx+\int_{\mathbb{R}^d}o\left(\|\frac{\widehat{f}_{nh_n}}{f_j}-\frac{f}{f_j}\|\right)f_j(x)dx\right\}\\
&\quad +o\left(\|\int_{\mathbb{R}^d}\phi(\frac{\widehat{f}_{n,h_n}(x)}{f_j(x)})_j(x)dx-\int_{\mathbb{R}^d}\phi(\frac{f(x)}{f_j(x)})f_j(x)dx\|\right)
\end{aligned}
$$

$$
\begin{aligned}
\sqrt{nh^d}\quad & \left\{\widehat{\mathcal{D}}_\phi^\psi(\widehat{f}_{n,h_n},f_j)-\mathcal{D}_\phi^\psi(f,f_j)\right\}\\
&= \psi'\left(\int_{\mathbb{R}^d}\phi(\frac{f(x)}{f_j(x)})f_j(x)dx\right)\int_{\mathbb{R}^d}\left[\sqrt{nh^d}(\widehat{f}_{nh_n}(x)-f(x))\phi'(\frac{f(x)}{f_j(x)})dx+o(1)\right]+o(1)\\
&\approx \int_{\mathbb{R}^d}\left[\psi'\left(\mathcal{D}_\phi(f,f_j)\right)\phi'(\frac{f(x)}{f_j(x)})\right]\sqrt{nh^d}(\widehat{f}_{nh_n}(x)-f(x))dx
\end{aligned} \tag{13}
$$

131 replacing Eq.13 in Eq.12

$$
\begin{aligned}
\sqrt{nh_n^d}\widehat{\mathcal{DI}}_n &= \int_{\mathbb{R}^d}\left[\psi'\left(\mathcal{D}_\phi(f,f_1)\right)\phi'(\frac{f(x)}{f_j(x)})\right]\sqrt{nh^d}(\widehat{f}_{nh_n}(x)-f(x))dx\\
&\quad -\int_{\mathbb{R}^d}\left[\psi'\left(\mathcal{D}_\phi(f,f_2)\right)\phi'(\frac{f(x)}{f_2(x)})\right]\sqrt{nh^d}(\widehat{f}_{nh_n}(x)-f(x))dx\\
&= \int_{\mathbb{R}^d}\left[\psi'\left(\mathcal{D}_\phi(f,f_1)\right)\phi'(\frac{f(x)}{f_1(x)})-\psi'\left(\mathcal{D}_\phi(f,f_2)\right)\phi'(\frac{f(x)}{f_2(x)})\right]\times\sqrt{nh^d}(\widehat{f}_{nh_n}(x)-f(x))dx
\end{aligned}
$$

By Eq 11, we have

$$\sqrt{nh_n^d}\widehat{\mathcal{DI}}_n \longrightarrow \mathcal{N}\left(0, \Gamma^2\right)$$

where

$$\Gamma^2 = \left\{ \int_{\mathbb{R}^d} \left[ \psi^{'}\left(\mathcal{D}_\phi(f, f_1)\right) \phi^{'}(\frac{f(x)}{f_1(x)}) - \psi^{'}\left(\mathcal{D}_\phi(f, f_2)\right) \phi^{'}(\frac{f(x)}{f_2(x)}) \right] \sigma(x) dx \right\}^2 .$$

Note that in the case of the $\alpha$-divergence the asymptotic variance $\Gamma^2$ is

$$\Gamma^2 := \Gamma^2(\alpha) = \left\{ \int_{\mathbb{R}^d} \left[ \psi^{'}\left( \int_{\mathbb{R}^d} \phi(\frac{f(x)}{f_1(x)}) f_1(x) dx \right) \phi^{'}(\frac{f(x)}{f_1(x)}) - \psi^{'}\left( \int_{\mathbb{R}^d} \phi(\frac{f(x)}{f_2(x)}) f_2(x) dx \right) \phi^{'}(\frac{f(x)}{f_2(x)}) \right] \sigma(x) dx \right\}^2 .$$

with $\psi(x) = x$ and $\phi(x) = \frac{1}{\alpha(\alpha-1)}(x^\alpha - \alpha(x-1) - 1)$

$$\Gamma^2(\alpha) = \left\{ \frac{1}{\alpha - 1} \int_{\mathbb{R}^d} \left[ \left(\frac{f(x)}{f_1(x)}\right)^{\alpha-1} - \left(\frac{f(x)}{f_2(x)}\right)^{\alpha-1} \right] \sqrt{f(x)} dx \int_{\mathbb{R}^d} K^2(z) dz) \right\}^2$$

