A RANDOM PARAMETER APPROACH TO MODELING AND FORECASTING TIME SERIES

by

Deborah A. Guyton

Dissertation submitted to the Graduate Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

APPROVED:

Robert V. Foutz, Chairman

Raymond H. Myers

Richard G. Krutchkoff

Klaus Hinkelmann

N. T. Stephens

July, 1979

Blacksburg, Virginia

ACKNOWLEDGMENTS

The research reported in this dissertation was conducted at Virginia Polytechnic Institute and State University in Blacksburg, Virginia.

The author wishes to express her sincere appreciation to Dr. Robert V. Foutz for his patience, guidance, imaginative ideas and critical commentary during her years of graduate study. Appreciation is extended to the Department of Statistics for its financial support as well as the invaluable experience received as a GTA and GRA during her years of graduate study.

Thanks are expressed to other members of the committee, in particular to Dr. R. G. Krutchkoff and Dr. R. H. Myers for their helpful suggestions and constant encouragement. Thanks are also expressed to of the Statistics Consulting Center for her interest and assistance with computer programming problems, and to for her careful typing of this dissertation. Thanks are expressed to all friends interested in her progress, in particular to for the informative discussions with him.

Special thanks are extended to the author's parents, and , for their support and encouragement during her years spent in graduate school. And a special thank you is extended to W. S.

TABLE OF CONTENTS

CHAPTER I

A NEED FOR RANDOM PARAMETER MODELS

1.1 Introduction

A great deal of data in economics, the natural sciences and the social sciences is in the form of a time series where values of current observations are strongly dependent upon preceding observations. Economic behavior such as yearly gross national product, the monthly consumer price index, daily stock market prices, monthly sales and price figures, and data collection related to characteristics of the population are often quantified in time series. There are many examples of time series data in the natural sciences. Daily air temperature, monthly ozone levels in the stratosphere, the daily water level of a lake or stream, hourly yields of a chemical process are all recorded over time. The primary concern of a time series analysis is to study the nature of the dependence among the observations. Often the objectives of the study are to model the dependence structure of the time series and to use the model to develop optimal forecasts of future time series observations.

The ability to forecast optimally is of great practical importance. Just as optimal sales forecasts are needed for business planning, optimal forecasts of other time series are required for all types of scientific decision making. It should be emphasized though that the

precision of the forecast ultimately depends on the accuracy of the model in describing the dependence structure of the observations.

This dissertation investigates a new approach to the important problems of modeling and forecasting time series. A new class of models is presented which has promise for better fitting some time series with complicated dependence structures and for forecasting these time series more accurately than is possible with the currently used models from which the new class is derived. The theory and methodology needed to implement the new class of models for time series analysis and forecasting is investigated in depth. One particular model from this new class of models is chosen to illustrate the theory. A detailed description of an estimation and forecasting procedure for this specific model is given, and an evaluation is made of the promised superiority of this proposed model over standard models for time series analysis and forecasting.

Before going into further detail, it is necessary to present some basic definitions and important concepts for this study and to specify some restrictions which more clearly define the problem actually investigated. This can be accomplished while presenting a summary of current modeling and forecasting techniques. In section 1.2 a time series is discussed as a realization of a stochastic process. Stationarity and weak stationarity are defined with attention restricted to weakly stationary time series. The major tool for modeling a time series is the autocorrelation function when using the time domain approach to time series analysis. In section 1.3 the basic procedure for this type of analysis is given as well as the standard linear

stochastic models used in the analysis. An optimal predictor is also developed.

In section 1.4 a new random parameter model is motivated from a discussion of the standard AR(1) model and its characteristic behavior. Suggestions for using the new random parameter model to fit a set of data are discussed in section 1.5. Literature concerning this new model is scarce, but a review of literature discussing more general random parameter time series models is given in section 1.6.

