Independence Tests based on the Conditional Expectation

SERGIO ORTOBELLI
Dipartimento di Scienze aziendali, economiche e metodi quantitativi
University of Bergamo
Via dei Caniana 2, Bergamo
ITALY;
and
Department of Finance
VŠB-TU Ostrava
Sokolskà 33, Ostrava
CZECH REPUBLIC
sergio.ortobelli@unibg.it

TOMMASO LANDO
Dipartimento di Scienze aziendali, economiche e metodi quantitativi
University of Bergamo
Via dei Caniana 2, Bergamo
ITALY;
and
Department of Finance
VŠB-TU Ostrava
Sokolskà 33, Ostrava
CZECH REPUBLIC
tommaso.lando@unibg.it

Abstract: - In this paper we propose a new procedure for testing independence of random variables, which is based on the conditional expectation. As it is well known, the behaviour of the conditional expectation may determine a necessary condition for stochastic independence, that is, the so called mean independence. We provide a necessary and sufficient condition for independence in terms of conditional expectation and propose an alternative method to test independence based on this result. Consequently, we provide general class of tests. Observe that generally some non-parametric methods are needed to approximate the conditional expectation, since its exact expression (given the joint distribution) is usually unknown, except for few trivial cases (e.g. Gaussian): we generalize this well known result to the family of elliptical distributions. In order to obtain a sufficiently accurate approximation of the conditional expectation, we propose to use the kernel method or, alternatively, the recently introduced OLP method.

Key-Words: - Independence test, conditional expectation, Kernel, Non Parametric test

1 Introduction

The definition of stochastic independence states that two events are independent if and only if the probability of both occurring equals the product of the probabilities of each occurring. Hence, independence of random variables (r.v.'s) basically corresponds to independence between each couple of events they are related to. Based on this definition, many statistical tests that have been proposed in the literature are generally suitable for discrete and in particular categorical r.v.'s, rather than continuous r.v.'s. In this paper we attempt to bridge this gap by introducing an alternative necessary and sufficient condition for stochastic independence, thereby we propose a new methodology for testing independence.

Several well known procedures, such as the Pearson's Chi squared test [1], or the Hoeffding's test [2] are based on the main definition of independence, provided above. By the application of this criterion, it is possible to evaluate independence of random variables trough their realizations. Nevertheless, if some realizations are not independent, this approach is sufficient to guarantee that the random variables are not independent, which leads to rejection of the so called *null hypothesis* (i.e. independence). However, it is worth noting that methods that are based on this logic but could present some drawbacks when dealing with continuous r.v.'s. Indeed, in this case we are not able to guarantee that the random variables are independent only because a few events are independent. In other words, these tests may present a high probability of *type II error*, that is, the failure to reject a false null hypothesis, especially in the continuous case.

We argue that an alternative method for testing stochastic independence is needed, especially in those areas of study where distributions are generally assumed to be continuous. Speaking of which, in several financial applications, the Chi-squared test and other similar methods are typically used although the financial random variables are continuous. For instance, when we evaluate the risk interval forecasts, with reference to the information

available at each time, we may use the tests proposed by [3] and [4], or, with a Chi-squared test, we may also evaluate the time-independence.

The alternative method proposed in this paper is based on the conditional expectation between random variables. Speaking of which, it should be stressed that generally the conditional expectation between two random variables X and Y, i.e. $E(Y|X)$, needs to be approximated with some non parametric method, because its form is often unknown. Indeed, we recall that, for a given joint distribution, we are not generally able to derive an exact mathematical expression of the conditional expectation, except for some trivial text book examples. For instance, the distribution of $E(Y|X)$ is well known in the Gaussian case where, in particular, $E(Y|X)$ is a linear function of the r.v. X (this rule holds also for the Student's t distribution, see e.g. [5]). In this paper, we extend this result to a larger class of distribution, that is, the family of elliptical joint distributions.

As observed by [6], the conditional expectation between two random variables X and Y, i.e. $E(Y|X)$, can be approximated in a satisfactory way by the use of two different methodologies, namely the Kernel method and the OLP method, recently introduced by [7]. On the one hand, the kernel non-parametric regression (see [8] and [9]) allows to estimate $E(Y|X=x)$ as a locally weighted average, based on the choice of an appropriate kernel function: the method yields consistent estimators, provided that the kernel functions and the random variable Y satisfy some conditions. On the other hand, the OLP method is aimed at estimating the random variable $E(Y|X)$ and it has been proved to be consistent without requiring any regularity assumption.

The main result of the paper is a theorem which determines a necessary and sufficient condition for independence, based on the conditional expectation. Generally, the conditional expectation is related to independence because of the definition of mean independence, that is, $E(Y|X)=E(Y)$. However, mean independence is a weak condition, compared to stochastic independence, therefore we generally cannot analyze the independence structure between X and Y by observing $E(Y|X)$. Nevertheless, in section 4 we prove that two random variables X and Y are independent if and only if, given a non constant, continuous and positive (or negative) function h, $E(h(Y)|X)=E(h(Y))$. Hence, once we have chosen h, we are able to approximate the r.v. $E(h(Y)|X)$ and verify if this condition is verified. In particular, we can analyze the dependence structure between X and Y reducing the risk of misevaluations. We also observe that the behaviour of $E(h(Y)|X)$ gives stronger indications about independence compared to other well known dependence measures, such as correlation and mean dependence measures.

Finally, by using the aforementioned methods for approximating $E(h(Y)|X)$, we propose a decision rule aimed at testing independence. This method could be especially useful and appropriate in case of continuous r.v.'s.

