FINANCIAL OPTIONS RESEARCH CENTRE

University of Warwick

Multivariate Distributional Tests based on the Empirical Characteristic Function Approach: A Comparison

Mascia Bedendo
and
Stewart Hodges

December 2000

Financial Options Research Centre
Warwick Business School
University of Warwick
Coventry
CV4 7AL
Phone: (0)24 76 524118

FORC Preprint: 2000/112

Multivariate Distributional Tests based on the Empirical Characteristic Function Approach: A Comparison

Mascia Bedendo

Financial Options Research Centre
University of Warwick
CV4 7AL Coventry UK
e-mail: forcmb@razor.wbs.warwick.ac.uk

Stewart D. Hodges

Financial Options Research Centre
University of Warwick
CV4 7AL Coventry UK
e-mail: forcsh@razor.wbs.warwick.ac.uk

Abstract

With the purpose of identifying appropriate testing procedures for multivariate distributional forecasts, in this paper we compare the power of two versions of multivariate goodness-of-fit tests based on the Empirical Characteristic Function (ECF) in detecting deviations of the true distribution of the data from the forecast. Various Monte Carlo experiments carried out for dimensions up to 16 suggest the superiority of the continuous version of the test over the discrete one, in terms of both computational feasibility and statistical properties. The applicability of this testing procedure to the evaluation of density forecasts of financial asset returns generated in the context of risk management and Value at Risk models is carefully investigated.

Keywords: multivariate density forecasting, distributional tests, empirical characteristic function, Value at Risk, mixture of normal.

Funding from the ESRC (award number R022250134) and FORC corporate members is gratefully acknowledged.

1. Introduction

A distributional forecast is an estimate of the density function that, according to the forecaster, will characterise the future evolution of a random variable. Therefore, unlike point or interval forecasts, which restrict the attention to a single point or percentile of the forecasted distribution, a density forecast provides an accurate and complete representation of the uncertainty associated with the changes in the values of the variable of interest. As Tay and Wallis (2000) state in their survey, density forecasting constitutes an important issue in finance: given the non-normality of financial asset returns, the estimation of only the first two moments is insufficient and a complete characterisation of the distribution of the returns is needed.

Density forecasting finds a natural application in the field of risk management, since the estimation of the joint model of asset returns represents the first step in the computation of the Value at Risk,[1] the second step being the translation of that model into a distribution for the bank portfolio. In this context an inappropriate forecast will adversely affect the usefulness of VaR estimates for quantifying the market risk exposure of a financial institution and, consequently, the adequacy of the amount of capital addressed to meet the capital requirements for market risk. Therefore, besides the creation of the distributional forecast itself, the *ex post* assessment of the "goodness" of the predicted distribution is of crucial importance in risk management for providing directions in which the forecasting process could be improved.

However, little attention has been dedicated to the study of this second aspect, in particular for high-dimensional problems such as those that frequently occur in finance. The analysis generally becomes very complicated in a multidimensional context. The main issue here is the "curse of dimensionality": since a high-dimensional space is mostly empty, it is particularly difficult to implement accurate testing procedures that are able to pick up small deviations from the null hypothesis, unless the sample size is gigantic. Moreover, the concept of ordering used in many univariate goodness-of-fit tests is not obvious in a multidimensional space. For all these reasons, many testing procedures commonly adopted in the univariate case (as the ones based on the empirical distribution function) do not easily generalise to the multivariate case, unless the dimensionality is very small. This is clearly not the case for VaR models,

[1] Value at Risk (VaR) represents a specified percentile of the distribution of the change in the value of a portfolio over a certain time horizon, due to price variations in the underlying risk factors.

where the number of the relevant risk factors to be modelled could be around 200-300, and the sample size is not large enough to avoid the "curse of dimensionality" problem.
Methods other than the ones adopted in the unidimensional framework need therefore to be identified for an appropriate appraisal of multivariate density forecasts. The present study constitutes an attempt in this direction, by investigating whether multivariate goodness-of-fit tests based on the empirical characteristic function (ECF) might be good candidates for the assessment of multidimensional distributional forecasts.

The paper is structured as follows. Section 2 briefly reviews the existing relevant literature. Section 3 outlines the features and properties of the two testing procedures based on the ECF approach contrasted in the present work. In section 4 we provide and discuss the outcomes from some Monte Carlo experiments carried out to evaluate and compare the relative performance of the two tests. Section 5 concludes the paper.

2. The Literature on Tests of Distributional Forecasts

2.1. Assessment of univariate distributional forecasts.

The existing literature on the evaluation of distributional forecasts is rather recent and has focused mainly on univariate distributions. In most cases the probability integral transform approach, originally suggested by Rosenblatt (1952), or modifications of it (Berkowitz, 1999), are used (Dawid, 1984; Diebold, Gunther and Tay, 1998; Diebold, Hahn and Tay, 2000). The uniformity of the transformed variable is then tested either with graphical approaches (Diebold, Gunther and Tay, 1998) or with formal goodness-of-fit tests based on the empirical distribution function (Noceti, Hodges and Smith, 2000).

2.2. Assessment of multivariate distributional forecasts.

Much less research has been devoted to the assessment of multivariate density forecasts. As stated earlier, the univariate distributional testing procedures usually cannot be extended to the multivariate case.

Some attempts to develop a multivariate version of Kolmogoroff-Smirnov and Cramér-von Mises statistics have been made (Kotz and Johnson, 1985; Justel, Peña and Zamar, 1997). A multidimensional application of the probability integral transform approach has been suggested by Diebold, Gunther and Tay (1998), and Clements and Smith (1999). Unfortunately, for more than two dimensions, the computational burden becomes unsustainable and the properties of the test statistics themselves are unknown. These problems rule out the applicability of those tests in risk management, where a much larger number of risk factors is involved.

The difficulty in practical implementation in truly high-dimensional problems also constitutes the main drawback of a recent, promising approach aimed at modelling and evaluating the dependence structure between variables, based on the use of copula functions.[2] The attractive feature of this approach, that has contributed to its increasing popularity amongst the researchers, is the possibility to model more general concepts of dependence than the basic linear correlation, which is an adequate measure of dependence only for distributions belonging to the elliptical family. However, since the formulations commonly proposed for the copula functions have very few parameters, they do not seem suitable for describing satisfactorily the complicated dependence structure amongst hundreds of variables.

Projection-pursuit type testing techniques (Huber, 1985) have also been suggested as multivariate goodness-of-fit tests. The underlying idea is to re-map multivariate data set into a unidimensional framework along interesting low-dimensional linear projections (see Zhu, Fang, and Bhatti, 1997, for an application). The disadvantage of this method is the large number of projections that have to be generated to guarantee good statistical power.

