Chapter 3

TEST OF THE PREDICTABILITY OF A LINEAR MODEL AND COMPARISON OF THE PREDICTABILITY OF TWO LINEAR MODELS BASED ON THE χ^2 AND THE CORRELATED GAMMA-RATIO DISTRIBUTIONS

3.1 ESTIMATION OF PREDICTIONS.

Consider the linear model :

$$\mathbf{Y_t} = \mathbf{X_t b} + \mathbf{e_t} \qquad (3.1.1)$$

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, where $\left(\ell_t \geq m, \left|\mathbf{X}'_\mathbf{t}\mathbf{X_t}\right| \neq 0\right)$
- $\boldsymbol{\beta}$ is an mx1 vector of regression coefficients and
- $\mathbf{e}_t$ is an $(\ell_t \times 1)$ vector of normal error random variables with $E(\mathbf{e}_t)=0$ and $V(\mathbf{e}_t)=\sigma^2 I_t$, where I_t is the $\ell_t \times \ell_t$ identity matrix.

The prediction of the (t+1) time-point is given by:

$$\hat{Y}^{\circ}_{t+1} = \mathbf{X}^{\circ}_{\mathbf{t+1}}\hat{\mathbf{b}}_\mathbf{t} \qquad (3.1.2)$$

where

$\hat{\mathbf{b}}_t$ is the least squares estimator of β at time t, given by:

$$\hat{\mathbf{b}}_t = (\mathbf{X}_t'\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{Y}_t \qquad (3.1.3)$$

and

$\mathbf{X}^{\circ}_{t+1}$ is a 1xm vector of values of the regressors for the (t+1)time-point.

The variance of the prediction $\hat{Y}^{\circ}_{t+1}$ is, then, given by:

$$V(\hat{Y}^{\circ}_{t+1}) = \sigma_t^2\left\{\mathbf{X}^{\circ}_{t+1}(\mathbf{X}_t'\mathbf{X}_t)^{-1}\mathbf{X}^{\circ\,'}_{t+1} + 1\right\} \qquad (3.1.4)$$

where

$$\hat{\sigma}^2 = S_t^2 = \frac{\left[\mathbf{Y}_t - \mathbf{X}_t\,\hat{\mathbf{b}}_t\right]'\left[\mathbf{Y}_t - \mathbf{X}_t\,\hat{\mathbf{b}}_t\right]}{\left[\ell_t - m\right]} \qquad (3.1.5)$$

After the value Y°_{t+1} for the (t+1) time-point has been observed, the model to be used for predicting the value of the (t+2)time-point becomes:

$$\mathbf{Y}_{t+1} = \mathbf{X}_{t+1}\mathbf{b} + \mathbf{e}_{t+1}$$

where now the matrices X_{t+1} and Y_{t+1} are defined as :

$$\mathbf{X}_{t+1} = \begin{pmatrix}\mathbf{X}_t \\ \mathbf{X}^{\circ}_{t+1}\end{pmatrix} \quad \text{and} \quad \mathbf{Y}_{t+1} = \begin{pmatrix}\mathbf{Y}_t \\ Y^{\circ}_{t+1}\end{pmatrix}$$

with dimensions $(\ell_t + 1)$ x m and $(\ell_t + 1) \times 1$ respectively.

3.2 TESTING THE PREDICTABILITY OF A LINEAR MODEL

Xekalaki and Katti (1984) and Panaretos et al. (1997), used the difference between the observed and the predicted value of the dependent variable on every time point, to evaluate the predictability of a linear model.

Let

Y°_{t+1} : the observed value of the dependent variable for the (t+1) time-point.

$\hat{Y}^{\circ}_{t+1}$: the predicted value of the dependent variable for the (t+1) time-point.