In the special case where $\alpha = 1/2$, this asymptotic variance does not depend to the unknown density $f$ and it is expressed by :

$$\Gamma^2(1/2) = \left\{ 2 \int_{\mathbb{R}^d} \left( \sqrt{f_1(x)} - \sqrt{f_2(x)} \right) dx \int_{\mathbb{R}^d} K^2(z) dz \right\}^2 .$$

But the case $\alpha \neq 1/2$, $\Gamma^2(\alpha)$ is unknown because it depends on $f$ which is also unknown. In practice, one way solve this problem is to substitute $f$ with its consistency kernel estimator $\widehat{f_{nh_n}}$ and to plug it in $\Gamma^2(\alpha)$.

## 5. Computational Results

### *5.1. Example*

To illustrate the model procedure discussed in the preceding section. I rely on a simple specification such that:

$$\begin{cases} W_t = \varepsilon_t, & (M1), \\ W_t = -0.2 W_{t-1} + \varepsilon_t - \varepsilon_{t-1} & (M2). \end{cases}$$

with $\varepsilon_t \sim \mathcal{N}(0, 1)$, It was in this case the densities under $M1$ and $M2$ respectively:

$$f_1(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \qquad f_2(x) = \frac{1}{\sqrt{2\pi \times 2.5}} \exp\left(-\frac{x^2}{2 \times 2.5}\right)$$

We consider various sets of experiments in which data are generated from the mixture of a Normal $\mathcal{N}(0, 1)$ and Normal $\mathcal{N}(0, 2.5)$ distributions. Hence the DGP (Data Generating Process) is generated from $m(\pi)$ with the density

$$m(\pi) = \pi \mathcal{N}(0, 1) + (1 - \pi) \mathcal{N}(0, 2.5)$$

where $\pi(\pi \in [0, 1])$ is specific value to each set of experiments. In each set of experiments severals random sample are drawn from this mixture of distributions. The sample size varies from 100 to 2000, and for each sample size the number of replication is 1000. we choose value of the parameter $\alpha = 0.5$ , that corresponds to the Hellinger distance(this choice provided to the known asymptotic variance) . The aim is to compare the distance between true density and the density $\mathcal{N}(0, 1)$, and the distance between the true density and the density $\mathcal{N}(0, 2.5)$.
We choose different values of $\pi$ which are $0.00, 0.25, 0.43, 0.75, 1.00$.. Although our proposed model selection procedure does not require that the data generating process belong to either of the competing models, we consider the two limiting cases $\pi = 1.00$ and $\pi = 0.00$ for they correspond to the correctly specified cases. To investigate the case where both competing models are misspecified but not at equal distance from the DGP, we consider the case $\pi = 0.25$, $\pi = 0.75$ and $\pi = 0.43$. Second case is interpreted similarly as a $\mathcal{N}(0, 2.5)$ slightly contaminated by a $\mathcal{N}(0, 1)$ distribution. The former case correspond to a *DGP* which is $\mathcal{N}(0, 1)$ but slightly contaminated by a $\mathcal{N}(0, 2.5)$

distribution. In the last case, $\pi = 0.43$ is the value for which the $\widehat{\mathcal{D}}_\alpha(\widehat{f_n}, f_1)$ and the $\widehat{\mathcal{D}}_\alpha(\widehat{f_n}, f_2)$ family are approximately at equal distance to the mixture $m(\pi)$ according to the $\alpha$-divergence with the above cells. Thus, this series of experiments approximates the null hypothesis of our proposed model selection test $\widehat{\mathcal{DI}}_\alpha$. The results of our different sets of experiments are presented in **Tables 1-5** .