The Empirical Characteristic Function Approach

An alternative, promising approach for assessing multivariate distributional forecasts involves the implementation of multivariate goodness-of-fit tests based on the empirical characteristic function (ECF). The area of statistical inference where the ECF has been more extensively used is parameter estimation. Examples of employment of this technique in hypotheses testing can be found, among others, in Epps and Pulley (1983), Feigin and Heathcote (1976), Feuerverger and Mureika (1977), Feuerverger (1993), Hall and Welsh (1983), Heathcote (1972), Baringhaus and Henze (1988), Fan (1997). However, only in the last two works the approach is extended to a multidimensional framework.

The use of the ECF in statistical inference is justified by the asymptotic convergence of the ECF to the theoretical one, and the one-to-one correspondence between distribution and characteristic function of a variable (Heathcote, 1972). Employing the characteristic function in testing procedures has important advantages over the use of the distribution function. Since the former is the Fourier transformation of the latter, it allows the retention of all the

[2] For a good survey on copulas, see Nelsen (1998).

information contained in the sample, while presenting a more convenient formulation for computation (Feuerverger and Mureika, 1977).[3] The asymptotic properties of the ECF have been largely investigated by Feuerverger and Mureika (1977), Feuerverger and McDunnough (1981a, b).

Goodness-of-fit tests based on the ECF are usually computed from the integrated squared difference between the ECF $C_n(t)$ of the data and the characteristic function $C_0(t)$ of the multivariate null distribution, evaluated:

- at a single fixed value of t (Heathcote, 1972). In this case the advantage of a simplified analysis is generally offset by the lack of statistical consistency of the testing procedure;
- at a set of values $t_1, t_2, \ldots, t_m$. This approach has been adopted by Fan (1997) to develop a general multivariate goodness-of-fit test;
- over a continuous range of t described by a weight function. Hall and Welsh (1983) and Epps and Pulley (1983) suggested this procedure to test for univariate normality. The multivariate case has been investigated by Baringhaus and Henze (1988), Henze and Zirkler (1990), Henze and Wagner (1997).

In the present work, testing procedures belonging to the two latter categories have been implemented, and their relative performances in detecting deviations of various nature in the actual distribution of the data from the forecasted distribution, have been compared in different dimensions, up to 16. Most of the literature on the ECF approach to multivariate hypotheses testing focuses on the theoretical properties of the tests, and very few papers have investigated the practical application of those procedures through sampling experiments, especially for a truly "high" number of dimensions. Therefore we think that our results may contribute to identify which version (if any) of this testing approach seems more appropriate for an accurate, yet computationally feasible, assessment of multivariate density forecasts developed in the context of VaR models.

3. The Empirical Characteristic Function Goodness-of-Fit Techniques

In this section we describe the features of the two multivariate goodness-of-fit testing procedures based on the ECF that have been contrasted in the present study.[4]

---

[3] The characteristic function behaves friendly under shifts, scale changes, and summation of independent variables.

[4] To avoid excessive length of the main body of the text, only the main formulas have been reported here. The remaining details have been included in the Appendix.

3.1. The "Fixed Grid Points" (FGP) approach.

One of the tests we considered is a simplified version of the goodness-of-fit procedure suggested by Fan (1997). The test is expressed as a quadratic measure between real and imaginary components of the ECF and of the characteristic function under the null distribution, evaluated at m fixed grid points $t_1, t_2, \ldots t_m$ (where m is a positive integer and $\rightarrow \infty$ as $n \rightarrow \infty$ to ensure consistency of the test).

Let $X_1, X_2, \ldots. X_n$ represent a sample of independent observations from a random d-dimensional vector. The empirical characteristic function (ECF) $C_n(t)$ is defined as follows:

$$C_n(t) = \frac{1}{n}\sum_{j=1}^{n} \exp(it'X_j) = \left[\frac{1}{n}\sum_{j=1}^{n} \cos(t'X_j)\right] + i\left[\frac{1}{n}\sum_{j=1}^{n} sin(t'X_j)\right] \tag{1}$$

and the characteristic function (CF) $C_0(t)$ as:

$$C_0(t) = E[\exp(it'X)] \tag{2}$$

The test statistic introduced by Fan (1997) has the following formula:

$$T_n = [Z_n(\bar{t}_m) - Z_0(\bar{t}_m)]'\Omega_0^{-1/2}(\bar{t}_m)W(\bar{t}_m)\Omega_0^{-1/2}(\bar{t}_m)[Z_n(\bar{t}_m) - Z_0(\bar{t}_m)] \tag{3}$$

where $Z_n(t_m)$ is the vector of real and imaginary components of the ECF, $Z_0(t_m)$ is the corresponding vector for the hypothetical characteristic function,

$$Z_n(\bar{t}_m) = \begin{bmatrix} \operatorname{Re} C_n(t_1) \\ \operatorname{Re} C_n(t_2) \\ \\ \operatorname{Re} C_n(t_m) \\ \operatorname{Im} C_n(t_1) \\ \operatorname{Im} C_n(t_2) \\ \\ \operatorname{Im} C_n(t_m) \end{bmatrix}, Z_0(\bar{t}_m) = \begin{bmatrix} \operatorname{Re} C_0(t_1) \\ \operatorname{Re} C_0(t_2) \\ \\ \operatorname{Re} C_0(t_m) \\ \operatorname{Im} C_0(t_1) \\ \operatorname{Im} C_0(t_2) \\ \\ \operatorname{Im} C_0(t_m) \end{bmatrix} \tag{4}$$

$$Z_n(\bar{t}_m) - Z\ (\bar{t}_m) = \frac{1}{n}\sum_{j=1}^{n} \begin{bmatrix} \cos(t_1'X_j) - E[\cos(t_1'X_j)] \\ . \\ \cos(t_m'X_j) - E[\cos(t_m'X_j)] \\ \sin(t_1'X_j) - E[\sin(t_1'X_j)] \\ . \\ \sin(t_m'X_j) - E[\sin(t_m'X_j)] \end{bmatrix} \equiv \frac{1}{n}\sum_{j=1}^{n} K_j(\bar{t}_m) \tag{5}$$

$\Omega_0(\bar{t}_m) = Var[K_j(\bar{t}_m)]$, represents the ($2mx2m$) covariance matrix of $K_j(t_m)$, whose components are computed as specified in Fan's paper. W is a diagonal weight matrix aimed at directing the power of the test statistic towards different frequencies. In our analysis, we assumed $W=I_{2m}$, which assured both computational convenience and more powerful tests against high-frequency alternatives. The formula of our test then simplifies to:

$$T_n = [Z_n(\bar{t}_m) - Z_0(\bar{t}_m)]' \Omega_0^{-1}(\bar{t}_m)[Z_n(\bar{t}_m) - Z_0(\bar{t}_m)] \qquad (6)$$

In principle, this approach can be widely used for testing the goodness of any type of distributional forecast in any dimension, under both simple and composite null hypotheses, provided that consistent estimators of the unknown parameters are given. As it will be shown later on in the analysis, the statistical properties of this testing procedure are strongly affected by the choice of the grid points. Therefore, the main issues in this context regard the criteria that drive the choice of m and, consequently, of the different t's, and their location.[5]

In the general case, the asymptotic distribution of the test statistic under the null is a weighted sum of independent χ^2 with one degree of freedom. Given our specific choice for the matrix W, the distribution of the test simplifies to χ^2 with $2m$ degrees of freedom.

