The Correlated Gamma-Ratio Distribution in Model Evaluation and Selection

Panaretos, John and Psarakis, Stelios and Xekalaki, Evdokia and Karlis, Dimitris

8 September 2005

THE CORRELATED GAMMA-RATIO DISTIRIBUTION IN MODEL EVALUATION AND SELECTION

J. Panaretos, S. Psarakis, E. Xekalaki and D. Karlis
Department of Statistics
Athens University of Economics & Business
76, Patission St. 104 34
Athens, GREECE

1. INTRODUCTION

A problem common to researchers is that of selecting a model among several competing alternatives. Model selection procedures have been an attractive field for research and various criteria have been proposed (see, for example, Weisberg, 1985, Linhart and Zuchini, 1986, Burnham and Anderson, 1998 and Cook and Weisberg, 1999), especially for nested models. In assessing the quality of a fitted model the researcher must keep in mind two different and, in many cases competing, issues: the goodness of fit and the usefulness of the model for prediction.

In trying to decide which of two competing models to use one must discern between two lines of thought. The first pertains to viewing the problem as a "best-fitting" model determination problem in the sense that a model is sought which is closest to the observed data. This, being the most common approach in regression model selection, considers some measure of the adequacy of the model to describe existing observations for constructing a model selection procedure. The second approach is entirely different from the first in that it takes account of the predictive adequacy of a model and aims at comparing it with that of an alternative model, thus leading to a comparative model evaluation test. (In the case of nested models this would correspond to a model specification test in the sense of MacKinnon (1983) and Royston and Thompson (1995)).

While the statistical literature abounds in model selection procedures based on the descriptive ability of the competing models, it appears to be rather scant in compative evaluation tests based on the predictive ability of the models despite the fact that evaluating the forecasting potential of a model before this can be used for planning and decision making is of great importance. Some predictive approaches have been considered in various Bayesian contexts (see, for example, Gelfand and Dey, 1994, Laud and Ibrahim, 1995, Ibrahim and Laud, 1996). From a classical viewpoint, however, assessing the predictive abilities of contending models and using them for comparative evaluation appears to be a rather ignored issue. Such approaches have been considered by Geisser (1975), Snee (1977), Butler and Rothman (1980), Xekalaki and Katti (1984), Psarakis (1993), West and Cho (1995), West (1996), and Greenberg and Parks (1997), among others. The procedures proposed are based on some criteria which are evaluated for all the competing models. The "selection" of the "best" model is based on optimizing the appropriate criterion.

Xekalaki and Katti (1984) introduced an evaluation scheme of a sequential nature, based on the idea of scoring rules for rating the predictive behaviour of competing models in which the researcher's subjectivity plays an important role. Its effect is reflected through the rules according to which the performance of the model is scored and rated.

In this paper, an evaluation method is suggested which is again based on the use of a scoring rule but is free of the element of subjectivity. A scoring rule is suggested to rate the behaviour of a linear forecasting model for each of a series of n points in time. A final rating which embodies the step by step scores is then used as a statistic for testing the predictive adequacy of the model. This leads to a procedure for selecting between non-nested models, based on testing the hypothesis that the two models have the same forecasting potential. This procedure is somewhat different from the existing methods in the sense that it aims at selecting one of two competing models on the basis of their predictive performance and not of their descriptive ability. Further, it is shown that the theoretical setup of the new procedure, appropriately modified, can form the framework for the development of a model selection procedure based on the comparison of the descriptive potential of the competing models.

The plan of the paper is as follows: In section 2, a statistic is proposed for comparing the predictive ability of two competing models. It is shown that its distribution is a generalized form of the F distribution which is obtained as a limiting case of the former. Properties of the derived distribution are also examined. Section 3 considers this statistic for constructing a test of the hypothesis that two models are equivalent in their predictive ability thus introducing a procedure for model selection on the basis of the predictive abilities of the competing models. The procedure is illustrated by a real data application. Simulation results concerning the ability of the procedure to identify the model with the best predictive ability are also given. Section 4 proposes the use of a statistic of a similar nature for testing whether two models are equivalent in their descriptive ability leading to a goodness of fit model selection procedure for non-nested models. Simulation results reflecting the behaviour of the new procedure are reported.

Section 5 examines the robustness of the procedure when using an estimate of the correlation between the prediction errors or the residuals of the two models. Finally, concluding remarks and a short discussion are given in section 6.

2. THE CORRELATED GAMMA RATIO DISTRIBUTION

2.1 RATING THE PREDICTIVE PERFORMANCE OF REGRESSION MODELS

Consider the linear model

$$\mathbf{Y}_t = \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \quad t = 0, 1, 2, ...$$

where $\mathbf{Y}_t$ is an $\ell_t \times 1$ vector of observations on the dependent random variable, $\mathbf{X}_t$ is an $\ell_t \times m$ matrix of known coefficients $\left(\ell_0 > m, \left|\mathbf{X}_t'\mathbf{X}_t\right| \neq 0\right)$, $\boldsymbol{\beta}$ is an m×1 vector of regression coefficients and $\boldsymbol{\varepsilon}_t$ is an $\ell_t \times 1$ vector of normal error random variables with $E(\boldsymbol{\varepsilon}_t)=0$ and $V(\boldsymbol{\varepsilon}_t)=\sigma^2 I_t$. Here I_t is the $\ell_t \times \ell_t$ identity matrix. Therefore, a prediction for the value of the dependent random variable for time t+1 will be given by the statistic

$$\hat{Y}^0_{t+1} = \mathbf{X}^{0'}_{t+1} \hat{\boldsymbol{\beta}}_t,$$

where $\hat{\boldsymbol{\beta}}_t = (\mathbf{X'}_t \mathbf{X}_t)^{-1} \mathbf{X'}_t \mathbf{Y}_t$ is the least squares estimator of $\boldsymbol{\beta}$ at time t and $\mathbf{X}^0_{t+1}$ is an $m \times 1$ vector of values of the regressors at time $t+1, t = 0, 1, 2, ...$ Obviously,

$$\mathbf{X}_{t+1} = \begin{bmatrix} \mathbf{X}_t \\ \mathbf{X}^{0'}_{t+1} \end{bmatrix} \text{ and } \mathbf{Y}_{t+1} = \begin{bmatrix} \mathbf{Y}_t \\ Y^0_{t+1} \end{bmatrix}$$

are of dimension $\ell_{t+1} \times m$ and $\ell_{t+1} \times 1$ respectively, where $\ell_{t+1} = \ell_t + 1, \quad t = 0, 1, 2, ...$

The predictive behaviour of the model would naturally be evaluated by a measure that would be based on a statistic reflecting the degree of agreement of the observed actual value $\hat{Y}^0_{t+1}$ to the predicted value $\hat{Y}^0_{t+1}$. Such a statistic may be the statistic $\left|r_{t+1}\right|$, where

$$r_{t+1} = \frac{\hat{Y}^0_{t+1} - Y^0_{t+1}}{S_t\sqrt{\left(1 + \mathbf{X}^{0'}_{t+1}\left(\mathbf{X}'_t\mathbf{X}_t\right)^{-1}\mathbf{X}^0_{t+1}\right)}}, \quad t = 0, 1, \ldots \tag{1}$$

Obviously, $|r_{t+1}|$ is merely an estimate of the standardized distance between the predicted and the observed value of the dependent random variable when σ^2 is estimated on the basis of the preceding ℓ_t observations available at time t. S_t^2 is given by

$$S_t^2 = \mathbf{e'}_t\,\mathbf{e}_t/(\ell_t - m)$$

where

$$\mathbf{e}_t = \mathbf{Y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}_t$$

$$\text{i.e.,}\quad S_t^2 = \frac{(\mathbf{Y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}_t)'(\mathbf{Y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}_t)}{(\ell_t - m)},\quad t = 0, 1, 2, \ldots,$$

So, a score based on $|r_{t+1}|$ can provide a measure of the predictive adequacy of the model for each of a series of n points in time. Then, as a final rating of the model one can consider the average of these scores, or any other summary statistic that can be regarded as reflecting the forecasting potential of the model.

In the sequel, we consider using r_i^2 as a scoring rule to rate the performance of the model at time t for a series of n points in time, (t =1, 2, ..., n) and we define

$$R_n = \sum_{t=1}^{n} r_t^2 \Big/ n \tag{2}$$

the average of the squared recursive residuals, to be the final rating of the model.

It has been shown (Brown *et al*., 1975, Kendall *et al.,* 1983) that if $\boldsymbol{\varepsilon}_t$ is a vector of normal error variables with $E(\boldsymbol{\varepsilon}_t)=0$ and $V(\boldsymbol{\varepsilon}_t)=\sigma^2\,\mathbf{I}_t$, the quantities

$$w_{t+1} = \frac{\hat{Y}^0_{t+1} - Y^0_{t+1}}{\sqrt{1 + \mathbf{X}^{0'}_{t+1}\left(\mathbf{X}'_t\mathbf{X}_t\right)^{-1}\mathbf{X}^0_{t+1}}},\quad t = 0, 1, 2, \ldots$$

are independently and identically distributed normal variables with mean 0 and variance σ^2. Then, according to Kotlarski's (1966) characterization of the normal distribution by the t distribution, the quantities $r_{t+1} = w_{t+1}/s_t$, t = 0, 1, 2, ... constitute a sequence of independent t variables with $\ell_t - m$ degrees of freedom, t = 0, 1, 2, ... Hence, by the assumptions of the model considered and for large ℓ_0, the variables r_{t+1}, t = 0, 1, 2, ... constitute a sequence of approximately standard normal variables which are mutually independent. This implies that

$$nR_n = \sum_{t=1}^{n} r_t^2$$

is a chi-square variable with n degrees of freedom. (Psarakis and Panaretos, 1990).

Consider now A and B to be two competing linear models that have been used for prediction purposes for a number n_A and n_B of years, respectively. Then, a decision on whether model A is equivalent to model B would naturally be based on the ratio of the average scores of the two models as given by the statistic

$$R_{n_A,n_B} = \frac{R_{n_A}^A}{R_{n_B}^B} \tag{3}$$

where $R_{n_A}^A$, $R_{n_B}^B$, are given by (2) for n = n_A and n = n_B and refer to model A and model B, respectively.

For large ℓ_0^A, ℓ_0^B the distribution of the rating ratio statistic R_{n_A,n_B} can be approximated by that of an F variable with n_A and n_B degrees of freedom whenever the ratings of the two models are independent. Hence, values of R_{n_A,n_B} in the right tail of the F distribution with n_A and n_B degrees of freedom will indicate a higher performance by model A. However, under the model selection setting, the assumption of independence does not seem to be satisfied. Determining the exact distribution of R_{n_A,n_B} in the case of dependent ratings would, therefore, be desirable as in practice data on ratings are often matched. (In the latter case, $n_A = n_B = n$.) As shown in the sequel, this is a generalized form of the F distribution.