Then it is known that (**Xekalaki and Katti (1984))** :

$$\hat{Y}^{\circ}_{t+1} - Y^{\circ}_{t+1} \sim N\left(0,\ S^2_t\left(\mathbf{X}^{\circ}_{t+1}\left(\mathbf{X}'_t\mathbf{X}_t\right)^{-1}\mathbf{X}^{\circ\,\prime}_{t+1} + 1\right)\right) \quad (3.2.1)$$

A statistical function that could be used for the evaluation of the predictability of the model is the function:

$$r_{t+1} = \frac{\hat{Y}^{\circ}_{t+1} - Y^{\circ}_{t+1}}{S_t\sqrt{\left(\mathbf{X}^{\circ}_{t+1}\left(\mathbf{X}'_t\mathbf{X}_t\right)^{-1}\mathbf{X}^{\circ\,\prime}_{t+1} + 1\right)}} \quad t = 0,1,2... \quad (3.2.2)$$

where S_t is given by (3.1.5).
Because of the (3.2.1) and (3.2.2) it's obvious that:

$$r_{t+1} \sim N(0,1) \quad (3.2.3)$$

Xekalaki and Katti (1984) considered as a scoring rule for the performance of the model the average of r_t^2 :

$$R_n = \frac{\sum_{t=1}^{n} r_t^2}{n} \quad (3.2.4)$$

If the r_t were independent and because of (3.2.3) then it is known that :

$$nR_n = \sum_{t=1}^{n} r_t^2 \sim \chi_n^2 \qquad (3.2.5)$$

Theorem (Brown (1975), Kendall (1983)).

If $e(t) \sim N(0,\sigma^2 I_t)$ then the quantities :

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

are i.i.d normal variables with mean 0 and variance σ^2. Also, the quantities :

$$r_{t+1} = {}^{W_{t+1}}\!/_{S_t} \qquad t = 0,1,2,\ldots$$

are independent variables, t-distributed, with $(\ell_t - m)$ degrees of freedom. **For large ℓ_t, the variables r_{t+1} t=0,1,2,... given by (3.2.2) are approximately standard normal variables which are mutually independent.**

According to the above theorem the assumptions are met for (3.2.5) to hold. So, the predictability of a linear model can be tested as follows:

$$\begin{Bmatrix} H_0\text{: the model is appropriate for predictions} \\ H_A\text{: the model presents lack of predictability} \end{Bmatrix}$$

The null hypothesis is rejected for large values of $\sum_{t=1}^{n} r_t^2$, that is when the value of $\sum_{t=1}^{n} r_t^2$ is at the right tail of the χ^2-distribution.

This hypothesis test, has been studied by **Box and Jeckins (1970)** and is called «Portmanteau test» (see paragraph 2.19) as well by **Spanos (1986)** taking into consideration a specific number of residuals.

3.3 COMPARING THE PREDICTABILITY OF TWO LINEAR MODELS

Let A and B two linear models which are given by (3.1.1). Suppose that we have observations for n_1 , n_2 time points respectively and we want to choose the more adequate model (the one that has the greater predictability). A statistical function appropriate for testing the hypothesis :

H_0 : the two models give equivalent predictions

H_A : model A predicts better than model B

based on the ratio of the average scores of the two models was given by **Panaretos et al. (1997):**

$$R_{n_1,n_2} = \frac{R_{n_1}(A)}{R_{n_2}(B)} \quad (3.3.1)$$

It is known that if $n_1 R_{n_1}$, $n_2 R_{n_2}$ were independent and χ^2-distributed with n_1, n_2 d.f, respectively, then the ratio R_{n_1,n_2} is F-distributed with n_1, n_2 d.f. On the other hand, the residuals of predictions $r_A(t), r_B(t)$ of the two linear models A, B given by (3.2.2) are not independent since they come from the same response. The following theorem is instrumental in developing the test.

Theorem. (Kibble(1941) Patil(1984)).

Let X_i , Y_i , $i=1,2,..,n$ be standard normal distributed random variables following jointly, the bivariate standard normal distribution. Then, the joint distribution of :

$$X = \sum_{i=1}^{n} X_i^2 \Big/ 2 \quad \text{and} \quad Y = \sum_{i=1}^{n} Y_i^2 \Big/ 2 \quad (3.3.2)$$

is Kibble’s bivariate Gamma type (1941) with probability density function :

$$f(X,Y) = \frac{e^{-\frac{X+Y}{1-\rho^2}}}{\Gamma(\kappa)\,\left(1-\rho^2\right)^{\kappa}} \sum_{i=0}^{\infty} \frac{\left(\rho/1-\rho^2\right)^{2i}}{\Gamma(i+1)\;\Gamma(i+\kappa)} (XY)^{\kappa-1+i} \quad (3.3.3)$$

where $\kappa=n/2$.

