CHAPTER 1

INTRODUCTION AND SUMMARY

The method of time series analysis introduced in Box and Jenkins (1970) has become an often applied tool in empirical dynamic modeling in economics. The influential studies by Zellner and Palm (1974) and Prothero and Wallis (1976), in which links between theoretical economic models and time series models are established, seem to have provided convincing arguments for the usefulness of the analysis of univariate and multivariate time series for economic modeling. A recent example may be a textbook on macroeconomics by Blanchard and Fisher (1989), of which the first pages are dedicated to the inspection of the time series properties of some macroeconomic aggregates. Furthermore, a widely held view is now that insights in the behavior of univariate time series can be useful for the construction of models relating two or more variables, see, e.g., Haugh and Box (1977) and Davidson *et al.* (1978).

The Box-Jenkins approach to autoregressive integrated moving average (ARIMA) modeling consists of three stages, i.e. identification, estimation and diagnostic checking. It appears however that sometimes this approach may be not be appropriate. For example, their iterative model building process can yield too large or too small a model, and a series is often considered to be integrated even when it is not. It is readily imaginable that a wrong decision can have implications for forecasting or econometric model building. Things can get even more complicated in time series with seasonality, for which it is common practice to assume first order as well as seasonal integration.

In this thesis I will treat the issue of model selection and seasonality in time series in more detail. The main focus will be on appropriate strategies for selecting models for univariate time series with and without seasonality, which are alternatives or modifications to the Box-Jenkins method. Illustrations will be provided by numerous empirical applications, and also by Monte Carlo simulations. A subset of the data used in the applications is displayed in Appendix A.

Chapter 2 is dedicated to some technical preliminaries concerning the analysis of time series in the time domain. The literature on aspects of time series models is by now too extensive to survey in several pages, and hence I limit the preliminaries to discussing only those concepts which will be used in later chapters. Moreover, I will restrict the analysis to the framework used in the standard Box-Jenkins approach, not only for convenience but also because of its widespread use. I will therefore not explicitly consider alternative time series concepts such as those proposed in, e.g., Harvey (1989), who interprets the time series in terms of components like trends and cycles. The first part of chapter 2 deals with concepts in univariate time series. Since parts of this thesis make use of multivariate time series models, I discuss them in the second part of chapter 2. Most attention will go out to the Johansen (1988) cointegration approach, for it will be often applied in later chapters.

Chapter 3 considers model selection strategies for univariate time series without seasonality. First, I will discuss some current strategies. It will be argued that the Box-Jenkins approach and related methods may not be convenient. For example, Monte Carlo evidence shows that using the Hannan and Rissanen (1982) approach, one will often not be able to detect the true data generating process. Other simulation results show that the choice between two incorrect models with e.g., the Akaike (1974) and the Schwarz (1978) information criteria may largely depend on the unknown true model. This phenomenon may be explained using 'distance' concepts, see Sneek (1984) and, for analytic results in the linear regression context, Franses (1989a). Furthermore, it appears that the rejection rate of both incorrect models on the basis of checks on residual autocorrelation does not seem to be affected by the unknown data generating process. Together with some philosophical arguments, this leads to the formulation of an alternative, more robust, model selection strategy. The three successive steps in this strategy are to tentatively specificy a set of models, then to test these models for possible deficiencies, and finally to select a model which passes these tests using several criteria. The subsequent sections in this chapter 3 will consider each of these steps.

The determination of the set of models necessitates the sometimes difficult decision on the stationarity of a time series. Broadly speaking, stationarity implies that a time series fluctuates around a constant mean with constant variance, and it is an assumption for ARMA time series analysis. One remedy for nonstationarity is to transform the original series to first differences and base further model building on this transformed series. A drawback of automatically differencing a time series, however, is that overdifferencing results in a noninvertible model, or that a unit root is introduced in the moving average part of the model. This unit root causes an estimation bias with respect to the moving average parameters. This implies that testing for noninvertibility should preferably not be based on their estimates. In section 3.3, I therefore propose to use the sample autocorrelations of a possibly noninvertible moving average process, which can be consistently estimated without bias. Standard normal test statistics for noninvertibility of moving average processes can now easily be constructed. Application to several of the series in Schwert (1987) shows that his conclusion that many macroeconomic time series should be doubly differenced may not be valid.

The stationarity property of a time series is usually considered in a univariate context. It is however conceivable that a certain variable is related to other variables, and that its time series can be contained in a multivariate system. Given a multivariate time series model, it is possible to construct univariate time series models, the orders of which are generally larger than those of the multivariate model, see Zellner and Palm (1974). The test statistics for stationarity of the univariate series are often based on auxiliary regressions which use an approximative autoregressive model. In section 3.3, it will be analytically argued that applying too short such an approximation can result in the suggestion of a unit root in univariate series while their bivariate data generating system is stationary. Applications of the Johansen (1988) method to test for multivariate unit roots and of a univariate test to the well-known Lydia Pinkham data on sales and advertising illustrate this phenomenon.

After deciding on the appropriate transformation it is proposed to construct a set of models. This may be based on inspecting graphs and (recursive) autocorrelation functions. Once a set has been established it necessary to test for several of the underlying assumptions. Examples of these are that the residuals are white noise, that they are normally distributed, and that there are no significant outliers. This second step will be illustrated with modeling the annual unemployment rate series for the United States.

A formal treatment of statistical model selection criteria is given in the final section of chapter 3. It is argued that these criteria can be divided in, at least, two categories related to model use, i.e. forecasting and description. Generally, the forecasting criteria may differ from the, say, descriptive criteria, which are those for models which should only fit within the estimation period. A new forecasting model selection criterion will be proposed. Again, the unemployment rate series, for which there emerge three reasonably adequate rival autoregressive models, serves as an illustration. The detection of influential observations possibly effecting model selection is done with the method developed in Franses (1990a).