3.2. The "Weight Function" (WF) approach.

The second testing procedure we considered is based on the integrated squared difference between the empirical and the theoretical characteristic function, weighted by the function $G(t)$, which is itself a distribution function, whose corresponding density $g(t)$ is symmetric around the origin:

$$D_{n,\beta} = \int |C_n(t) - C_0(t)|^2 \, dG(t) \qquad (7)$$

$dG(t)=g(t)dt$

This approach was originally suggested by Baringhaus and Henze (1988) for testing multivariate normality. This test statistic presents some appealing properties of consistency and good power against any fixed alternative to the null distribution, but cannot be easily extended to test any kind of multivariate null distribution, as that would require the evaluation of a d-dimensional integral.

The main issue here regards the choice of the weight function. According to Epps and Pulley (1983), the following requirements should be met:

1. high weight must be assigned where the distance between the CFs of the alternative and of the null hypothesis is larger. For many continuous distributions, this applies for values of t between 0 and 3;
2. more weight should be put in some interval around the origin, where most information conveyed by the CF is contained, and the precision of the sample characteristic function is greater;
3. the weight function must present a convenient formulation, such that the integral has a closed form, to ensure the computational feasibility of the test statistic. This point represents the main limitation to extend the application of this approach to a wide range of distributions.

In our analysis, we follow the common practice of selecting as weight function the density of a multivariate normal distribution $N_d(0, \beta^2 I_d)$:

$$g(t) = (2\pi\beta^2)^{-\frac{d}{2}} \exp\left(-\frac{\|t\|^2}{2\beta^2}\right), t \in R^d \qquad (8)$$

that, for an appropriate choice of β, satisfies the above-mentioned requirements.[6]
The test statistic assumes the following representation:

$$T_{n,\beta}(X_1,X_2,\ldots X_n) = n\ (4I\{S_n \text{ is singular}\} + D_{n,b}I\{S_n \text{ is nonsingular}\}) \qquad (9)$$

For the purposes of our study, an exact expression for $D_{n,b}$ has been derived under the null of 1) multivariate normality on the original variables; 2) multivariate normality on standardised variables; 3) mixture of multivariate normal distributions, on standardised variables.[7]

Since this testing procedure can be interpreted as the continuous version of the previous one, where the computation was made over a fixed number of points, we would expect its

[5] Some directions on this matter are provided only for applications of the ECF technique to parameter estimation. Feuerverger and McDunnough (1981b) suggest that the t's should be equally spaced and sufficiently fine and extended.
[6] This choice is particularly attractive, for both its statistical properties and its high tractability, but suitable for few specification forms for the null distribution, close to the multivariate normal. In the present work we restrict the analysis to distributional forecasts of this type (multivariate normal and mixures of multivariate normal distributions).
[7] For the detailed formulas, as well as the empirical critical values computed at 95% confidence interval, see Appendix.

employment to present some advantages. First of all, by using a continuous weight function, all the moments of the ECF and the theoretical CF are matched continuously and, hence, more information from the sample is exploited. Secondly, in this case there is no need to worry about the selection of the different t's, as they are simply integrated out, provided that the weight function is chosen appropriately.

4. Some Monte Carlo Experiments

4.1. General framework.

The main purpose of the present analysis has been to assess the power of the ECF testing procedures outlined in the previous section to detect deviations of the actual distribution of the data from the forecasted one (while maintaining a good level of tractability in a high-dimensional context) and, consequently, to identify which statistic of the two represents the most appropriate candidate as a test for distributional forecasts. Since, for finite samples, theoretical results regarding the relative power of these two classes of tests are not available, we provided and discussed the evidence obtained from some Monte Carlo simulations. We investigated how the two testing procedures compare, in terms of power, to each other (for the purpose of ranking between them) and, when appropriate, to alternative techniques whose good statistical properties are widely recognised (for the purpose of a more general appraisal of the performance of the ECF tests). In the attempt of identifying a nested sequence of tests for the density forecasts, that would enable us to diagnose the nature of the misspecification, controlled experiments have been designed to test separately for biases of various type in the forecast. Samples have been generated according to a distribution different, in some aspects, from the forecasted one. Various choices for the possible specification for the densities under the null and the alternative hypotheses have been contemplated, according to the specific nature of the bias in the forecast we were interested in detecting. More specifically, we developed tests for bias: 1) in the mean; 2) in the covariance structure; 3) in higher moments when the forecasted distribution is a multinormal, which then provides a test for misspecification of the form of the distribution; 4) in higher moments when the forecasted distribution is fat-tailed.

The power of the two testing procedures has been expressed as percentage of rejections out of 1,000 replications. In order to investigate how the performance of the tests changes with the dimensionality, we considered problems in dimensions $d = 2$, 4 and 16. The sample size was n=100 for $d = 2$, 4, and n=200 for $d = 16$. The tests of type 1) and 2) have been computed on the original variables, whereas scaled residuals have been used for tests 3) and 4), to guarantee invariance of the testing procedures to transformations of location and scale.

A further aim of our experiments was to gain useful insights on the extent to which the value of the parameters of the two tests (β and grid points) affects their power and, consequently, on the most appropriate choice that would ensure the best statistical properties in both cases. For this purpose, each analysis has been replicated for various choices of the relevant parameters. The values selected for β in the WF approach were $\beta = 0.5$, 1, 2, in line with those suggested by Henze and Zirkler (1990).[8] The simulations for the FGP test were carried on by using 5, 10, and 20 grid points for problems in 2 and 4 variables and 20, 30 and 50 points in a 16-dimensional context. The choice for the grid points was made by equally spacing an interval ranging approximately from 0 to 5,[9] according to the values chosen for m (more details are provided in the headings of the tables included in the Appendix).

4.2. Size of the tests.

The size of the two tests under comparison was evaluated by computing the frequency of rejection (at the 95% confidence interval) out of 10,000 replications when the forecasted distribution (assumed to be, for simplicity, $N(\mathbf{0}, \mathbf{I}_d)$) coincides with the true distribution of the data. The results, reported in Table 1, seem to indicate that for both approaches the size decreases as the dimensionality of the problem increases. In particular, for d=16 the WF tests displays percentages of rejection above 5% whatever choice is made for the parameter β. This suggests a tendency to over-reject the null in high dimensions that should be taken into account when assessing the power of this procedure.[10] The alternative approach generally exhibits better size for most of the possible choices in terms of number of grid points.