**Table 1.** $DGP = \mathcal{N}(0, 1)$

| $n$ | | 20 | 100 | 300 | 500 | 1000 | 1500 | 2000 |
|---|---|---|---|---|---|---|---|---|
| $\widehat{\mathcal{D}}_1$ | | -0.05 | 0.007 | -0.002 | 0.016 | -0.004 | 0.012 | 0.006 |
| $\widehat{\mathcal{D}}_2$ | | 0.16 | 0.12 | 0.14 | 0.16 | 0.14 | 0.14 | 0.14 |
| $\widehat{\mathcal{DI}}_\alpha$ | | -0.21 | -0.11 | -0.15 | -0.14 | -0.146 | -0.12 | -0.14 |
| | Correct | 8.4% | 8% | 26.4% | 57.8% | 95.6% | 100% | 100% |
| | Indecisive | 91.6% | 92% | 73.6% | 42.2% | 4.4% | 0% | 0% |
| | Incorrect | 0% | 0% | 0% | 0% | 0% | 0% | 0% |

**Table 2.** $DGP = \mathcal{N}(0, 2.5)$

| $n$ | | 20 | 100 | 300 | 500 | 1000 | 1500 | 2000 |
|---|---|---|---|---|---|---|---|---|
| $\widehat{\mathcal{D}}_1$ | | 0.26 | 0.14 | 0.22 | 0.28 | 0.23 | 0.24 | 0.24 |
| $\widehat{\mathcal{D}}_2$ | | -0.039 | -0.016 | -0.008 | -0.006 | -0.004 | -0.002 | -0.001 |
| $\widehat{\mathcal{DI}}_\alpha$ | | 0.30 | 0.16 | 0.23 | 0.29 | 0.23 | 0.24 | 0.24 |
| | Correct | 30.8% | 68.4% | 94.2% | 99% | 100% | 100% | 100% |
| | Indecisive | 69% | 31.6% | 5.6% | 1% | 0% | 0% | 0% |
| | Incorrect | 0.2% | 0% | 0.2% | 0% | 0% | 0% | 0% |

**Table 3.** $DGP = .75 * \mathcal{N}(0, 1) + .25 * \mathcal{N}(0, 2.5)$

| $n$ | | 20 | 100 | 300 | 500 | 1000 | 1500 | 2000 |
|---|---|---|---|---|---|---|---|---|
| $\widehat{\mathcal{D}}_1$ | | -0.014 | 0.015 | -0.001 | 0.01 | -0.002 | 0.01 | 0.01 |
| $\widehat{\mathcal{D}}_2$ | | 0.19 | 0.19 | 0.16 | 0.13 | 0.13 | 0.11 | 0.12 |
| $\widehat{\mathcal{DI}}_\alpha$ | | -0.21 | -0.17 | -0.16 | -0.12 | -0.13 | -0.1 | -0.11 |
| | $\mathcal{N}(0, 1)$ | 1.6% | 5.4% | 34.4% | 67.4% | 99% | 100% | 100% |
| | Indecisive | 98.4% | 94.6% | 64.4% | 32.6% | 1% | 0% | 0% |
| | $\mathcal{N}(0, 2.5)$ | 0% | 0% | 0% | 0% | 0% | 0% | 0% |

**Table 4.** $DGP = .43 * \mathcal{N}(0, 1) + .57 * \mathcal{N}(0, 2.5)$

| $n$ | | 20 | 100 | 300 | 500 | 1000 | 1500 | 2000 |
|---|---|---|---|---|---|---|---|---|
| $\widehat{\mathcal{D}}_1$ | | 0.1 | 0.05 | 0.04 | 0.05 | 0.04 | 0.053 | 0.057 |
| $\widehat{\mathcal{D}}_2$ | | 0.08 | 0.02 | 0.06 | 0.04 | 0.05 | 0.056 | 0.058 |
| $\widehat{\mathcal{DI}}_\alpha$ | | 0.02 | 0.03 | -0.02 | 0.01 | -0.01 | -0.002 | -0.01 |
| | $\mathcal{N}(0, 1)$ | 1.4% | 0.2% | 0.2% | 0% | 0% | 0% | 0% |
| | Indecisive | 98.4% | 99.8% | 99.8% | 100% | 100% | 100% | 100% |
| | $\mathcal{N}(0, 2.5)$ | 0.2% | 0% | 0% | 0% | 0% | 0% | 0% |

