# Testing Multivariate Distributions

Jushan Bai*

Professor

Department of Economics, New York University

Zhihong Chen†

Assistant Professor

School of International Trade and Economics,

University of International Business and Economics

*Department of Economics, NYU, 269 Mercer St, New York, NY 10003 Email: Jushan.Bai@nyu.edu.

†School of International Trade and Economics, University of International Business and Economics, Beijing 100029, China Email: zhihong.chen@uibe.edu.cn

# Testing Multivariate Distributions

Jushan Bai* Zhihong Chen †

Jan 17, 2006

**Abstract**

In this paper, we consider testing distributional assumptions based on residual empirical distribution functions. The method is stated for general distributions, but attention is centered on multivariate normal and multivariate t-distributions, as they are widely used, especially in financial time series models such as GARCH. Using the fact that joint distribution carries the same amount of information as the marginal together with conditional distributions, we first transform the multivariate data into univariate independent data based on the marginal and conditional cumulative distribution functions. We then apply the Khmaladze's martingale transformation (K-transformation) to the empirical process in the presence of estimated parameters. The K-transformation purges the effect of parameter estimation, allowing a distribution free test statistic to be constructed. We show that the K-transformation takes a very simple form for testing multivariate normal and multivariate t distributions. For example, when testing normality, we show that K-transformation for multivariate data coincides with that of univariate data. For multivariate t, the transformation depends on the dimension of the data but in a very simple way. We also extend the test to serially correlated observations, including multivariate GARCH models. Finally, we present a practical application of our test procedure on a real multivariate financial time series data set.

*Department of Economics, NYU, 269 Mercer St, New York, NY 10003 Email: Jushan.Bai@nyu.edu.

†School of International Trade and Economics, University of International Business and Economics, Beijing, China 100029 Email: zhihong.chen@uibe.edu.cn

The first author acknowledge financial support from the NSF (grants SES-0137084)

# 1 Introduction

This paper considers the problem of testing multivariate distributions with a focus on the multivariate normal distribution and the multivariate $t$ distribution. This focus is largely motivated by our empirical analysis, which in turns stems from recent developments in the statistical analysis of financial data. When modelling conditional volatility for financial variables as in generalized conditional heteroskedasticity (GARCH), the two most frequently used distributions are multivariate normal and multivariate t, see Tsay (2002). Quite often, it is not clear which distribution provides a better description of the financial variables. Both distributions under GARCH can generate heavy tails and time varying volatility. Both can do a good job in terms of predicting the future conditioning variance. However, when computing the value at risk (VaR) of a portfolio, there could be a huge difference. Normality assumption is likely to underreport the value at risk when the data do not fit the assumption. Therefore, it is useful to know which distribution provides a better characterization for the portfolio's return distribution.

Many tests exist in the literature for multivariate normality, although tests on multivariate t are relatively scant. For multivariate normality, Mecklin and Mundfrom (2004) provided a thorough survey. They classified the tests into four groups: graphic approaches, skewness and kurtosis approaches (e.g. Mardia (1970)), goodness-of-fit approaches (e.g. chi-square test, Kolmogorov and Simirnov test) and finally consistency approaches (e.g. Epps and Pulley (1983), Baringhaus and Henze(1988), Henze and Zirker (1990)). The literature is huge and we have to omit many important contributions; but readers are referred to the comprehensive survey article by Mecklin and Mundfrom (2004). Each procedure has its own advantages and disadvantages. For example, the skewness and kurtosis test is easy to use and performs well against asymmetry. The well known Jarque-Bera (1981,1987) normality test in the econometrics literature is based on symmetry and kurtosis. The chi-square test is widely used for distributional assumptions and has intuitive appeal. When the dimension is high, however, the number of cells required may be large and the number of observations in each cell will be small. The Kolmogorov test is difficult to apply in the presence of estimated parameters, particularly for multivariate data, where the number of estimated parameters is large. When the estimated parameters are ignored, the inference will be invalid. And iid is the usual assumption in most of the existing tests.

In this paper, we propose an alternative procedure. This procedure combines the Kolmogorov test and the K-transformation in Khmaladze (1981). The K-transformation aims

to purge the effect of parameter estimation, yielding a distribution-free test. The procedure is particularly suited for testing multivariate normality and multivariate t. These two classes of distributions enjoy similar properties. Both the marginal distributions and the conditional distributions are in the same family of distributions, enabling simple computation. One appealing property of the proposed procedure is its applicability to time series observations with time-varying means and time-varying covariance matrices. Our monte carlo simulation shows the procedure is easy to implement and has good finite size and power. We use asymptotic critical values, and no specialized tables or simulations are needed.

The paper is organized as follows. In section 2, we start out by outlining the idea of the procedure. The outline is applicable for any multivariate distribution. In section 3, we specialize the general principle to multivariate normality. Section 4 considers time series data such as vector autoregressive models and GARCH processes. In Section 5, we further elaborate the procedure for multivariate t distributions. Section 6 provides monte carlo simulations to assess the finite sample performance of the procedure. Section 7 applies the procedure to a real financial data set by testing the joint conditional distribution of IBM stock's return and the S&P 500 index. And section 8 concludes.

# 2 Description of the method

## 2.1 Preliminary

To introduce the idea, we first consider a bivariate distribution. Suppose the joint density function of $(X, Y)$ is given by

$$f_{XY}(x, y)$$

From

$$f_{XY}(x, y) = f_X(x) f_{Y|X}(y|x)$$

where $f_X$ is the marginal density function of $X$ and $f_{Y|X}$ is the conditional density function of $Y$ conditional on $X$. It is clear that the knowledge of the joint distribution is equivalent to the knowledge of both marginal and conditional distributions. Similarly, from the joint cdf $F_{XY}(x, y)$, one can obtain the marginal cdf $F_X(x)$ and the conditional cdf $F_{Y|X}(y|x)$, and vice versa. As a result, instead of directly testing specifications on the joint distribution $F_{XY}(x, y)$, we test specifications on both the marginal distribution $F_X(x)$ and the conditional distribution $F_{Y|X}(y, x)$.

One key step is to use the integral transformation to obtain uniformly distributed random variables. This transformation allows us to handle nonidentically distributed random variables as well as joint dependence and serial dependence. While $X$ and $Y$ are dependent, a key insight is that $F_X(X)$ and $F_{Y|X}(Y|X)$ are two independent uniform random variables. This can be seen from the following argument.

For an arbitrary random variable $Z$ , if its cdf $F_Z(z)$ is continuous, then $F_Z(Z)$ is $U(0,1)$ random variable. Now the conditional distribution of $Y$ conditional on $X = x$ is $F_{Y|X}(y|x)$, it follows that the conditional distribution of $F_{Y|X}(Y|X)$ (conditional on $X = x$) is $U(0,1)$. Since this conditional distribution does not depend on $x$, the unconditional distribution of $F_{Y|X}(Y|X)$ is also $U(0,1)$ and is independent of $X$. Thus, $F_{Y|X}(Y|X)$ is also independent of any functions of X, in particular, of $F_X(X)$.

The above argument shows that we can turn the multivariate hypothesis testing into testing univariate uniform random variables. This is possible because knowing the joint cdf implies knowing the marginal cdf and the conditional cdf and vice versa. Using these cdf's we can transform the random variables into uniform random variables. What is the most interesting and useful is that these uniform random variables are also iid. This allows for constructing empirical processes that has a Brownian bridge and its limiting process. We will discuss this further below.

Extending this argument to general multivariate distributions is straightforward. Suppose we want to test the joint distribution of $Y = (Y_1, ..., Y_m)$ is

$$F(y_1, ..., y_m).$$

From this joint distribution, one can obtain the marginal distribution $F(y_1)$ and the conditional distributions $F(y_2|y_1), F(y_3|y_1, y_2),\ ...,\ F(y_m|y_1, ..., y_{m-1})$. Conversely, from these marginal and conditional distributions, we can also obtain the joint distribution. Thus testing the random vector $Y$ having a joint cdf $F(y_1, ..., y_m)$ is equivalent to testing

$$F(Y_1),\ F(Y_2|Y_1),\ F(Y_3|Y_1, Y_2), ...,\ F(Y_m|Y_1, ..., Y_{m-1})$$

are $m$ iid $U(0,1)$ random variables.

Now suppose we have a random sample of size $n$ on the random vector $Y$, denoted by (with some a abuse of notation) $Y_1, Y_2, ..., Y_n$ such that $Y_i = (Y_{i1}, ..., Y_{im})$. Then

$$F(Y_{i1}), F(Y_{i2}|Y_{i1}), F(Y_{i3}|Y_{i1}, Y_{i2}), ..., F(Y_{im}|Y_{i1}, ..., Y_{i,m-1}) \qquad i = 1, 2, ..., n$$

form $nm$ number of iid $U(0,1)$ random variables. Now define an empirical process

$$V_{nm}(r) = \frac{1}{\sqrt{nm}} \sum_{i=1}^{n} \sum_{k=1}^{m} [I(U_{ik} \leq r) - r]$$

where $U_{ik} = F(Y_{ik}|Y_{i1}, ..., Y_{i,k-1})$. Then as $n \to \infty$, it is well known that

$$V_{nm} \Rightarrow B(r)$$

where $B(r)$ is a Brownian bridge on [0,1], a zero mean Gaussian process with covariance function $EB(r)B(s) = r \wedge s - rs$. From this weak convergence, one can easily construct test statistic such that

$$S = \max_{r} V_{nm}(r)$$

then by the continuous mapping theorem

$$S \xrightarrow{d} \max_{0 \leq r \leq 1} B(r).$$

## 2.2 When parameters are estimated

The preceding argument assumes that the distribution is fully specified. In practice, however, the joint distribution is only specified up to a vector of unknown parameters. In general, Let $\theta$ be the underlying parameter vector so that we may write $Y \sim F(y_1, ..., y_m; \theta)$. For example, for a normal distribution, we have $Y \sim N(\mu, \Sigma)$. Here $\theta$ consists of $\mu$ and the non-redundant elements of $\Sigma$. Both the marginal and conditional distributions depend on $\theta$. In the bivariate case $Y = (Y_1, Y_2)$, the marginal distribution $Y_1$ can be written as $F_{Y_1}(y_1; h_1(\theta))$ and the conditional distribution as $F_{Y_2|Y_1}(y_2|y_1; h_2(\theta))$, where $h_1$ and $h_2$ are two functions. Let $\widehat{\theta}$ be the MLE of $\theta$. It is clear that $h_1(\widehat{\theta})$ and $h_2(\widehat{\theta})$ are MLE of $h_1(\theta)$ and $h_2(\theta)$, respectively. We mention that MLE is not necessary. Any root-n consistent estimator for $\theta$ is sufficient. This is an advantage of the method proposed, as MLE can be difficult to compute for some distributions. In addition, direct estimators for $\tau_1 = h_1(\theta)$ and $\tau_2 = h_2(\theta)$ instead of the plug-in estimators can also be used. The point is that the parameters of the marginal and conditional distributions can be obtained in various ways.