Additionally, the selection between a first order autoregressive and a first order moving average model will be treated in more detail. This decision can be of relevance in dynamic econometric model building. I will propose a new test statistic which is based on the estimated multiple correlation coefficient when an autoregressive model is fitted. This test statistic appears to be somewhat distinct from a rival statistic, and its empirical size and power seem to be more satisfactory. Finally, the conclusions of this chapter 3, as well as those of the next chapter, will be summarized in chapter 5.

Chapter 4 is entirely dedicated to aspects of model selection in case of time series with seasonality. Although the content of chapter 3 is of course also relevant to seasonal time series, the main focus in this chapter will be on the model for seasonality. To be more specific, consider a time series which is measured s times a year. Suppose further that the series is nonstationary. There are now several types of models which can deal with this seasonality and nonstationarity. The approach adopted in the traditional Box-Jenkins literature is to doubly difference the original series, i.e. to apply a seasonal and a first order difference filter. A second model requires the annual differencing filter only. A third model uses the first order filter to remove nonstationarity, while it includes a constant and s-1 seasonal dummy variables to cope with seasonality. In a fourth model the series is not transformed, and, e.g., trend, constant and seasonal dummies are incorporated.

Selection between these types of models is rather difficult, because most of them are not nested. This entails that the available model selection test statistics vary with the assumed null hypothesis. As with most tests for nonstationarity, a second problem is that the null hypothesis usually is a nonstationary model, and hence most standard statistical techniques are not valid. This implies that tables with critical values for the test statistics have to be generated via Monte Carlo simulations. In section 4.1, I propose an alternative model selection strategy which tries to overcome these difficulties. It will be argued that some simple, though often applied, models for time series with seasonality and nonstationarity are nested in the so-called airline model which is initially advocated in Box and Jenkins (1970). In this model, which appears to be appropriate in many practical occasions, a doubly differenced variable is modeled with a multiplicative moving average process in which the orders 1 and s are represented. When this moving average model is not invertible, one or both of the differencing filters is not necessary. The test statistics will be based on the sample autocorrelations of the moving average process, which causes that the statistics asymptotically follow standard normal distributions.

The application of this test procedure is however limited by the adequacy of the airline model. An approach which allows more complicated models is developed in Hylleberg *et al.* (1990). In their approach the null model is the model in which the series is annually differenced, and the alternative model can contain a variety of deterministic terms. The basic idea is that the annual differencing filter implies that the time series contains s unit roots, s-1 of which are so-called seasonal unit roots. In Hylleberg *et al.* (1990) their procedure is applied to quarterly time series. In section 4.2, I present extensions of their method to time series consisting of monthly and bimonthly observations. The tables with critical values for the test statistics are displayed in Appendix B.

The relevance of an appropriate model for seasonality and nonstationarity, as opposed to automatically doubly differencing, is reviewed in section 4.3. This relevance is largely illustrated with empirical examples. The first case considers forecasting, and its exposition is based on Franses (1990d). It is argued that with residual autocorrelations it may be difficult to distinguish between a multiplicative seasonal Box-Jenkins model (MSBJ) in which the variable is doubly differenced, and a model in which the variable is first order differenced and where seasonal dummies (FDSD) are included. Furthermore, it is shown with simulation evidence and with three empirical series that using an MSBJ model while an FDSD model is more appropriate results in systematic forecast errors, in the sense that the forecasts often will be either too high or too low.

The second case shows that establishing a relationship between two variables may be facilitated by appropriate seasonal models. In Heuts and Bronckers (1988) it is argued that the monthly industrial production index, when doubly differenced, does not help in forecasting and in explaining new heavy truck sales in the Netherlands. When the apparently suitable FDSD model for this index is used however, it does help in forecasting.

The third case considers the effect on linearity in case the MSBJ model is used instead of an adequate FDSD model. It can be shown that a small amount of additive outlying observations in the original series can provide that a doubly differenced time series shows characteristics typical for bilinear time series. A stunning example is given by the unemployment in Germany series which, when doubly differenced, is successfully modeled with a bilinear model in Subba Rao and Gabr (1984). Testing for seasonal unit roots shows that an FDSD model may be more appropriate. Application of a test for linearity indicates that now the model can be linear, albeit that there are several outliers.

An entirely new approach to model selection in seasonal time series is proposed, and illustrated with empirical examples, in the final section of chapter 4. The basic idea is that several univariate models imply the presence or absence of cointegration relationships between the seasons when their observations are treated as separate series. A side benefit is that a new way of drawing graphs of seasonal time series is introduced. For quarterly time series such a figure shows four patterns, each of which representing a certain quarter. One has now the opportunity to visually check whether it applies that 'summer becomes winter', as it is likely to occur for a seasonally integrated process. When the four lines behave rather similar and without intersections, this may be an indication that a model with seasonal dummies is appropriate.

Representing a univariate quarterly series with a multivariate series containing four series of annual observations per quarter provides that the Johansen (1988) method can be applied to test for cointegrating relationships. It can be shown that several univariate models are nested within such a multivariate model. An additional advantage is that this approach also specifies models between the seasonally and first order differenced models. This may be convenient because studies like Osborn (1990a) have shown that seasonal differencing may often not be appropriate, but that estimated parameters corresponding to seasonal dummies do not always follow constant patterns over time. In the second part of the final section of chapter 4, I extend this multivariate approach to the selection of bivariate models. An application is given by the selection of a United Kingdom consumption function.

Chapter 5 contains the conclusions and a brief summary of the major results.