**Table 5.** $DGP = .25 * \mathcal{N}(0, 1) + .75 * \mathcal{N}(0, 2.5)$

| $n$ | | 20 | 100 | 300 | 500 | 1000 | 1500 | 2000 |
|---|---|---|---|---|---|---|---|---|
| $\widehat{\mathcal{D}}_1$ | | 0.69 | 0.83 | 1.006 | 0.86 | 1.08 | 1.04 | 0.99 |
| $\widehat{\mathcal{D}}_2$ | | -0.024 | 0.039 | 0.02 | 0.06 | 0.05 | 0.046 | 0.06 |
| $\widehat{\mathcal{DI}}_\alpha$ | | 0.67 | 0.79 | 1.04 | 0.8 | 1.03 | 0.99 | 0.92 |
| | $\mathcal{N}(0, 1)$ | 0.6% | 0% | 0% | 0% | 0% | 0% | 0.1% |
| | Indecisive | 21% | 17% | 0.4% | 0.2% | 0.2% | 0.2% | 0.1% |
| | $\mathcal{N}(0, 2.5)$ | 78.4% | 83% | 99.6% | 99.8% | 99.8% | 99.8% | 99.9% |

Thus this set of experiments corresponds approximately to the null hypothesis of our proposed model selection test $\widehat{\mathcal{DI}}$ . The results of our different sets of experiments are presented in **Tables 1-5**. The first half of each table gives the distance between the true density $f$ and $f_1$ sample take density model 1 $\mathcal{D}_1$, the distance between $f$ and $f_2$ Model 2 $\mathcal{D}_2$ and the difference between the two distance. The second half of each table gives in percentage the number of times our proposed model selection procedure based on $\widehat{\mathcal{DI}}$ favors the model 1, the model 2, and indecisive. The tests

are conducted at 5% nominal significance level. In the first two sets of experiments ($\pi = 0.00$ and $\pi = 1.00$) where one model is correctly specified, we use the labels "correct, incorrect" and "indecisive" when a choice is made. The first halves of **Tables 1-5** confirm our asymptotic results.
In **Tables 4**, we observed a high percentage of bad decisions. This is because both models are now specified incorrectly. In contrast, turning to the second halves of the **Tables 1** and **2**, we first note that the percentage of correct choices using $\mathcal{DI}$ statistic steadily increases and ultimately converges to 100%
The preceding comments for the second halves of **Tables 1** and **2** also apply to the second halves of **Tables 3** and **5**.

In *Figures 1*,3, 5, 7 and 9 we plot the histograms of data sets and overlay the curves for $\mathcal{N}(0,1)$ and $\mathcal{N}(0,2.5)$ distributions. When the *DGP* is correctly specified *Figure 1*, the $\mathcal{N}(0,1)$ distribution has reasonable chance of being distinguished from $\mathcal{N}(0,1)$ distribution.
Similarly, in *Figure 3*, as can be seen, the $\mathcal{N}(0,2.5)$ distribution closely approximates the data sets. In *Figures 5* and 7 two distributions are close but the $\mathcal{N}(0,1)$ (*Figure 5*) and the $\mathcal{N}(0,2.5)$ distributions (*Figure 7*) does appear to be much closer to the data sets. When $\pi = 0.43$, the distribution for both (*Figure 9*) $\mathcal{N}(0,1)$ distribution and $\mathcal{N}(0,2.5)$ distribution are similar.

As expected, our statistic divergence $\sqrt{nh_n^d}\widehat{\mathcal{DI}}_\alpha$ diverges to $-\infty$ (*Figures 2* and 6) and to $+\infty$ ( *Figures 4* and 8) more rapidly symmetrical about the axis that passes through the mode of data distribution. This follows from the fact that these two distributions are equidistant from the *DGP* and would be difficult to distinguish from data in practice.
*Figure 10* allows a comparison with the asymptotic $\mathcal{N}(0,\Gamma)$ approximation under our null hypothesis of equivalence. *Figure 11*, Hence the density indicator $\widehat{\mathcal{DI}}_\alpha$ is very closer to the $\mathcal{N}(0,\Gamma)$.