Now introduce

$$U_{ik} = F(Y_{ik}|Y_{i1}, ..., Y_{i,k-1}; h_k(\theta))$$

and

$$\widehat{U}_{ik} = F(Y_{ik}|Y_{i1}, ..., Y_{i,k-1}; h_k(\widehat{\theta}))$$

Analogous to the definition of $V_{nm}(r)$, we define

$$\widehat{V}_{nm}(r) = \frac{1}{\sqrt{nm}} \sum_{i=1}^{n} \sum_{k=1}^{m} [I(\widehat{U}_{ik} \leq r) - r] \tag{1}$$

Owing to the estimation of the parameters, the limit process of $\widehat{V}_{nm}(r)$ is no longer a Brownian bridge; an extra term will be present in the limit process. In general, we have the representation,

$$\widehat{V}_{nm}(r) = V_{nm}(r) + \bar{g}(r)' \sqrt{mn}(\widehat{\theta} - \theta) + o_p(1)$$

where $\bar{g}(r)$ is a vector of deterministic functions (depends on the actual distribution $F$). Clearly, the limiting process of $\widehat{V}_{nm}$, unlike that of $V_{nm}$, is not distribution-free and depends on the distribution of $\sqrt{mn}(\widehat{\theta} - \theta)$. As a consequence, a test directly based on $\widehat{V}_{nm}(r)$ is difficult to use. This is a well known problem for the Kolmogorov test.

Khmaladze (1988) proposed a transformation method (K-transformation thereafter) that can remove the effect of extra term. The idea of this transformation is to project the process $\widehat{V}_{nm}(r)$ onto $\bar{g}(r)$ and then use the projection residuals. The projection residuals no longer contain the extra term. Because the limiting process $V_{nm}(r)$ is a Brownian bridge, which can be represented as $W(r) - rW(1)$, where $W(r)$ is a standard Brownian motion on $[0, 1]$, the K-transformation will also need to eliminate the drift term $rW(1)$. Therefore, instead of projecting $\widehat{V}_{nm}(r)$ on $\bar{g}(r)$ alone, the K-transformation projects it on $g(r) = (r, \bar{g}(r)')'$. Furthermore, because $W(r)$ is a (continuous-time) random walk, it is more efficient to project $d\widehat{V}_{nm}(r)$ (the counterpart of the difference operator in discrete time) onto the derivative of $g$, $\dot{g}(r)$. The corresponding residuals are the (continuous-time) generalized least squares residuals. We state the K-transformation

$$\widehat{W}_{mn}(r) = \widehat{V}_{mn}(r) - \int_0^r \Big[ \dot{g}(s)' C^{-1}(s) \int_s^1 \dot{g}(\tau) d\widehat{V}_{mn}(\tau) \Big] ds \tag{2}$$

where $C(s) = \int_s^1 \dot{g}(r)\dot{g}'(r)dr$ and $\dot{g}$ is the derivative $g$.

The transformation has an intuitive interpretation. Note that on the interval $[s, 1]$, the least squares estimator when regressing $dV_{nm}(r)$ on $\dot{g}(r)$ is given by

$$\Big( \int_s^1 \dot{g}(\tau)\dot{g}'(\tau)d\tau \Big)^{-1} \int_s^1 \dot{g}(\tau) d\widehat{V}_{mn}(\tau)$$

This is analogous to the discrete-time least squares formula. Denoting this estimated coefficient by $\beta(s)$, the predicted value for the differential $d\widehat{V}_{nm}(s)$ will be the regressor $\dot{g}(s)$ multiplied by the estimated regression coefficient $\beta(s)$, that is, $\dot{g}(s)'\beta(s)$. The predicted

value for $\widehat{V}_{nm}(r)$ is simply the integration of the predicted value for the differential $d\widehat{V}_{nm}(s)$, integrated over the interval $[0, r]$, i.e., $\int_0^r \dot{g}(s)'\beta(s)ds$. This expression is exactly the second term on the right hand side of (2). Finally the projection residual is the difference between $\widehat{V}_{nm}(r)$ and its predicted value. This difference gives the right-hand side of (2). Bai (2003) shows that the K-transformation is in fact calculating the continuous-time counterpart of the recursive residuals in Brown, Durbin and Evans (1976). It is well known that sum of recursive residuals leads to a Brownian motion process. Here the same result holds. That is

$$\widehat{W}_{nm}(r) \Rightarrow W(r).$$

where $W(r)$ is a standard Brownian motion on [0,1]. Now define the test statistic

$$S_{nm} = \max_r |\widehat{W}_{nm}(r)|$$

the continuous mapping theorem implies

$$S_{nm} \xrightarrow{d} \max_r |W(r)|.$$

Therefore, employing the K-transformation, we are able to obtain distribution-free test statistic again. The limiting distribution is the extreme value of a standard Brownian motion instead of Brownian bridge.

The asymptotic critical values can be obtained analytically, and can also be obtained via simulation easily. For convenience, we provide the percentiles of the distribution in Table 1 via simulation. From the table, we see that the critical values at 1significance are 2.787, 2.214, and 1.940, respectively.

We will show subsequently that the K-transformation for testing multivariate normality is very simple. Regardless the value of $m$ or the dimension of $\theta$, the K-transformation takes the same form. In fact we will show that under the assumption of $Y_i \sim N(\mu, \Sigma)$, then

$$\widehat{V}_{nm}(r) \quad = \quad V_{nm}(r) - \phi(\Phi^{-1}(r))a_{nm} - \phi(\Phi^{-1}(r))\Phi^{-1}(r)b_{nm} + o_p(1) \tag{3}$$

where $a_{nm}$ and $b_{nm}$ are random quantities that do not depend on $r$; $\phi(x)$ and $\Phi(x)$ are the density and cdf of $N(0, 1)$. The K-transformation does not need to known $a_{nm}$ and $b_{nm}$. In fact, the transformation implicitly estimates these quantities. A very useful fact to be shown later is that the dimension of $g$ is fixed when testing normality. More specifically,

$$g(r) = (r, \phi(\Phi^{-1}(r)), \phi(\Phi^{-1}(r))\Phi^{-1}(r))'$$

which is a $3 \times 1$ vector. This is the same $g$ as that for testing univariate normality, see Bai (2003). This shows that K-transformation is extremely simple for testing multivariate normality.

Table 1: The distribution of $X = \sup_r |W(r)|$

| $P(X \le x)$ | $x$ | $P(X \le x)$ | $x$ | $P(X \le x)$ | $x$ | $P(X \le x)$ | $x$ | $P(X \le x)$ | $x$ |
|---|---|---|---|---|---|---|---|---|---|
| 1.00 | $\infty$ | 0.80 | 1.625 | 0.60 | 1.260 | 0.40 | 1.011 | 0.20 | 0.799 |
| 0.99 | 2.787 | 0.79 | 1.602 | 0.59 | 1.245 | 0.39 | 0.999 | 0.19 | 0.787 |
| 0.98 | 2.551 | 0.78 | 1.578 | 0.58 | 1.231 | 0.38 | 0.988 | 0.18 | 0.776 |
| 0.97 | 2.407 | 0.77 | 1.556 | 0.57 | 1.218 | 0.37 | 0.978 | 0.17 | 0.765 |
| 0.96 | 2.303 | 0.76 | 1.534 | 0.56 | 1.205 | 0.36 | 0.967 | 0.16 | 0.754 |
| 0.95 | 2.214 | 0.75 | 1.514 | 0.55 | 1.192 | 0.35 | 0.956 | 0.15 | 0.742 |
| 0.94 | 2.146 | 0.74 | 1.494 | 0.54 | 1.178 | 0.34 | 0.945 | 0.14 | 0.730 |
| 0.93 | 2.083 | 0.73 | 1.476 | 0.53 | 1.165 | 0.33 | 0.935 | 0.13 | 0.718 |
| 0.92 | 2.028 | 0.72 | 1.457 | 0.52 | 1.153 | 0.32 | 0.924 | 0.12 | 0.705 |
| 0.91 | 1.982 | 0.71 | 1.440 | 0.51 | 1.140 | 0.31 | 0.914 | 0.11 | 0.692 |
| 0.90 | 1.940 | 0.70 | 1.421 | 0.50 | 1.129 | 0.30 | 0.904 | 0.10 | 0.679 |
| 0.89 | 1.898 | 0.69 | 1.403 | 0.49 | 1.116 | 0.29 | 0.893 | 0.09 | 0.664 |
| 0.88 | 1.860 | 0.68 | 1.386 | 0.48 | 1.104 | 0.28 | 0.882 | 0.08 | 0.650 |
| 0.87 | 1.825 | 0.67 | 1.368 | 0.47 | 1.093 | 0.27 | 0.872 | 0.07 | 0.634 |
| 0.86 | 1.790 | 0.66 | 1.352 | 0.46 | 1.080 | 0.26 | 0.861 | 0.06 | 0.617 |
| 0.85 | 1.759 | 0.65 | 1.336 | 0.45 | 1.069 | 0.25 | 0.851 | 0.05 | 0.600 |
| 0.84 | 1.730 | 0.64 | 1.320 | 0.44 | 1.057 | 0.24 | 0.841 | 0.04 | 0.578 |
| 0.83 | 1.703 | 0.63 | 1.305 | 0.43 | 1.045 | 0.23 | 0.830 | 0.03 | 0.556 |
| 0.82 | 1.676 | 0.62 | 1.290 | 0.42 | 1.034 | 0.22 | 0.819 | 0.02 | 0.527 |
| 0.81 | 1.651 | 0.61 | 1.275 | 0.41 | 1.022 | 0.21 | 0.809 | 0.01 | 0.487 |

# 3 Testing multivariate normality

For ease of exposition, we focus on the bivariate normality. Extension to the multivariate normality is straightforward. Let $Y = (Y_1, Y_2)$ be a bivariate normal vector such that

$$Y \sim N(\mu, \Sigma)$$

where

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \text{ and } \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

It follows that

$$Y_1 \sim N(\mu_1, \sigma_1^2)$$

and the conditional distribution of $Y_2$ is

$$Y_2|Y_1 \sim N(\mu_{2|1},\, \sigma^2_{2|1})$$

where

$$\mu_{2|1} = \mu_2 + \sigma_{21}\sigma_1^{-2}(Y_1 - \mu_1)$$

and

$$\sigma^2_{2|1} = \sigma^2_2 - \sigma^2_{12}\sigma_1^{-2}.$$

Therefore, the marginal cdf of $Y_1$

$$F_1(y_1;\theta) = \Phi\Big(\frac{y_1 - \mu_1}{\sigma_1}\Big)$$

and the conditional cdf of $Y_2$ conditional on $Y_1 = y_1$ is

$$F_{2|1}(y_2|Y_1;\theta) = \Phi\Big(\frac{y_2 - \mu_{2|1}}{\sigma_{2|1}}\Big)$$

As argued in the previous section, replacing $y_1$ and $y_2$ by $Y_1$ and $Y_2$, respectively, the following two random variables

$$U_1 = \Phi\Big(\frac{Y_1 - \mu_1}{\sigma_1}\Big) \text{ and } U_2 = \Phi\Big(\frac{Y_2 - \mu_{2|1}}{\sigma_{2|1}}\Big)$$

are independent $U(0,1)$.