The paper is organized as follows. In section 2 we establish the condition under which the conditional expectation $E(Y|X)$ can be expressed as a linear function of X. Then, in section 3 we review two alternative methods for approximating $E(Y|X)$, namely the OLP and the kernel methods. In section 4 we prove that stochastic independence can be expressed in terms of conditional expectation, and thereby we provide a new alternative definition of independence, based on which we can propose a class of tests of independence. Future work and possible applications are eventually discussed in the conclusion.

2 The determination of a conditional expectation

As is stated in the introduction, the main objective of this paper is to present an alternative approach for testing independence of random variables, based on the concept of conditional expectation. In this section, we provide some definitions and, especially, an important result that makes it possible to determine the distribution of a conditional expectation when the random variables are jointly elliptically distributed, generalizing the well know rule that holds e.g. in the Gaussian case.

Let Y be an integrable random variable on the probability space $(\Omega, \Im, P)$ and let $\Im'$ be a sub-sigma-algebra of $\Im$ (i.e. $\Im' \subseteq \Im$). The conditional expectation of Y given $\Im'$ is the unique (P a.s.) random variable $E(Y|\Im')$ such that:

i) $E(Y|\Im')$ is $\Im'$-measurable;

ii) for any $A \in \Im'$, $\int_A E(Y|\Im')dP = \int_A YdP$.

Let $X: \Omega \to \mathbb{R}$ and $Y: \Omega \to \mathbb{R}$ be integrable random variables defined on the probability space $(\Omega, \Im, P)$. If $\Im' = \sigma(X)$ is the sigma algebra generated by X we can write $E(Y|\sigma(X)) = E(Y|X) = g(X)$. Generally, the distribution of $g(X)$ is unknown, unless the joint

distribution of the random vector (X,Y) follows some special distribution, e.g. the Gaussian distribution or the multivariate *t* distribution. However, if we assume that X and Y are jointly normally distributed, i.e. $(X,Y)\sim N(\mu,\Sigma)$, (where obviously $\mu=(\mu_X,\mu_Y)$ is the vector of means and $\Sigma=\left((\sigma_X{}^2,\rho_{XY}\sigma_X\sigma_Y),(\rho_{XY}\sigma_X\sigma_Y,\sigma_Y{}^2)\right)$ is the variance-covariance matrix[1]) we can obtain the distribution of the random variable $E(Y|X)$ quite easily. Indeed, it is well known that $g(x)=E(Y|X=x)=\mu_Y+\rho_{XY}\frac{\sigma_Y}{\sigma_X}(x-\mu_X)$, and thereby

$$E(Y|X)=\mu_Y+\rho_{XY}\frac{\sigma_Y}{\sigma_X}(X-\mu_X) \qquad (1)$$

is normally distributed with mean μ_Y and variance $\rho_{XY}^2\sigma_Y^2$. Clearly, if $\rho_{XY}=\pm1$, then $Y=\mu_Y+\rho_{XY}\frac{\sigma_Y}{\sigma_X}(X-\mu_X)$ *P*- almost surely and equation (1) holds for any joint distribution of the vector (X,Y). Equation (1) holds also for joint Student's *t* bivariate vector, as pointed out by [5]. In this paper, we extend the application of equation (1) to a larger class of distributions, as stated in the following theorem.

Theorem 1 *Assume that the random vector* $Z=(X,Y)$ *with dispersion matrix* $\Sigma=\left((\sigma_X{}^2,\sigma_{XY}),(\sigma_{XY},\sigma_Y{}^2)\right)$ *and mean* $\mu=(\mu_X,\mu_Y)$ *belongs to a family of jointly elliptically distributed vectors* $Ell(\mu,\Sigma)$. *Moreover, assume that* $Z=\mu+AG$, *where* A *is a continuous positive random variable which is independent from the Gaussian vector* G *that has null mean and variance covariance matrix* Σ *(i.e.* $E(G)=(0,0)$, $Cov(G)=\Sigma$*). Then,*

$$E(Y|X)=\mu_Y+\frac{\sigma_{XY}}{\sigma_X^2}(X-\mu_X)\sim Ell\left(\mu_Y,\frac{\sigma_{XY}^2}{\sigma_X^2}\right).$$

Proof

Without loss of generality we can prove the result when the mean vector (μ_X,μ_Y) is a null vector. Let f_A be the density distribution of the random variable A and denote with $G=(X_1,Y_1)$ the bivariate Gaussian vector. Then $X=AX_1$ and $Y=AY_1$. Then the cumulative distribution of $Y|X=x$ is given by:

$$F_{Y|X=x}(y)=P(Y\le y|X=x)=$$
$$=\int_0^{+\infty}P\left(Y_1\le\frac{y}{a}|X_1=\frac{x}{a}\right)f_A(a)da$$

and its density $\frac{\partial F_{Y|X=x}(y)}{\partial y}$ is simply given by:

$$\frac{f_{X,Y}(x,y)}{f_X(x)}=\int_0^{+\infty}\frac{f_{X_1,Y_1}\left(\frac{x}{a},\frac{y}{a}\right)}{af_{X_1}\left(\frac{x}{a}\right)}f_A(a)da.$$