2.2 DETERMINING THE DISTRIBUTION OF THE RATING RATIO STATISTIC

Kotlarski (1964) has shown that, under certain conditions, the quotient X/Y, where X,Y are positive valued random variables not necessarily independent, follows the F distribution. According to him, a necessary and sufficient condition for the ratio of two variables to follow an F distribution can be established through the form of the Mellin transform of their joint distribution. In particular, Kotlarski has shown that for a distribution function F(x, y) to belong to the set of joint distribution functions of two not necessarily independent positive valued random variables X and Y, whose quotient X/Y follows the F distribution with parameters p_1 and p_2, is necessary and sufficient that its Mellin transform

$$h(u,v) = \int_0^\infty \int_0^\infty x^u y^v \, dF(x,y)$$

satisfies the condition $h(u,-u) = \Gamma(p_1 + u)\Gamma(p_2 - u) / (\Gamma(p_1)\Gamma(p_2))$.

As shown in the Appendix, the Mellin transform of the joint distribution of R_n^A, R_n^B is

$$h(u,v) = \frac{(1-\rho^2)^{-k+u+v+2k}}{\Gamma(k)} \sum_{i=0}^{\infty} \frac{\rho^{2i}\Gamma(u+k+i)\Gamma(v+k+i)}{i!\Gamma(k+i)} \tag{4}$$

and does not satisfy Kotlarski's condition. Hence, the distribution of the rating ratio statistic $R_{n,n} = R_n^A / R_n^B$ cannot be of the F form. As shown in the sequel, it is distributed according to a more general form of distribution which leads to the F distribution as a limiting case.

To specify the distribution of the test statistic, consider the random variables $X_i = r_i^A$, $Y_i = r_i^B$, i =1, 2, ..., n obtained from relationship (1), for model A and model B, respectively. Each of the variables X_i, Y_i follows the standard normal distribution. The joint distribution is therefore the bivariate standard normal distribution with a correlation coefficient denoted by ρ. Under these conditions, the joint distribution of the random variables

$$X = \frac{\sum_{i=1}^{n} X_i^2}{n} = R_n^A \qquad \text{and} \qquad Y = \frac{\sum_{i=1}^{n} Y_i^2}{n} = R_n^B$$

is Kibble's (1941) bivariate Gamma distribution as defined by the probability density function

$$f(x,y) = \frac{\rho^{-(k-1)}}{\Gamma(k)(1-\rho)^2} (xy)^{\frac{k-1}{2}} e^{-\frac{x+y}{1-\rho^2}} I_{k-1}\left(\frac{2\rho\sqrt{xy}}{1-\rho^2}\right), \tag{5}$$

for k=n/2 . Here $I_k(x$ is the modified Bessel function of the first kind of order k given by

$$I_k(x) = \sum_{i=0}^{\infty}\left(\frac{x}{2}\right)^{k+2i} \frac{1}{\Gamma(i+1)\Gamma(i+k+1)}. \qquad (6)$$

(see Abramowitz and Stegun, 1974)

Substituting (6) in (5) we can rewrite the joint probability density function of Kibble's bivariate Gamma distribution as

$$f(x,y) = \frac{\rho^{-(k-1)}}{\Gamma(k)(1-\rho)^2} e^{-\frac{x+y}{1-\rho^2}} \sum_{i=0}^{\infty}\left(\frac{\rho}{1-\rho}\right)^{k+2i-1} \frac{x^{k+i-1} y^{k+i-1}}{\Gamma(i+1)\Gamma(i+k)}.$$

It is known that for any dependent random variables X, Y the distribution function of $Z = X/Y$ is given by

$$F_Z(z) = P(X/Y \le z) = \int_0^{\infty} P(X \le zy | Y = y)\ f_Y(y)\,dy,$$

where $F_U(\cdot)$ and $f_U(\cdot)$ denote the distribution function and the probability density function of a random variable U, respectively. Then, the density function of the quotient $Z = X/Y$ can be written as

$$f_Z(z) = \int_0^{\infty} f_{X|Y=y}(zy)\, f_Y(y)dy = \int_0^{\infty} \frac{f_{X,Y}(zy,y)}{f_Y(y)}\, y\, f_Y(y)\,dy$$

$$= \int_0^{\infty} y\, f_{X,Y}(zy,y)\,dy.$$

In our case, this leads to

$$f_{X/Y}(z) = \int_0^{\infty} y\, f_{x,y}(zy,y)\,dy$$

$$= \frac{1}{(1-\rho^2)^k \Gamma(k)} \sum_{i=0}^{\infty} \frac{\rho^{2i}}{(1-\rho^2)^{2i} i!\Gamma(i+k)} \int_0^{\infty} \exp\left(-\frac{zy+y}{1-\rho^2}\right) z^{k+1-i} y^{2(k+i)-1}\,dy$$

$$= \frac{z^{k-1}}{(1-\rho^2)^k \Gamma(k)} \sum_{i=0}^{\infty} \frac{\rho^{2i} z^i}{(1-\rho^2)^{2i} i!\Gamma(i+k)} \int_0^{\infty} \exp\left(-y\frac{z+1}{1-\rho^2}\right) y^{2(k+i)-1}\,dy.$$

In fact, the integral in the expression of $f_{X/Y}(z)$ is a gamma integral and hence one obtains

$$f_{X/Y}(z) = \frac{z^{k-1}}{\left(1-\rho^2\right)^k \Gamma(k)} \left(1+z\right)^{-2k} \sum_{i=0}^{\infty} \frac{\Gamma(2k+2i)}{\Gamma(i+k)} \left(\frac{\rho}{1+z}\right)^{2i} \frac{z^i}{i!} . \tag{7}$$

Furthermore, it holds that

$$\frac{\Gamma(2k+2i)}{\Gamma(k)\Gamma(i+k)} = \frac{\Gamma(2k+2i)\Gamma(2k)\Gamma(k)}{\Gamma(k)\Gamma(i+k)\Gamma(2k)\Gamma(k)} = \frac{(2k)_{(2i)}}{k_{(i)}} \left[B(k,k)\right]^{-1} =$$

$$= \left[B(k,k)\right]^{-1} \frac{2^{2i}\left(\frac{2k}{2}\right)_{(i)} \left(\frac{2k+1}{2}\right)_{(i)}}{k_{(i)}}$$

where $B(\alpha,\beta)$ is defined as

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

and

$$\alpha_{(mn)} = n^{nm} \left(\frac{\alpha}{n}\right)_{(m)} \left(\frac{\alpha+1}{n}\right)_{(m)} \cdots\cdots \left(\frac{\alpha+n-1}{n}\right)_{(m)}$$

with $\alpha_{(n)}$ denoting the ascending factorial.

Letting α=2k, m=i , n=2 we have

$$\frac{\Gamma(2k+2i)}{\Gamma(k)\Gamma(i+k)} = \frac{2^{2i}\left(\frac{2k+1}{2}\right)_{(i)}}{B(k,k)} .$$

Substituting in (6) we obtain

$$f_{X/Y}(z) = \left(1-\rho^2\right)^k \frac{z^{k-1}(1+z)^{-2k}}{B(k,k)} \sum_{i=0}^{\infty} \left(\frac{2k+1}{2}\right)_{(i)} \frac{\left[4\rho^2(z+1)^{-2}\right]^i z^i}{i!}$$

$$= \left(1-\rho^2\right)^k \frac{z^{k-1}(1+z)^{-2k}}{B(k,k)} \left[1 - 4\frac{\rho^2 z}{(z+1)^2}\right]^{-\frac{2k+1}{2}}$$

Therefore,

$$\boxed{f_{X/Y}(z) = \frac{\left(1-\rho^2\right)^k}{B(k,k)} z^{k-1}(1+z)^{-2k} \left[1 - \left[\frac{2\rho}{z+1}\right]^2 z\right]^{-\frac{2k+1}{2}}} . \tag{8}$$

The density function in (8) defines the distribution of the quotient X/Y when the joint distribution of (X,Y) is Kibble's bivariate gamma. In the sequel, we refer to this distribution as **the correlated gamma - ratio (CGR) distribution with parameters ρ and k.** This distribution does not appear to have been studied and used in the literature. In our search, we only found that Izawa (1965) was led to a reparameterized form of this distribution. In order to apply it later, we have derived its percentage points. A sample of them is provided in the Appendix for p=0.0(0.1)0.9.

One can see that in the case where X and Y are independent, whence ρ=0, the probability density function of the quotient X/Y takes the form

$$f_{X/Y}(z) = \frac{1}{B(k,k)} z^{k-1}(1+z)^{-2k}.$$

This is the probability density function of the Beta type II distribution and, because of the relationship of the F distribution with the Beta type II distribution, it is an F distribution with 2k and 2k degrees of freedom. In the next section properties of the CGR distribution will be examined.

2.3 PROPERTIES OF THE CORRELATED GAMMA RATIO DISTRIBUTION

The moments of the CGR distribution can be found using its relationship to Kibble's bivariate Gamma distribution and, in particular, through its Mellin transform as given by (4) since the r-th simple moment of the GCR distribution can be written as

$$E(Z^r) = E(X^r Y^{-r}) = h(r,-r). \tag{9}$$

Using this result, we obtain, after tedious algebraic manipulation

$$E(X/Y) = E(Z) = \frac{(1-\rho^2)+k-1}{k-1}, \quad k>1 \tag{10}$$

and

$$\mathrm{Var}(Z) = \frac{(5k-4)(1-\rho^2)^2 + 2(k-1)(k-2)(1-\rho^2)}{(k-1)^2(k-2)}, \quad k>2 .$$

(The proofs of these results can be found in the Appendix).

Note that if ρ=0, i.e. X and Y are uncorrelated, the expected value and the variance reduce to the expected value and the variance of a F distribution with 2k and 2k

degrees of freedom, respectively. The expected value does not exist for values of $k \leq 1$. Moreover, the expected value is always greater than 1 and tends to 1 as k tends to infinity.
Also, as k increases the variance of Z decreases. This shows that, for the CGR distribution, as the value of k increases the expected value tends to 1 and the variance tends to 0, i.e. the distribution is more concentrated towards the value 1. In general, the distribution tends to a degenerate distribution at 1, as k increases. The same is true when ρ^2 tends to 1 in absolute value. This should be expected since a unit correlation implies a perfect agreement between X and Y and thus a unit value for their ratio.