4.3. Bias in the first moment.

We contrasted the power to detect a bias in the mean of the distribution displayed by the two testing procedures under investigation, as well as by the Hotelling-T^2 statistic, which

[8] Since the tail behaviour of a probability distribution is reflected by the behaviour of the CF at the origin, a small value for β that concentrates the mass of the weight function near the origin, should improve the sensitivity of the test against alternative distributions with fatter tails.
[9] Other choices for the number of grid points and the criteria for selecting the points themselves (for instance by sampling from a standard multivariate normal with low-discrepancy Faure sequences) were attempted, but overall they proved to be less powerful than the ones adopted here.
[10] Similar figures for the size of this test were reported by Henze and Zirkler (1990).

represents a good benchmark for testing this type of misspecification under the null hypothesis of multivariate normality. We assume that the distributional forecast coincides with a standard multivariate normal $N(\mathbf{0},\mathbf{I}_d)$, when the data follow a multivariate normal $N(\mu,\mathbf{I}_d)$. The mean μ of the true distribution is set equal to the mean of the process under the null (a vector of zeros in our case) increased by a vector of equal components, which represent the contours of the multivariate normal for a given choice for the constant $c>0$:[11]

$$(x-\mu)'\Sigma^{-1}(x-\mu)=c^2$$

The choices for c we considered to explore a wide range of alternatives are: $c = 2, 4, 6, 8$, and the value that defines the contours relative to the 95% of probability density.

The results, in terms of percentage of rejections of the null hypothesis, are shown in Table 2. As expected, all the tests exhibit a loss in power as the dimensionality of the problem increases. Hotelling-T^2 statistic turns out to be the most powerful test: a satisfactory percentage of rejections of the null is reached for biases in the mean when $c\geq4$ (for d=2, 4) or $c\geq5$ (for d=16). The performance of the two tests based on the ECF is quite similar, with the approach that uses the weight function performing slightly better in low dimensions and slightly worse in higher dimensions. For this testing procedure, the highest power is obtained for a small β (β=0.5), and deviations from the forecasted mean relative to a choice of $c\geq4$ (for d=2) or $c\geq6$ (for d=4, 16). The loss in power consequent to a wrong choice for β is important, especially when the problem involves many variables: for d=16, the loss from choosing β=2 instead of β=0.5 is around 70-80 percentage points for the two largest biases. For the test based on fixed grid points, the number of points associated with better statistical properties seems to be around m=5 when the dimensionality of the problem is low and m=20 for a 16-dimensional problem. As for the other approach, an inappropriate selection for the parameters of the test (which, in this case, translates into selecting an excessive, or too small, number of grid points) would substantially reduce the power of the test statistic, in particular as the dimensionality increases.

4.4. Bias in the second moment.

We investigated the power of the tests based on the ECF to detect a bias in the covariance matrix when the forecasted distribution is $N(\mathbf{0}, \mathbf{I}_d)$, while the true DGP is $N(\mathbf{0},\Sigma)$. The following measure of the "distance" between two matrices Σ and $\mathbf{I}_d$:

[11] A contours consists of the set of points of equal probability density of a multivariate normal.

$$\delta = trace[(\Sigma^{-1} - I_d)\,(\Sigma^{-1} - I_d)]$$

has been employed to "quantify" how different the true and the hypothesised covariance structures are. The second moments of the alternative, real DGP have been selected such that a similar range of distances ($\delta \cong 0.03, 0.5, 1.5$) is replicated for all the dimensions considered in the analysis.

As a benchmark, we proposed a procedure suggested by Nagao (1973) to test for the equality of a covariance matrix to a given matrix, based on the eigenvalues of a combination of the two matrices.

Table 3 displays the outcomes of the testing experiment. As expected, given the ability of the eigenvalues to capture the covariance structure of a distribution, Nagao's test is the most powerful in detecting a bias in the second moment of the forecasted distribution. At a 95% confidence interval, a good level of power is attained for biases corresponding approximately to $\delta \cong 0.5$.

Between the testing criteria based on the ECF, the FGP approach is preferable for low dimensions, but the order of preference reverts as the dimensionality increases, when the WF approach exhibits a level of power very close to the one displayed by Nagao's test. The difference in power between the two testing procedures also increases along with the dimensionality of the problem, rising from few percentage points for the largest bias in d=2, 4 to 15-20 points in 16 dimensions. In this latter case, a percentage of rejections of the null above 95 cannot be attained for the discrete version of the test even when the difference between the true and the hypothesised covariance matrix is the largest amongst the alternatives investigated, and the choice made for the number of grid points is the one that shows the highest power (m=30).

This seems to suggest that the test based on the WF is more appropriate for problems involving several variables. For this testing procedure, small values for β again seem to guarantee better power, especially in a high-dimensional context. As before, an inappropriate selection for β or for the number of grid points strongly affects the power of the tests, though less drastically than when a bias occurs in the mean.

4.5. Misspecification of the distributional forecast.

An experiment was designed to assess the performance of the two tests in detecting a misspecification in the form of the distributional forecast although the first two moments were correctly specified. The null hypothesis was that the data follow a standard multivariate

normal. The alternative specifications for the true distribution we considered were: 1) two mixtures of multivariate normal distributions with same mean and covariance matrix as the null, but different degrees of kurtosis;[12] 2) a multivariate standard t with 5 degrees of freedom; 3) a multivariate standard t with 10 degrees of freedom.
For this experiment, the power of the ECF tests has been compared to that of the well-known test for multivariate normality proposed by Mardia (1970), based on multivariate measures of skewness and kurtosis.

The outcomes of the analysis have been reported in Table 4. The counterintuitive increase in the power to spot a misspecification in the form of the distribution that all the testing procedures seem to exhibit is due to the particular choice made for the distributions used for the sample generation. The WF approach clearly outperforms the FGP approach for all the alternatives under investigation, and in each dimension. At the 95% confidence level, the WF approach (for β=0.5 or β=1) in a 16-dimensional context, represents a powerful test against all the alternatives we specified. For the case of 2 and 4 variables, this test possesses good power only against few alternative specifications, namely the mixture with the highest level of excess kurtosis, and the multivariate t with 5 degrees of freedom. On the other hand, the FGP version of the test hardly shows a reasonable power against any of the alternatives considered, in any dimension, even for an appropriate choice of the number of grid points (m=10 for d=2, m=20 for d=4, and m=30 for d=16). The gap in the power with respect to the continuous version increases dramatically along with the number of variables of the problem: for d=16 the gap varies from 30 to 80 percentage points, according to the different alternative distributions. Again, the power of the WF test against distributions that are more fat-tailed than the null proves to be higher for small values of β, and the loss in power consequent to a wrong choice for β becomes more relevant in higher dimensions. Our results show a substantial loss from choosing β=2 when d=16, that reaches a maximum of 50 percentage points when the true DGP is the mixture with the highest kurtosis.

Quite surprisingly, the WF procedure seems also to outperform Mardia's test when the alternative hypotheses are represented by the two mixtures of normal distributions. However, when the data are distributed according to a multivariate *t*, the test suggested by Mardia displays the best statistical performance.

[12] As an indicator of excess kurtosis for discriminating between the degree of fat-tailness of the alternative specifications, we used the measure introduced by Mardia (1970).