Therefore, the conditional expectation is given by:

$$E(Y|X=x)=\int_{-\infty}^{+\infty}y\frac{f_{X,Y}(x,y)}{f_X(x)}dy$$
$$=\int_0^{+\infty}f_A(a)\left(\int_{-\infty}^{+\infty}\frac{y}{a}\frac{f_{X_1,Y_1}\left(\frac{x}{a},\frac{y}{a}\right)}{af_{X_1}\left(\frac{x}{a}\right)}dy\right)da=$$
$$=\int_0^{+\infty}af_A(a)\left(\int_{-\infty}^{+\infty}t\frac{f_{X_1,Y_1}\left(\frac{x}{a},t\right)}{f_{X_1}\left(\frac{x}{a}\right)}dt\right)da$$
$$=\int_0^{+\infty}af_A(a)E\left(Y_1|X_1=\frac{x}{a}\right)da$$
$$=\int_0^{+\infty}af_A(a)\frac{\sigma_{XY}}{\sigma_X^2}\frac{x}{a}da=\frac{\sigma_{XY}}{\sigma_X^2}x.$$

The above equalities are a consequence of the Fubini's theorem and the change of variable $t=y/a$ Thus, in the most general case, the equality $E(Y|X)=\mu_Y+\frac{\sigma_{XY}}{\sigma_X^2}(X-\mu_X)$ holds and $E(Y|X)$ is elliptical, because elliptical distributions are scalar and translation invariant.

Some examples of jointly elliptical distributions that admit the decomposition $Z=\mu+AG$ are: the *t*-Student multivariate distribution or all distributions used for symmetric Lévy processes where *A* is the subordinator such as Normal Inverse Gaussian (NIG) symmetric vectors, Variance Gamma (VG) symmetric vectors, alpha stable sub-Gaussian vectors (see, among others, [10], [11]). As a matter of fact, any semi-martingale can be written as a time changed Brownian Motion. Moreover, observe that when the vector *Z* admits finite covariance matrix, then $\rho_{XY}=\frac{\sigma_{XY}}{\sigma_Y\sigma_x}$ is the Pearson's correlation measure. However, this property of the conditional expected value holds even for some elliptical distributions which do not admit finite variance, for example α –stable sub-Gaussian vector with $\alpha\in(1,2)$, where $\frac{\sigma_{XY}}{\sigma_Y\sigma_x}$ represents an alternative measure of correlation. Basically, if we know that the bivariate random vector $Z=(X,Y)$ admits the decomposition $Z=\mu+AG$, we also know the general form of the distribution of $g(X)$, and therefore we can estimate it quite easily. For instance, we could approximate $g(X)$ by estimating

[1] With a little abuse of notation, in this paper we write the dispersion matrix as

$$\Sigma=\begin{bmatrix}\sigma_X{}^2 & \sigma_{XY}\\ \sigma_{YX} & \sigma_Y{}^2\end{bmatrix}=\left((\sigma_X{}^2,\sigma_{XY}),(\sigma_{YX},\sigma_Y{}^2)\right).$$

the unknown parameters μ_Y, σ_{XY} and σ_X^2 respectively with the sample mean, the sample covariance (covariation) coefficient and the sample variance (dispersion). Furthermore, Theorem 1 can be easily extended to an n-dimensional ($n>2$) elliptical distributed vector as suggested in the following corollary.

Corollary 1 *Let $Z=[X,Y]'$ be an n-dimensional elliptically distributed $Ell(\mu,\Sigma)$ vector where X and Y are respectively m-dimensional and n-m dimensional vectors $(n>m\geq 1)$, the dispersion matrix is $\Sigma=\left((\Sigma_X,\Sigma_{XY}),(\Sigma_{YX},\Sigma_Y)\right)$ and the mean is $\mu=(\mu_X,\mu_Y)$. Moreover, assume that $Z=\mu+AG$, where A is a continuous positive random variable which is independent from the Gaussian vector G that has null mean and variance covariance matrix Σ. Then,*

$$E(Y|X)=\mu_Y+\Sigma_{YX}\Sigma_X^{-1}(X-\mu_X)\sim Ell\left(\mu_Y,\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\right).$$

Therefore, for several symmetric distributions used in different branches of engineering (i.e., elliptical distributions that admits the decomposition $Z=\mu+AG$) the conditional expectation depends on a particular linear correlation measure.
Unfortunately, the non linearity cannot be contemplated by equation (1) and its extension of Theorem 1 and Corollary 1. In other words, with equation (1) we cannot distinguish whether two random variables are independent or uncorrelated and dependent, because in both cases equation (1) yields $g(X)=E(Y|X)=E(Y)$. Let us consider this elementary counterexample.
Example. Let X be Gaussian distributed with null mean and $Y=X^2$. Then X and Y are dependent and uncorrelated random variables, that is, $\rho_{XY}=0$ and $E(Y|X)=Y\neq E(X^2)$.
In the most general case, when the assumptions of Theorem 1 are not verified, we need to approximate the conditional expectation by the use of some non parametric methods: in the next section we illustrate two alternative approaches.

3 Approximation methods

Since in several cases we do not know the distribution of $g(X)$, we generally cannot estimate it with parametric methods. Moreover, we generally do not have available a random sample drawn from the random variable $g(X)$, and therefore the estimation of the conditional expectation may not be simple. However, there are two alternative methods for estimating $g(X)$ and its distribution based on a standard bivariate random sample

$$(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)$$

of independent observations from the bi-dimensional variable (X,Y). In what follows, we briefly summarize the OLP method, recently introduced by, and the kernel non parametric regression method, as proposed in [6].