For small values of k, the CGR distribution is very skewed to the left and has a large coefficient of kurtosis. As the value of k increases, the distribution tends to become symmetric. The same is true whenever the value of the parameter ρ^2 is near 1. The skewness decreases with ρ^2. This can be seen in Figures 1 through to 3 in the Appendix where the probability density function of the CGR distribution is depicted for various values of ρ and k.

It is interesting to note that because of the symmetry of Kibble's bivariate Gamma distribution, if Z follows the CGR distribution, the same is true for 1/Z.

In the sequel, it is shown that a limiting case of the CGR distribution is the t distribution.

Let Z follow the CGR distribution with density function given by (3). Consider the variable

$$T = \frac{\rho}{\sqrt{1-\rho^2}} \frac{Z-1}{Z+1}.$$

Then,

$$F_T(t) = P(T \leq t) = P\left(Z \leq \frac{\rho + t\sqrt{1-\rho^2}}{\rho - t\sqrt{1-\rho^2}} \right) = F_Z\left(\frac{\rho + t\sqrt{1-\rho^2}}{\rho - t\sqrt{1-\rho^2}} \right),$$

where $-\frac{\rho}{\sqrt{1-\rho^2}} < t < \frac{\rho}{\sqrt{1-\rho^2}}$.

We have, therefore, for the probability density function of T that

$$f_T(t) = f_Z\left(\frac{\rho + t\sqrt{1-\rho^2}}{\rho - t\sqrt{1-\rho^2}}\right)\frac{2\rho\sqrt{1-\rho^2}}{\left(\rho - t\sqrt{1-\rho^2}\right)^2},$$

where $-\frac{\rho}{\sqrt{1-\rho^2}} < t < \frac{\rho}{\sqrt{1-\rho^2}}$.

This reduces to

$$f_T(t) = \frac{1}{\rho}\frac{2^{1-2k}}{B(k,k)}\left[1 - \left(\frac{\sqrt{1-\rho^2}}{\rho}t\right)^2\right]^{k-1}\left(1+t^2\right)^{-\frac{2k+1}{2}},$$

where $-\frac{\rho}{\sqrt{1-\rho^2}} < t < \frac{\rho}{\sqrt{1-\rho^2}}$.

Taking the limit as $\rho \to 1$ we obtain

$$\lim_{\rho\to 1} f_T(t) = \frac{2^{1-2k}}{B(k,k)}\left(1+t^2\right)^{-\frac{2k+1}{2}}, \quad -\infty < t < +\infty .$$

But, this is the probability density function of the t distribution. Hence it was shown that the t-distribution arises as a limiting case of the CGR distribution.

3. COMPARING THE PREDICTIVE ABILITY OF TWO LINEAR MODELS

Consider now the problem of comparing the predictive ability of two competing models, say A and B, that have been used for prediction purposes for n_A and n_B years, respectively. The dependent variable is denoted by Y. Denote by $\hat{Y}_{t+1}^{0\ A}$ and $\hat{Y}_{t+1}^{0\ B}$ the one step ahead predicted values of Y obtained by models A and B respectively at times $t = 0, 1, 2, ..., n$ on the basis of the first $\ell_t^A = \ell_0^A + t$ and $\ell_t^B = \ell_0^B + t$ observations, where ℓ_0^A, ℓ_0^B denote the numbers of observations used at time t=0 for estimating the regression coefficients of the models. Obviously $\ell_n^A \equiv \ell_0^A + n = n_A$ and $\ell_n^B \equiv \ell_0^B + n = n_B$. Let w_{t+1}^A, w_{t+1}^B denote the prediction errors corresponding to models A and B. As noted before, the quantities w_{t+1}^A and w_{t+1}^B, $t = 0, 1, 2, ..., n$ are independently and identically distributed normal random variables with means 0 and variances σ_A^2 and σ_B^2, respectively. Both of these quantities are linear combinations of the first ℓ_t^A and ℓ_t^B observations on Y, respectively which, under the null hypothesis of the equivalence of the models in their predictive ability, are independently and identically normally distributed variables. Therefore, under the null hypothesis, any linear combination of w_{t+1}^A and w_{t+1}^B, $t = 0, 1, ...$ will be a normal variable which in turn implies that the joint distribution of w_{t+1}^A/σ_A and w_{t+1}^A/σ_B is a bivariate standard normal distribution with some correlation coefficient denoted by ρ and so is the joint distribution of r_{t+1}^A and r_{t+1}^B for large ℓ_0^A and ℓ_0^B, respectively. Then, the statistics nR_n^A, nR_n^B jointly follow Kibble's bivariate gamma distribution and hence their ratio has the CGR distribution with parameters ρ and n/2 as defined by (5).

In the sequel, the performance of the new statistic is examined via simulation and a real data application is provided.

3.1 THE PERFORMANCE OF THE PROPOSED TEST

A small simulation experiment was carried out in order to examine the performance of the new criterion. Samples of given sample size n were simulated from an assumed model. The experiment concerned classical trend models in classical time series analysis. So, data were generated from a linear trend model and the null hypothesis that the predictive ability of the linear trend model is equivalent to that of some alternative models was tested. Two alternative models were considered: the exponential trend model $\ln Y = \alpha_1 + \beta_1 t + \varepsilon_1$ and the quadratic trend model $Y = \alpha_2 + \beta_2 t^2 + \varepsilon_2$, where ε_1, ε_2 denote the respective normal error terms. The values used for the coefficients were $\alpha_1 = 5$, $\beta_1 = 10$, while the variance for the error term was $\sigma^2 = 1$. For each sample size, a number *k* of observations were used for building the model before starting obtaining predictions. Several different values were used, corresponding to different proportions of values kept for prediction purposes. 5000 replications were used for each sample size.

The parameter ρ was estimated as the correlation coefficient of the standardized residuals of the predictions of the two competing models. In the next section the robustness of this approach is examined.

From the results reported in Table 1, one can see that the test performs very well for distinguishing the linear model from the polynomial model. For the exponential model, the performance of the test improves when more observations are used. This is a consequence of the fact that the exponential model can be very close to a linear one by appropriately selecting the parameters. However, when more data are collected, the curvature of the exponential model can produce significant differences. The entries of Table 1 are the mean p-values calculated from the replications considered and the proportions of times the equivalence hypothesis was rejected.

Table 1 about here

The test performs well when the models are non-nested. For nested models, the model with added components seldom performs worse than the simpler model. Moreover, since the added variables may have a coefficient close to 0, the differences can be very small, and almost non-detectable. In order to check this, a simulation was run on data generated from an exponential trend model with a moderate curvature. As was expected, in testing the exponential model against a linear one, the null hypothesis of equivalence was rejected in all cases. Further investigation is needed for determining the difference that it is detectable via the new procedure.

3.2 AN APPLICATION TO CROP-YIELD DATA

For the purpose of illustrating the procedure described in this section, a problem presented in Xekalaki and Katti (1984) is re-examined. The problem referred to the selection of a linear model among several competing ones that were used by the United States Department of Agriculture (USDA) to predict the corn yield for 10 Crop Reporting Districts (CRD 10, 20, …, 100), based on several sets of real data for the State of Iowa for the years 1956 to 1980. The competing models use information about the weather conditions (e.g. temperature, rainfall, etc) for the previous time periods as well general trend factors for predicting the crop yield. (Linardis (1998), presents a detailed description of the models).

The aim of the application is to compare the predictability of these models for every district, using the CGR distribution. Table 2 gives the estimated correlations between the standardized residuals of the predictions for the two models used by the USDA, for each crop reporting district.

To compare the two crop yield models we need to test a hypothesis of the form:

H_0: model A is equivalent to model B,

versus

H_1: the two models are not equivalent.

Rejection of the null hypothesis indicates that one of the models performs better. In performing the test one can employ a one sided alternative, in a manner similar to that used when testing for equality of variances via the F-test. However, when rejecting the null hypothesis, the value of the test statistic can indicate which of the models is the "best". For a two sided alternative, the test statistic must be inverted if its value is less than one.

The values of the ratios

$$R_{n,n} = \frac{nR_n^A}{nR_n^B} = \frac{\sum_{t=1}^{n}(r_t^A)^2}{\sum_{t=1}^{n}(r_t^B)^2} \quad \text{or} \quad R_{n,n}^{-1} = \frac{nR_n^B}{nR_n^A} = \frac{\sum_{t=1}^{n}(r_t^B)^2}{\sum_{t=1}^{n}(r_t^A)^2}$$

for the data, and the results of the test procedure are reported in Table 2. One can see from Table 3 that, for six districts, the models are equivalent. Model A performs better in 3 cases while only in one case model B is superior.

Table 2 about here

The method presented in this section to compare the predictive ability of two linear models differs from those based on the notion of nesting in that the two models may have completely different sets of independent variables. Moreover, the selection of the 'best' model takes into account random variability and is not deterministic, as for example in the case where only the value of the statistic would be used.

4. MODEL SELECTION FOR NON-NESTED MODELS IN THE CLASSICAL REGRESSION FRAMEWORK

4.1 THE PROBLEM OF MODEL SELECTION BASED ON THE DESCRIPTIVE ABILITY

Let us now consider to the possibility of developing a procedure leading to the selection of a model as a result of comparing the descriptive potentials of two competing models.

The aim is to be able to select a model that is closest to the mechanism that produced the observed data among several competing models that provide a satisfactory fit. The literature abounds in model selection criteria constructed on the basis of goodness of fit considerations. A common element to all these procedures is that they are not based on probabilistic arguments, i.e. model selection is based on deterministic criteria that ignore random fluctuations. (See, for example, the review by Hocking, 1980). The equivalence of two models has not been tested, except through asymptotic results based on the change in likelihood. This fact can be seen in the classical regression model, where the addition of a new variable improves the coefficient of determination even in the case of a totally inadequate variable. Whether this improvement indicates a statistically significant contribution of the new variable to the coefficient of determination is not known.

Comparing nested models has attracted much interest and a variety of procedures for selecting the most appropriate model have been proposed in the literature. The case of non-nested models has received much less attention in the statistical literature. In fact, econometricians appear to be more concerned with non-nested models, since the majority of related publications has appeared in econometric journals. The reason is that in econometrics a difference in model assumptions can lead to entirely different models (e.g. a linear trend versus an exponential trend).

Two models are non-nested when they have separate parametric families and none can be obtained from the other as a limiting case. Hoel (1947), initiated the discussion about non-nested models. For detailed reviews the interested reader is referred to Mc Aleer (1987, 1995), White (1983). Mc Aleer (1995) also provides an extensive literature on econometric applications of non-nested models classified with respect to the subject, the type of models considered and the methodology used.