Considering the same number of time-points, according to the above theorem, we regard the correspondence :

$$X_i \longrightarrow r_A(t) \quad , \quad Y_i \longrightarrow r_B(t)$$

the joint distribution of the variables:

$$X = \sum_{i=1}^{n} X_i^2 \Big/ 2 \quad \text{and} \quad Y = \sum_{i=1}^{n} Y_i^2 \Big/ 2$$

is Kibble's bivariate Gamma distribution.

Panaretos et al. (1997) have shown (see appendix) that the ratio of X and Y :

$$Z = X/Y$$

follows a Correlated gamma ratio distribution with p.d.f :

$$f_{X/Y}(Z) = \frac{\left(1 - \rho^2\right)^{\kappa}}{B(\kappa, \kappa)} Z^{\kappa-1}(1 + Z)^{-2\kappa}\left(1 - \left(\frac{2\rho}{Z + 1}\right)^2 Z\right)^{-\frac{2\kappa+1}{2}} ,0 < Z < +\infty \quad (3.3.4)$$

Using the Correlated Gamma Ratio distribution, we may compare the predictability of two linear models:

$$H_0 : M_A \simeq M_B$$

$$H_A : M_A > M_B$$

meaning that

H_0: model A is equivalent to model B

H_A: model A is better than model B .

The advantage of model selection using the Correlated Gamma Ratio distribution, compared with the other methods, is that we do not have to know the functional form of the models that we compare.

In the following pages we present some plots of the Correlated Gamma Ratio distribution for different values of κ and correlation coefficient ρ.

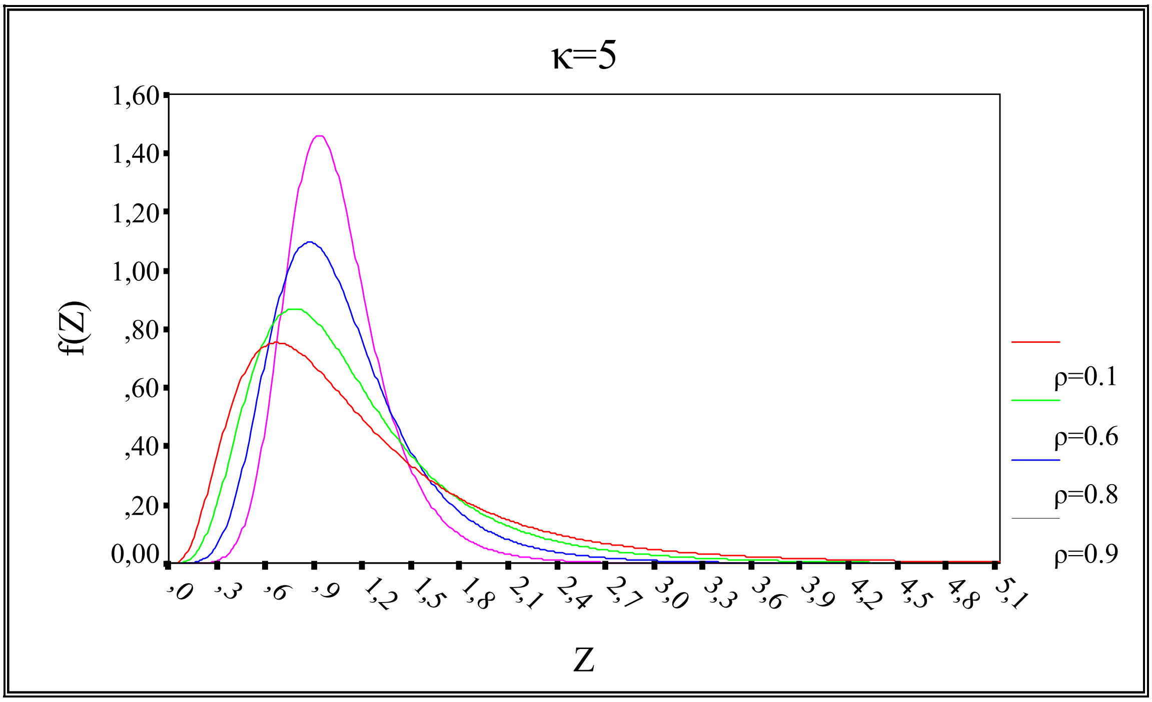

***Figure 3** : P.d.f of the Correlated Gamma Ratio Distribution for κ=5 and different values of the correlation coefficient .*