Now suppose $Y_1, ..., Y_n$ are iid with the same distribution as $Y$. Analogous to the above

$$U_{i1} = \Phi\Big(\frac{Y_{i1} - \mu_1}{\sigma_1}\Big) \text{ and } U_{i2} = \Phi\Big(\frac{Y_{i2} - \mu_{2|1,i}}{\sigma_{2|1}}\Big) \qquad i = 1, 2, ..., n$$

form $2n$ iid $U(0,1)$ random variables, where $\mu_{2|1,i} = \mu_2 + \sigma_{21}\sigma_1^{-2}(Y_{i1} - \mu_1)$, which depends on $Y_{i1}$. These uniform random variables are unobservable because the parameters are unknown. Let $\widehat{\mu} = \frac{1}{n}\sum_{i=1}^n Y_i$ and $\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^n (Y_i - \widehat{\mu})(Y_i - \widehat{\mu})'$ be the MLE of $(\mu, \Sigma)$. Replacing the unknown parameters by their estimators, we obtain

$$\widehat{U}_{i1} = \Phi\Big(\frac{Y_{i1} - \widehat{\mu}_1}{\widehat{\sigma}_1}\Big) \text{ and } \widehat{U}_{i2} = \Phi\Big(\frac{Y_{i2} - \widehat{\mu}_{2|1,i}}{\widehat{\sigma}_{2|1}}\Big)$$

where $\widehat{\mu}_{2|1,i}$ is equal to $\mu_{2|1,i}$ with unknown parameters replaced by their estimators. And $\widehat{\sigma}_{2|1}$ is similarly defined. Thus define

$$\widehat{V}_{2n}(r) = \frac{1}{\sqrt{2n}} \sum_{i=1}^n \Big[ I(\widehat{U}_{i1} \le r) - r + I(\widehat{U}_{2i} \le r) - r \Big] \tag{4}$$

$\widehat{V}_{2n}(r)$ is an easily computable process. For example, for each given $r$, it is equal to the number of $\widehat{U}_{i1}$ less than or equal to $r$ plus the corresponding number of $\widehat{U}_{i2}$ minus $2r$ then divided by $\sqrt{2n}$.

The following theorem gives the representation of $\widehat{V}_{2n}(r)$.

**Theorem 3.1** *Under assumption of normality,*

$$\widehat{V}_{2n}(r) = V_{2n}(r) - \phi(\Phi^{-1}(r))\, a_n - \phi(\Phi^{-1}(r))\Phi^{-1}(r)\, b_n + o_p(1) \tag{5}$$

*where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of a standard univariate normal r.v., and*

$$\begin{aligned} a_n &= \tfrac{1}{\sqrt{2}}\Big[(\tfrac{1}{\sigma_1})\sqrt{n}(\widehat{\mu}_1 - \mu_1) + n^{-1/2}\sum_{i=1}^{n}(\tfrac{1}{\sigma_{2|1}})(\widehat{\mu}_{2|1,i} - \mu_{2|1,i})\Big] \\ b_n &= \tfrac{1}{\sqrt{2}}\Big[(\tfrac{1}{2\sigma_1^2})\sqrt{n}(\widehat{\sigma}_1^2 - \sigma_1^2) + (\tfrac{1}{2\sigma_{2|1}^2})\sqrt{n}(\widehat{\sigma}_{2|1}^2 - \sigma_{2|1}^2)\Big] \end{aligned}$$

From this asymptotic representation, we see that the limiting process of $\widehat{V}_{2n}(r)$ will be a Brownian bridge plus extra terms. These extra terms make the Kolmogorov-Smirnov test difficult to use. The actual expression of $a_n$ and $b_n$ would become very important when using Kolmogorov-Smirnov test. For the K-transformation, the actual expressions of $a_n$ and $b_n$ are irrelevant, all needed is that they are stochastically bounded. In our case, they are each $O_p(1)$. The K-transformation only needs deterministic quantities that are functions of $r$, any variable that is not a function $r$ will be flushed into $a_n$ and $b_n$. The K-transformation implicitly estimates $a_n$ and $b_n$ and then forms a prediction of $\widehat{V}_{2n}(r)$ based on the predictor $g(r)$. The K-transformation then uses the prediction residuals so that terms involving $g$ will be eliminated.

With respect to testing normality, a striking feature is that the $g$ function is very simple. This function is identical to that for testing univariate normality, see Bai (2003). This remains true for general multivariate normality other than bivariate normality. The only changes are the expressions of $a_n$ and $b_n$. As pointed out earlier, the expressions of $a_n$ and $b_n$ are immaterial with respect to the K-transformation. This fact makes this procedure very appealing. Let

$$g(r) = (r, \phi(\Phi^{-1}(r)), \phi(\Phi^{-1}(r))\Phi^{-1}(r))'$$

and its derivative

$$\dot{g}(r) = (1, -\Phi^{-1}(r), 1 - \Phi^{-1}(r)^2)$$

From these we obtain the transformed process

$$\widehat{W}_{2n}(r) = \widehat{V}_{2n}(r) - \int_0^r \left[\dot{g}(s)'C^{-1}(s)\int_s^1 \dot{g}(\tau)d\widehat{V}_{2n}(\tau)\right]ds$$

where $C(s) = \int_s^1 \dot{g}(r)\dot{g}'(r)dr$. Now let

$$S_n = \max_{0\le r\le 1} |\widehat{W}_{2n}(r)|$$

we have

**Corollary 3.2** *Under the assumption of Theorem 1*

$$S_n \xrightarrow{d} \max_{0\le r\le 1} |W(r)|$$

The asymptotic critical values of this test statistic can be found in table 1.

# 4 Serially correlated multivariate data

## 4.1 Vector autoregression

We can extend the preceding argument to allow observations to be serially correlated. For concreteness, we consider a vector autoregressive models (VAR). We assume the data are generated from a VAR(p).

$$Y_t = A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + \varepsilon_t$$

and consider testing the null hypothesis that $\varepsilon_t$ are iid $\sim N(0,\Sigma)$. We assume observations $Y_{-p+1}, ..., Y_0, Y_1, ...., Y_n$ are available and the entire analysis will be conditional on the first $p$ observations $Y_{-p+1}, ..., Y_0$. Define the information set at time $t$ as $I_t = \{Y_t, I_{t-1}\}$, $t = 1, 2, ..., n$ with $I_0 = \{Y_{-p+1}, ..., Y_0\}$,then under the null hypothesis that $\varepsilon_t$ are iid $N(0,\Sigma)$,

$$Y_t | I_{t-1} \sim N(\mu_t, \Sigma)$$

where $\mu_t = A_1 Y_{t-1} + \cdots + A_p Y_{t-p}$. The only difference from the previous section is that we have a time-varying mean. Furthermore, this time-varying mean is also stochastic. In order to obtain iid uniform random variables, we need the cdf of each of the following conditional distributions

$$\begin{array}{lcl} Y_{t1} & | & I_{t-1} \\ Y_{t2} & | & (Y_{t1}, I_{t-1}) \\ \vdots & & \\ Y_{tm} & | & (Y_{t1}, ..., Y_{t,m-1}, I_{t-1}) \end{array} \tag{6}$$

So in the time series setting, the conditional information includes both contemporaneous and past information. For example, for the conditional distribution of $Y_{t2}$, the conditional information includes $Y_{t1}$ (contemporaneous information, the same $t$) and past information $I_{t-1}$.

All these conditional distributions are normal. It is straightforward to express the conditional means and conditional variances in terms of $\mu_t$ and $\Sigma$. Let $\mu_{k|k-1,t}$ be the conditional mean and $\sigma^2_{k|k-1,t}$ be the conditional variance of the above $k$th random variable, that is,

$$E[Y_{tk} \,|\, Y_{t1}, ..., Y_{t,k-1}, I_{t-1}] = \mu_{k|k-1,t}$$

$$Var[Y_{tk} \,|\, Y_{t1}, ..., Y_{t,k-1}, I_{t-1}] = \sigma^2_{k|k-1,t}$$

The conditional variance $\sigma^2_{k|k-1,t}$ is in fact time invariant, but we keep $t$ here in order to incorporate GARCH models to be considered later. Then

$$U_{tk} = \Phi\Big(\frac{Y_{tk} - \mu_{k|k-1,t}}{\sigma_{k|k-1,t}}\Big)$$

for $k = 1, ..., m$ and $t = 1, ..., n$ form $n \cdot m$ number of iid uniform random variables. Replacing unknown parameters by the estimated parameters (e.g., least squares estimators), we obtain $\widehat{U}_{tk}$ for $t = 1, ..., n$ and $k = 1, ..., m$. Then we can construct $\widehat{V}_{nm}$ as in (1).

**Theorem 4.1** *Under the assumption that $\varepsilon_t$ are iid $N(0, \Sigma)$, Theorem 3.1 holds with new expressions for $a_n$ and $b_n$ such that $a_n = O_p(1)$ and $b_n = O_p(1)$.*

An equivalent way to compute $U_{tk}$ and $\widehat{U}_{tk}$ is to use $\varepsilon_{tk}$ and $\widehat{\varepsilon}_{tk}$. The latter is simpler. From

$$Y_t = \mu_t + \varepsilon_t$$

and because $\mu_t$ is in the information set $I_{t-1}$, $Y_t$ and $\varepsilon_t$ have the same amount of information once being conditional on $I_{t-1}$. Thus, taking conditional expectation on each side of the following,

$$Y_{tk} = \mu_{tk} + \varepsilon_{tk}$$

we have

$$\begin{aligned} E[Y_{tk} \,|\, Y_{t1}, ..., Y_{t,k-1}, I_{t-1}] &= \mu_{tk} + E[\varepsilon_{tk} \,|\, Y_{t1}, ..., Y_{t,k-1}, I_{t-1}] \\ &= \mu_{tk} + E[\varepsilon_{tk} \,|\, \varepsilon_{t1}, ..., \varepsilon_{t,k-1}, I_{t-1}] \\ &= \mu_{tk} + E[\varepsilon_{tk} \,|\, \varepsilon_{t1}, ..., \varepsilon_{t,k-1}] \end{aligned}$$

The last equality follows because $\varepsilon_t$ is independent of $I_{t-1}$. In summary

$$\mu_{k|k-1,t} = \mu_{tk} + \mu^{\varepsilon}_{k|k-1,t}$$

where

$$\mu^{\varepsilon}_{k|k-1,t} = E[\varepsilon_{tk} \,|\, \varepsilon_{t1}, ..., \varepsilon_{t,k-1}].$$

It follows that

$$Y_{tk} - u_{tk|k-1} = Y_{tk} - u_{tk} - u^{\varepsilon}_{k|k-1,t} = \varepsilon_{tk} - u^{\varepsilon}_{k|k-1,t}$$

Furthermore, the conditional variance of $Y_{tk}$, $\sigma^2_{k|k-1,t}$, is equal to the conditional variance of $\varepsilon_{k|k-1,t}$, and thus

$$U_{tk} = \Phi\Big(\frac{Y_{tk} - \mu_{k|k-1,t}}{\sigma_{k|k-1,t}}\Big) = \Phi\Big(\frac{\varepsilon_{tk} - \mu^{\varepsilon}_{k|k-1,t}}{\sigma_{k|k-1,t}}\Big)$$

Replacing the unknown parameters by the estimated ones, we have

$$\widehat{U}_{tk} = \Phi\Big(\frac{\widehat{\varepsilon}_{tk} - \widehat{\mu}^{\varepsilon}_{k|k-1,t}}{\widehat{\sigma}_{k|k-1,t}}\Big) \tag{7}$$

where $\widehat{\varepsilon}_t$ is the estimated residuals.