A first obvious procedure would be to set up a comprehensive model that includes both of the competing models as special cases and subsequently apply standard loglikelihood tests. If the likelihood functions of two competing models are denoted by L_0 and L_1 respectively, the enlarged model can be of the form $L_0^{\theta} L_1^{1-\theta}$ or of the form $\theta L_0 + (1-\theta) L_1$. The former was suggested by Cox (1962) and Atkinson (1970), while the latter which considers a mixture setting was discussed by Davidson and MacKinnon (1981) and McKinnon (1983). Both models have the major advantage of leading to an estimation procedure as well to significance testing. Estimation is usually not a problem since the mixture setting is quite useful. But, in many cases, the sample evidence may indicate that both models are as plausible or they may call for the rejection of both models failing thus to lead to some conclusive result. The idea has been further exploited by Royston and Thompson (1995).

Alternative procedures are based on the Bayesian paradigm and use the posterior odds for selecting the model that is better supported by the data. The dependence of the results on the chosen prior distribution is the main disadvantage of such procedures.

Cox (1961) proposed a test statistic based on the Neyman-Pearson likelihood ratio principle, further developed for regression models by Pesaran (1974). The approach was further investigated by Pace and Salvan (1990). Other test procedures are described by Mc Aleer (1995) and include tests constructed for specific models and information criteria.

Usually, the hypotheses in testing for non-nested regression models take the form

$$H_0 : Y = Xb_0 + u_0, \quad u_0 \sim N(0, \sigma_0^2 I)$$

$$H_1 : Y = Zb_1 + u_1, \quad u_1 \sim N(0, \sigma_1^2 I).$$

Note that in comparing non-linear regression models, two models are considered to be non-nested if at least one of the columns of each of the regressor matrices cannot be expressed as a linear combination of the columns of the other.

4.2 A NEW TEST STATISTIC FOR COMPARING THE DESCRIPTIVE ABILITY OF TWO MODELS

Consider two competing models, say A and B, and denote the dependent random variable by Y. The problem of comparing two models A and B on the basis of their descriptive ability is equivalent to that of comparing the goodness of fit of two normal distributions, namely $Y \sim N(\mu_A, \sigma_A^2)$ or $Y \sim N(\mu_B, \sigma_B^2)$. This requires a null hypothesis formulation different from the one that is usually considered by many authors. In particular, one would want to test the null hypothesis that the two models are equivalent against the alternative that they are not equivalent. By contrast, the usual null hypothesis in model selection for non-nested models has been that one model is valid against an alternative hypothesis that another model is valid. Hence, the testing hypothesis procedure leads to the selection of one of the models.

The mean of the dependent variable can be expressed as usual, as a linear function of the independent variables. The purpose is to examine whether the two models fit the data "equally satisfactorily". Note that the models are not assumed to be non-nested and, therefore, models with entirely different regressors can be considered.

Suppose that the two competing models A and B have been used for n_A and n_B years respectively. The dependent variable is again denoted by Y. According to whether model A or model B is used and for $i = 1, 2, ..., n$, let $\widehat{Y}_i^A$ and $\widehat{Y}_i^B$ denote the fitted values using the first $\ell_t^A = \ell_0^A + t - 1$ and $\ell_t^B = \ell_0^B + t - 1$ observations, where ℓ_0^A, ℓ_0^B denote the numbers of observations used at time t=1 for estimating the regression coefficients of the models. Obviously, $\ell_n^A \equiv \ell_0^A + n = n_A$ and $\ell_n^B \equiv \ell_0^B + n = n_B$. Let $e_i^A = Y_i - \widehat{Y}_i^A$ and $e_i^B = Y_i - \widehat{Y}_i^B$, $i = 1, 2, ..., n$ denote the corresponding residuals. By the assumptions of the regression models considered, these residuals follow normal distributions with zero means and variances σ_A^2 and σ_B^2, respectively. All of these residuals are linear combinations of the first ℓ_t^A and ℓ_t^B observations on Y which, under the hypothesis of the equivalence of the two models in their descriptive ability, are independently and identically distributed normal variables with zero means and variances equal to, say, σ^2. Therefore, any linear combination of e_i^A and e_i^B will be a normal variable. Hence, under the null hypothesis each of the pairs $\left(e_i^{\,A}/\sigma,\ e_i^{\,B}/\sigma\right)$ follows a bivariate standard normal distribution with a correlation parameter equal to, say, ρ. Thus, the quantities

$$T_A = \frac{\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i^{\,A}\right)^2}{n\sigma^2} \quad \text{and} \quad T_B = \frac{\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i^{\,B}\right)^2}{n\sigma^2}$$

follow jointly Kibble's bivariate gamma distribution. Their ratio, given by

$$T = \frac{T_A}{T_B} = \frac{\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i^{\,A}\right)^2}{\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i^{\,B}\right)^2},$$

follows the CGR distribution with parameters ρ and n/2 as shown in section 2.

One can see that the statistic T is the ratio of the sums of the squared residuals corresponding to the two models. Consider the null hypothesis that the two models are equivalent versus the alternative hypothesis that model A provides a better fit. Obviously, values of the statistic at the right tail of the GCR distribution will call for the rejection of the null hypothesis.

It would be interesting to examine the relationship of the procedure described above to the classical ones based on the coefficient of determination defined by

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}.$$

As is well known (e.g. Draper and Smith, 1981), the coefficient of determination follows a Beta type I distribution. Therefore, the quantity

$$\varphi = 1 - R^2 = \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}$$

also follows a Beta distribution with interchanged parameters. Note that the statistic φ is always less than 1. The pertinence of the statistic φ to the statistic T defined above in the context of the problem of comparing the linear model against the constant model becomes obvious taking into account the following points. The beta form of the distribution of the statistic φ is deduced on the basis of the independence of the statistics $\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$ and $\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2$ and the fact that each of them is gamma distributed. (For a proof of the independence the reader is referred to Rao and Toutenbourg, 1997). It is also known that the Beta type II distribution arises as the distribution of the random variable Y=X/(1-X), when X follows the Beta type I distribution. Hence, the statistic T is of a nature similar to that of the coefficient of determination but with the additional advantage that it takes account of the dependence between the competing models. In the paper by Miller (1984) and the discussion that followed, the fact that the F-test is used for selecting the variables to enter or to remove in stepwise algorithms was strongly criticized. The reason was that these tests (known as F-change tests) assume independence, which seems to be a rather arbitrary and unlikely to be fulfilled assumption. The test proposed in this paper clearly deals with the dependence between the two statistics.

As is known, if one model is nested in another, the coefficient of determination of the extended model is always greater than the coefficient of determination of the initial one. Therefore, in the case of two competing models A and B, the ratio of the φ statistics associated with the two models leads to the statistic T defined above, since the denominators common to both models vanish. Thus, the proposed statistic T measures the relative improvement of the complement of the coefficient of determination for the two competing models.

It becomes obvious from the above that the CGR distribution takes into consideration the dependence between the two models, and especially of the dependence between the residuals of the models for the same value of the response variable.

4.3 A SIMULATION EXPERIMENT

A simulation experiment was carried out in order to study the behaviour of the proposed procedure. Samples of the specific sample sizes were generated from a simple model of the form $Y = \alpha_1 X + \beta_1 + \varepsilon_1$, where ε_1 denotes the error term. The values considered for the coefficients were $\alpha_1 = 5, \beta_1 = 10$, while that of the variance of the error term was $\sigma^2 = 1$. The explanatory random variable X was generated from a N(5,3) distribution. For each sample size, half of the observations were kept for building the model and the rest for prediction purposes. 5000 replications were used for each sample size.

Three alternative models were considered, namely $Y = \alpha_2 \frac{1}{X} + \beta_2 + \varepsilon_2$, $Y = \alpha_3 X^2 + \beta_3 + \varepsilon_3$ and $Y = \alpha_4 X_2 + \beta_4 + \varepsilon_4$, where ε_i, denotes the error term of model i and $\varepsilon_i \sim N(0, \sigma^2)$, i =1, ..., 4. One can see that the models are not nested. In addition, the third model contains a different variable as an explanatory variable. In order to get a better insight into the ability of the proposed procedure to select the true model, the variable X_2 was set equal to the true value of X plus a noise term generated from a normal distribution with a zero mean and variance equal to 0.5. In this case, the resulting variable, X_2, is quite close to X. The results reported in Table 3, reveal that as the sample size increases the use of the CGR distribution provides a powerful tool for model selection even in the case where the alternative model is quite close to the true one. The entries of the table are the proportions of times the null hypothesis that the two models are equivalent was rejected at a level of significance equal to 5%.

Table 3 about here

5. ROBUSTNESS OF THE PROCEDURE WITH REGARD TO THE USE OF AN ESTIMATE OF ρ

In both of the model selection procedures described above, the value of the parameter ρ is estimated from the data as the sample correlation between the standardized residuals or prediction errors of the two competing models. As is well known, the sample correlation coefficient is only asymptotically unbiased as an estimator of ρ and coincides with maximum likelihood estimator of ρ under normality (see, e.g. Giri, 1996). In order to examine the implication of using an estimate of ρ, we calculated an approximate confidence interval for the correlation coefficient using the method based on Fisher's transformation.

Suppose that the observed correlation takes the value r. Then, Fisher's transformation is calculated as

$$z = 0.5\ln\left(\frac{1+r}{1-r}\right).$$

A 95% confidence interval for z can be calculated by $z \pm 1.96\sqrt{\frac{1}{n-3}}$, where n is the sample size on which the correlation has been estimated. Thus a 95% confidence interval for the population correlation coefficient is $(\rho_L, \rho_U) = \left(\frac{\exp(2z_L)-1}{\exp(2z_L)+1}, \frac{\exp(2z_U)-1}{\exp(2z_U)+1}\right)$, where $z_{U,}$ z_L are the lower and upper limits of the confidence interval for z.

A simulation experiment was conducted for investigating this issue. A linear trend model was tested against a quadratic trend model, using the experimental setup described in section 3.1. For each replication, the p-value of the test was calculated for the lower limit of ρ, the estimated value of ρ and the upper limit of ρ. The convention was used that if the upper limit exceeded 1, the limit was set equal to 1 and similarly for the lower limit. The results showed that the change in the mean p-value based on 10000 replications was usually small, and in particular, not greater than ± 0.05 (Table 4). This indicates that the error committed due to the use of an estimate of ρ, is not crucial for the decision. However, a warning must be given for p-values close to the significance level, as small perturbations can result in misleading decisions.