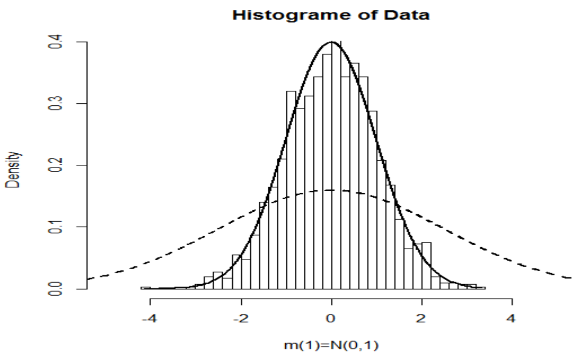

Figure 1: Histogram of ($DGP = \mathcal{N}(0,1)$)

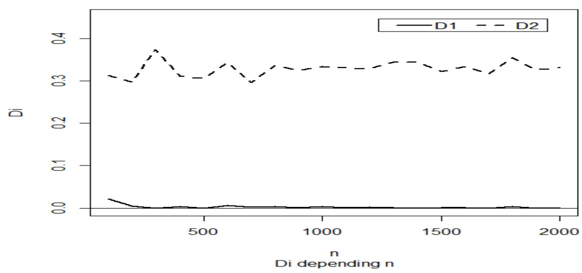

Figure 2: $\widehat{\mathcal{D}}_1$ and $\widehat{\mathcal{D}}_2$ depending $n$

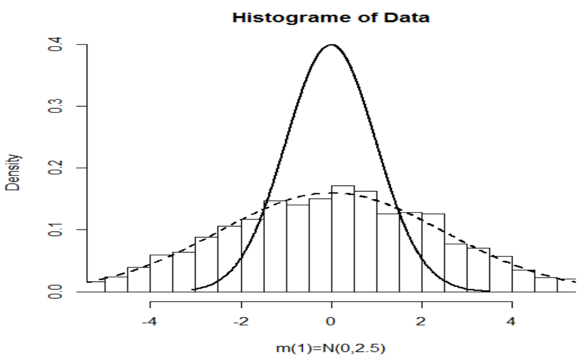

Figure 3: Histogram of ($DGP = \mathcal{N}(0,2.5)$)

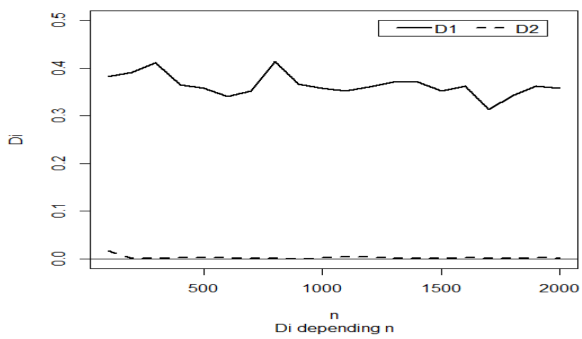

Figure 4: $\widehat{\mathcal{D}}_1$ and $\widehat{\mathcal{D}}_2$ depending $n$

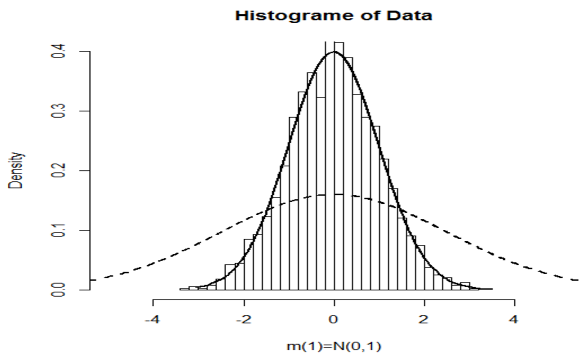


Figure 5: Histogram of $(DGP = .75 * \mathcal{N}(0, 1) + .25 * \mathcal{N}(0, 2.5))$