We now summarize the procedure:

1. Estimate the parameters in VAR(p) process to obtained the residuals:

$$\widehat{\varepsilon}_t = Y_1 - \widehat{A}_1 Y_{t-1} - \cdots - \widehat{A}_p Y_{t-p}$$

and compute

$$\widehat{\Sigma} = \frac{1}{n}\sum_{t=1}^{n} \widehat{\varepsilon}_t \widehat{\varepsilon}_t'$$

2. Compute $\widehat{U}_{tk}$ using $\widehat{\varepsilon}_t$ and $\widehat{\Sigma}$ according to (7).

3. Construct the process $\widehat{V}_{nm}(r)$ and $\widehat{W}_{nm}(r)$ and compute $S_{nm}$.

We can see that after obtaining the residuals $\widehat{\varepsilon}_t$, the remaining steps are identical to the previous section, that is, we treat $\widehat{\varepsilon}_t$ as an observable variable.

For an illustration, consider the bivariate case. Note that $\mu^{\varepsilon}_{t,1|0}$ is simply the unconditional mean of $\varepsilon_{t1}$, so it is zero; this is true whether we have bivariate or multivariate distribution. Next the conditional mean of $\varepsilon_{t2}$ conditional on $\varepsilon_{t1}$ is $\mu^{\varepsilon}_{t,2|1} = \sigma_{12}\sigma_1^{-2}\varepsilon_{t1}$. Thus

$$\widehat{U}_{t1} = \Phi\Big(\frac{\widehat{\varepsilon}_{t1}}{\widehat{\sigma}_1}\Big) \text{ and } \widehat{U}_{t2} = \Phi\Big(\frac{\widehat{\varepsilon}_{t2} - \widehat{\sigma}_{12}\widehat{\sigma}_1^{-2}\widehat{\varepsilon}_{t1}}{\widehat{\sigma}_{2|1}}\Big)$$

where $\widehat{\sigma}_{2|1} = [\widehat{\sigma}_2^2 - \widehat{\sigma}_{12}^2/\widehat{\sigma}_1^2]^{1/2}$.

## 4.2 Multivariate GARCH models

The preceding section assumes that $\varepsilon_t$ is independent of $I_{t-1}$, we now relax this assumption and consider a particular dependence process for $\varepsilon_t$, which is the multivariate GARCH, see Bollerslev (1986) and Tsay (2002). Let us again consider the VAR(p) model

$$Y_t = A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + \varepsilon_t$$

but instead of assuming $\varepsilon_t$ are iid normal, we assume

$$\varepsilon_t | I_{t-1} \sim N(0, \Sigma_t)$$

so that the conditional distribution has time varying covariance matrix. The GARCH model assumes $\Sigma_t$ is random but depends on the past $\varepsilon_t$ with a fixed number of unknown parameters. This implies that

$$Y_t | I_{t-1} \sim N(\mu_t, \Sigma_t)$$

where, as before, $\mu_t = A_1 Y_{t-1} + \cdots + A_p Y_{t-p}$. Thus the random variables described in equation (6) are all normally distributed. The only difference is that the conditional variances are also time varying. Once the unknown parameters are estimated from the GARCH model, we can again compute $\widehat{U}_{tk}$ easily. It is important to note that while the conditional distribution of $Y_t$ is normal, the unconditional distribution is not normal under GARCH. In fact, the distribution of $Y_t$ has heavy tails, and it may not even have finite variance depending on the parameter values in the GARCH process, see Bollerslev (1987).

Again, as in Section 4.1, it is more convenient to work with disturbances. For concreteness, we focus on bivariate GARCH. Let us write

$$\Sigma_t = \begin{bmatrix} \sigma^2_{11,t} & \sigma_{12,t} \\ \sigma_{21,t} & \sigma^2_{22,t} \end{bmatrix}$$

where the conditional variance $\Sigma_t$ is time-varying. Throughout, we also write

$$\sigma^2_{1,t} = \sigma_{11,t}, \text{and } \sigma^2_{2,t} = \sigma_{22,t}$$

as part of conventional notation. This means that

$$\varepsilon_{t1} | I_{t-1} \sim N(0, \sigma^2_{1,t})$$

$$\varepsilon_{t2} | (\varepsilon_{t1}, I_{t-1}) \sim N(\sigma_{12,t}\sigma^{-2}_{1,t}\varepsilon_{t1}, \sigma^2_{2|1,t})$$

where $\sigma^2_{2|1,t} = \sigma^2_{22,t} - \sigma^2_{21,t}/\sigma^2_{11,t}$. Therefore, this means that

$$U_{t1} = \Phi\Big(\frac{\varepsilon_{t1}}{\sigma_{1,t}}\Big) \text{ and } U_{t2} = \Phi\Big(\frac{\varepsilon_{t2} - \sigma_{12,t}\sigma^{-2}_{1,t}\varepsilon_{t1}}{\sigma_{2|1,t}}\Big)$$

are iid $U(0,1)$. Replacing $\varepsilon_t$ and $\{\sigma_{ij,t}\}$ by $\widehat{\varepsilon}_t$ and $\{\widehat{\sigma}_{ij,t}\}$, obtained from a multivariate GARCH model, we will obtain $\widehat{U}_{t1}$ and $\widehat{U}_{t2}$ for $t = 1, 2..., n$. Therefore, the procedure is identical to VAR(p) in previous section. The only difference is the conditional variances are time varying.

We next consider the modelling of $\Sigma_t$. Due to its symmetry, $\Sigma_t$ contains three distinct processes. Instead of directly modelling the three processes $(\sigma^2_{1,t}, \sigma_{21,t}, \sigma^2_{2,t})$ of $\Sigma_t$, Tsay (2002) suggested a reparametrization that turns out convenient. Tsay suggested modelling the triple:

$$(\sigma^2_{1,t}, q_{21,t}, \sigma^2_{2|1,t})$$

where

$$q_{21,t} = \sigma_{21,t}/\sigma^2_{1,t} \qquad \text{and}$$
$$\sigma^2_{2|1,t} = \sigma^2_{2,t} - \sigma^2_{21,t}/\sigma^2_{1,t}$$

Introduce

$$\eta_{t1} = \varepsilon_{t1}, \text{ and } \eta_{t2} = \varepsilon_{t2} - \sigma_{21,t}\sigma^{-2}_{1,t}\varepsilon_{t1}$$

Clearly, $\sigma^2_{2|1,t}$ is the conditional variance of $\eta_{t2}$, conditional on $I_{t-1}$. With these reparametrizations, the likelihood function takes a very simple form as shown by Tsay (2002). In addition, $U_{t2}$ is simply $\Phi(\eta_{t2}/\sigma_{2|1,t})$. After estimating a GARCH process, it is straightforward to compute these $\widehat{U}_{tk}$. Further details on GARCH modelling are given in our empirical applications.

# 5 Testing multivariate t-distribution

The entire analysis for testing multivariate normality is readily extended to multivariate t-distributions. A standard (univariate) t-distribution with degree of freedom $\nu$ has the density

$$q_\nu(x) = c(\nu + x^2)^{-(1+\nu)/2}$$

where $c$ is a constant making the integral of $q_\nu(x)$ on the real line being 1. Let $Q_\nu(x)$ denote the cdf. A random variable $Y$ is said to have a generalized $t$ distribution with parameters $(u, h^2, \nu)$ if

$$t = (Y - u)/h$$

has a standard t-distribution with $\nu$ degrees of freedom. We denote $Y \sim t(u, h^2, \nu)$. It is clear that

$$P(Y \leq y) = Q_\nu\Big(\frac{y-u}{h}\Big) \tag{8}$$

so that we can easily compute probabilities of a generalized t random variable in terms of a standard t random variable, much like the normal distribution. This is convenient because most statistical packages such as SPLUS and MATLAB have $q_v(x)$ and $Q_v(x)$ built in. Note that $h$ is not the standard deviation because $Var(Y) = \frac{\nu}{\nu-2}h^2$.

A random vector $X = (X_1, ..., X_m)$ is said to have (generalized) multivariate t distribution with parameters $(u, \Omega, \nu, m)$ if its density

$$f_X(x; u, \Omega, \nu, m) = C[\nu + (x-u)'\Omega^{-1}(x-u)]^{-(\nu+m)/2}$$

where $x = (x_1, ..., x_m)'$, $u = (u_1, ..., u_m)'$, and $C$ is a normalizing constant (depending on the parameters). We denote $X \sim t(u, \Omega, \nu)$. It is known that $E(X) = u$ and $Var(X) = \frac{\nu}{\nu-2}\Omega$.

Analogous to multivariate normality, when $X$ has a multivariate t distribution, any subvector of $X$ is also multivariate t. In particular, each random variable $X_i$ $(i = 1, ..., m)$ is a univariate generalized $t$ random variable. [1]Furthermore, conditional distributions are also multivariate $t$. In particular, the conditional distributions $X_k|(X_1, ..., X_{k-1})$ for $k = 2, 3, ..., m$ are all univariate generalized $t$; see, e.g. Zellner (1971). However, unlike the normality case, the conditional variance is no longer a constant, but a function of the conditional variables. Further properties on multivariate $t$ can be found in Kotz and Nadarajah (2004).

