A Predictive Model Evaluation and Selection Approach - The Correlated Gamma Ratio Distribution

Xekalaki, Evdokia; Panaretos, John and Psarakis, Stelios

2003

A PREDICTIVE MODEL EVALUATION AND SELECTION APPROACH — THE CORRELATED GAMMA RATIO DISTIRIBUTION

Evdokia Xekalaki, John Panaretos & Stelios Psarakis
Department of Statistics
Athens University of Economics & Business, Greece

1. Introduction

Evaluating the forecasting potential of a model before it can be used for planning and decision making has been the concern of many statistical workers. A number of evaluation techniques has thus been considered and much theory has been developed, especially for nested models based mainly on goodness of fit considerations.

Predictive evaluation appears to have received less attention, despite the fact that the predictive ability of a model is a very important characteristic of the model. Xekalaki and Katti (1984) introduced an evaluation scheme of a sequential nature that can be used for models that are not necessarily nested. It is based on the idea of scoring rules for rating the predictive behavior 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. (see, also Panaretos et al., 1997, Psarakis, 1993, Psarakis & Panaretos, 1990).

Model comparison problems have also attracted much interest. The selection procedures that have been developed are mainly based on criteria for testing the null hypothesis that one model is valid against an alternative hypothesis that another model is valid. Such testing procedures lead to the selection of one of two competing models. The problem of testing whether two models can be considered as "*equivalent*" in some sense requires a different hypothesis formulation and has only been approached indirectly through the concept of encompassing (see, e.g., Gouriéroux et al., 1993, Gouriéroux & Monfort, 1996) and through asymptotic results based on the change in likelihood.

In this chapter, an evaluation method is proposed that is based on Xekalaki and Katti's idea of using a scoring rule but is free of the element of subjectivity. In particular, a scoring rule is suggested to rate the behavior 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. The problem of comparative evaluation is also considered and a test procedure is suggested for testing whether two linear models that are not necessarily nested can be considered to be "*equivalent*" in their predictive abilities. In this case, a distribution which is a generalized form of the F distribution arises as the distribution of the sample statistic is considered. This distribution and the scoring rule associated with it are used for comparing two linear models on real data. In particular, in section 2, the regression model setting considered in the sequel is presented and the scheme suggested for evaluating the predictive ability of a linear model is described. Section 3 deals with the problem of comparatively evaluating two competing linear models in their predictive abilities. The distribution of the test statistic used is derived and studied is sections 4 and 5 while selected percentage points of it are provided in the Appendix. The procedure is illustrated on several crop yield data sets (section 6).

2. Rating the Predictive Ability of a Linear Model

Consider the linear model

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

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 $\left(\ell_0 > m,\ |\mathbf{X}_t'\mathbf{X}_t| \neq 0\right)$, $\boldsymbol{\beta}$ is an $m \times 1$ 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 it 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 $t+1$, $t = 0, 1, 2, \ldots$ 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 $\ell_{t+1} = \ell_t + 1, \quad t = 0, 1, 2, \ldots$.

The predictive behavior 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 $|r_{t+1}|$, where

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

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

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

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

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

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.

3. Comparative Evaluation of the Predictive Ability of Two Linear Models With the Use of a Generalized Form of the F Distribution

Consider now A and B to be two competing linear models that have been used for prediction purposes for a number n_1 and n_2 of years, respectively. A null hypothesis that is interesting to test is whether two models have "*equivalent*" forecasting abilities. This is a hypothesis that can be defined only implicitly, but it exists as a mathematical entity. The closest description of it is *"H_0: models A and B have equal mean squared prediction errors."* This is a hypothesis that can be tested formally using conventional methods, in all cases in which neither, one, or both models are correctly specified using the average standardized distances between the observed value of the dependent variable and its predicted values by models A and B. Then, a decision on whether models A and B are "*equivalent*" in their predictive ability would naturally be based on the ratio of the average scores of the two models as given by the statistic

$$R_{n_1,n_2} = \frac{R_{n_1}(A)}{R_{n_2}(B)} \tag{3}$$

where $R_{n_1}(A)$, $R_{n_2}(B)$, are given by (2) for $n=n_1$ and $n=n_2$ and refer to model A and model B, respectively.

For large ℓ_{t_1}, ℓ_{t_2} the distribution of the statistic R_{n_1,n_2} can be approximated by the F distribution with n_1 and n_2 degrees of freedom whenever the ratings of the two models are independent. Hence, values of R_{n_1,n_2} in the right tail of the F distribution with n_1 and n_2 degrees of freedom will indicate a higher performance by model A.