1.2 Time Series and Stochastic Processes

A time series is a set of observations generated sequentially through time. If the set of observations is continuous, the time series is continuous; if the set of observations is discrete, the time series is discrete. For the purposes of this dissertation only discrete time series where observations are made at some fixed interval will be considered. One can number the equally spaced time intervals and denote the observations in the time series by the sequence $\{y_t | t \in I\}$ where $I = \{0, \pm 1, \pm 2, \ldots\}$ is the set of all integers. Discrete time series usually arise by sampling a continuous time series or accumulating a variable over a period of time.

In many situations a time-dependent phenomenon is due in part to unknown factors and cannot be described by a completely deterministic model which allows exact calculation of the future behavior of the phenomenon. For these phenomena, it is often possible to derive a model that can be used to calculate the probability of a future value lying between two specified limits. Such a probability model or stochastic

process is used for phenomena where future values can be described only in terms of a probability mechanism. An observed time series to be analyzed may be regarded as one sample realization from an infinite population of such time series that could have been generated by the underlying probability mechanism, or stochastic process.

The notion of dependence among observations in the time series is formalized by treating the observations as values taken on by the corresponding sequence of correlated random variables $\{Y_t | t \in I\}$ which make up a stochastic process. In particular, a finite set of observations $\{y_{t_1}, y_{t_2}, \ldots, y_{t_n}\}$ making up an equispaced time series can be described by an n dimensional random variable $\{Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n}\}$ from the collection $\{Y_t | t \in I\}$. This n dimensional random variable has a joint probability distribution $p(y_{t_1}, y_{t_2}, \ldots, y_{t_n})$.

A special class of stochastic processes, stationary stochastic processes, is based on the assumption that the process exhibits a state of statistical equilibrium about some constant mean level. A stochastic process is said to be strictly stationary if its behavior is unaffected by a change in time origin in the sense that the joint distribution of variables $Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n}$ associated with the n observations $y_{t_1}, y_{t_2}, \ldots, y_{t_n}$ is the same as the joint distribution of variables $Y_{t_1+k}, Y_{t_2+k}, \ldots, Y_{t_n+k}$ associated with the n observations $y_{t_1+k}, y_{t_2+k}, \ldots, y_{t_n+k}$ for all intervals k, all choices of indices $t_1, t_2, \ldots, t_n$ and all $(y_{t_1}, y_{t_2}, \ldots, y_{t_n})$ in the range of the random variable Y_t. One should note that the indices $t_1, t_2, \ldots, t_n$ are not necessarily consecutive. Thus for a discrete stochastic process to be strictly

stationary, the joint distribution of any set of observations must be unaffected by shifting all the times of observation forward or backward by any integer amount k.

Assume the mean and variance of these random variables exist. Then stationarity implies that the mean values are constant over time:

$$Ey_t = \mu \qquad \forall t. \qquad (1.1)$$

Without loss of generality, assume $\mu = 0$ for this study. One can always subtract the mean from each observation in the set and obtain a set of observations centered around zero. Also, stationarity implies that the covariances depend upon the time displacement or lag k but are independent of time t:

$$Cov(y_t, y_{t+k}) = Cov(y_0, y_k) = C(k) \qquad \forall t. \qquad (1.2)$$

If one only assumes that the random variables of the time series have the property

$$Var(y_t) = C(0) < \infty$$

and satisfy conditions (1.1) and (1.2) above, then the time series is said to be weakly stationary.

Most approaches to time series analysis only make the assumption of weakly stationary time series. If a process does not satisfy these conditions, it is nonstationary. It does not have an affinity for a mean or its variance is not necessarily finite. But many nonstationary processes can be transformed into weakly stationary ones and the usual

analysis done on the transformed observations. Therefore, for this study attention is restricted to weakly stationary time series.

The relationship between observations y_t and y_{t+k} from a weakly stationary time series is measured by the correlation coefficient ρ_k between the random variables Y_t and Y_{t+k}, where ρ_k is given by

$$\rho_k = \frac{C(k)}{C(0)} = \frac{Ey_t y_{t+k}}{Var(y_t)} \qquad \forall k.$$

A specification of the entire dependence structure in the time series is given in terms of the autocorrelation function, the collection $\{\rho_k | k \in I\}$. Once the autocorrelation structure of a series is known or has been estimated, a time series analysis proceeds to model this structure as accurately as possible.