Partition $X = (X_1', X_2')'$ where $X_1$ is $p \times 1$, and $X_2$ is $(m-p) \times 1$, and let

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \text{ and } \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix} \tag{9}$$

be partitioned conformably. Then both the marginal $X_1$ and the conditional $X_2|X_1$ are generalized multivariate t such that

$$X_1 \sim t(u_1, \Omega_{11}, \nu)$$

and

$$X_2|X_1 \sim t(u_{2|1},\ \Omega_{2|1},\ \nu + m - p)$$

where

$$u_{2|1} = u_2 + \Omega_{21}\Omega_{11}^{-1}[X_1 - u_1] \tag{10}$$

[1]Conditional distributions associated with a standard multivariate $t$(where $\Omega$ is the correlation matrix) are not necessarily standard multivariate $t$, but are generalized multivariate $t$.

$$\Omega_{2|1} = a[\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}] \tag{11}$$

with

$$a = [\nu + (X_1 - u_1)'\Omega_{11}^{-1}(X_1 - u_1)]/(\nu + m - p)$$

Therefore, if both $X_1$ and $X_2$ are scalars ($m = 2, p = 1$), it follows immediately from (8)

$$P(X_1 \le x) = Q_\nu\left(\frac{x - u_1}{\Omega_{11}^{1/2}}\right)$$

$$P(X_2 \le x|X_1) = Q_{\nu+1}\left(\frac{x - u_{2|1}}{\Omega_{2|1}^{1/2}}\right)$$

because $\nu + m - p = \nu + 1$. Thus

$$U_1 = Q_\nu\left(\frac{X_1 - u_1}{\Omega_{11}^{1/2}}\right) \text{ and } U_2 = Q_{\nu+1}\left(\frac{X_2 - u_{2|1}}{\Omega_{2|1}^{1/2}}\right) \tag{12}$$

are two independent uniform random variables.

For testing multivariate $t$, we focus on the case of bivariate distribution. The extension to general multivariate case follows quite naturally. Now suppose $Y_1, Y_2, ..., Y_n$ form a random sample from a bivariate $t(u, \Omega, \nu)$ with parameters given in (9). Denote $Y_t = (Y_{t1}, Y_{t2})$. The previous analysis shows

$$Y_{t1} \sim t(u_1,\, \Omega_{11},\, \nu), \quad \text{and} \quad Y_{t2}|Y_{t1} \sim t(u_{2|1,t},\, \Omega_{2|1,t},\, \nu + 1)$$

where $u_{2|1,t}$ and $\Omega_{2|1,t}$ are given in (10) and (11), respectively, with $X_1$ replaced by $Y_{t1}$. The subscript $t$ in $u_{2|1,t}$and $\Omega_{2|1,t}$ signify their dependence on $Y_{t1}$. Therefore,

$$U_{t1} = Q_\nu\left(\frac{Y_{t1} - u_1}{\Omega_{11}^{1/2}}\right) \text{ and } U_{t2} = Q_{\nu+1}\left(\frac{Y_{t2} - u_{2|1,t}}{\Omega_{2|1,t}^{1/2}}\right) \tag{13}$$

$(t = 1, 2, ..., n)$ form $2n$ iid uniform random variables.

Next consider the case that $u$ and $\Omega$ are estimated. Because $EY_t = u$, and $E(Y_t - u)(Y_t - u)' = \frac{\nu}{\nu-2}\Omega$. Let $\Sigma$ be the variance, i.e., $\Sigma = \frac{\nu}{\nu-2}\Omega$. Consider the moment estimator

$$\widehat{u} = \frac{1}{n}\sum_{t=1}^{n} Y_t, \text{ and } \widehat{\Sigma} = \frac{1}{n-1}\sum_{t=1}^{n}(Y_t - \widehat{u})(Y_t - \widehat{u})'$$

then $(\widehat{u}, \widehat{\Sigma})$ is unbiased for $(u, \Sigma)$. Thus

$$\widehat{\Omega} = [(\nu - 2)/\nu]\widehat{\Sigma}$$

is unbiased for $\Omega$. These estimators are also $\sqrt{n}$ consistent. We assume $\nu$ is known. The case of unknown $\nu$ requires a separate analysis, which we will omit. In the case that $\nu$ is assumed to take on integer values, a consistently estimated $\nu$ can be treated as known because consistency implies $P(\widehat{\nu} \neq \nu) = 0$ in view of the discreteness of $\nu$.

Given the estimated parameters, we can construct $\widehat{U}_{t1}$ and $\widehat{U}_{t2}$ as in (13) with the unknown parameters replaced by their estimators, for example, $\widehat{u}_{2|1,t} = \widehat{u}_2 + \widehat{\Omega}_{21}\widehat{\Omega}_{11}^{-1}(Y_{t1} - \widehat{u}_1)$ and $\widehat{\Omega}_{2|1,t}$ is similarly obtained using (11).

Let $\widehat{V}_{2n}$ be defined in (4) with newly constructed $\widehat{U}_{tk}$ $(k = 1, 2; t = 1, ..., n)$, we have

**Theorem 5.1** *Under assumptions of bivariate t, we have*

$$\widehat{V}_{2n}(r) = V_{2n}(r) - \bar{g}(r)'\xi_n + o_p(1)$$

*where*

$$\bar{g}(r) = \begin{bmatrix} q_\nu(Q_\nu^{-1}(r)) \\ q_\nu(Q_\nu^{-1}(r))Q_\nu^{-1}(r) \\ q_{\nu+1}(Q_{\nu+1}^{-1}(r)) \\ q_{\nu+1}(Q_{\nu+1}^{-1}(r))Q_{\nu+1}(r) \end{bmatrix}, \ \xi_n = \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{n}(\widehat{u}_1 - u_1)/\Omega_{11}^{1/2} \\ \frac{1}{2}\sqrt{n}(\widehat{\Omega}_{11} - \Omega_{11})/\Omega_{11} \\ \frac{1}{\sqrt{n}}\sum_{t=1}^n (\widehat{u}_{2|1,t} - u_{2|1,t})/\Omega_{2|1,t}^{1/2} \\ \frac{1}{\sqrt{n}}\frac{1}{2}\sum_{t=1}^n (\widehat{\Omega}_{2|1,t} - \Omega_{2|1,t})/\Omega_{2|1,t} \end{bmatrix}$$

*where $q_v$ and $Q_v$ are, respectively, the density and cdf of a standard univariate t random variable with $v$ degrees of freedom.*

The actual expression for $\xi_n$ plays no role in the martingale transformation, but the expression of $\bar{g}(r)$ is important. Let $g(r) = (r, \bar{g}(r)')'$, then $g$ is $5 \times 1$ vector. Given $g$, the K-transformation is straightforward. It is interesting to note that for multivariate t distribution, the $g$ function has higher dimension than its counterpart in the normal distribution case. For a normal distribution, the dimension of $g$ does not depend on $m$, but for multivariate t, the dimension of $g$ is $2m + 1$. The K-transformation is $\widehat{W}_{2n}(r) = \widehat{V}_{2n}(r) - \int_0^r \left[\dot{g}(s)'C^{-1}(s)\int_s^1 \dot{g}(\tau)d\widehat{V}_{2n}(\tau)\right]ds$ and the test statistic is $S_n = \max_{0\leq r\leq 1}|\widehat{W}_{2n}(r)|$.

An alternative strategy is to perform two separate tests. The first is to test $U_{t1}(t = 1, ..., n)$ are iid uniform and the second is to test $U_{t2}(t = 1, ..., n)$ are iid uniform. The first test uses

$$g = (r,\, q_\nu(Q_\nu^{-1}(r)),\, q_\nu(Q_\nu^{-1}(r))Q_\nu^{-1}(r))'$$

in the K-transformation and the second test uses

$$g = (r,\, q_{\nu+1}(Q_{\nu+1}^{-1}(r)),\, q_{\nu+1}(Q_{\nu+1}^{-1}(r))Q_{\nu+1}^{-1}(r))'$$

in the transformation. So in each test, the $g$ function is $3 \times 1$ and has the same form, but the second $g$ uses the pdf and cdf of $t$ with one more degree of freedom. Let $S_{n1}$ and $S_{n2}$ be the corresponding two test statistics. Asymptotically, $S_{n1}$ and $S_{n2}$ are independent and have the same distribution. Let

$$T_n = \max\{S_{n1}, S_{n2}\}$$

Let $F_S(s)$ denote the cdf of the limiting random variable of $S_{n1}$, clearly, the limiting distribution of $T_n$ has a cdf $F_S(s)^2$.

We next consider extending the iid sample to time series observations.

**VAR with GARCH errors.** For simplicity, we consider the bivariate case. Suppose

$$Y_t = A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + \varepsilon_t$$

Again, let $I_t = (Y_t, I_{t-1})$ with $I_0 = (Y_{-p+1}, ..., Y_0)$. We test the hypothesis that

$$\varepsilon_t | I_{t-1} \sim t(0, \Omega_t, \nu)$$

where

$$\Omega_t = \begin{bmatrix} \Omega_{11,t} & \Omega_{12,t} \\ \Omega_{21,t} & \Omega_{22,t} \end{bmatrix}$$

This is equivalent to $Y_t | I_{t-1} \sim t(u_t, \Omega_t, \nu)$ where $u_t = A_1 Y_{t-1} + \cdots + A_p Y_{t-p}$. Instead of constant mean and constant variance, $Y_t$ now has time-varying (conditional) mean and variance. But these time-varying random parameters pose no new difficulty. Replacing the time-invariant triple $(\widehat{u}_1, \widehat{u}_2, \widehat{\Omega}_{ij})$ by the time-varying triple $(\widehat{u}_{t1}, \widehat{u}_{t2}, \widehat{\Omega}_{ij,t})$, all preceding arguments go through, except that the expression of $\xi_n$ in Theorem 5 is different. But the expression of $\xi_n$ plays no role in the K-transformation.