However, under the conditions of the problem, the assumption of independence does not seem to be satisfied.

Determining the exact distribution of R_{n_1,n_2} in the case of dependent ratings would, however, be desirable as in practice data on ratings are often matched. (In the latter case, $n_1=n_2=n$.)

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 Kotlarski (1964), 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 (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 (Kotlarski, 1964): For a distribution function F(x,y) to belong to the set Ψ it is necessary and sufficient that its Mellin transform

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

satisfies the condition

$$h(u,-u) = \frac{\Gamma(p_1+u)}{\Gamma(p_1)} \frac{\Gamma(p_2-u)}{\Gamma(p_2)} .$$

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

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

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

$$f(x,y) = \frac{\rho^{-(k-1)}}{\Gamma(k)(1-\rho^2)} (xy)^{\frac{k-1}{2}} e^{-\frac{x+y}{1-\rho^2}} I_{k-1}\left[\frac{2\rho\sqrt{xy}}{1-\rho^2}\right], \qquad (4)$$

where k=n/2 and $I_k(x)$ is the modified Bessel function of the first kind of order k given by (see Abramowitz & Stegun, 1972)

$$I_k(x) = \sum_{i=0}^{\infty} \left(\frac{x}{2}\right)^{k+2i} \frac{1}{\Gamma(i+1)\Gamma(i+k+1)} . \qquad (5)$$

Therefore,

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

So, finally, the probability density function of the bivariate gamma distribution of $(R_n(A), R_n(B))$ is given by

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

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 $R_n(A)$, $R_n(B)$.

For Kibble's bivariate Gamma distribution, we obtain, by the definition of the Mellin transform

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

Definition by I, the double integral in the right-hand side of the above relationship, we have

$$I = \frac{\Gamma(u+k+i)\ \Gamma(v+k+i)}{(1-\rho^2)^{-(u+v+2k+2i)}}.$$

This, in turn, implies that

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

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

or, equivalently that

$$h(u,v) = \frac{\Gamma(k+u)}{\Gamma(k)} \frac{\Gamma(k+v)}{\Gamma(k)} \left(1-\rho^2\right)^{u+v+k} {}_2F_1\left(k+u, k+v; k; \rho^2\right), \qquad (6)$$

where

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

is the hypergeometric series with $\alpha_{(r)}$ denoting the ascending factorial (see Abramowitz & Stegun, 1972).

One can see that the Mellin transform of Kibble's distribution given (6) 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.

In the next section, it is shown that the distribution of $R_{n,n}$ is a generalized form of the F distribution.

4. The Distribution of the Ratio X/Y When X and Y Follow Kibble's Bivariate Gamma Distribution

It is known that if X and Y are dependent random variables, the distribution function of Z=X/Y is given by

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

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

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

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

This leads to

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

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

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

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

Furthermore,

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

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

Here, we made use of the identities

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

and

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

Letting α=2k, m=i , n=2 one obtains

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

Hence (7) can be written as

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

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

Therefore,

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

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**.** (A reparameterized form of this distribution was arrived at by Izawa (1965)).

Note: 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 with parameters k and R or, equivalently of the F distribution with 2k and 2k degrees of freedom.

5. The t Distribution as a Limiting Case of the Correlated Gamma Ratio Distribution

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

Let Z follow the CGR distribution with density function given by (8). Consider the variable

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

Then,

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

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

Using (8), 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

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

But this is the probability density function of the t distribution.

In the Appendix, some graphs of the probability density function of the correlated gamma-ratio distribution are provided for different values of k and ρ. Also, Tables A1, A2 and A3 provide percentage points of the distribution for selected values of the parameter k (k=1(1) 30, 40, 50, 60) and of the correlation coefficient ρ (ρ=0.0(0.1) 0.9).

6. An Application to Crop-Yield Data

For the purpose of illustrating the model selection procedure, a problem presented in Xekalaki and Katti (1984), concerning the selection of a linear model among several competing ones considered by the United States Department of Agriculture (USDA) to predict the corn yield for 10 Crop Reporting Districts (CRD 10, 20, …,100), was re-examined 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 as general trend factors for predicting the crop yield. A detailed description of the models can be found in Linardis (1998).

The aim of the application is to compare the predictability of these models for every district, using the Correlated Gamma - Ratio distribution.

