Chapter 5

DISCUSSION

In this dissertation, we studied methods that test the predictive ability of one linear model as well as methods that compare the predictive abilities of two competing linear models based on the χ^2 and the Correlated Gamma-Ratio Distribution. The main statistic for these tests is the Σr^2_t, the sum of squared standardized residuals of predictions.

Comparing the criteria reviewed in the second chapter and the one based on the Σr^2_t it seems that Σr^2_t is more appropriate for data that come from a time series.

Firstly, the criteria that test the descriptive ability and the predictive ability of a model use a different type of discrepancy. By estimating a model, the goal is to find out if it fits the data and make predictions. In data that come from a time series, the target is to make good predictions. So, it is more effective and logical to use these statistics that are informative of the effectiveness of the model in a straightforward way and not in an indirect way. Consequently, the type of discrepancy that is used for the evaluation of the predictability is more informative for the appropriateness of predictions than the one used by the criteria that evaluate the descriptive ability of a model.

Secondly, the criteria that test the descriptive ability, just test the goodness of fit of the linear model with reference to all the time points, while Σr^2_t

takes into consideration the predictions of all the linear models at each point in time.

Ideally, one may claim that the results between the comparable criteria should be close because if a linear model fits the data well then the predictions have to lie on the line estimated by the least squares method. However, in many cases, in practice, models that describe the data well, do not give satisfactory predictions.

From another point of view it is claimed that a non parsimonious model (the model that contains more than the appropriate number of predictors) does not give satisfactory predictions despite possible indications of the opposite given by the criteria that evaluate the descriptive ability of the models. This happens because by adding a great number of variables in a model, the "least squares method" obliges the data to fit better in the estimated line and the model has 'lack of fortuity'. Consequently, models that are parsimonious give better predictions. That is why in many cases the χ^2-test indicates 'appropriateness of predictions' for models containing a small number of predictors (see e.g. tables 5 and 9).

An additional advantage of the criterion based on Σr^2_t is that it does not take into consideration the functional form of the model while the other criteria do. Another advantage is that Σr^2_t is appropriate for all kinds of linear models (nested, non-nested and overlapping models) and it takes into consideration the correlations between the residuals through the Correlated Gamma Ratio-Distribution.

Two hypothesis tests that concern the statistical function Σr^2_t, were presented in the third chapter. The first one evaluates the predictability of one linear model. From the numerical results of the applications, it seems that this test has to be examined more from a

theoretical point of view. Besides, standardization of the residuals of predictions should be studied.

The other hypothesis test, compares the ratios of Σr^2_t of two linear models according to a new distribution, the Correlated Gamma-Ratio distribution which is a generalized form of the F distribution. Based on the data sets considered, this test appears to frequently result in equivalence of the predictive abilities of the compared linear models.

It seems that the results of the hypothesis tests would be more accurate if a greater number of time-points were available. This is justified because the residuals of predictions will be approximately standard normal variables when we have a large number of time-points.

Generally, a further theoretical analysis of the hypothesis tests and the Correlated Gamma-Ratio distribution may lead to answers to the above queries.

It seems that these criteria may work well for the evaluation of the predictability of the linear models.