4.6. Bias in higher moments.

To identify the ECF testing procedure that performs better in detecting biases in higher moments of the forecasted distribution, we also implemented our tests in a context where the null distribution has fat tails and the true process presents higher kurtosis. As null hypothesis we chose the mixture of two multivariate normal distributions with less excess kurtosis from the previous analysis; as alternatives we considered the other, more leptokurtic, mixture, and the multivariate t distributions, with 5 and 10 degrees of freedom. The choice of a mixture of multivariate normal distributions as a possible specification for the joint distributional forecast of financial asset returns guarantees a high tractability, and seems a plausible assumption to make given its ability to model the fat-tailness exhibited by financial returns. Under this particular assumption for the null, the WF procedure consists of an adaptation of the original test suggested by Henze and Zirkler (1990) to the case when the null hypothesis is a standardised mixture of multivariate normal distributions (see Appendix).

The results of the experiment, displayed in Table 5, mainly recall those from the previous analysis. Again the WF approach seems to be the most powerful test, provided that an appropriate, small value is selected for the parameter β. The performance exhibited by the continuous version of the test overcomes the one shown by the discrete version for any choice of the alternative distribution and any dimension. However, the gap in power between the two procedures, especially for high-dimensional problems, is narrower than the one experienced in the previous case. None of the testing procedures seems to display a satisfactory power against two out of the three alternative specifications chosen for the true DGP. A percentage of rejections of the null above 95 can be obtained only from the WF approach, for d=16, and for a highly leptokurtic (multivariate t with 5 degrees of freedom) alternative, whereas much less power has been observed when the bias in the excess kurtosis is less significant. As before, a wrong choice for the parameters has a strong adverse impact on the performance of both the techniques under investigation, especially when we consider problems with a large number of variables.

5. Conclusions

In the present paper we have investigated the statistical performance in testing for misspecifications in the null distribution, of two multivariate goodness-of-fit tests based on the ECF (a continuous and a discrete version), that we deemed promising procedures for appraising distributional forecasts in a high-dimensional framework, typical of VaR models.

The outcomes from the overall analysis seem to indicate that the continuous procedure possesses good power to detect misspecifications of different nature in the null distribution and, therefore, could represent a good candidate in the evaluation of multivariate distributional forecasts.

In particular, the superiority of the continuous version of the ECF testing criteria over its discrete analogue seems to become more evident as the dimensionality of the problem increases; this might suggest its appropriateness for testing density forecasts developed within VaR problems, where hundreds of variables are involved.

We should acknowledge the fact that the particular choice made for the grid points in the FGP test could not be the most appropriate, and that a different one could have produced better power. However, in any case, the excessive degree of arbitrariness involved in the process of selection of the number and the magnitude of the grid points would seriously affect the reliability and usefulness of the FGP approach, especially in high-dimensional problems.

Much less arbitrariness is involved in the WF approach. Here the power of the test strongly depends on the constant β, but the results from all our experiments agree in recommending a small value for β, particularly in a multidimensional context.

However, the WF procedure presents a serious drawback: its applicability is limited to few specifications for the null distribution, whose characteristic function possesses a tractable form that consents a closed form solution to the integral in (7).

In the present analysis we assumed two forms for the density forecast that satisfy that requirement: 1) a multivariate normal distribution; 2) a mixture of multivariate normal distributions.

For the first specification, we are obviously aware of the fact that the joint distribution of financial returns clearly deviates from multivariate normality. However, it is a matter of fact that in practice many forecasts are still based only on estimates of the mean vector and the covariance matrix of asset returns and, hence, are specified as a multivariate Gaussian. Besides that, some studies (Sornette, Simonetti and Andersen, 1999) indicate that most of the plausible specifications chosen to describe the distribution of financial asset returns, support a transformation to normality that maintains the fat-tailed nature of the marginal distributions and the existence of nonlinear dependence between the assets. Therefore, the assumption of multivariate Gaussian can still represent a sensible choice for the distribution of those transformed variables. In both cases the procedure suggested here can be applied successfully to test for biases in the forecasted distribution.

The second specification seems a plausible assumption for the multivariate distribution of asset returns for, at least, two reasons. Firstly such an assumption implies that the each single variable follows itself a mixture of normal distributions which, given its ability to model the characteristic of fat-tailness exhibited by financial returns, has been recommended by the relevant literature (Hull and White 1998) for this purpose. Secondly this specification allows us to model the dependence structure amongst the variables in a more flexible way than it would be possible under the assumption of multivariate normality. Simple scatter plots of observations generated from a bivariate normal (Fig. 1.a.) and a bivariate mixture with the same first two moments (Fig. 1.b.) suggest that the latter specification can account for some dependence in the tails of the distribution.

However, in some cases it might be necessary to relax the previous assumptions on the form of the distribution, and consider a more general specification for the joint distribution of asset returns, that enables us to model other, less specific, concepts of dependence amongst the assets. In such circumstances, the implementation of the WF statistic can become highly problematic, or even impossible. We might need to consider other testing procedures, perhaps on a case by case basis, especially when the form chosen for the distributional forecast is particularly complicated. Future research might also throw light on whether the WF approach can be extended and applied to other classes of multivariate distributions, provided that different, appropriate choices are made for the weight functions.

References

Baringhaus, L., and Henze, N., 1988, A consistent test for multivariate normality based on the empirical characteristic function, *Metrika*, **35**: 339-348.

Berkowitz, J., 1999, Evaluating the forecasts of risk models, mimeo, Federal Reserve Board, Washington D.C.

Clements, M.P., and Smith, J., 1999, Evaluating the forecast densities of linear and non-linear models: Applications to output growth and unemployment, mimeo, University of Warwick.

Dawid, A.P., 1984, Statistical theory, the prequential approach, *Journal of the Royal Statistical Society A*, **147**: 278-290.

Diebold, F.X., Gunther T.A., Tay, A.S., 1998, Evaluating density forecasts with applications to financial risk management, *International Economic Review*, **39**: 863-883.

Diebold, F.X., Hahn J., Tay, A.S., 2000, Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns on foreign exchange, *Review of Economics and Statistics*, **81**: 661-673.

Epps, T.W., and Pulley, L.B., 1983, A test for normality based on the empirical characteristic function, *Biometrika*, **70**: 723-726.

Fan, Y., 1997, Goodness-of-Fit Tests for a Multivariate Distribution by the Empirical Characteristic Function, *Journal of Multivariate Analysis*, **62**: 36-63.

Fang, K.T., and Wang, Y., 1994, *Number-theoretic methods in statistics*, Chapman & Hall.

Feigin, P.D., and Heathcote, C.R., 1976, The empirical characteristic function and the Cramér-von Mises statistic, *Sankhya Ser. A*, **38**: 309-325.

Feuerverger, A., 1993, A consistent test for bivariate dependence, unpublished manuscript, University of Toronto.