Let m_A and m_B denote these two models respectively. To compare the two crop yield models we need to test a hypothesis of the form:

H_0: *Models* m_A *and* m_B *are of "equivalent" predictive ability* (symbolically, $m_A \sim m_B$)versus an alternative

H_1: *The two models differ in their predictive ability, i.e.,* m_A *is of higher predictive ability (symbolically,* $m_A \succ m_B$*) or of lower predictive ability (symbolically,* $m_A \prec m_B$*)* ,

where the term "equivalent" is used in the sense defined in section 3.

Rejection of the null hypothesis indicates that one of the models performs differently. With a one-sided alternative, one may proceed in a manner similar to that used when testing for equality of variances via the F-test. The results of testing the predictive equivalence of models m_A and m_B on the crop yield data and considered together with the estimated values of the correlations between the standardized prediction errors for the two models are summarized in Table16.1.

Table 16.1: Results of testing the null hypothesis of predictive equivalence of models $\mathbf{m_A}$ and $\mathbf{m_B}$ H_0: $\mathbf{m_A} \sim \mathbf{m_B}$ on the crop yield data of the 10 reporting districts the state of Iowa (n=24).

Crop reporting district	H_1	Sums of squared recursive residuals: Model m_A $(n R_n(A))$	Model m_B $(n R_n(B))$	$R_{n,n}$	Estimated value of ρ	p-value	model to be selected (*"best" model*)
CRD 10	$m_A \succ m_B$	58.844	92.798	0.634	0.803	0.0355	model A
CRD 20	$m_A \succ m_B$	58.681	59.595	0.985	0.908	0.4656	" equivalent"
CRD 30	$m_A \succ m_B$	24.638	35.354	0.697	0.885	0.0337	model A
CRD 40	$m_A \prec m_B$	69.677	66.691	1.044	0.449	0.453	" equivalent"
CRD 50	$m_A \succ m_B$	49.005	51.028	0.961	0.620	0.45	" equivalent"
CRD 60	$m_A \prec m_B$	55.949	32.789	1.706	0.155	0.0963	model B
CRD 70	$m_A \succ m_B$	39.933	49.012	0.815	0.561	0.275	" equivalent"
CRD 80	$m_A \prec m_B$	57.396	52.232	1.098	0.796	0.353	" equivalent"
CRD 90	$m_A \prec m_B$	61.461	41.810	1.470	0.669	0.1068	" equivalent"
CRD 100	$m_A \succ m_B$	46.515	73.943	0.629	0.593	0.0868	model A

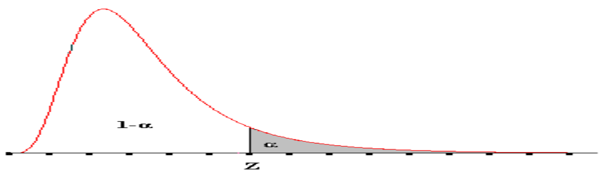
From this table, one may see that for six districts, the models are of equivalent predictive ability. Model m_A performs "*better*" in 3 cases while only in one case model m_B is "*superior.*"

In all the cases considered, the parameter ρ was estimated from the data as the sample correlation between the standardized prediction errors of the two competing models. The extent to which the use of an estimate of ρ may affect the selection procedure has to be investigated. Of course, asymptotically, it is not expected to have any impact because ρ is estimated consistently. The first investigation results for small to moderate sample sizes are not indicative of any appreciable effect either.

APPENDIX

Table A1: Percentage points of the Correlated Gamma Ratio distribution for α=0.1

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

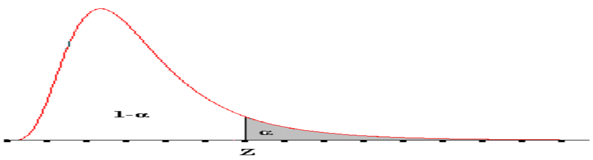


k \ ρ	**0.0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
1	9	8.93	8.72	8.36	7.85	7.2	6.4	5.45	4.33	3.02
2	4.11	4.08	4.01	3.88	3.71	3.48	3.2	2.85	2.44	1.93
3	3.055	3.04	3.00	2.92	2.81	2.67	2.49	2.27	2.00	1.66
4	2.59	2.58	2.55	2.49	2.41	2.3	2.17	2.00	1.8	1.53
5	2.32	2.31	2.29	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	