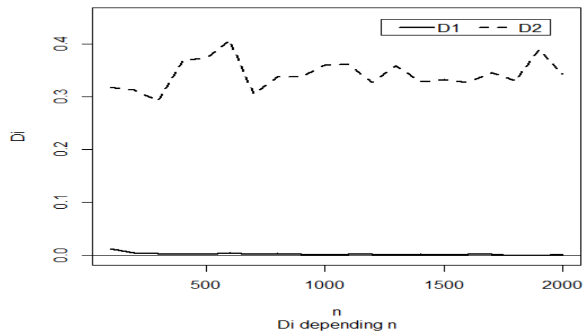


Figure 6: $\widehat{\mathcal{D}}_1$ and $\widehat{\mathcal{D}}_2$ depending $n$

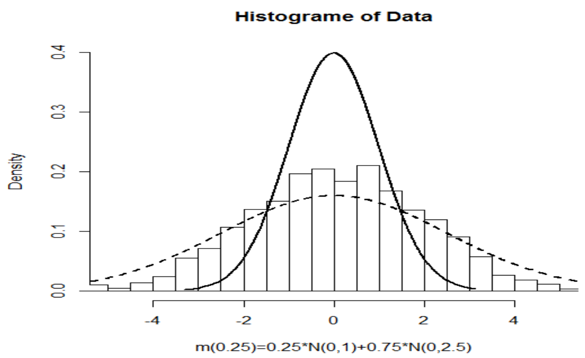


Figure 7: Coparaison barplot of Di depending n $(DGP = .25 * \mathcal{N}(0, 1) + .75 * \mathcal{N}(0, 2.5))$

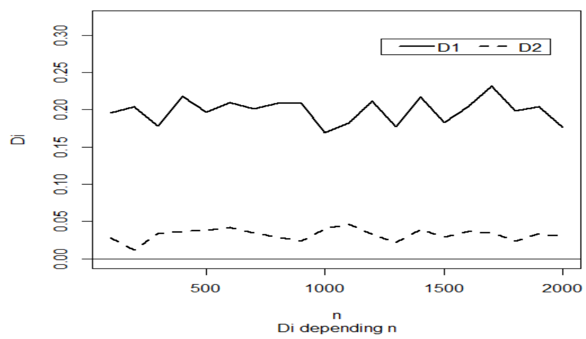


Figure 8: $\widehat{\mathcal{D}}_1$ and $\widehat{\mathcal{D}}_2$ depending $n$

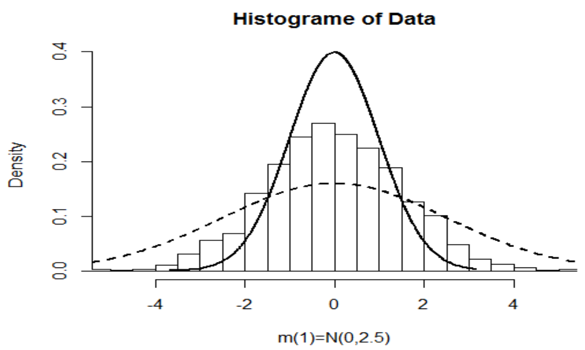


Figure 9: Coparaison barplot of Di depending n $(DGP = .43 * \mathcal{N}(0, 1) + .57 * \mathcal{N}(0, 2.5))$