The OLP method

The OLP method, recently introduced by [7] (see also [6] and [12]), is aimed at approximating the conditional expectation, based on an appropriate partition of the sample space.
Define by $\sigma(X)$ the σ-algebra generated by X (that is, $\sigma(X)=X^{-1}(\mathcal{B})=\{X^{-1}(B):B\in\mathcal{B}\}$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$). The σ-algebra $\sigma(X)$ can be approximated by a σ-algebra generated by a suitable partition of Ω. In particular, for any $k\in\mathbb{N}$, we consider the partition $\{A_j\}_{j=1}^k=\{A_1,\dots,A_k\}$ of Ω in k subsets, described as follows:

$$A_1=\left\{\omega:X(\omega)\leq F_X^{-1}\left(\frac{1}{k}\right)\right\},$$
$$A_h=\left\{\omega:F_X^{-1}\left(\frac{h-1}{k}\right)<X(\omega)\leq F_X^{-1}\left(\frac{h}{k}\right)\right\},$$
$$for\ h=2,\dots,k-1$$
$$A_k=\Omega-\cup_{j=1}^{k-1}A_j=\{\omega:X(\omega)>F_X^{-1}\left(\frac{k-1}{k}\right)\}.$$

Note that the partition $\{A_j\}_{j=1}^k$ is determined by a number $(k-1)$ of percentiles of X. Moreover, by definition of percentile, each interval A_j has equal probability, that is, $P(A_j)=1/k$, for $j=1,\dots,k$. Starting with the trivial sigma algebra $\Im_1=\{\emptyset,\Omega\}$, we can obtain a sequence of sigma algebras generated by these partitions, for different values of k. Generally:

$$\Im_k=\sigma\left(\{A_j\}_{j=1}^k\right),k\in\mathbb{N}. \qquad (2)$$

Hence, it is possible to approximate the random variable $E(Y|\Im_X)$ by

$$E(Y|\Im_k)(\omega)=\sum_{j=1}^k\frac{1_{A_j}(\omega)}{P(A_j)}\int_{A_j}YdP=$$

$$\sum_{j=1}^k E(Y|A_j)1_{A_j}(\omega), \qquad (3)$$

where $1_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$. Indeed, by definition of the conditional expectation, observe that $E(Y|\Im_k)$ is a $\Im_k$-measurable function such that, for any set $A \in \Im_k$, (that can be seen as a union of disjoint sets, in particular $A = \bigcup_{A_j \subseteq A} A_j$) we obtain the equality

$$\int_A E(Y|\Im_k)dP = \int_A Y(\omega)dP(\omega) \quad (4)$$

It is proved in [7] that $E(Y|\Im_{k(h)})$ converges almost certainly to the random variable $E(Y|X)$, that is:

$$\lim_{k\to\infty} E(Y|\Im_k) = E(Y|X) \text{ a.s..} \quad (5)$$

Hence, if we approximate $E(Y|\Im_k)$, then we also approximate $g(X)$, for sufficiently large *k*. However, in practical situations, we do not know the probability measure *P*, which is necessary in order to approximate $E(Y|A_j)$ in equation (3). In these cases, we are able to approximate the random variable $E(Y|\Im_k)$, which in turns approximates $E(Y|X)$, using the large number law based on the observations of a random sample. Let $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ be a random sample of independent observations from the bi-dimensional variable (X, Y). First, as we generally do not know the marginal distribution of X, we can determine the partition $\{\hat{A}_j\}_{j=1}^k$ using the percentiles of the empirical distribution, obtained from the observations $(x_1, \dots, x_n)$. The number of intervals k should be basically an increasing function of the number of observations n, as discussed below. Then, if we assume to know the probability p_i, corresponding to the *i*-th outcome y_i, we obtain:

$$E(Y|\hat{A}_j) = \sum_{x_i \in \hat{A}_j} y_i p_i / P(\hat{A}_j). \quad (6)$$

Otherwise, we can give uniform weight to each observation, and thus we can use the following estimator of $E(Y|A_j)$:

$$\hat{a}_j = \frac{1}{n_{\hat{A}_j}} \sum_{x_i \in \hat{A}_j} y_i, \quad (7)$$

where n_{A_j} is the number of observations in $\hat{A}_j$, that is, $n_{A_j} = \#\{x_i : x_i \in A_j, i = 1, .., n\} \cong n/k$ (to clarify the explanation, for $k = 4$ we obtain the three quartiles, and therefore $n_{A_j} \cong \frac{n}{4}$ and similarly $P(A_j)$ can be estimated by $\frac{1}{4}$). Note that, fixed k, as the number of observations n grows, $P(\hat{A}_j) \xrightarrow{n\to\infty} P(A_j) = 1/k$ and $\hat{a}_j$ is an asymptotically unbiased estimator of $E(Y|A_j)$. Therefore, we are always able to approximate $E(Y|\Im_k)$ and the conditional expectation $E(Y|X)$ by using the following estimator :

$$\hat{g}_n^{OLP}(X) = \sum_{j=1}^k 1_{X\in\hat{A}_j} \sum_{x_i\in\hat{A}_j} y_i \frac{1}{n_{\hat{A}_j}} = \\ = \sum_{j=1}^k 1_{X\in\hat{A}_j} \hat{a}_j. \quad (8)$$

where X is assumed to be independent from the i.i.d. observations (x_i, y_i). Note that $\hat{g}_n^{OLP}$ is a simple $\Im_k$ measurable function, and it is conceptually different from the classical estimators, which are generally aimed at estimating an unknown parameter rather than a random variable. Observe that, given a bivariate sample of size *n*, $\hat{g}_n^{OLP}$ yields k distinct values, where each one has frequency $n_{\hat{A}_j} \cong n/k$, for $j = 1, \dots, k$. These outcomes can be used to estimate the unknown distribution function of $g(X)$.