Feuerverger, A., and McDunnough, P., 1981a, On some Fourier methods for inference, *Journal of American Statistical Association*, **76**: 379-386.

Feuerverger, A., and McDunnough, P., 1981b, On the efficiency of empirical characteristic function procedures, *Journal of the Royal Statistical Society B*, **43**: 147-156.

Feuerverger, A., and Mureika, R.A., 1977, The empirical characteristic function and its applications, *Annals of Statistics*, **5**: 88-97.

Hall, P., and Welsh, A.H., 1983, A test for normality based on the empirical characteristic function, *Biometrika*, **70**: 485-489.

Heathcote, C.R., 1972, A test of goodness-of-fit for symmetric random variables, *Australian Journal of Statistics*, **14**: 172-181.

Henze, N., and Wagner, T., 1997, A new approach to the BHEP tests for multivariate normality, *Journal of Multivariate Analysis*, **62**: 1-23.

Henze, N., and Zirkler, B., 1990, A class of invariant consistent tests for multivariate normality, *Commun. Statistics Theory Methods*, **19**: 3595-3617.

Huber, P., 1985, Projection Pursuit (with discussion), *Annals of Statistics*, **13**: 435-535.

Hull, J.C., and White, A., 1998, Value at risk when daily changes in market variables are not normally distributed, *Journal of Derivatives*, **5**(3): 9-19.

Justel, A., Peña, D., and Zamar, R., 1997, A Kolmogorov-Smirnov test of goodness-of-fit, *Statistics & Probability Letters*, **36**: 251-265.

Kotz, S., and Johnson, N.L., 1985, *Encyclopedia of Statistical Sciences*, Wiley N.Y.

Mardia, K.V., 1970, Measures of multivariate skewness and kurtosis with applications, *Biometrika*, **57**: 519-530.

Nagao, H., 1973, On some test criteria for covariance matrix, *Annals of Statistics*, **1**: 700-709.

Nelsen, R.B., 1998, An introduction to copulas, *Lecture Notes in Statistics*, 139, Springer Verlag.

Noceti, P., Smith, J., and Hodges, S.D., 2000, An evaluation of tests of distributional forecasts, FORC Preprint 2000/102, University of Warwick.

Rosenblatt, M. 1952, Remarks on a multivariate transformation, *Annals of Mathematical Stat*istics, **23**: 470-472.

Sornette D., Simonetti P., and Andersen, J.V., 1999, "Nonlinear" covariance matrix and portfolio theory for non-Gaussian multivariate distributions, working paper, UCLA.

Tay, A.S., and Wallis, K.F., 2000, Density forecasting: a survey, *Journal of Forecasting*, **19**: 235-254.

Zhu, L., Fang, K., and Bhatti, M.I., 1997, On estimated Projection Pursuit-type Cramér-von Mises Statistics, *Journal of Multivariate Analysis*, **63**: 1-14.

Appendix

We included in this section the complete formulations for the test statistics based on the weight function approach that have been employed in our analysis.

The original statistic suggested by Henze and Zirkler (1990) was computed on the scaled residuals in place of the original variables:

$$Y_j = S_n^{-1/2}\left(X_j - \bar{X}_n\right) \tag{A.1}$$

and under the null hypothesis of multivariate normality. The formula is as follows:

$$D_{n,\beta} = \frac{1}{n^2}\sum_{j,k=1}^{n}\exp\left(-\frac{\beta^2}{2}\left\|Y_j - Y_k\right\|^2\right) - 2\left(1+\beta^2\right)^{-\frac{d}{2}}\frac{1}{n}\sum_{j=1}^{n}\exp\left(-\frac{\beta^2}{2\left(1+\beta^2\right)}\left\|Y_j\right\|^2\right) + \left(1+2\beta^2\right)^{-\frac{d}{2}} \tag{A.2}$$

and has been used in the present study to test for misspecification of the density forecast when the forecasted density was a standard multivariate Gaussian.

When the statistics is computed on the original, instead of the standardised, variables, the formulation becomes:

$$D_{n,\beta} = \frac{1}{n^2}\sum_{j,k=1}^{n}\exp\left(-\frac{\beta^2}{2}\left\|X_j - X_k\right\|^2\right) + \beta^{-d}\left|2\Sigma + \frac{1}{\beta^2}I\right|^{-1/2} - 2\beta^{-d}\frac{1}{n}\left|S\right|^{-1/2}\sum_{j=1}^{n}\exp\left(-\frac{1}{2}\left\|S^{-1/2}\left(X_j - \mu\right)\right\|^2\right) \tag{A.3}$$

This statistic has been used to test for biases in the mean and in the covariance matrix, under the null of standard multivariate normal.

To test for biases in higher moments when the forecasted distribution assumes the form of a mixture of multivariate normal distributions, an extended version of the first statistic, computed on standardised variables, was derived. Its expression is the following:

$$\begin{aligned} D_{n,\beta} = {} & \frac{1}{n^2}\sum_{j,k=1}^{n}\exp\left(-\frac{\beta^2}{2}\left\|Y_j - Y_k\right\|^2\right) + p^2\beta^{-d}\left|2\Sigma_1^* + \frac{1}{\beta^2}I\right|^{-\frac{1}{2}} + \left(1-p\right)^2\beta^{-d}\left|2\Sigma_2^* + \frac{1}{\beta^2}I\right|^{-\frac{1}{2}} + \\ & -2\frac{p}{n}\beta^{-d}\left|S_1\right|^{-\frac{1}{2}}\sum_{j=1}^{n}\exp\left(-\frac{1}{2}\left\|S_1^{-\frac{1}{2}}Y_j\right\|^2\right) - 2\frac{\left(1-p\right)}{n}\beta^{-d}\left|S_2\right|^{-\frac{1}{2}}\sum_{j=1}^{n}\exp\left(-\frac{1}{2}\left\|S_2^{-\frac{1}{2}}Y_j\right\|^2\right) + \\ & + 2p\left(1-p\right)\beta^{-d}\left|\Sigma_1^* + \Sigma_2^* + \frac{1}{\beta^2}I\right|^{-\frac{1}{2}} \end{aligned} \tag{A.4}$$

where:

$$S_j = \Sigma_j^* + \frac{1}{\beta^2} I$$

$\Sigma_{mix} = p\Sigma_1 + (1-p)\Sigma_2$ represents the covariance structure of the mixture,[1]

$\Sigma_j^* = \Sigma_{mix}^{-1/2} \Sigma_j \left(\Sigma_{mix}^{-1/2}\right)'$ is the standardised covariance matrix of the j-th component, such that the mixture based on the scaled components has zero mean and identity covariance matrix.

[1] Since the mean is the same for both components of the mixture, all the higher moments around the mean are obtained as average of the moments of the single components, weighted by their respective proportions in the mixture.