Table A2: Percentage points of the Correlated Gamma Ratio distribution for α=0.05

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

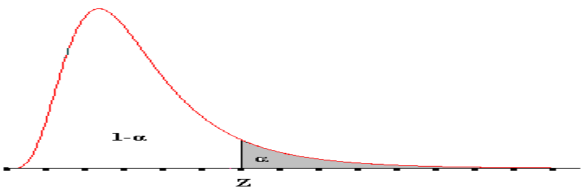


k \ ρ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	19	18.80	18.3	17.4	16.27	14.73	12.84	10.60	8.02	5.04
2	6.39	6.34	6.20	5.97	5.64	5.22	4.7	4.07	3.34	2.46
3	4.284	4.26	4.18	4.04	3.85	3.61	3.31	2.945	2.51	1.97
4	3.44	3.42	3.36	3.27	3.145	2.96	2.74	2.48	2.16	1.76
5	2.98	2.96	2.92	2.84	2.74	2.6	2.43	2.22	1.965	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	

Table A3: Percentage points of the Correlated Gamma Ratio distribution for α=0.01

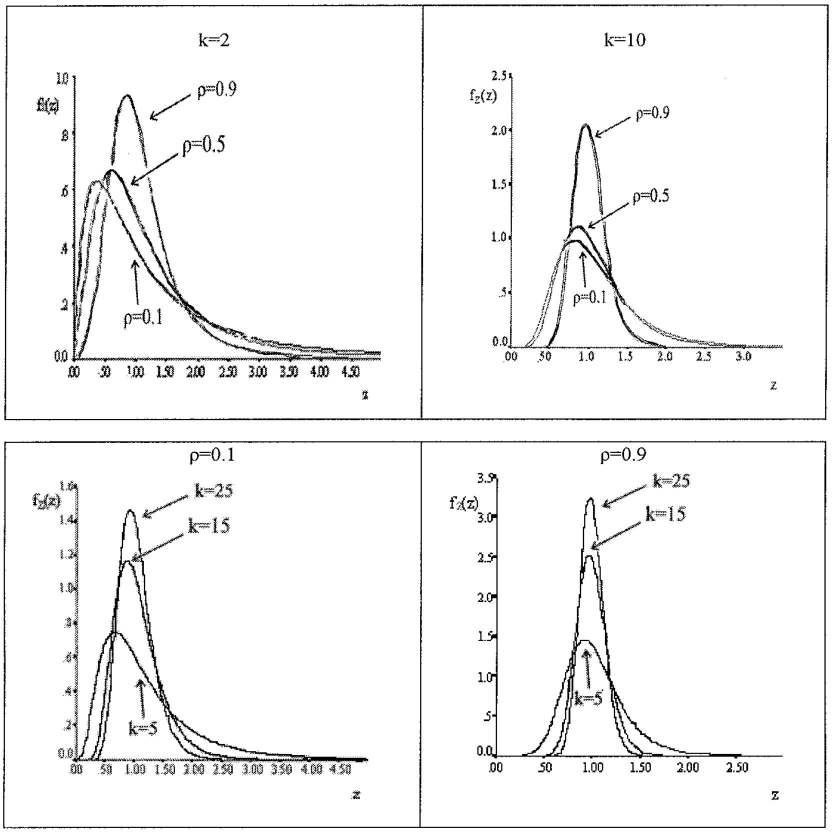
$$\int_0^z \frac{\left(1-\rho^2\right)^k}{B(k,k)} t^{k-1}(1+t)^{-2k}\left[1-\left[\frac{2\rho}{t+1}\right]^2 t\right]^{-\frac{2k+1}{2}} dt = 1-\alpha = 0.99$$



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	-

The probability density function of the Correlated Gamma Ratio Distribution



The probability density function of the Correlated Gamma-Ratio distribution for selected values of k and ρ

Acknowledgment

The authors thank Professors D.R. Cox and C.R. Rao for their comments on an earlier version of the chapter.

7. References

Abramowitz, M. and Stegun, I.A. (1972), *Handbook of Mathematical Functions.* New York, Dover.

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.

Gouriéroux, C. and Monfort A. (1996). *Simulation-Based Econometric Methods,* CORE Lectures, Oxford University Press.

Gouriéroux, C., Monfort A., and Renault E. (1993). "Indirect Inference". *Journal of Applied Econometrics*, 8, 85-118.

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.

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.*

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.

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)

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

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

Received: January 1999, Revised: May 1999