Obviously, the selected number of intervals *k* can vary between 1 and *n* and, in order to improve the accuracy of the estimates, it must generally be an increasing function of *n*. In particular, for $k = 1$ we approximate the random variable $g(X)$ with a number, i.e. the sample mean $\bar{y}$, which is obviously not appropriate. On the other hand, for $k = n$ we approximate $g(X)$ with the marginal distribution of Y, given by $y_1, .., y_n$, which is also generally inappropriate. As shown by [6], in order to maximize i) the number of intervals, and ii) the number of observations in each interval ($n_{\hat{A}_j}$) and to enhance the performance of the method as well, we can choose:

$$k = \lceil \sqrt{n} \rceil, \quad (9)$$

where $\lceil x \rceil$ is the integer part of x. By doing so, we obtain k intervals containing (approximately) k observations.

The kernel method

The kernel method is typically used to estimate the regression function $g(x) = E(Y|X = x)$. In particular, if we do not know the general form of $g(x)$, except that it is a continuous and smooth function, then we can consider the following kernel estimator:

$$\hat{g}_n^{ker}(x) = \frac{\sum_{i=1}^n y_i K\left(\frac{x-x_i}{h(n)}\right)}{\sum_{i=1}^n K\left(\frac{x-x_i}{h(n)}\right)}, \quad (10)$$

where $K(x)$, denoted by *kernel*, is a density function (typically unimodal and symmetric around zero) such that i) $K(x) < C < \infty$; ii) $\lim_{x\to\pm\infty} |xK(x)| = 0$ (see [8] and [9]). Moreover, $h(n)$ is the smoothing parameter, often referred to as the *bandwidth* of the kernel, and it is a positive number such that $h(n) \to 0$ when $n \to \infty$. When the kernel K is the probability density function of a standard normal distribution, then the bandwidth is the standard deviation. It was proved in [8] that if Y is quadratically integrable (see also [13]) then $\hat{g}_n^{ker}(x)$ is a consistent estimator for $g(x)$. As a consequence, we know that $\hat{g}_n^{ker}(X) \to_{a.s.} g(X)$. Then, from a practical point of view, if we apply the kernel estimator to a bi-variate random sample $(x_1, y_1), \dots, (x_n, y_n)$ we obtain the vector $(g_1, \dots, g_n) = (\hat{g}_n^{ker}(x_1), .., \hat{g}_n^{ker}(x_n))$. In other words, each value g_i is a weighted average of kernels, centered at each sample observation x_i. Since we know that $g_i \to E(Y|X = x_i)$ when $n \to \infty$, then we can also estimate the distribution function $P(g(X) \le x)$ of $g(X)$ with any parametric or non-parametric method, based on the outcomes $(g_1, \dots, g_n)$.

With regard to the choice of the kernel function and the bandwidth, we recall that many several sophisticated techniques have been proposed in the literature. For instance, under assumptions of normality we can simply use the normal kernel, and, as for the bandwidth parameter, we can use the Sturge's or the Scott's rule (see [14] and [15]).

In order to clarify the interpretation of Theorem 1 as well as the application of the Kernel method, we make use of some figures. Let us consider a bivariate random sample from a joint Student's t random vector (X, Y), where X and Y are strongly correlated. As Theorem 1 holds, we know that $E(Y|X) = \mu_Y + \frac{\sigma_{XY}}{\sigma_X^2}(X - \mu_X)$. This is apparent from Fig. 1, which shows the observations and the regression function, estimated via the kernel method. However, the best approximation of the relationship between X and $|Y|$ cannot obviously be linear, and the conditional expectation estimator is able to identify it, as it is shown in Fig. 2.

Therefore, since the proposed methodology to approximate the relationships among random variables is non parametric, we do not need to assume any formal or distributional relationships among the random variables. We can simply computing it and this represents one of the principal advantages of using conditional expectation estimators in statistical approximations. Moreover, in the next section we show that the application of a positive and non-constant function is indeed a key point for identifying a dependence structure between random variables.

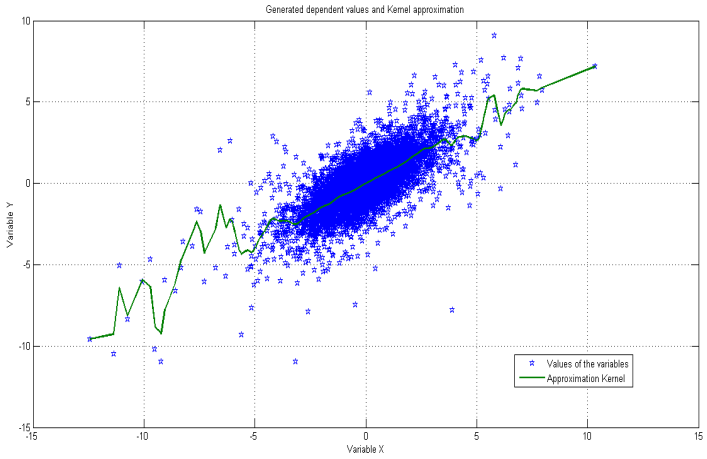

Fig. 1. Random generated values of X and Y and kernel approximation.