Another interesting feature is that when the two models are equivalent, the correlation coefficient is quite high. On the other hand, when the models differ, the correlation coefficient is low. The confidence intervals considered above are bounded in the admissible values and this explains the great asymmetry in the values of Table 4 for values near 0. Further investigation with p-values close to other values (different from 0 or 1) revealed a similar behaviour.

Table 4 about here

6. DISCUSSION

In this paper model selection procedures have been proposed in a hypothesis testing framework. The appealing feature of these procedures is that they can detect whether the improvement provided by a model is attributed to random fluctuations. They can be used for selecting a model either on the basis of its goodness of fit or on the basis of its predictive ability. These two issues are rather contradictory in practice as parsimonious models with a satisfactory descriptive ability can lead to poor predictors of future observations.

The theory beyond the derivation of the CGR distribution allows for constructing test statistics for several other purposes. Since the distribution can naturally be derived from standard normal theory, the asymptotic normality of many procedures seems to be a start for developing similar comparative procedures.

The fact that the distribution arises as the ratio of two dependent random variables allows for constructing comparison procedures. For example, the traditional chi-square for goodness of fit is based on a large sample normal approximation of binomial probabilities. Model selection procedures between alternative models can be derived in a similar manner. This enhances the application potential of the defined distribution to other fields.

Finally, in this paper exact hypothesis testing is introduced. This procedure can be extended so as to yield stepwise methodologies since it provides the exact distribution under the null hypothesis when adding or removing variables.

The sequential nature of both procedures developed in this paper implies their applicability to time series data.

Acknowledgements

The authors would like to thank Professor D. R. Cox for his comments and suggestions on an earlier version of this paper. They would also like to thank Mr. Apostolos Linardis for performing the numerical calculations of section 3.

REFERENCES

Abramowitz, M., Stegun, I.A. (1974). *Handbook of Mathematical Functions.* Dover, New York.

Atkinson, A. C. (1970). A Method for Discriminating Between Models (with discussion). *Journal of the Royal Statistical Society*, **B 32,** 323-353.

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

Burnham, K. P. and Anderson, D. R. (1998). *Model Selection and Inference: A Practical Information-Theoretic Approach*. Springer Verlag.

Butler, R. and Rothman, E. D. (1980). Predictive Intervals based on the Reuse of the Sample. *Journal of the American Statistical Association*, **75**, 881-889.

Cook, R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics.* Wiley, New York.

Cox, D. R. (1961). Tests of Separate Families of Hypotheses. *Proceeding of the 4th Berkeley Symposium on Mathematics, Probability and Statistics, vol 1., 105-123*. University of California Press, Berkeley.

Cox, D. R. (1962). Further Results on Tests of Separate Families of Hypotheses. *Journal of the Royal Statistical Society*, **B 24**, 406-424.

Davidson, R. and MacKinnon, J. G. (1981). Several Tests for Model Specification in the Presence of Alternative Hypotheses. *Econometrica,* **49**, 781-793.

Draper, N. and Smith, H. (1981). *Applied Regression Analysis (2nd Edition).* Wiley, New York.

Geisser, S. (1975). The Predictive Sample Reuse Method with Applications. *Journal of the American Statistical Association*, **70**, 320-328.

Gelfand, A.E. and Dey, D.K. (1994). Bayesian Model Choice: Asymptotics and Exact Calculations. *Journal of Royal Statistical Society*, **B 56**, 501-514.

Giri, N. C. (1996). *Multivariate Statistical Analysis*. Marcel and Decker, New York.

Greenberg, E. and Parks, R. P. (1997). A Predictive Approach to Model Selection and Multicollinearity. *Journal of Applied Econometrics*, **12**, 67-75.

Hocking, R. R. (1976). The Analysis and Selection of Variables in Linear Regression. *Biometrics,* **32**, 1-49.

Hoel, P. G. (1947). On the Choice of Forecasting Formulas. *Journal of the American Statistical Association*, **42**, 605-611.

Ibrahim, J. and Laud, P. W. (1996). Predictive Approaches to Bayesian Model Selection. *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner (*D.A Berry, K. M. Chaloner and J. K. Geweke, eds.), 349-358. Wiley, New York.

Izawa, T. (1965). Two or Multi-dimensional Gamma-type Distribution and Its Application to Rainfall Data. *Papers in Meteorology and Geophysics,* **XV** (3 ~ 4), 167-200.

Kendall, M. Stuart, A. and Ord, J. K. (1983). *The Advanced Theory of Statistics: Vol. 1 (4th edition).* Griffin, London.

Kibble, W. F. (1941). A Two Variate Gamma Type Distribution. *Sankhya*, **5**, 137-150.

Kotlarski, I. (1964). On Bivariate Random Variables Where the Quotient of Their Coordinates Follows Some Known Distributions. *Annals of Mathematical Statistics*, 1673-1684.

Kotlarski, I. (1966). On Characterizing the Normal Distribution by the Student's Law. *Biometrika*, **53**, 603-606.

Laud, P. W. and Ibrahim, J. (1995). Predictive Model Selection. *Journal of the Royal Statistical Society*, **B 57**, 247-262.

Linardis, A. (1998). A Comparative Study of Test Procedures Used in Assessing the Forecasting Ability of Linear Models on Crop Yield Data. *M.Sc. Thesis,* *Department of Statistics, Athens University of Economics and Business, ISBN: 960-7929-13-6.*

Linhart, H., Zuchini, W. (1986). Model Selection. Wiley, New York

Mc Aleer, M. (1987). Specification Tests for Separate Models: a survey. In *Specification Analysis in the Linear Model, M.L. King and D.E.A. Giles (eds), 149-196.* Routledge and Kegan Paul, London.

Mc Aleer, M. (1995). The Significance of Testing Non-Nested Models. *Journal of Econometrics*, **67**, 149-171.

McKinnon, J. G. (1983). Model Specification Tests against Non-Nested Alternatives.(with reply). *Econometric Reviews*, **2**, 85-110, 151-158.

Miller, A. J. (1984). Selection of Subsets of Regression Variables. *Journal of the Royal Statistical Society,* **A 147**, 389-425.

Pace, L. and Salvan, A. (1990). Best Conditional Tests for Separate families of Hypotheses. *Journal of the Royal Statistical Society*, **B 52**, 125-134.

Panaretos, J., Psarakis, S. and Xekalaki, E. (1997). The Correlated Gamma-Ratio Distribution in Model Evaluation and Selection. Technical Report no 33, Department of Statistics, Athens University of Economics and Business.

Pesaran, M. H. (1974). On the General Problem of Model Selection. *Review of Economic Studies*, **41**, 153-171.

Psarakis, S. and Panaretos, J. (1990). On an Approach to Model Evaluation and Selection, *XV Biometric Conference. Budapest, Hungary, July 1990.*

Psarakis, S. (1993). Methods of Evaluating the Predictive Ability of Linear Models. *Ph.D. Thesis*, *Department of Statistics, Athens University of Economics and Business.* (In Greek)

Rao, C. R. and Toutenburg, H. (1997). *Linear Models: Least Squares and Alternatives*. Springer.

Royston, P. and Thompson, S.G. (1995). Comparing Non-Nested Regression Models. *Biometrics,* **51**, 114-127.

Snee, R. D. (1977). Validation of Regression Models: Methods and Examples, *Technometrics*, **19**, 415-428.

Thompson, M. L. (1978a). Selection of variables in multiple Regression: Part I. A review and evaluation. *International Statistical Review*, **46**, 1-19.

Thompson, M. L. (1978b). Selection of Variables in Multiple Regression: Part II. Chosen Procedures, Computations and Examples. *International Statistical Review*, **46**, 129-146.

Weisberg, S. (1985). *Applied Linear Regression*. Wiley, New York.

West, K. D. (1996). Asymptotic Inference About Predictive Ability. *Econometrica*, **64**, 1067-1084.

West, K. D. and Cho, D. (1995). The Predictive Ability of Several Models of Exchange Rate Volatility. *Journal of Econometrics*, **69,** 367-391.

White, H. (1983). Non-nested Models. *Journal of Econometrics* (special issue) **21**, 1-160.

Xekalaki, E. and Katti, S. K. (1984). A Technique for Evaluating Forecasting Models. *Biometrical Journal*, **26(1)**, 173-184.

Table 1.

Proportions of times the null hypothesis that the linear trend model is equivalent to an exponential (respectively a quadratic) trend model was rejected and the associated mean p-values (simulation results).

		Alternative models			
		Exponential		**Quadratic**	
n	**k**	mean p-value	proportion of rejections	mean p-value	proportion of rejections
40	10	0.124	0.415	0.014	0.927
40	15	0.134	0.396	0.017	0.896
40	20	0.155	0.327	0.025	0.853
40	25	0.183	0.274	0.038	0.759
40	30	0.198	0.285	0.054	0.663
80	20	0.078	0.556	0.000	0.999
80	30	0.104	0.454	0.001	0.998
80	40	0.128	0.367	0.002	0.995
80	50	0.157	0.307	0.006	0.973
80	60	0.190	0.232	0.016	0.920
120	30	0.066	0.594	0.000	0.999
120	45	0.092	0.494	0.000	0.999
120	60	0.122	0.380	0.000	0.999
120	75	0.163	0.288	0.001	0.996
120	90	0.206	0.194	0.006	0.978
160	40	0.061	0.631	0.000	0.999
160	60	0.088	0.511	0.000	0.999
160	80	0.121	0.387	0.000	0.999
160	100	0.147	0.326	0.000	0.999
160	120	0.194	0.219	0.002	0.995

Table 2.

Results of testing the predictive equivalence of models A and B on the crop yield data of the 10 reporting districts the state of Iowa (n=24).

	Sums of squared recursive residuals						
	model A ($n R_n^A$)	model B ($n R_n^B$)	ρ	$R_{n,n}$	$R_{n,n}^{-1}$	p-value	model to be selected (“best”model)
CRD 10	58.844	92.798	0.803		1.577	0.0355	model A
CRD 20	58.681	59.595	0.908		1.015	0.4656	“equivalent”
CRD 30	24.638	35.354	0.885		1.434	0.0337	model A
CRD 40	69.677	66.691	0.449	1.044		0.453	“equivalent”
CRD 50	49.005	51.028	0.620		1.041	0.45	“equivalent”
CRD 60	55.949	32.789	0.155	1.706		0.0963	model B
CRD 70	39.933	49.012	0.561		1.227	0.275	“equivalent”
CRD 80	57.396	52.232	0.796	1.098		0.353	“equivalent”
CRD 90	61.461	41.810	0.669	1.470		0.1068	“equivalent”
CRD 100	46.515	73.943	0.593		1.589	0.0868	model A

Table 3.