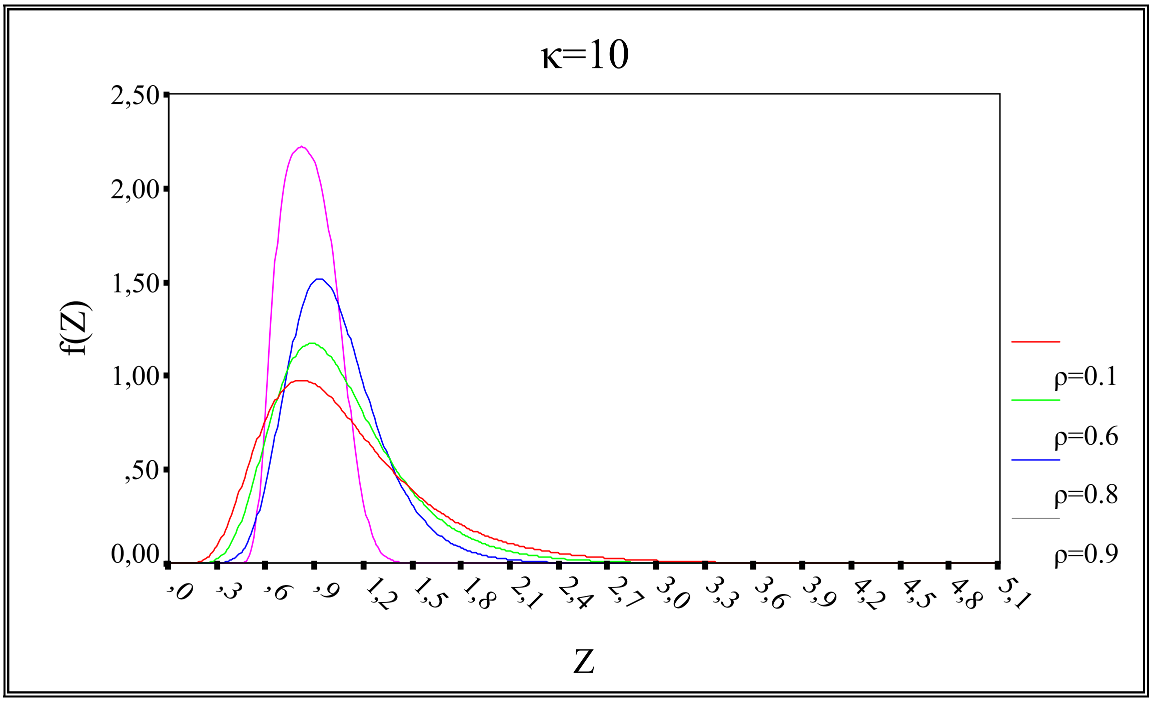

***Figure 4** : P.d.f of the Correlated Gamma Ratio Distribution for κ=10 and different values of the correlation coefficient .*