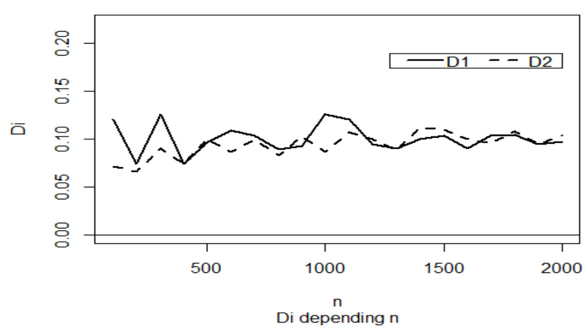


Figure 10: $\widehat{\mathcal{D}}_1$ and $\widehat{\mathcal{D}}_2$ depending $n$

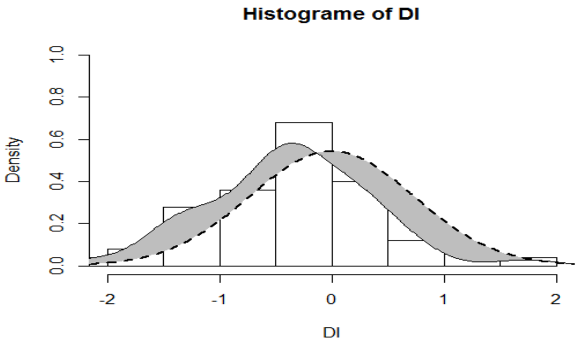


Figure 11: Coparaison of density depending $n$ density data (continuous curve), density $\mathcal{N}(0,\Gamma)$(dashed curve)

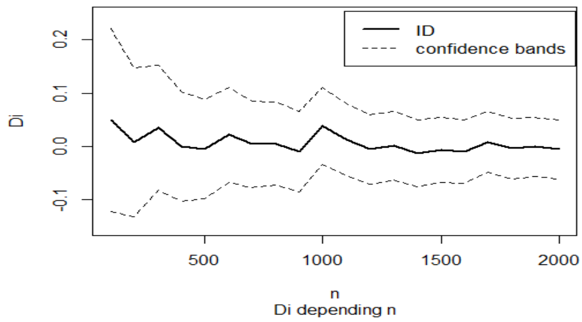


Figure 12: $\widehat{\mathcal{ID}}$ depending $n$ aut its confidence bande at the levels 95%

## 6. Concluding remarks and future works

We have formulated the $\mathcal{DI}$ method and applied it to the problem of choosing between a random walk and a stationary frist order autoregressive model, using $(\phi,\psi)$-divergence type statistics. In this context, we have considered some convenient asymptotically standard tests based on $(\phi,\psi)$-divergence type statistics that use estimators in non parametric case. The results of the numerical experiments are most encouraging and show that $\mathcal{DI}$ method performs very well and can be considered as a useful tool for addressing problems in model selection. these test allow to determine whether the competing model is as close to true distribution against the alternative hypothesis that one model is closer. Here closeness is evaluated according to the discrepancy implicit in the $(\phi,\psi)$-divergence type statistic considered.

## References


[1] Akaike. H, An Approximation to the Limit Theorems For Associated Random Fields and Related Systems.. *Math., Tokyo.* **6** (1954) 127–132.

[2] Bulinski. A and Shashkin. A, Limit Theorems For Associated Random Fields And Related Systems, Advanced Series on Statistical Science and Applied Probability, Vol. 10.

[3] Basu, A.; Harris, I.R.; Hjort, N.L.; Jones, M.C. Robust and efficient estimation by minimising a density power divergence. Biometrika. 1998, 85, 549559.

[4] S. Bouzebda and I. Elhattab, Uniform-in-bandwidth consistency for kernel-type estimators of Shannon's entropy. *Electronic Journal of Statistics.* **5** (2011) 440–459.

[5] D. Bosq and J. P. Lecoutre. Théorie de l'estimation fonctionnelle. *Économie et Statistiques Avancées. Economica, Paris.* (1987)

[6] Cressie, N.; Read, T.R.C. Multinomial goodness of fit tests. J. R. Stat. Soc. Ser. B. 1984, 46, 440464.

[7] P. Deheuvels. Uniform limit laws for kernel density estimators on possibly unbounded intervals. *Stat. Ind. Technol.*(2000) 477–492.

[8] P. Deheuvels and D. M. Mason. General asymptotic confidence bands based on kernel-type function estimators. *Stat. Inference Stoch. Process.* **7**(2004) 225–277.

[9] L. Devroye and L. Gyorfi. Nonparametric density estimation. *Wiley Series in Probability and Mathematical Statistics*: Tracts on Probability and Statistics. John Wiley & Sons Inc., New York. The L1 view. (1985)

[10] L. Devroye and G. Lugosi. Combinatorial methods in density estimation. *Springer Series in Statistics. Springer-Verlag, New York.* (2001)

[11] U. Einmahl and D. M. Mason. An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.* **13** (2000) 1–37.