Proportions of times the null hypothesis that the two models are equivalent was rejected at the 5% level of significance (simulation results)

	Alternative model		
Sample size	$Y = \alpha_2 \frac{1}{X} + \beta_2 + \varepsilon_2$	$Y = \alpha_3 X^2 + \beta_3 + \varepsilon_3$	$Y = \alpha_4 X_2 + \beta_4 + \varepsilon_4$
20	0.547	0.928	0.209
30	0.727	0.984	0.278
50	0.908	1.000	0.398
100	0.994	1.000	0.619
200	1.000	1.000	0.873
250	1.000	1.000	0.931
500	1.000	1.000	0.997
1000	1.000	1.000	1.000

Table 4.

The mean p-value of testing the equivalence of a linear trend model to a quadratic trend model when the upper (respectively the lower) limit of a 95% confidence interval for the correlation coefficient is used

n	k	lower limit ρ_L	estimated value of ρ	upper limit ρ_U
40	10	0.010	0.014	0.014
40	15	0.011	0.017	0.018
40	20	0.014	0.025	0.027
40	25	0.018	0.038	0.043
40	30	0.021	0.054	0.069
80	20	0.000	0.000	0.000
80	30	0.000	0.001	0.000
80	40	0.001	0.002	0.002
80	50	0.003	0.006	0.006
80	60	0.008	0.016	0.017
120	30	0.000	0.000	0.000
120	45	0.000	0.000	0.000
120	60	0.000	0.000	0.000
120	75	0.000	0.001	0.001
120	90	0.003	0.006	0.006
160	40	0.000	0.000	0.000
160	60	0.000	0.000	0.000
160	80	0.000	0.000	0.000
160	100	0.000	0.000	0.000
160	120	0.001	0.002	0.001

APPENDIX I

- **The Mellin Transform of the Correlated Gamma-Ratio Distribution**
- **The moments of the Correlated Gamma Ratio distribution**

The Mellin Transform of the Correlated Gamma-Ratio Distribution

According to Kotlarski (1964), a necessary and sufficient condition for the ratio of two variables to follow a F distribution can be established through the form of the Mellin transform of their joint distribution. In particular, Kotlarski (1964) has shown that if Ψ is the set of joint distribution functions F(x,y) of two not necessarily independent positive valued random variables X and Y, whose quotient X/Y follows the F distribution with parameters p_1 and p_2, then the following result holds.

Theorem 1 (Kotlarski (1964)): For a distribution function F(x,y) to belong to the set Ψ it is necessary and sufficient that its Mellin transform $h(u,v)=\int_0^\infty\int_0^\infty x^u y^v d\,F(x,y)$ satisfies the condition $h(u,-u)=\frac{\Gamma(p_1+u)}{\Gamma(p_1)}\frac{\Gamma(p_2-u)}{\Gamma(p_2)}$.

In the sequel is shown that the Mellin transform of the joint distribution of R_n^A and R_n^B defined in section 2.1 does not satisfy Kottlarski's condition for the distribution of their ratio to be of the F form.

For our problem, consider the random variables $X_i=r_i^A$, $Y_i=r_i^B$, $i=1,2,...,n$ obtained from (1) for model A and model B, respectively. Each of the variables X_i, Y_i follows the standard normal distribution. The joint distribution is therefore the bivariate standard normal distribution with a correlation coefficient denoted by ρ. Under these conditions, the joint distribution of the random variables

$$X=\frac{\sum_{i=1}^n X_i^2}{n}=R_n^A \qquad \text{and } Y=\frac{\sum_{i=1}^n Y_i^2}{n}=R_n^B$$

is Kibble's (1941) bivariate Gamma distribution as defined by the probability density function

$$f(x,y)=\frac{\rho^{-(k-1)}}{\Gamma(k)(1-\rho)^2}(xy)^{\frac{k-1}{2}}e^{-\frac{x+y}{1-\rho^2}}I_{k-1}\left(\frac{2\rho\sqrt{xy}}{1-\rho^2}\right).$$

Here $I_k(x)$ is the modified Bessel function of the first kind of order k given by

$$I_k(x) = \sum_{i=0}^{\infty} \left(\frac{x}{2}\right)^{k+2i} \frac{1}{\Gamma(i+1)\Gamma(i+k+1)}.$$

(see Abramowitz and Stegun, 1974)

Combining the last two relationships, the joint probability density function of Kibble's bivariate Gamma distribution can be rewritten as

$$f(x,y) = \frac{\rho^{-(k-1)}}{\Gamma(k)(1-\rho)^2} e^{-\frac{x+y}{1-\rho^2}} \sum_{i=0}^{\infty} \left(\frac{\rho}{1-\rho}\right)^{k+2i-1} \frac{x^{k+i-1} y^{k+i-1}}{\Gamma(i+1)\Gamma(i+k)}.$$

To determine whether an F form can be deduced for the distribution of $R_{n,n}$, one needs to examine if Kotlarski's theorem applies for the joint distribution of R_n^A and R_n^B.

For Kibble's bivariate Gamma distribution, the Mellin transformation is given as

$$h(u,v) = E\left(X^u Y^v\right)$$

$$= \frac{\left(1-\rho^2\right)^{-k}}{\Gamma(k)} \sum_{i=0}^{\infty} \left(\frac{\rho}{\left(1-\rho^2\right)}\right)^{2i} \frac{1}{\Gamma(i+1)\Gamma(i+k)} \int_0^{\infty}\int_0^{\infty} e^{-\frac{x+y}{1-\rho^2}} x^{u+k-1+i} y^{v+k+i-1}\, dx\, dy.$$

The double integral in the right hand side term is a double gamma integral. Hence, after algebraic manipulation one can find that the Mellin transform can be written as

$$h(u,v) = \frac{\left(1-\rho^2\right)^{-k+u+v+2k}}{\Gamma(k)} \sum_{i=0}^{\infty} \frac{\rho^{2i}\Gamma(u+k+i)\Gamma(v+k+i)}{i!\Gamma(k+i)}$$

$$= \frac{\left(1-\rho^2\right)^{u+v+k}}{\Gamma(k)} \frac{\Gamma(k+u)\Gamma(k+v)}{\Gamma(k)} \sum_{i=0}^{\infty} \frac{(k+u)_{(i)}(k+v)_{(i)}}{k_{(i)}} \frac{\rho^{2i}}{i!}.$$

The sum in the second term is a hypergeometric series and thus the Mellin transform can be written as

$$h(u,v) = \frac{\Gamma(k+u)}{\Gamma(k)} \frac{\Gamma(k+v)}{\Gamma(k)} \left(1-\rho^2\right)^{u+v+k} {}_2F_1\left(k+u, k+v; k; \rho^2\right)$$

where

$${}_2F_1(a,b;c;z) = \sum_{r=0}^{\infty} \frac{a_{(r)} b_{(r)}}{c_{(r)}} \frac{z^r}{r!}$$

denotes the hypergeometric series (see, Abramowitz and Stegun, 1974).

One can see that the Mellin transform of Kibble's distribution does not satisfy the conditions of Theorem 1. Hence, the quotient R_n^A/R_n^B does not follow the F distribution when R_n^A and R_n^B are dependent.

The moments of the Correlated Gamma Ratio distribution

The moments of a random variable Z which has the CGR distribution are related to the Mellin transform of the Bivariate Kibble's distribution by the formula

$$E(Z^r) = E(X^r Y^{-r}) = h(r,-r)$$

Setting r=1, we obtain $E\left(\frac{X}{Y}\right)$ which is the first simple moment of the CGR distribution.

So, we have that

$$E\left(\frac{X}{Y}\right) = \frac{\Gamma(k+1)\Gamma(k-1)}{\Gamma(k)\Gamma(k)}\left(1-\rho^2\right)^k {}_2F_1(k+1,k-1;k;\rho^2)\,. \tag{A.1}$$

From Abramowitz and Stegun (1974), (relation 15.2.18, p. 558), we have that

$$(c-a-b)\,{}_2F_1(a,b;c;z) = (c-a)\,{}_2F_1(a-1,b;c;z) + b(1-z)\,{}_2F_1(a,b+1;c;z)\,. \tag{A.2}$$

Using (A.2) with α=k+1, b=k-1, z=ρ^2 and c=k we obtain

$$k\ {}_2F_1(k+1,k-1;k;\rho^2) = {}_2F_1(k,k-1;k;\rho^2) + (k-1)(1-\rho^2)\ {}_2F_1(k+1,k;k;\rho^2)\,. \tag{A.3}$$

Also, from relation (15.2.18) of Abramowitz and Stegun (1974) it holds that

$${}_2F_1(a,b;b;z) = {}_2F_1(b,a;b;z) = \frac{1}{(1-z)^a}\,. \tag{A.4}$$

Substituting (A.4) in (A.3) we obtain

$${}_2F_1(k+1,k-1;k;\rho^2) = \frac{1}{k(1-\rho^2)^k}\left[(1-\rho^2)+(k-1)\right].$$

Substituting back to (A.1) yields

$$E\left(\frac{X}{Y}\right) = E(Z) = \frac{(1-\rho^2)+k-1}{k-1}, \quad k>1.$$

For the derivation of the variance we need to calculate the second moment $E\left(Z^2\right)$. Again, using (4) we can write

$$E\left(Z^2\right) = E\left[\left(\frac{X}{Y}\right)^2\right] = \frac{\Gamma(k+2)\Gamma(k-2)}{\Gamma(k)\Gamma(k)}\left(1-\rho^2\right)^k {}_2F_1(k+2, k-2; k; \rho^2)\,. \qquad \text{(A.5)}$$

Clearly, this is quite complicated. A recurrence relation is needed for ${}_2F_1(k+2, k-2; k; \rho^2)$. Using (A.2) successively we find that

$${}_2F_1(a,b;c;z) = \frac{(c-a)\,{}_2F_1(a-1,b;c;z) - b(1-z)\,{}_2F_1(a,b+1;c;z)}{(c-a-b)}$$

$$= \frac{(c-a)(c-a+1)}{(c-a-b)(c-a-b+1)^2}\,{}_2F_1(a-2,b;c;z)$$

$$+ \frac{2b(1-z)(c-a)}{(c-a-b+1)(c-a-b-1)^2}\,{}_2F_1(a-1,b+1;c;z)$$

$$+ \frac{b(b+1)(1-z)^2}{(c-a-b)(c-a-b-1)^2}\,{}_2F_1(a,b+2;c;z). \qquad \text{(A.6)}$$

Substituting α=k+2, b=k-2, z=ρ^2 and c=k and using (A.4), we obtain

$${}_2F_1(k+2, k-2; k; \rho^2) = \frac{2}{k(k-1)^2}\,{}_2F_1(k, k-2; k; \rho^2)$$

$$+ \frac{(k-2)(k-1)(1-\rho^2)^2}{k(k+1)}\,{}_2F_1(k+2,k;k;\rho^2) + \frac{4(k-2)(1-\rho^2)}{(k-1)(k+1)}\,{}_2F_1(k+1,k-1;k;\rho^2)$$

$$= \frac{2}{k(k-1)}\frac{1}{(1-\rho^2)^{k-2}} + \frac{(k-2)(k-1)(1-\rho^2)^2}{k(k+1)}\frac{1}{(1-\rho^2)^{k+2}}$$

$$+ \frac{4(k-2)(1-\rho^2)}{(k-1)(k+1)}\frac{1}{k(1-\rho^2)^k}\left[(1-\rho^2) + (k-1)\right]$$

Then, substituting in (A.5) we obtain

$$E\left[\left(\frac{X}{Y}\right)^2\right] = \frac{(k+1)k}{(k-1)(k-2)}\left(1-\rho^2\right)^k {}_2F_1(k+2, k-2; k; \rho^2) =$$

$$= \frac{6(1-\rho^2)^2 + 4(k-2)(1-\rho^2)}{(k-1)(k-2)} + 1 \qquad .$$

Note that the second moment does not exist for k<2.