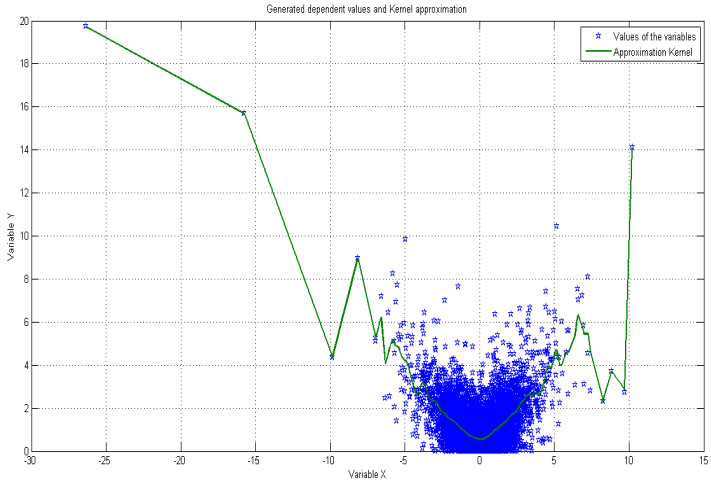

Fig. 2. Random generated values of X and $|Y|$ and kernel approximation

4 New tests of independence

In this section, we describe a new method for testing independence of random variables. The method is based on the conditional expectation and it could be a useful alternative to the well-known Pearson Chi squared test and other tests which have been proposed for continuous random variables. We recall that many well-known tests, such as the Chi squared test and the Hoeffding's test, are based on the main definition of stochastic independence, that is as follows.

Definition. *Two random variables X and Y are independent if, for any couple of Borel sets A and B, we obtain*

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B). \quad (11)$$

The Chi squared statistics is actually a probability distance (or divergence, see e.g. [16], [17]) which measures the dissimilarity between the true distribution and a theoretical one (e.g. under the condition of independence defined above). Therefore, if $X = \sum_{i=1}^{n} a_i I_{[X \in A_i]}$ and $Y = \sum_{j=1}^{m} b_j I_{[Y \in B_j]}$ are discrete random variables (where the collections $\{A_i\}_{i=1,\dots,n}$ and $\{B_j\}_{j=1,\dots,m}$ are partitions of the real line), the Chi squared test can be performed properly. As a matter of fact, in order to verify the independence it is sufficient to show that, for any *i*=1,…,*n* and *j*=1,…,*m*, we have

$$p_{ij} = P\big(X \in A_i, Y \in B_j\big) = \\ = P(X \in A_i)P\big(Y \in B_j\big) = p_i p_j. \quad (12)$$

The Chi squared statistic is given by

$$\chi^2 = N \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{(p_{ij} - p_i p_j)^2}{p_i p_j} \quad (13)$$

where *N* is the sample size. As is well known, a small value of χ^2 corresponds to a situation of "closeness" to independence. Observe that χ^2 can also be used to test independence of continuous random variables although, in this case, it can only guarantee that the random variables are not independent. In other words, $\chi^2 \neq 0$ is a sufficient condition for non-independence while $\chi^2 = 0$ is just a necessary condition for independence.

Nevertheless, the following theorem identifies a necessary and sufficient condition for independence and thereby it establishes an alternative definition of stochastic independence. It is well known that when two integrable random variables *X* and *Y* are independent then $E(Y|X)=E(Y)$ but generally the converse is not true. However, if we apply a positive and non-constant function to the r.v. *Y* we obtain the following result.

Theorem 2. *Let X, Y be real-valued and continuous random variables defined on the probability space $(\Omega, \Im, P)$. Let h be a non-constant, continuous positive function defined on the support of Y, such that $E[h(Y)] < \infty$. Then X, Y are stochastically independent if and only if, $E[h(Y)|X] = E[h(Y)]$.*

Proof.
If X, Y are independent the results is well known.

Assume that $E[h(Y)|X] = E[h(Y)]$, let us consider the set

$$D = \{(x, y): f_{X,Y}(x, y) > f_X(x) f_Y(y)\}$$

and the projection $i_2(x, y) = x$. Then, given $B = i_2(D)$, we obtain $P(X \in B) \geq P((X, Y) \in D)$. Moreover:

$$\int_{X \in B} E[h(Y)|X] dP - P(X \in B) E[h(Y)] =$$

$$= \int_B f_X(x) dx \int_{-\infty}^{\infty} h(y) \frac{f_{X,Y}(x, y)}{f_X(x)} dy \\ - \int_B f_X(x) dx \int_{-\infty}^{\infty} h(y) f_Y(y) dy \\ = 0$$

therefore

$$\int_B \int_{-\infty}^{\infty} h(y)(f_{X,Y}(x, y) - f_X(x) f_Y(y)) dx\, dy = 0.$$

As, by definition, the integrand is a positive function, the equation holds only if $\lambda(B \times \mathbb{R}) = 0$ (where λ is the Lebesgue measure) which implies that $P(X \in B) = P\big((X, Y) \in D\big) = 0$ as $P \ll \lambda$.

Similarly we prove that, if

$$D' = \big\{(x, y): f_{X,Y}(x, y) < f_X(x) f_Y(y)\big\},$$

then $P\big((X, Y) \in D'\big) = 0$. We conclude that $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ q.e.d..

Now, let us consider the following hypotheses

H_0: *X* and *Z* are independent;

H_1: *X* and *Z* are not independent.