Similar to the normal case, it is more convenient to construct the test in terms of residuals. Let $\widehat{\varepsilon}_t = Y_t - \widehat{A}_1 Y_{t-1} - \cdots - \widehat{A}_p Y_{t-p}$. The GARCH process provides a model for $\Sigma_t = \frac{\nu}{\nu-2}\Omega_t$, see Tsay (2002). After obtaining $\widehat{\Sigma}_t$ from a GARCH, we define $\widehat{\Omega}_t = \frac{\nu-2}{\nu}\widehat{\Sigma}_t$. Next define

$$\widehat{U}_{t1} = Q_\nu \left( \frac{\widehat{\varepsilon}_{t1}}{\widehat{\Omega}_{11,t}^{1/2}} \right) \text{ and } \widehat{U}_{t2} = Q_{\nu+1} \left( \frac{\widehat{\varepsilon}_{t2} - E(\widehat{\varepsilon_{t2}|\varepsilon_{t1}})}{\widehat{\Omega}_{2|1,t}^{1/2}} \right) \tag{14}$$

where

$$E(\widehat{\varepsilon_{t2}|\varepsilon_{t1}}) = \widehat{\Omega}_{21,t}\widehat{\Omega}_{11,t}^{-1}\widehat{\varepsilon}_{t1}$$

$$\widehat{\Omega}_{2|1,t} = \widehat{a}_t[\widehat{\Omega}_{22,t} - (\widehat{\Omega}_{21,t})^2\widehat{\Omega}_{11,t}^{-1}] \tag{15}$$

with

$$\widehat{a}_t = [\nu + (\widehat{\varepsilon}_{t1})^2 \widehat{\Omega}_{11,t}^{-1}]/(\nu+1)$$

The expression of $\xi_n$ in VAR+GARCH model is different from that in theorem 5 because $\varepsilon_t$ has time-varying (conditional) mean and variance. We have the following corollary:

**Corollary 5.2** *under assumptions of bivariate t, we have*

$$\widehat{V}_{2n}(r) = V_{2n}(r) - \bar{g}(r)'\xi_n + o_p(1)$$

$$\textit{where } \bar{g}(r) = \begin{bmatrix} q_\nu(Q_\nu^{-1}(r)) \\ q_\nu(Q_\nu^{-1}(r))Q_\nu^{-1}(r) \\ q_{\nu+1}(Q_{\nu+1}^{-1}(r)) \\ q_{\nu+1}(Q_{\nu+1}^{-1}(r))Q_{\nu+1}(r) \end{bmatrix}, \ \xi_n = \frac{1}{\sqrt{2n}} \begin{bmatrix} \sum_{t=1}^{n}(\widehat{u}_{1t} - u_{1t})/\Omega_{11,t}^{1/2} \\ \frac{1}{2}\sum_{t=1}^{n}(\widehat{\Omega}_{11,t} - \Omega_{11,t})/\Omega_{11,t} \\ \sum_{t=1}^{n}(\widehat{u}_{2|1,t} - u_{2|1,t})/\Omega_{2|1,t}^{1/2} \\ \frac{1}{2}\sum_{t=1}^{n}(\widehat{\Omega}_{2|1,t} - \Omega_{2|1,t})/\Omega_{2|1,t} \end{bmatrix}$$

The $\bar{g}(r)$ function is the same as in Theorem 5.1, only $\xi_n$ has a different expression. So the K-transformation is identical to the case of iid sample.

# 6 Simulations

We use simulations to assess the size and power properties of the suggested test statistic.

## 6.1 Testing conditional normality

To see the size of our test for conditional normality, random variables $Y_t$ are generated from bivariate normal distribution

$$Y_t \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}\right), \ t = 1, 2, ..., n$$

for various sample sizes. We let $\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}_{11}^2, \widehat{\sigma}_{22}^2, \widehat{\sigma}_{12}$ denote the sample means and sample variances and covariance. We then compute $\widehat{U}_{t1} = \Phi([Y_{t1} - \widehat{\mu}_1]/\widehat{\sigma}_{11})$ and $\widehat{U}_{t2} = \Phi((Y_{t2} - \widehat{\mu}_{t,2|1})/\widehat{\sigma}_{2|1})$, where $\widehat{\mu}_{t,2|1} = \widehat{\sigma}_{12}\widehat{\sigma}_{11}^{-2}(Y_{t1} - \widehat{\mu}_1)$ and $\widehat{\sigma}_{2|1}^2 = \widehat{\sigma}_{22}^2 - \widehat{\sigma}_{12}^2/\widehat{\sigma}_{11}^2$ are the estimated conditional mean and conditional variance of $Y_{2t}$, conditional on $Y_{t1}$. Once the $\widehat{U}_t$'s are obtained, the remaining computation becomes standard and is automated. For each sample, we compute the test statistic $S_n$. This is done with 5000 repetitions. The critical values at 10%, 5% and 1% are 1.940, 2.214 and 2.787 respectively. The results from 5000 repetitions are reported in Table2.

Table 2. Size of the Test for Multivariate Normal Distribution

| | 5000 repetitions | | |
|---|---|---|---|
| $n$ | 10% | 5% | 1% |
| 100 | 0.106 | 0.063 | 0.024 |
| 200 | 0.108 | 0.063 | 0.020 |
| 500 | 0.107 | 0.059 | 0.019 |

From Table 2, we see that the size appears to be reasonable.

For power, two different symmetric distributions from the elliptically contoured family are considered: multivariate uniform distribution and multivariate t distribution (with $df = 5$). The latter departs from normality with heavy tail. We also consider multivariate lognormal and multivariate chi-square distribution (with $df = 1$) that depart from normality with heavy skewness. We then proceed as if the data were generated from a bivariate normal distribution. We perform exactly the same computation as in testing bivariate normality with 5000 repetitions. The power of the test is shown in the following table.

Table 3. Power of the Test for Multivariate Normal Distribution

| multivariate uniform | | | | multivariate $t$ | | |
|---|---|---|---|---|---|---|
| $n$ | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 1.00 | 1.00 | 1.00 | 0.69 | 0.63 | 0.53 |
| 200 | 1.00 | 1.00 | 1.00 | 0.90 | 0.87 | 0.78 |
| 500 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | multivariate lognormal | | | multivariate $\chi^2$ | | |
| $n$ | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 200 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 500 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

Overall, the power is satisfactory. As $n$ increases, power gets larger, as expected.

## 6.2 Testing conditional t

If the null hypothesis is conditional t-distribution, for size, random variables $Y_t$ are generated from a bivariate t-distribution with degree of freedom 5:

$$Y_t \sim t\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}, 5\right), \ t = 1, 2, ..., n$$

After estimating $\widehat{\mu}_1$, $\widehat{\mu}_{2|1,t}$, $\widehat{\Omega}_{11}$ and $\widehat{\Omega}_{2|1,t}$ as described in section 5, we can transform $Y_{t1}$ and $Y_{t2}$ into two independent uniform random variables by setting $\widehat{U}_{t1} = Q_\nu\left(\frac{Y_{t1}-\widehat{u}_1}{\widehat{\Omega}_{11}^{1/2}}\right)$ and $\widehat{U}_{t2} = Q_{\nu+1}\left(\frac{Y_{t2}-\widehat{u}_{2|1,t}}{\widehat{\Omega}_{2|1,t}^{1/2}}\right)$, where $\widehat{u}_{2|1,t}$ and $\widehat{\Omega}_{2|1,t}^{1/2}$ are conditional on $Y_{t1}$. Once $\widehat{U}_t$'s are obtained, we compute the test statistic $T_n$ in the standard way. The results from 5000 repetitions are reported in Table 4.

Table 4. Size of the Test for Multivariate t Distribution

| | 5000 repetitions | | |
|---|---|---|---|
| $n$ | 10% | 5% | 1% |
| 100 | 0.100 | 0.059 | 0.023 |
| 200 | 0.097 | 0.057 | 0.026 |
| 500 | 0.098 | 0.064 | 0.025 |

To see the power, we generate the following alternative: multivariate uniform, multivariate Cauchy, multivariate lognomal and multivariate $\chi^2_{(1)}$. The results are reported in Table5.

Table 5. Power of the Test for Multivariate $t$ Distribution

| Multivariate uniform | | | | Multivariate Cauchy | | |
|---|---|---|---|---|---|---|
| $n$ | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 0.99 | 0.99 | 0.96 | 0.87 | 0.81 | 0.64 |
| 200 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.96 |
| 500 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | Multivariate lognormal | | | Multivariate $\chi^2$ | | |
| $n$ | 10% | 5% | 1% | 10% | 5% | 1% |
| 100 | 0.90 | 0.85 | 0.73 | 0.83 | 0.74 | 0.50 |
| 200 | 0.99 | 0.98 | 0.94 | 0.97 | 0.96 | 0.88 |
| 500 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

The size and the power are satisfactory.

# 7 Empirical Applications

In this section, we apply the test procedure to a pair of financial time series, namely, the monthly log returns of IBM stock and the S&P 500 index. The sample range is from January

1926 to December 1999 with 888 observations. The returns include dividend payments and are in percentages. Let $Y_{1t}$ =the returns of IBM stock and $Y_{2t}$ =the return of the S&P 500 index. Figure 1 shows that the two return series are concurrently correlated.

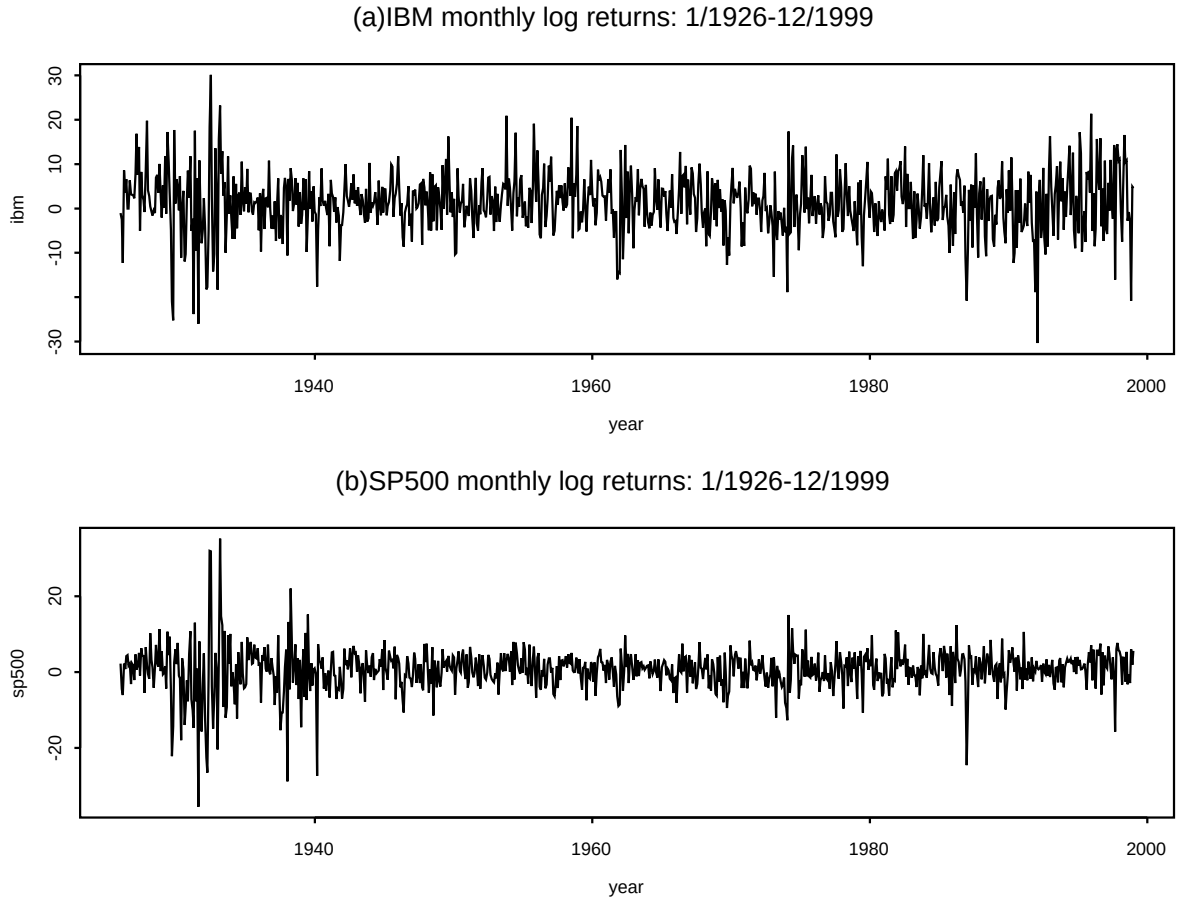


Figure 1: IBM and SP500 Monthly Log Returns, January 1926 to December 1999

The objective is to see which multivariate conditional distribution, conditional normal or conditional t-distribution, provides a better characterization of this bivariate financial series. In portfolio management, it is the conditional distribution that the managers care about the most. For example, to update the value at risk periodically, the conditional distribution, conditional on the given information, is the most relevant. When conditional normality is assumed, but the actual conditional distribution has heavy tails, it would be likely to under estimate the value at risk.