The variance of the CGR distribution is, therefore

$$\mathrm{Var}(Z) = E(Z^2) - [E(Z)]^2 = \frac{6(1-\rho^2)^2 + 4(k-2)(1-\rho^2)}{(k-1)(k-2)} + 1 - \left[\frac{(1-\rho^2)+k-1}{k-1}\right]^2$$

$$= \frac{(5k-4)(1-\rho^2)^2 + 2(k-1)(k-2)(1-\rho^2)}{(k-1)^2(k-2)}, \quad \text{for } k>2.$$

It is hard to derive closed form expressions but it is relatively easy to obtain the moments numerically using recurrence relation (A.2) for the hypergeometric function. Note that programming languages supporting recursion can be very helpful.

The coefficient of variation is

$$\mathrm{CV}(Z) = \frac{\sqrt{\dfrac{(5k-4)(1-\rho^2)^2 + 2(k-1)(k-2)(1-\rho^2)}{(k-2)}}}{(1-\rho^2)+(k-1)}$$

$$= \sqrt{\frac{2(k-2)(1-\rho^2)(k-1) + (5k-4)(1-\rho^2)^2}{2(k-2)(1-\rho^2)(k-1) + (k-2)(1-\rho^2)^2 + (k-2)(k-1)^2}} \quad .$$

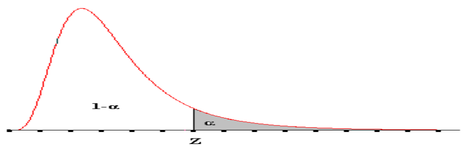
It can be seen that for k>4 the coefficient of variation is smaller than 1. Moreover, as the value of k increases the coefficient tends to be very small.

APPENDIX II

Percentage points of the Correlated Gamma Ratio distribution

Percentage points of the Correlated Gamma Ratio distribution for α=0.1

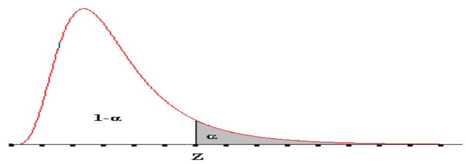
$$\int_0^z \frac{\left(1-\rho^2\right)^k}{B(k,k)}\, t^{k-1}(1+t)^{-2k}\left[1-\left[\frac{2\rho}{t+1}\right]^2 t\right]^{-\frac{2k+1}{2}} dt = 1 - \alpha = 0.90$$



k \ ρ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	9	8.93	8.72	8.36	7.85	7.2	6.4	5.45	4.33	3.02
2	4.11	4.08	4.01	3.88	3.71	3.48	3.2	2.85	2.44	1.93
3	3.055	3.04	3.00	2.92	2.81	2.67	2.49	2.27	2.00	1.66
4	2.59	2.58	2.55	2.49	2.41	2.3	2.17	2.00	1.8	1.53
5	2.32	2.31	2.29	2.24	2.18	2.09	1.98	1.84	1.67	1.46
6	2.15	2.14	2.12	2.08	2.02	1.95	1.85	1.74	1.59	1.41
7	2.02	2.01	2.00	1.96	1.91	1.85	1.76	1.66	1.54	1.37
8	1.93	1.92	1.90	1.87	1.83	1.77	1.70	1.61	1.49	1.34
9	1.85	1.846	1.83	1.80	1.76	1.71	1.64	1.56	1.455	1.315
10	1.79	1.785	1.775	1.75	1.71	1.665	1.6	1.525	1.425	1.295
11	1.745	1.74	1.725	1.705	1.67	1.62	1.565	1.49	1.4	1.277
12	1.705	1.70	1.685	1.665	1.63	1.59	1.535	1.465	1.38	1.265
13	1.665	1.664	1.65	1.63	1.60	1.56	1.51	1.44	1.36	1.253
14	1.635	1.63	1.62	1.6	1.57	1.53	1.485	1.423	1.345	1.24
15	1.605	1.604	1.59	1.575	1.546	1.51	1.465	1.405	1.33	1.31
16	1.585	1.58	1.57	1.55	1.525	1.49	1.445	1.39	1.32	1.225
17	1.56	1.553	1.546	1.53	1.505	1.471	1.43	1.376	1.307	1.216
18	1.54	1.535	1.525	1.510	1.486	1.455	1.415	1.364	1.297	1.207
19	1.52	1.519	1.51	1.495	1.471	1.44	1.402	1.351	1.287	1.203
20	1.505	1.504	1.495	1.48	1.456	1.426	1.39	1.341	1.28	1.197
21	1.49	1.489	1.48	1.465	1.44	1.415	1.377	1.331	1.274	1.193
22	1.475	1.474	1.466	1.451	1.43	1.404	1.379	1.323	1.353	1.187
23	1.465	1.460	1.455	1.440	1.567	1.391	1.358	1.315	1.259	1.183
24	1.454	1.450	1.442	1.428	1.408	1.382	1.35	1.306	1.252	1.178
25	1.442	1.44	1.432	1.418	1.4	1.374	1.34	1.3	1.246	1.174
26	1.432	1.43	1.422	1.408	1.39	1.366	1.344	1.292	1.240	1.17
27	1.422	1.42	1.412	1.4	1.382	1.356	1.326	1.286	1.238	1.166
28	1.412	1.410	1.402	1.39	1.372	1.35	1.32	1.28	1.23	1.163
29	1.404	1.402	1.394	1.382	1.366	1.342	1.312	1.274	1.226	1.16
30	1.396	1.394	1.386	1.375	1.358	1.336	1.306	1.27	1.222	1.157
40	1.333	1.332	1.326	1.316	1.302	1.284	1.259	1.228	1.189	1.134
50	1.293	1.291	1.287	1.279	1.267	1.249	1.229	1.203	1.168	1.119
60	1.265	1.264	1.259	1.252	1.24	1.226	1.207	1.183	1.152	

Percentage points of the Correlated Gamma Ratio distribution for α=0.05

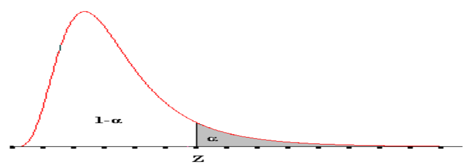
$$\int_0^z \frac{\left(1-\rho^2\right)^k}{B(k,k)}\, t^{k-1}(1+t)^{-2k}\left[1-\left[\frac{2\rho}{t+1}\right]^2 t\right]^{-\frac{2k+1}{2}} dt = 1 - \alpha = 0.95$$



k \ ρ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	19	18.80	18.3	17.4	16.27	14.73	12.84	10.60	8.02	5.04
2	6.39	6.34	6.20	5.97	5.64	5.22	4.7	4.07	3.34	2.46
3	4.284	4.26	4.18	4.04	3.85	3.61	3.31	2.945	2.51	1.97
4	3.44	3.42	3.36	3.27	3.145	2.96	2.74	2.48	2.16	1.76
5	2.98	2.96	2.92	2.84	2.74	2.6	2.43	2.22	1.965	1.64
6	2.687	2.675	2.65	2.57	2.485	2.37	2.23	2.06	1.835	1.56
7	2.49	2.47	2.44	2.39	2.31	2.21	2.09	1.935	1.75	1.51
8	2.335	2.325	2.29	2.25	2.18	2.1	1.985	1.85	1.675	1.46
9	2.22	2.21	2.19	2.14	2.18	2	1.95	1.775	1.63	1.427
10	2.125	2.115	2.095	2.055	2	1.93	1.837	1.725	1.585	1.4
11	2.05	2.04	2.02	1.983	1.935	1.87	1.783	1.677	1.55	1.375
12	1.983	1.977	1.955	1.925	1.876	1.815	1.735	1.635	1.515	1.355
13	1.93	1.922	1.905	1.875	1.83	1.775	1.697	1.605	1.49	1.338
14	1.884	1.876	1.86	1.83	1.787	1.733	1.663	1.577	1.47	1.324
15	1.843	1.835	1.82	1.794	1.752	1.7	1.63	1.552	1.453	1.31
16	1.805	1.798	1.783	1.757	1.72	1.675	1.61	1.527	1.427	1.297
17	1.775	1.767	1.753	1.727	1.697	1.644	1.582	1.508	1.414	1.287
18	1.745	1.74	1.723	1.697	1.667	1.620	1.563	1.493	1.397	1.277
19	1.717	1.711	1.697	1.678	1.644	1.59	1.543	1.472	1.387	1.27
20	1.695	1.69	1.676	1.653	1.624	1.576	1.527	1.46	1.375	1.262
21	1.672	1.667	1.654	1.633	1.604	1.564	1.511	1.447	1.362	1.254
22	1.654	1.647	1.635	1.613	1.584	1.549	1.498	1.434	1.353	1.247
23	1.633	1.629	1.617	1.597	1.567	1.531	1.484	1.424	1.344	1.242
24	1.615	1.612	1.6	1.581	1.553	1.516	1.469	1.412	1.336	1.236
25	1.6	1.596	1.585	1.566	1.54	1.504	1.458	1.401	1.328	1.229
26	1.585	1.581	1.57	1.552	1.526	1.491	1.447	1.390	1.320	1.224
27	1.57	1.566	1.558	1.54	1.514	1.48	1.437	1.383	1.314	1.22
28	1.558	1.556	1.544	1.528	1.502	1.47	1.426	1.374	1.307	1.215
29	1.546	1.543	1.532	1.516	1.492	1.459	1.418	1.367	1.302	1.211
30	1.534	1.531	1.522	1.505	1.482	1.45	1.41	1.359	1.296	1.207
40	1.447	1.445	1.437	1.423	1.404	1.378	1.346	1.303	1.249	1.175
50	1.391	1.390	1.382	1.37	1.355	1.332	1.304	1.267	1.22	1.156
60	1.353	1.35	1.345	1.334	1.319	1.299	1.274	1.241	1.199	