Given a bivariate random sample of n observations from (X, Z), let $Y = h(Z)$, where h satisfies the assumptions of Theorem 2. In particular, for simplicity, we propose to use the absolute value,

that is, $h(Z) = |Z|$. Then we can approximate the conditional expectation of Y given X, i.e. $E(h(Z)|X)$, by the use of the approximation methods described in section 3. In other words, we approximate $g(X) = E(h(Z)|X)$ with the estimator $\hat{g}_n(X)$, where $\hat{g}_n(X)$ may be obtained alternatively with the Kernel or the OLP methods. Theorem 2 states that X and Z are independent if $E(h(Z)|X) = E(h(Z))$ or, equivalently, when the r.v. $E(h(Z)|X)$ is actually a contant. Thus, as $\hat{g}_n(X)$ converges to $E(h(Z)|X)$, we choose H_0 when $\hat{g}_n(X)$ presents a null (or small) variance, otherwise we choose H_1. Therefore, this procedure makes it possible to determine whether two random variables are independent, based on the approximation of $E(h(Z)|X)$. Indeed, the variance of $E(h(Z)|X)$ is null only in case of independence and not, for instance, in case of uncorrelation. Fig. 3 confirms that, for two independent Student's t distributions, $\hat{g}_n(X)$ tends to a constant, while the same result does not hold when the two random variables are only uncorrelated.

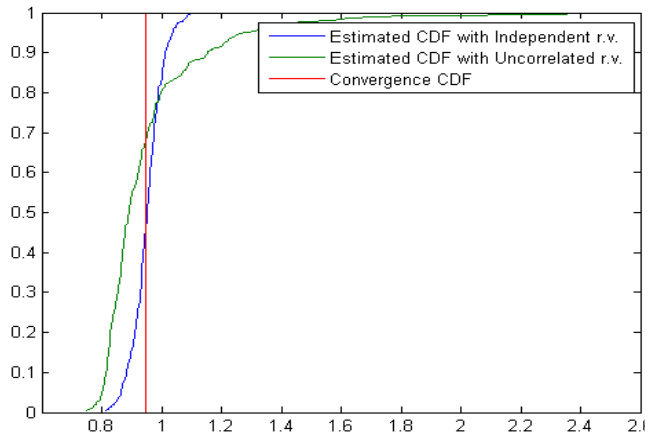

Fig. 3. Distributions of $E(|Y||X)$ for i) independent; and ii) uncorrelated Student's t random variables.

As specified in the following proposition, under some particular assumptions we can apply the Z-test. Consider the statistic

$$Z_n = \sqrt{n}\frac{\hat{\sigma}^2_{\hat{g}_n(X)}}{\sqrt{\hat{k}_{\hat{g}_n(X)}}} \qquad (14)$$

$\hat{\sigma}^2_{\hat{g}_n(X)}$ is the variance and $\hat{k}_{\hat{g}_n(X)}$ is the central moment of order four of the outcomes $(\hat{g}_n(x_1),..,\hat{g}_n(x_n))$ (obtained with the kernel or OLP methods). Thus, we propose the following decision rule:

$$\text{Reject } H_0 \text{ if } Z_n = \sqrt{n}\frac{\hat{\sigma}^2_{\hat{g}_n(X)}}{\sqrt{\hat{k}_{\hat{g}_n(X)}}} \geq c_\alpha \qquad (15)$$

where $c_\alpha = F^{-1}_{N(0,1)}(1-\alpha)$. Indeed, the following proposition holds.

Proposition 1. *Assume that* $Y_n = (\hat{g}_n(X))\text{^}2$ *is a stationary process that admits finite the third moment (for any n). Suppose the process* $\{Y_n | n \in \mathbb{N}\}$ *is an* α *-mixing process with mixing coefficients* α_n *such that* $\alpha_n = O(n^{-a})$ *for some* $a > 3$ *for large n. Then the following limits are verified for the test (15)*

$$\lim_{n\to\infty} P(reject\ H_0\ |H_0\ is\ true) \leq \alpha \text{ and}$$

$$\lim_{n\to\infty} P(reject\ H_0\ |H_0\ is\ false) = 1$$

Proof

We know that, under these conditions, the statistic Z_n converges to a Gaussian distribution (see [18]). q.e.d..

Since test (15) can be applied for any non constant continuous positive function h, then we can consider a parametric function $h(x) = h(x, \vartheta)$ with $\vartheta \in \Theta$ and enhance the performance of the with this alternative decision rule:

$$\text{Reject } H_0 \text{ if } min_{\vartheta\in\Theta}\sqrt{n}\frac{\hat{\sigma}^2_{\hat{g}_n(X)}}{\sqrt{\hat{k}_{\hat{g}_n(X)}}} \geq c_\alpha. \qquad (16)$$

For instance, we can approximate the proposed statistic for the conditional expectation $E(h(Z)|X)$ with $h(Z) = |Z|^\vartheta$ and then apply test (15) in order to verify independence.

In what follows we provide a simple example of application of the proposed decision rule.

Example

As one of the several possible applications of these tests, we may propose a particular comparison of financial variables (portfolios). Indeed, in a financial context the r.v.'s are generally assumed to be continuous.