**Testing bivariate conditional normality**. It is well known that financial data have heavy tail distributions. GARCH models under conditional normality may describe the heavy tail property, see Bollerslev (1987). We test conditional normality first.

As in Tsay (2002), we use maximum likelihood method to estimate this bivariate GARCH(1,1) model as

$$\begin{aligned} Y_{1t} &= 1.364 + 0.075Y_{1,t-1} - 0.058Y_{2,t-2} + \varepsilon_{1t} \\ Y_{2t} &= 0.643 + \varepsilon_{2t} \end{aligned}$$

where $(\varepsilon_{1t}, \varepsilon_{2t})$ is conditionally normal. The fitted conditional volatility model is

$$\begin{aligned} \sigma^2_{11,t} &= 3.714 + 0.113\varepsilon^2_{1,t-1} + 0.804\sigma^2_{11,t-1} \\ q_{21,t} &= 0.003 + 0.992q_{21,t-1} - 0.004\varepsilon_{2,t-1} \\ \sigma^2_{2|1,t} &= 1.023 + 0.021\varepsilon^2_{1,t-1} + 0.052\eta^2_{2,t-1} - 0.040\sigma^2_{11,t-1} + 0.937\sigma^2_{2|1,t-1} \end{aligned}$$

where $\eta_{2t}$ and $\sigma^2_{2|1,t}$ are defined in section (4.2). This GARCH(1,1) process allows us to compute $\widehat{\varepsilon}_{t1}$, $\widehat{\varepsilon}_{t2}$, $\widehat{\eta}_{t2}$, $\widehat{\sigma}^2_{11,t}$, and $\widehat{\sigma}^2_{2|1,t}$. Then we compute $\widehat{U}_{t1} = \Phi(\widehat{\varepsilon}_{t1}/\widehat{\sigma}_{11,t})$ and $\widehat{U}_{t2} = \Phi(\widehat{\eta}_{2t}/\widehat{\sigma}_{2|1,t})$. Given $\widehat{U}_{tk}$ $(k = 1, 2; t = 1, ..., n)$, the value of the test statistic is found to be $S_n = 4.8945$. However, the critical values of the test statistic at significance levels 10%, 5% and 1% are 1.940, 2.214 and 2.787, respectively. Panel (a) in figure 2 shows we should reject conditional normality assumption. In this figure, the dotted curve represents the original process $\widehat{V}_{2n}$; the solid curve represents the transformed process $\widehat{W}_{2n}$. And the horizontal dashed and dash-dotted lines give 90% and 99% confidence bands for a standard Brownian motion on $[0, 1]$ , respectively. In panel (a), $\widehat{W}_{2n}$ reaches out of the 99% confidence band. Therefore, we easily reject the conditional normality assumption. A GARCH model with conditional normality is still likely to underestimate the tail probabilities. This may have practical consequence if value at risk is computed using a conditional normal distribution. We then test if the conditional t-distribution is appropriate.

**Testing bivariate conditional $t$ distribution**. No additional model estimation is needed after we have estimated parameters in the GARCH-normal distribution because the $\Omega_t$ matrix in the conditional $t$-distribution is equal to $\frac{\nu-2}{\nu}\Sigma_t$, where $\Sigma_t$ is the conditional variance matrix in the normal case.

The conditional normality estimation provides an estimate $\widehat{\Sigma}_t$. It follows that $\widehat{\Omega}_t = [(\nu - 2)/\nu]\widehat{\Sigma}_t$. The value of $\nu$ is taken to be $\nu = 5$; this is the value that is shown to be appropriate for financial data and is used widely in empirical analysis; See, e.g., Engle and Gonzalez-Rivera (1991). Then we compute $\widehat{U}_{t1} = Q_\nu(\frac{\widehat{\varepsilon}_{t1}}{\widehat{\Omega}^{1/2}_{11,t}})$ and $\widehat{U}_{t2} = Q_{\nu+1}\left(\frac{\widehat{\eta}_{t2}}{\widehat{\Omega}^{1/2}_{2|1,t}}\right)$, according to (14) and (15)

Given $\widehat{U}_{tk}$ $(k = 1, 2; t = 1, ..., n)$, the value of the test statistic is found to be $S_n = 1.1805$. While the critical values of the test statistic at significance levels 10%, 5% and 1% are 1.940, 2.214 and 2.787, respectively. Panel (b) of figure 2 shows that $\widehat{W}_{2n}$, the solid curve, stays within 90% confidence band for a standard Brownian motion on $[0, 1]$. Therefore, the conditional $t$ distribution cannot be rejected.

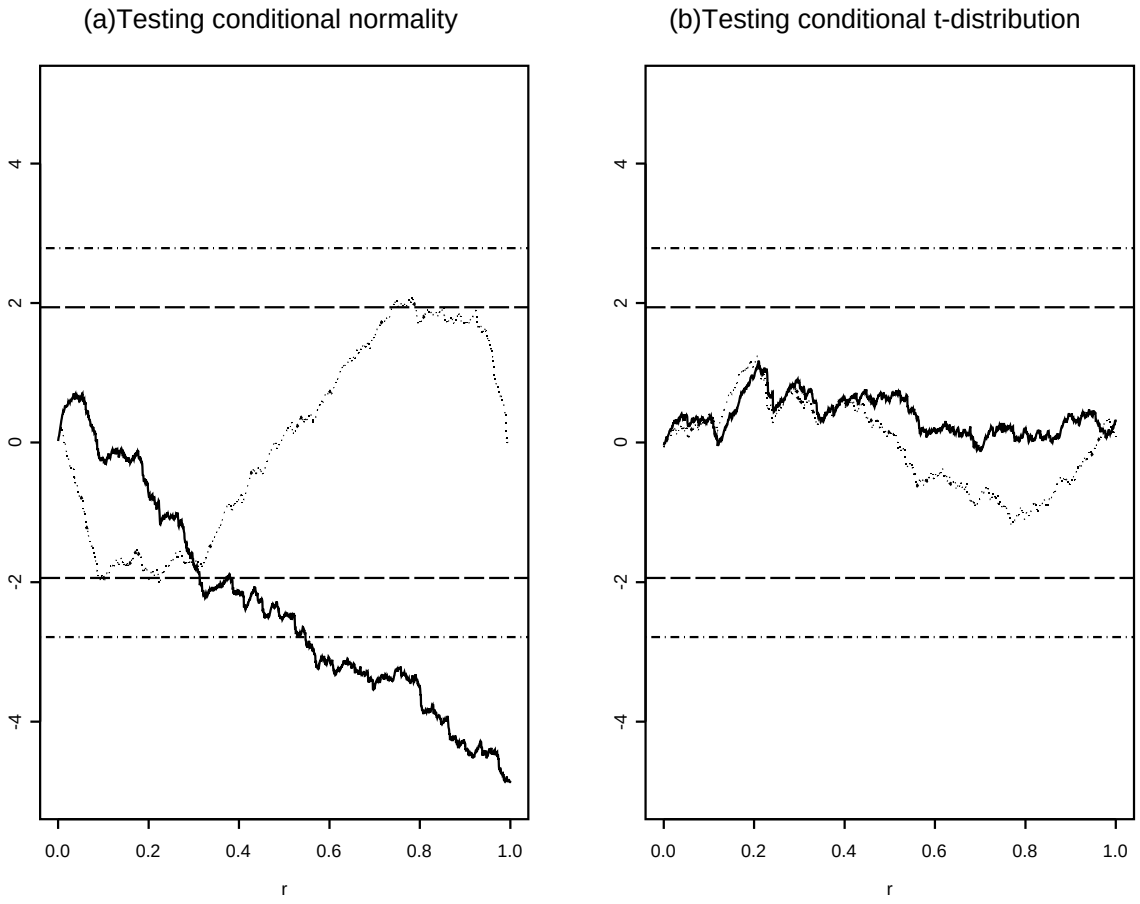


Figure 2: Testing the bivariate distribution of monthly log returns for IBM stock and the SP500 index fitted to a GARCH (1,1) process. The solid curve is the transformed process $\widehat{W}_{2n}$, and the dotted curve is the original process $\widehat{V}_{2n}$. The dashed horizonal lines give the 90 percent confidence band. And the dot-dashed lines give the 99 percent confidence band. Bivariate normality is rejected as $\widehat{W}_{2n}$ meanders outside the confidence band (a). While bivariate t-distribution cannot be rejected as $\widehat{W}_{2n}$ stays inside the confidence band (b).

It is well known that conditional normal GARCH model can generate heavy tail distributions. This test shows that the heavy tailedness generated by GARCH effect alone is not enough. We need a heavy tail conditional distribution (like $t$) combined with the GARCH

effect to capture the heavy tails of financial data.

# 8 Conclusion

This paper considers testing multivariate distributions, with a focus on multivariate normal distributions and multivariate $t$ distributions. Using Khmaladze's martingale transformation, we construct an asymptotically distribution free test. We show that the K-transformation takes a very simple form for testing multivariate normal and multivariate $t$ distribution. The method is applicable for vector time series models, including vector autoregressive and vector GARCH processes. We apply the method to testing multivariate conditional normality and multivariate conditional t distribution for some financial data. The empirical results have useful implications on computing the value at risk (VaR) of portfolio returns.

# References


Andrews,D. (1997), "A Conditional Kolmogorov Test," *Econometrica*, 65, 1097–1128.

Bai, J. (2003), "Testing Parametric Conditional Distributions of Dynamic Models," *Review of Economics and Statistics*, 85, 531-549.

Bontemps, C. and Meddahi, N. (2002), Testing normality: A GMM approach, Technical Report 2002s-63, CIRANO, Montral, Canada.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, 31, 307-327.