[12] H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences.* 186(1007) (1946) 453-461.

[13] Karagrigoriou, A.; Mattheou, K.; Vonta, F. On asymptotic properties of AIC variants with applications. Open J. Stat. 2011, 1, 105109.

[14] A. Lynda and F. Hocine. On the stability of the unit root test. *Journal Afrika Statistika.* **5** (2010) 228–237.

[15] J. Marriott and P. Newbold. Bayesian comparison of ARIMA and stationary ARMA models. *International Statistical Review.* **3** (1998) 323–336.

[16] Maronna, R.A.; Martin, R.D.; Yohai, V.J. Robust Statistics: Theory and Methods; Wiley: New York, NY, USA, 2006.

[17] Mattheou, K.; Lee, S.; Karagrigoriou, A. A model selection criterion based on the BHHJ measure of divergence. J. Stat. Plan. Inference 2009, 139, 228235.

[18] Menéndez, M. L., Morales, D., Pardo, L. and Salicru, M. (1995). Asymptotic behaviour and statistical applications of divergence measures in multinomial populations: A unified study. Statistical Papers, 36, 1-29.

[19] Menéndez, M. L., Morales, D., Pardo, L. and Vajda, I. (1997b). Testing in stationary models based on f-divergences of observed and theoretical frequencies. Kybernetika, 33, 465-475.
[20] Menéndez, M. L., Pardo, J. A., Pardo, L. and Pardo, M. C. (1997c). Asymptotic approximations for the distributions of the (h, f)-divergence goodness-of-fit statistics: Applications to Renyis statistic. Kybernetes, 26, 442-452.
[21] Morales, D., Pardo, L. and Vajda, I. (1997). Some new statistics for testing hypotheses in parametric models. Journal of Multivariate Analysis, 62, 1, 137-168.
[22] Nayak, T. K. (1985). On diversity measures based on entropy functions. Communications in Statistics (Theory and Methods), 14, 203-215.
[23] D. B. Owen. Statistical Inference Based on Divergence Measures. *Taylor & Francis Group, LLC, (2006).*
[24] Pardo et al. (1995)
[25] E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** (1962) 1065–1076.
[26] B. L. S. Prakasa Rao. Nonparametric functional estimation. *Probability and Mathematical Statistics. Academic Press Inc.* [Harcourt Brace Jovanovich Publishers], New York (1983).
[27] Read, T. R. C. and Cressie, N. A. C. (1988). Goodness of Fit Statistics for Discrete Multivariate Data. Springer-Verlag, New York.
[28] A. Rényi. On measures of entropy and information. *In Fourth Berkeley Symposium on Mathematical Statistics and Probability.* (1961)
[29] Ronchetti, E. Robust model selection in regression. Stat. Probab. Lett. 1985, 3, 2123.
[30] Ronchetti, E.; Staudte, R.G. A robust version of Mallows CP. J. Am. Stat. Assoc. 1994, 89, 550559.
[31] Salicru, M., Morales, D., Menendez, M. L. and Pardo, L. (1994). On the applications of divergence type measures in testing statistical hypotheses. Journal of Multivariate Analysis, 51, 372-391.
[32] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, **27** (1956) 832–837.
[33] Zografos, K., Ferentinos, K. and Papaioannou, T. (1990). ?-divergence statistics: Sampling properties, multinomial goodness of fit and divergence tests. Communications in Statistics (Theory and Methods), 19, 5, 1785-1802.
[34] Zografos, K. (1993). Asymptotic properties of f-divergence statistic and applications in contingency tables. International Journal of Mathematics and Statistical Sciences, 2, 5-21.
[35] Zografos, K. (1994). Asymptotic distributions of estimated f-dissimilarity between populations in stratified random sampling. Statistics and Probability Letters, 21, 147-151.
[36] Zografos, K. (1998a). f-dissimilarity of several distributions in testing statistical hypotheses. Annals of the Institute of Statistical Mathematics, 50, 295-310.