Chapter 1

INTRODUCTION

The choice of the appropriate linear model before this can be used for planning and decision making, has been the concern of many statistical workers. Primarily, because of the current availability of high-speed computers, this problem has received considerable attention in the recent statistical literature. Most of the methods in the literature evaluate the descriptive ability of the candidate models.

An alternative approach is to evaluate the predictability of the models.

Let's suppose that the data are taken as a time series and that data are given until the time point t. Then an appropriate linear model is estimated by the least square method. The methods that evaluate the descriptive ability of the candidate models are based on the discrepancy between the observed and the estimated value of the dependent variable, for all time points t.

With respect to methods that evaluate the predictive ability of a linear model, an alternative approach is the following. For every time-point a model is estimated by the least square method. Then, the predictability of the model is evaluated, based on the discrepancy of the predicted value for the t+1 time-point and the corresponding observed value of the dependent variable. Consequently, a different kind of discrepancy is used than the one used to evaluate the descriptive ability of the models. A kind of discrepancy that examines the

efficiency of the model in giving predictions in a straightforward way.

Under some conditions, the sum of the squared discrepancies that evaluate the predictability is χ^2-distributed. Based on this statistical function, one may test the predictability of a linear model. (Panaretos, Psarakis and Xekalaki (1997))

Considering the ratio of the squared discrepancies for two linear models, one is able to compare the predictability of two linear models. The discrepancies are dependent since they arise from the same response. According to Kibble (1941) and Patil (1984) the joint distribution of the squared discrepancies is Kibble's bivariate Gamma distribution. It is proved that under some conditions this ratio is distributed according to a generalized form of the F-distribution. (Panaretos et al, (1997))

This model evaluation method presents some advantages compared to other model selection procedures. It allows us to compare nested (the set of predictors of one model is a subset of the set of predictors of the other one) or non-nested (there are no common predictors) or overlapping models (there are common predictors but none of the sets of predictors is a subset of the other) without necessarily knowing their functional form. It also takes into consideration the possible correlations of the residuals. On the other hand, it seems that this evaluation method is more appropriate than other model selection procedures, for data that are taken as a time series, since in time series the main purpose of the investigation is the prediction and not the goodness of fit to the data.

In the second chapter of this dissertation the most frequently used methods to evaluate the descriptive ability of the candidate models are presented.

In the third chapter methods of testing the predictability of one and two linear models are studied based on the χ^2 and the generalization of the F-distribution .

In the fourth chapter, two applications with real data that concern corn crops in the states of Indiana and Iowa of the USA as well as a simulation study are considered.

In the fifth chapter the advantages and disadvantages of the methods that assess the predictive ability of one and two linear models are discussed. They are also compared with other model selection procedures.