Let us consider two portfolios of daily returns X and Y, taken from the NYSE (New York Stock Exchange), which are empirically uncorrelated. Suppose that we have available about three years of historical daily joint observations (750 trading

days). We need to test whether the losses and gains of the two portfolios are independent (see [19]).
Using the chi-squared test with one degree of freedom we fail to reject the null hypothesis (i.e. independence) at a significance level of 95%.
As an alternative approach, we apply the test given by (15) with $h(x) = |x|$ to the standardized random variables $\widetilde{X}$ and $\widetilde{Y}$. We obtain that the variance of $E\left(\left|\tilde{Y}\right||\tilde{X}\right)$ is equal to 0.0512, with the OLP method, and 0.0445, with the Kernel method. With regard to returns (i.e. percentages) such variances are apparently large. This is confirmed by the Z_n statistic as well as the Z test given by (15), that lead to reject the null hypothesis, differently from the chi-squared test. Hence, in this case we have verified that the newly introduced test is more powerful than the chi-squared test.

5 Conclusions

The new methodology proposed in this paper makes it possible to evaluate the independence structure of random variables based on an alternative definition of independence, where the approximation of the conditional expectation plays a key role. In this framework, the kernel method as well as the OLP method may provide sufficiently accurate approximations. Under the assumptions of Proposition 1, we can use a Gaussian approximation and apply a Z test in order to establish whether two r.v.'s are independent or dependent. This method for testing independence represents an alternative to other well known tests, such as the Chi squared test, and is especially suitable for dealing with continuous random variables. Moreover, it is worth noting that the method determines a general class of independence tests, in that, depending on the approximation method (OLP, kernel) and on the choice of the function h, we may obtain different decision rules with different characteristics. It should be stressed that these tests can be easily performed with the use of a computational software and are especially suitable for large samples, because of the convergence property of the test statistic (see Proposition 1).
As it is well known, tests of independence may have several applications in diverse areas of study, from medicine to engineering. However, we believe that the proposed class of tests could be especially suitable for dealing with financial data, that are often assumed to be continuous. The usefulness of this approach has been stressed at the end of section 4. We argue that the newly introduced test could generally yield an inferior probability of a type II error, in particular when random variables are continuous.
In future work, it would be interesting to investigate the properties of the different tests according to the form of the function h: this can be done with theoretical studies or with simulation studies as well. Furthermore, based on this method it would be interesting to provide an alternative approach to independent component analysis, that is particularly useful in several fields of research.

Acknowledgements
This paper has been supported by the Italian funds ex MURST 60% 2014, 2015 and MIUR PRIN MISURA Project, 2013–2015, and ITALY project (Italian Talented Young researchers). The research was supported through the Czech Science Foundation (GACR) under project 15-23699S and through SP2015/15, an SGS research project of VSB-TU Ostrava, and furthermore by the European Social Fund in the framework of CZ.1.07/2.3.00/20.0296.

References:

[1] Pearson K, X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Philosophical Magazine Series*, Vol. 5, No. 50, 1900, pp. 157–175.
[2] Hoeffding W, A non-parametric test of independence. *The Annals of Mathematical Statistics*, Vol. 19, 1948, pp. 546–557.
[3] Christoffersen P, Evaluating interval forecasts, *International Economic Review*, Vol. 39, 1998, pp. 841-862.
[4] Kupiec P, Techniques for verifying the accuracy of risk measurement models, *Journal of Derivatives*, Vol. 3, 1995, pp. 73-84.
[5] Roth M, *On the multivariate t distribution*, Technical report from Automatic Control at Linköpings universitet, 2013.
[6] Lando T, Ortobelli S, On the Approximation of a Conditional Expectation, *WSEAS Transactions on Mathematics*, Vol. 14, 2015, pp. 237-247.
[7] Ortobelli S, Petronio F, Lando T, A portfolio return definition coherent with the investors preferences, to appear in *IMA-Journal of Management Mathematics*.

[8] Nadaraya E A, On estimating regression, *Theory of Probability and its Applications*, Vol. 9, No. 1, 1964, pp. 141-142.

[9] Watson S G, Smooth regression analysis, *Sankhya, Series A*, Vol. 26, No. 4, 1964, pp. 359-372.

[10] Samorodnitsky G and Taqqu M S, *Stable Non Gaussian Random Processes*, Chapman & Hall, 1994.

[11] Shoutens W, *Lèvy processes in Finance, Pricing Financial Derivatives*, Wiley series in Probability and Statistics, 2003.

[12] Ortobelli S, Lando T, On the use of conditional expectation estimators, *New Developments in Pure and Applied Mathematics*, ISBN: 978-1-61804-287-3, 2015, pp. 244-246.

[13] Epanechnikov V A, Non-parametric estimation of a multivariate probability density, *Theory of Probability and its Applications*, Vol. 14, No. 1, 1965, pp. 153-158.

[14] Scott D V, On Optimal and Data-Based Histograms, *Biometrika*,Vol. 66, No. 3, 1979, pp. 605-610.

[15] Sturges H A, The Choice of a Class Interval, *Journal of the American Statistical Association*, Vol. 21, No. 153, 1926, pp. 65-66.

[16] Ali S M, Silvey S D, A general class of coefficients of divergence of one distribution from another, *Journal of the Royal Statistical Society*, Vol. 28, No. 1, 1966, pp. 131-142.

[17] Lando T, Bertoli-Barsotti L, Statistical Functionals Consistent with a Weak Relative Majorization Ordering: Applications to the Mimimum Divergence Estimation, *WSEAS Transactions on Mathematics*, Vol 13, 2014, pp. 666-675.

[18] Doukhan P, *Mixing: properties and examples*, Lecture notes in Statistics, 85. Berlin, Springer-Verlag, 1994.

[19] Ortobelli, S., Lando, T., Independence tests for financial variables, to appear in *Mathematical Methods and Computational Techniques II*.