Bollerslev, T. (1987), "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *Review of Economics and Statistics*, 69, 542-547.

Bowman, A. W., and Foster, P. J. (1993), "Adaptive smoothing and density-based tests of multivariate normality," *Journal of American Statistical Association*, 88, 529-537.

Brown, R.L., Durbin J., and Evans J.M. (1975), "Techniques for Testing the Constancy of Regression Relationships Over Time," *Journal of Royal Statistical Society*, Series B, 37,149-192.

Durbin, J., (1973), "Weak Convergence of Sample Distribution Functions When Parameters Are Estimated," *Annals of Statistics* 1, 279- 290.

Engle, R.F., and Gonzela-Rivera, G. (1991), "Semiparametric ARCH Models," *Journal of Business and Economic Statistics*, 9, 345-359.

Epps, T., and Pulley, L. (1983), "A Test for Normality Based on the Empirical Characteristic Function," *Biometrika,* 70, 723–726.

Jarque, C.M. and A.K. Bera (1980), "Efficient tests for normality, homoscedasticity and serial independence of regression residuals," *Economics Letters* 6, 255-259.

Jarque, C.M. and A.K. Bera (1987), "A test for normality of observations and regression residuals," *International Statistical Review*, 55, 163-172.

Khmaladze, E.V. (1981), "Martingale Approach in the Theory of Goodness-of-Fit Tests," *Theory of Probability and Its Applications*, XXVI, 240-257

Khmaladze, E.V. (1988), "An Innovation Approach to Goodness-of-Fit Tests," *The Annals of Statistics*, 16, 1503-1516

Kilian, L. and Demiroglu, U. (2000), "Residual-based tests for normality in autoregressions: Asymptotic theory and simulation evidence," *Journal of Business and Economic Statistics* 18, 40-50.

Kotz, S., and Nadarajah S. (2004), *Multivariate t Distributions and Their Applications*, Cambridge, UK ; New York : Cambridge University Press.

Lutkepohl, H. and Theilen, B. (1991), "Measures of multivariate skewness and kurtosis for tests of nonnormality," *Statistical Papers* 32, 179-193.

Mardia K.V. (1970), "Measures of Multivariate Skewness and Kurtosis with Applications," *Biometrika* 57, 519-530.

Mecklin, C.J., and Mundfrom, D.J. (2004), "An Appraisal and Bibliography of Tests for Multivariate Normality," *International Statistical Review*, 72, 123-138.

Richardson, M. and Smith, T. (1993), "A test for multivariate normality in stock returns," *Journal of Business* 66, 295-321.

Rosenblatt, M. (1952), "Remarks on a Multivariate Transformation," *Annals of Mathematical Statistics*, 23, 470-472

Tsay, R.S. (2002), *Analysis of Financial Time Series*, Wiley New York, 303-392

Zellner, A. (1971), *An introduction to Bayesian inference in econometrics*, J. Wiley New York, 366-389

# Appendix

Proof of Theorem 3.1. According to Theorem 1 in Bai(2002),

$$\widehat{V}_{2n}(r) = V_{2n}(r) - \overline{g}_1(r)'\sqrt{2n}(\widehat{\theta} - \theta) - \overline{g}_2(r)'\sqrt{2n}(\widehat{\theta} - \theta) + o_p(1)$$

where $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_{21}, \sigma_2)$, and

$$\overline{g}_1(r) = \text{plim}\frac{1}{2n}\sum_{i=1}^{n}\frac{\partial F_{i1}}{\partial\theta}(x|\theta)\Big|_{x=F_{i1}^{-1}(r|\theta)}$$

$$\overline{g}_2(r) = \text{plim}\frac{1}{2n}\sum_{i=1}^{n}\frac{\partial F_{i2}}{\partial\theta}(x|\theta)\Big|_{x=F_{i2}^{-1}(r|\theta)}$$

and

$$F_{i1}(x|\theta) = \Phi\Big(\frac{x-\mu_1}{\sigma_1}\Big), \quad F_{i2}(x|\theta) = \Phi\Big(\frac{x-\mu_{2|1,i}}{\sigma_{2|1}}\Big).$$

Note that $F_{i1}(x)$ does not depend on $i$ and $F_{2i}$ does. From

$$\frac{\partial F_{i1}(x)}{\partial\mu_1} = -\frac{1}{\sigma_1}\phi\Big(\frac{x-\mu_1}{\sigma_1}\Big), \quad \frac{\partial F_{i1}(x)}{\partial\sigma_1} = -\frac{1}{\sigma_1}\phi\Big(\frac{x-\mu_1}{\sigma_1}\Big)\Big(\frac{x-\mu_1}{\sigma_1}\Big)$$

and evaluating the preceding derivative at $x = F_{i1}^{-1}(x)$, or equivalently at $(x-\mu_1)/\sigma_1 = \Phi^{-1}(r)$, we obtain immediately that

$$\frac{\partial F_{i1}(x)}{\partial\mu_1}\Big|_{x=F_{i1}^{-1}(r)} = -\frac{1}{\sigma_1}\phi(\Phi^{-1}(r)), \quad \frac{\partial F_{i1}(x)}{\partial\sigma_1}\Big|_{x=F_{i1}^{-1}(r)} = -\frac{1}{\sigma_1}\phi(\Phi^{-1}(r))\Phi^{-1}(r).$$

Thus

$$\overline{g}_1(r)'\sqrt{2n}(\widehat{\theta}-\theta) = -\frac{1}{2}\Big[-\frac{1}{\sigma_1}\phi(\Phi^{-1}(r))\sqrt{2n}(\widehat{\mu}_1-\mu_1) - \frac{1}{\sigma_1}\phi(\Phi^{-1}(r))\Phi^{-1}(r)\sqrt{2n}(\widehat{\sigma}_1-\sigma_1)\Big]$$

Note that $(\widehat{\sigma}_1-\sigma_1)/\sigma_1 = (1/2)(\widehat{\sigma}_1^2-\sigma_1^2)/\sigma_1^2 + o_p(1)$. This leads to

$$\overline{g}_1(r)'\sqrt{2n}(\widehat{\theta}-\theta) = -\frac{1}{\sqrt{2}}\Big[-\frac{1}{\sigma_1}\phi(\Phi^{-1}(r))\sqrt{n}(\widehat{\mu}_1-\mu_1) - \frac{1}{2\sigma_1^2}\phi(\Phi^{-1}(r))\Phi^{-1}(r)\sqrt{n}(\widehat{\sigma}_1^2-\sigma_1^2)\Big].$$

Next,

$$\begin{aligned}\frac{\partial F_{i2}(x)}{\partial\theta} &= -\frac{1}{\sigma_{2|1}}\phi\Big(\frac{x-\mu_{2|1,i}}{\sigma_{2|1}}\Big)\Big(\frac{\partial u_{2|1,i}}{\partial\theta}\Big)\\ &\quad -\frac{1}{\sigma_{2|1}}\phi\Big(\frac{x-\mu_{2|1,i}}{\sigma_{2|1}}\Big)\Big(\frac{x-\mu_{2|1,i}}{\sigma_{2|1}}\Big)\Big(\frac{\partial\sigma_{2|1}}{\partial\theta}\Big)\end{aligned}$$

Evaluate the preceding derivatives at $x = F_{i2}^{-1}(r|\theta)$, or equivalently at $(x-\mu_{2|1,i})/\sigma_{2|1} = \Phi^{-1}(r)$, we obtain

$$\frac{\partial F_{i2}(x)}{\partial\theta}\Big|_{x=F_{i2}^{-1}(r|\theta)} = -\frac{1}{\sigma_{2|1}}\phi(\Phi^{-1}(r))\Big(\frac{\partial u_{2|1,i}}{\partial\theta}\Big) - \frac{1}{\sigma_{2|1}}\phi(\Phi^{-1}(r))\Phi^{-1}(r)\Big(\frac{\partial\sigma_{2|1}}{\partial\theta}\Big).$$

The last expression does not depend on $i$. Thus

$$\begin{aligned}\overline{g}_2(r)'\sqrt{2n}(\widehat{\theta}-\theta) &= -\phi(\Phi^{-1}(r))\Big[\frac{1}{\sigma_{2|1}}\frac{1}{2n}\sum_{i=1}^{n}\Big(\frac{\partial u_{2|1,i}}{\partial\theta}\Big)\sqrt{2n}(\widehat{\theta}-\theta)\Big]\\ &\quad -\phi(\Phi^{-1}(r))\Phi^{-1}(r)\Big[\frac{1}{2\sigma_{2|1}}\Big(\frac{\partial\sigma_{2|1}}{\partial\theta}\Big)\sqrt{2n}(\widehat{\theta}-\theta)\Big]\end{aligned}$$

However, it is easy to show that up to an $o_p(1)$ term

$$\frac{1}{2n}\sum_{i=1}^{n}\Big(\frac{\partial u_{2|1,i}}{\partial\theta}\Big)\sqrt{2n}(\widehat{\theta}-\theta) = \frac{1}{\sqrt{2n}}\sum_{i=1}^{n}(\widehat{\mu}_{2|1,i}-\mu_{2|1,i}).$$

In fact, the left hand side is a Taylor expansion of the right hand side, which is a more compact notation. Similarly, up to an $o_p(1)$ term, we can write

$$\frac{1}{2\sigma_{2|1}}\Big(\frac{\partial\sigma_{2|1}}{\partial\theta}\Big)\sqrt{2n}(\widehat{\theta}-\theta) = \frac{1}{2\sigma_{2|1}}\sqrt{2n}(\widehat{\sigma}_{2|1}-\sigma_{2|1}).$$

The right hand side can be further written as, up to an $o_p(1)$ term

$$\frac{1}{4\sigma_{2|1}^2}\sqrt{2n}(\widehat{\sigma}_{2|1}^2-\sigma_{2|1}^2),$$

obtained by multiplying and dividing $(\widehat{\sigma}_{2|1} + \sigma_{2|1})$. In summary, we have

$$\begin{aligned}\overline{g}_2(r)'\sqrt{2n}(\widehat{\theta} - \theta) &= -\phi(\Phi^{-1}(r))\Big[\frac{1}{\sqrt{2n}}\sum_{i=1}^{n}(\widehat{\mu}_{2|1,i} - \mu_{2|1,i})\Big] \\ &\quad -\phi(\Phi^{-1}(r))\Phi^{-1}(r)\Big[\frac{1}{2\sqrt{2}\sigma_{2|1}^2}\sqrt{n}(\widehat{\sigma}_{2|1}^2 - \sigma_{2|1}^2)\Big] + o_p(1).\end{aligned}$$

Theorem 3.1 is obtained after combining terms with $\overline{g}_1(r)'\sqrt{2n}(\widehat{\theta} - \theta)$.

The proofs for all other theorems and corollaries are omitted because the idea is the same as in the proof of Theorem 3.1, only the technical details are different. The proofs are available from the authors.