Percentage points of the Correlated Gamma Ratio distribution for α=0.01

$$\int_0^z \frac{\left(1-\rho^2\right)^k}{B(k,k)}\, t^{k-1}(1+t)^{-2k}\left[1-\left[\frac{2\rho}{t+1}\right]^2 t\right]^{-\frac{2k+1}{2}} dt = 1-\alpha = 0.99$$



k \ ρ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	99	98.10	95.2	90.3	83.5	74.8	64.1	51.7	36.7	20.4
2	15.98	15.84	15.42	14.71	13.72	12.45	10.90	9.05	6.91	4.45
3	8.47	8.40	8.20	7.87	7.40	6.8	6.05	5.17	4.13	2.91
4	6.03	5.99	5.86	5.64	5.34	4.95	4.47	3.89	3.2	2.38
5	4.85	4.82	4.73	4.57	4.34	4.05	3.69	3.25	2.73	2.11
6	4.155	4.13	4.06	3.93	3.75	3.52	3.23	2.88	2.46	1.94
7	3.7	3.68	3.62	3.51	3.36	3.16	2.92	2.62	2.27	1.83
8	3.37	3.36	3.30	3.21	3.08	2.91	2.7	2.45	2.14	1.75
9	3.13	3.12	3.07	2.99	2.87	2.72	2.53	2.31	2.03	1.68
10	2.94	2.93	2.88	2.81	2.705	2.565	2.405	2.2	1.95	1.63
11	2.785	2.775	2.735	2.67	2.575	2.45	2.3	2.11	1.88	1.59
12	2.66	2.65	2.61	2.55	2.465	2.35	2.21	2.04	1.825	1.555
13	2.555	2.545	2.51	2.455	2.375	2.27	2.135	1.975	1.78	1.525
14	2.465	2.455	2.425	2.37	2.295	2.195	2.075	1.925	1.74	1.497
15	2.39	2.38	2.35	2.3	2.23	2.135	2.025	1.88	1.705	1.475
16	2.32	2.31	2.285	2.235	2.17	2.08	1.975	1.84	1.675	1.46
17	2.26	2.25	2.225	2.18	2.117	2.035	1.935	1.805	1.645	1.437
18	2.208	2.195	2.172	2.13	2.07	1.99	1.895	1.773	1.62	1.418
19	2.16	2.15	2.127	2.086	2.03	1.955	1.86	1.744	1.599	1.41
20	2.115	2.105	2.085	2.046	1.994	1.92	1.83	1.72	1.58	1.395
21	2.075	2.07	2.049	2.01	1.956	1.89	1.801	1.695	1.56	1.384
22	2.04	2.034	2.01	1.976	1.925	1.86	1.775	1.675	1.544	1.374
23	2.005	2	1.98	1.946	1.897	1.835	1.754	1.654	1.53	1.364
24	1.978	1.972	1.952	1.918	1.872	1.810	1.732	1.634	1.512	1.352
25	1.95	1.944	1.924	1.892	1.848	1.788	1.712	1.618	1.5	1.344
26	1.924	1.918	1.90	1.868	1.824	1.766	1.694	1.602	1.488	1.336
27	1.9	1.894	1.876	1.846	1.804	1.748	1.676	1.588	1.476	1.328
28	1.878	1.872	1.854	1.826	1.784	1.73	1.66	1.574	1.464	1.32
29	1.856	1.852	1.834	1.806	1.766	1.712	1.645	1.561	1.455	1.314
30	1.838	1.832	1.816	1.788	1.748	1.696	1.632	1.55	1.446	1.308
40	1.69	1.685	1.672	1.65	1.619	1.578	1.525	1.458	1.374	1.259
50	1.597	1.594	1.583	1.565	1.538	1.502	1.456	1.4	1.327	1.229
60	1.536	1.532	1.522	1.506	1.48	1.449	1.409	1.359	1.294	-

APPENDIX III

The Probability density function of the Correlated Gamma Ratio distribution

Figure 1
The probability density function of the correlated gamma-ratio distribution
Curve 1 corresponds to the case k=2, ρ=0.2
Curve 2 corresponds to the case k=2, ρ=0.9

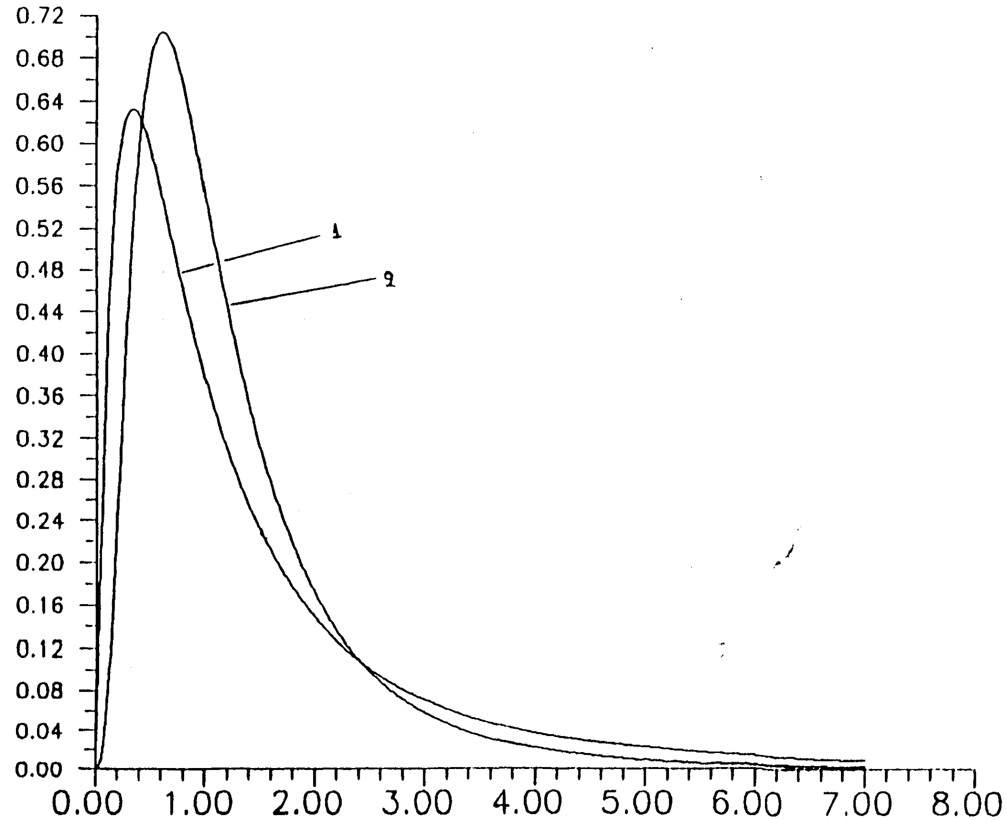


Figure 2
The probability density function of the
correlated gamma-ratio distribution
Curve 1 corresponds to the case $k=15$, $\rho=0.2$
Curve 2 corresponds to the case $k=15$, $\rho=0.9$

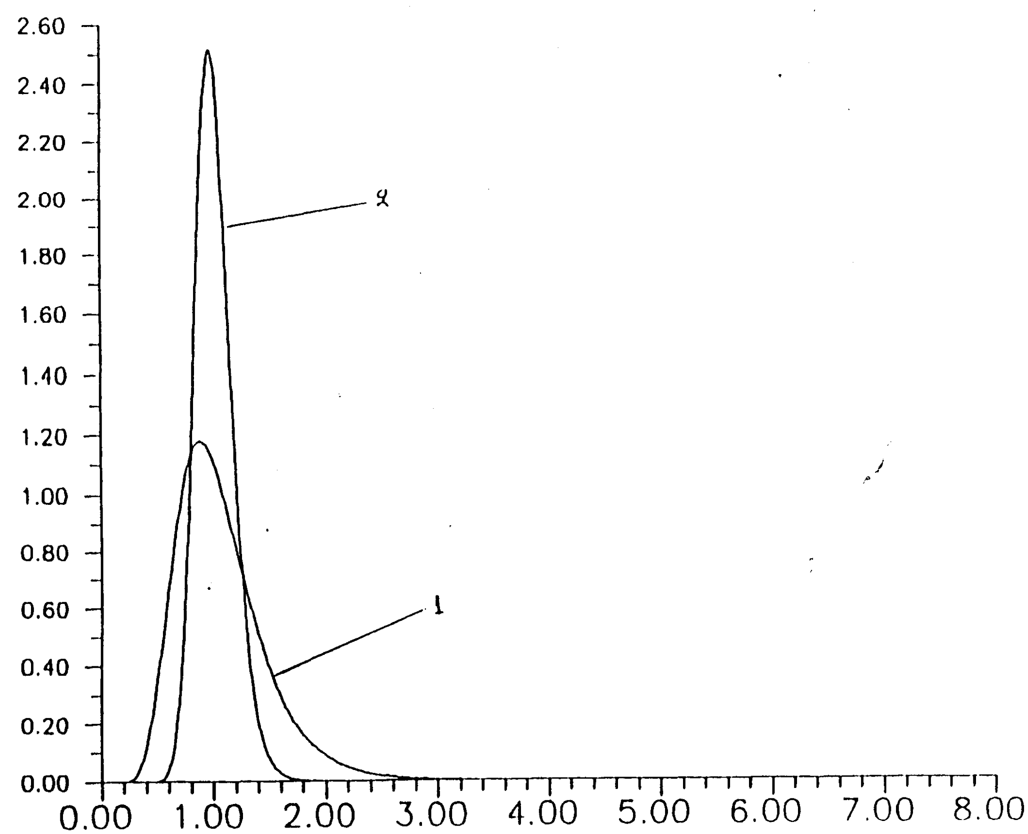


Figure 3
The probability density function of the correlated gamma-ratio distribution
Curve 1 corresponds to the case $k=2$, $\rho=0.0$
Curve 2 corresponds to the case $k=4$, $\rho=0.0$