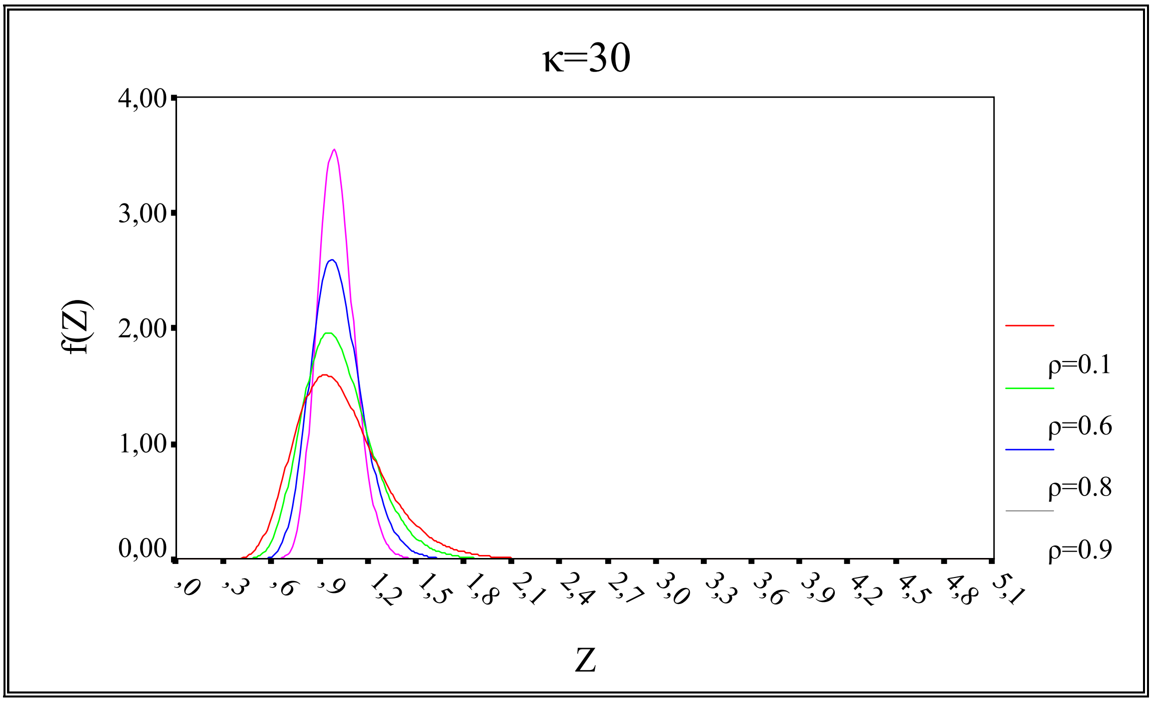

Figure 5 : *P.d.f of the Correlated Gamma Ratio Distribution for κ=30 and different values of the correlation coefficient.*

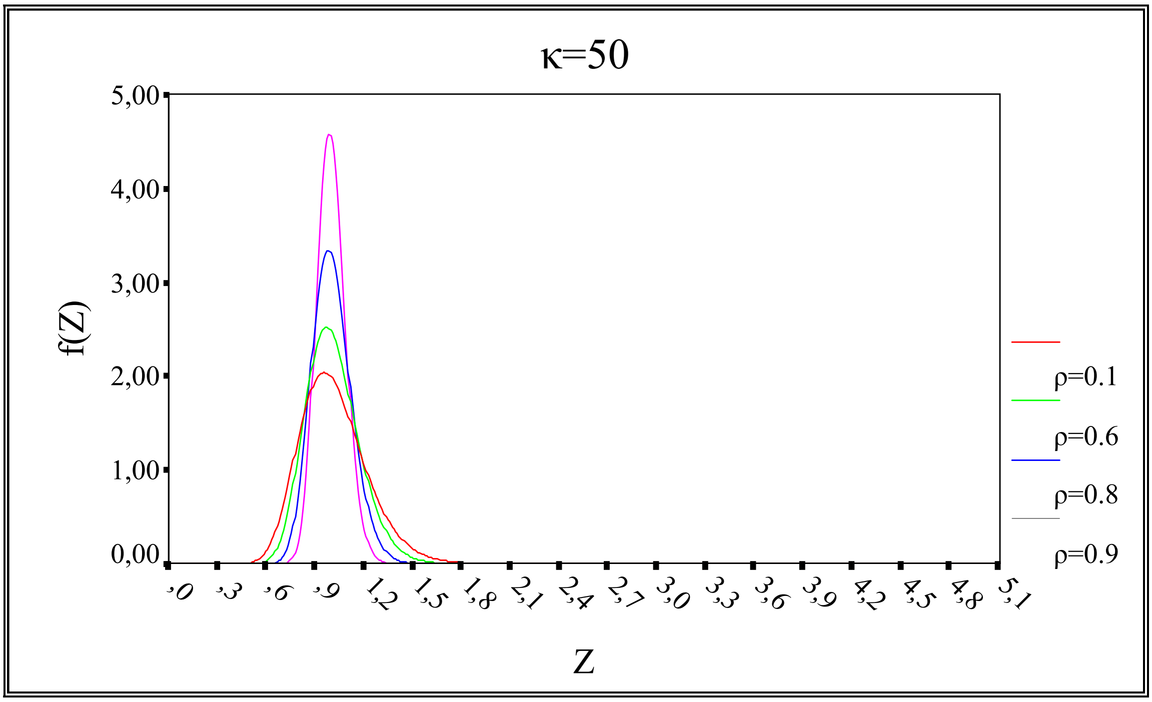

Figure 6 : *P.d.f of the Correlated Gamma Ratio Distribution for κ=50 and different values of the correlation coefficient.*