Table A.1.: Critical values for WF testing procedures

Critical values for test on standardised variables
H_0: multivariate normal

	$\beta=0.5$	$\beta=1$	$\beta=2$
$d=2$	0.0790	0.5542	1.0965
$d=4$	0.1670	0.7793	1.0557
$d=16$	0.7304	0.9999	1

Critical values for test on original variables
H_0: multivariate normal

	$\beta=0.5$	$\beta=1$	$\beta=2$
$d=2$	0.8440	1.3427	1.4189
$d=4$	1.0371	1.2751	1.1253
$d=16$	1.0622	1.0037	1

Critical values for test on standardised variables
H_0: mixture of multivariate normal distributions

	$\beta=0.5$	$\beta=1$	$\beta=2$
$d=2$	0.0971	0.5965	1.1375
$d=4$	0.1899	0.8121	1.0897
$d=16$	0.7937	1.0061	1.0001

Table 1: Size of the tests H_0: $N(0, I_d)$ vs H_1: $N(0, I_d)$

Dimension	Weight Function approach			Fixed Grid Points approach				
				0.5 - 2.5 Step 0.5 m=5	0.5 - 5 Step 0.5 m=10	0.25 - 5 Step 0.25 m=20	0.17 – 5.1 Step 0.17 m=30	0.125 - 5 Step 0.125 m=50
	β =0.5	β =1	β =2					
2	4.9	4.5	4.4	3.9	5.2	2.7		
4	3.4	4.7	4.9	1.7	2.6	3.7		
16	6.0	5.7	5.4			3.7	5.8	3.7

Table 2: Bias in the mean H_0: $N(0, I_d)$ vs H_1: $N(\mu, I_d)$

Dimension	c	μ	Weight Function approach			Fixed Grid Points approach					Hotelling-T^2
			β =0.5	β =1	β =2	0.5 - 2.5 Step 0.5 m=5	0.5 - 5 Step 0.5 m=10	0.25 - 5 Step 0.25 m=20	0.17 – 5.1 Step 0.17 m=30	0.125 - 5 Step 0.125 m=50	
2	2	**0.141**	40.5	28.0	17.3	25.3	20.4	7.2			41.8
	*2.45	**0.173**	54.0	42.0	25.6	36.5	24.9	8.8			56.6
	4	**0.283**	94.7	86.2	62.1	79.0	63.5	27.0			94.5
	6	**0.424**	100	99.7	96.9	99.4	97.1	76.3			100
	8	**0.566**	100	100	100	100.0	100.0	98.9			100
4	2	**0.100**	27.7	17.9	8.1	22.2	22.5	10.5			29.2
	*3.08	**0.154**	62.8	39.3	15.4	51.6	42.8	19.0			68.7
	4	**0.200**	87.1	67.9	27.7	80.0	68.4	33.0			93.2
	6	**0.300**	99.8	96.4	66.3	99.4	98.4	81.6			100
	8	**0.400**	100	100	93.6	100.0	100.0	99.7			100
16	2	**0.035**	11.7	5.7	18.3			14.6	15.3	6.3	17.4
	4	**0.071**	49.8	9.2	20.0			51.6	36.9	9.1	70.2
	*5.13	**0.091**	76.0	15.0	19.1			81.0	60.4	19.4	92.6
	6	**0.106**	91.0	21.3	20.5			92.0	78.3	29.5	98.3
	8	**0.141**	99.8	39.5	20.9			99.9	98.6	76.0	100

*Value corresponding to the contours relative to 95% of the probability density.

Table 3: Bias in the covariance H_0: $N(0, I_d)$ vs H_1: $N(0, \Sigma)$

Dimension	Σ	δ	Weight Function approach			Fixed Grid Points approach					Nagao's test
			$\beta=0.5$	$\beta=1$	$\beta=2$	0.5 - 2.5 Step 0.5 m=5	0.5 - 5 Step 0.5 m=10	0.25 - 5 Step 0.25 m=20	0.17 – 5.1 Step 0.17 m=30	0.125 - 5 Step 0.125 m=50	
2	$\Sigma=[\sigma_1=2, \sigma_2=3, \rho=0.9]$	1.519	100	100	100	100.0	100.0	100.0			100
	$\Sigma=[\sigma=1, \rho=0.4]$	0.526	18.5	40.6	33.4	66.2	66.5	39.2			93.7
	$\Sigma=[\sigma=1, \rho=0.1]$	0.021	5.8	6.8	5.9	18.1	16.8	8.3			12.6
4	$\Sigma=[\sigma=1, \rho=0.4]$	1.631	93.0	95.9	72.0	99.7	99.4	99.8			100
	$\Sigma=[\sigma=1, \rho_{12}=0.3, \rho_{13}=0.2, \rho_{14}=0.1]$	0.432	13.8	19.1	13.6	38.3	39.8	36.0			64.8
	$\Sigma=[\sigma^2=1.1, \rho=0]$	0.033	10.3	10.7	6.8	10.5	13.4	11.2			21.6
16	$\Sigma=[\sigma^2=1.44, \rho=0]$	1.494	100	100	100			83.7	85.8	62.1	100
	$\Sigma=[\sigma^2=1.21, \rho=0]$	0.482	100	83.3	73.9			40.2	44.9	21.8	99.8
	$\Sigma=[\sigma^2=1.05, \rho=0]$	0.036	17.0	6.4	19.1			12.6	16.2	7.3	23.5

Table 4: Misspecification of the density forecast H_0: $N(\mathbf{0}, I_d)$

Dimension	H_1	Weight Function approach			Fixed Grid Points approach					Mardia's test
		β =0.5	β =1	β =2	0.5 - 2.5 Step 0.5 m=5	0.5 - 5 Step 0.5 m=10	0.25 - 5 Step 0.25 m=20	0.17 – 5.1 Step 0.17 m=30	0.125 - 5 Step 0.125 m=50	
2	$0.3N(\mathbf{0}, \Sigma[\sigma=1, \rho=0.7]) + 0.7N(\mathbf{0},\Sigma[\sigma=1, \rho=-0.3])$	12.9	12.9	11.3	19.7	22.3	15.1			16.5
	$0.5N(\mathbf{0}, \Sigma[\sigma=1, \rho=0.9]) + 0.5N(\mathbf{0}, \Sigma[\sigma=1, \rho=-0.9])$	83.4	99.7	99.9	97.0	95.4	72.0			75.9
	t_5	79.4	73.4	61.2	51.7	56.3	49.1			87.7
	t_{10}	32.3	27.0	16.4	21.5	27.2	17.6			47.9
4	$0.3N(\mathbf{0}, \Sigma[\sigma=1, \rho_{12,13}= 0.7]) + 0.7N(\mathbf{0}, \Sigma[\sigma=1,\rho_{12,13}=-0.3])$	18.4	23.1	22.6	4.8	6.4	16.7			18.7
	$0.5N(\mathbf{0}, \Sigma[\sigma=1, \rho_{12,13}=0.7]) + 0.5N(\mathbf{0}, \Sigma[\sigma=1, \rho_{12,13}=-0.7])$	74.5	99.6	98.5	30.0	28.1	41.0			52.5
	t_5	97.7	95.8	83.9	14.0	16.9	49.3			98.5
	t_{10}	56.0	46.4	24.2	4.0	8.1	24.7			73.8
16	$0.7N(\mathbf{0}, \Sigma[\sigma^2 =1.21, \rho=0]) + 0.3N(\mathbf{0}, \Sigma[\sigma^2=0.5, \rho=0])$	100	99.1	60.1			6.8	20.1	9.0	99.1
	$0.6N(\mathbf{0}, \Sigma[\sigma^2=0.625, \rho=0]) + 0.4N(\mathbf{0}, \Sigma[\sigma^2=1.56125, \rho=0])$	100	99.9	49.3			12.1	36.7	19.4	100
	t_5	100	100	99.9			21.2	71.4	58.4	100
	t_{10}	100	99.7	66.7			10.2	42.6	26.4	100

Table 5: Bias in higher moments

Dimension	H_0 and H_1	Weight Function approach			Fixed Grid Points approach				
		$\beta=0.5$	$\beta=1$	$\beta=2$	0.5 - 2.5 Step 0.5 m=5	0.5 - 5 Step 0.5 m=10	0.25 - 5 Step 0.25 m=20	0.17 – 5.1 Step 0.17 m=30	0.125 - 5 Step 0.125 m=50
2	H_0: 0.3N(**0**, Σ[σ=1, ρ=0.7]) + 0.7N(**0**, Σ[σ=1, ρ=-0.3])								
	0.5N(**0**, Σ[σ=1, ρ=0.9]) + 0.5N(**0**, Σ[σ=1, ρ=-0.9])	42.8	83.8	94.2	90.6	83.4	35.3		
	t_5	58.2	56.0	40.9	16.4	26.1	23.4		
	t_{10}	18.3	14.0	8.9	4.4	9.5	6.8		
4	H_0: 0.3N(**0**, Σ[σ=1, $\rho_{12,13}$= 0.7]) + 0.7N(**0**, Σ[σ=1, $\rho_{12,13}$=-0.3])								
	0.5N(**0**, Σ[σ=1, $\rho_{12,13}$=0.7]) + 0.5N(**0**, Σ[σ=1, $\rho_{12,13}$=-0.7])	22.8	57.4	43.6	13.4	12.1	10.6		
	t_5	90.8	89.2	73.1	5.2	5.4	28.4		
	t_{10}	37.5	33.4	22.2	2.0	1.7	8.1		
16	H_0: 0.7N(**0**, Σ[σ^2 =1.21, ρ=0]) + 0.3N(**0**, Σ[σ^2=0.5, ρ=0])								
	0.6N(**0**, Σ[σ^2=0.625, ρ=0]) + 0.4N(**0**, Σ[σ^2=1.56125, ρ=0])	22.5	6.0	3.2			4.4	20.2	8.7
	t_5	100	99.2	57.1			8.1	57.4	42.6
	t_{10}	36.9	6.4	4.0			2.8	24.1	16.4

Fig. 1.a. Gaussian

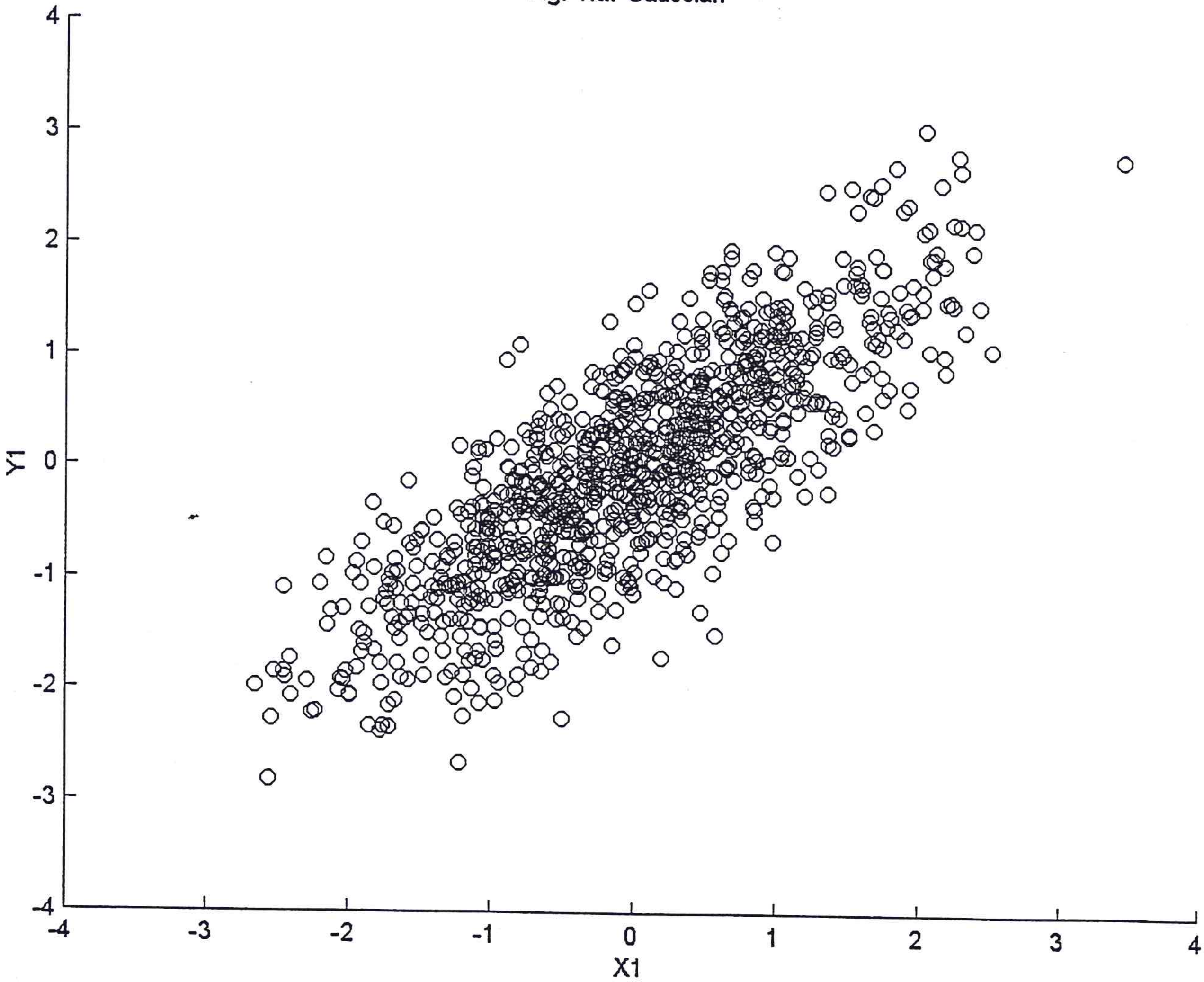

Fig. 1.b. Mixure of normal