1.3 Time Domain Approach to Time Series Analysis

The study of a stationary time series through the autocorrelation function ρ_k is the time domain approach to time series analysis. A basic procedure for the analysis is to

(i) estimate the autocorrelation function,

(ii) identify a model for the time series as accurately as possible by matching the estimated autocorrelation function behavior to the known characteristic autocorrelation function behavior of specific models,

(iii) estimate the parameters of the model,

(iv) and usually develop optimal forecasts for future observations of the time series using the chosen model.

One modern approach for modeling relies on the use of three classes of linear stochastic models. The class of autoregressive models of order p, denoted by AR(p), expresses the observation in the time series at time t as a linear combination of the past p observations plus a random disturbance u_t:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + u_t. \tag{1.3}$$

The coefficients $\alpha_1, \ldots, \alpha_p$ are fixed parameters, and the u_t's are often assumed to be independent identically distributed normal random variables with zero means and some common variance σ^2. The class of moving average models of order q, denoted by MA(q), expresses y_t as a linear combination of present and past random disturbances:

$$y_t = u_t - \psi_1 u_{t-1} - \ldots - \psi_q u_{t-q}. \tag{1.4}$$

The coefficients $\psi_1, \ldots, \psi_q$ in the model are the fixed parameters. A third class of mixed autoregressive, moving average models of order p,q, denoted by ARMA(p,q), expresses y_t both in terms of past y's and in terms of present and past random disturbances:

$$y_t = \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + u_t - \psi_1 u_{t-1} - \ldots - \psi_q y_{t-q}. \tag{1.5}$$

Note that models (1.3) and (1.4) are the special cases of (1.5) with $q = 0$ and $p = 0$, respectively. All three classes of models are discussed in depth in Box and Jenkins [1976, Ch. III].

The ARMA(p,q) models may be used to represent a great variety of autocorrelation structures. If one could include infinite order moving average models, MA(∞), then the correlation structure of every weakly

stationary time series (with continuous spectrum) could be exactly represented by a (possibly infinite order) moving average. This is discussed in detail in Koopmans [1974, p. 214]. However, since the parameters in the model must be estimated from the finite number of present and past observations of the time series, only finite parameter models are feasible; and a primary concern is in identifying a "parsimonious" model, a model with as few unknown parameters as possible.

Once a parsimonious model has been identified and the unknown parameters in the model have been estimated, the model may be used to forecast future time series observations. Consider the problem of forecasting an observation ℓ time units in the future, $y_{t+\ell}$, based on a (possibly infinite order) moving average model:

$$y_{t+\ell} = u_{t+\ell} - \psi_1 u_{t+\ell-1} - \cdots - \psi_\ell u_t - \psi_{\ell+1} u_{t-1} - \cdots$$

by linear combinations $\sum_{j=0}^{\infty} \psi_j u_{t+\ell-j}$ such that the expected mean squared error,

$$E[\hat{y}_t(\ell) - y_{t+\ell}]^2 \tag{1.6}$$

is minimized over all other forecasts that are linear functions of the present and past random disturbances, or equivalently the present and past observations, $y_t, y_{t-1}, \ldots$. The optimal forecast which minimizes (1.6) conditional on the data, $y_t, y_{t-1}, \ldots$, is

$$\hat{y}_t(\ell) = -\psi_\ell u_t - \psi_{\ell+1} u_{t-1} - \cdots \tag{1.7}$$

which is the projection of $y_{t+\ell}$ onto the linear space generated by the present and past random variables $u_t, u_{t-1}, \ldots$. The projection is

unique, neglecting zero probabilities, even with only a finite amount of data. In practice the forecast $\hat{y}_t(\ell)$ is computed by expressing the unobserved disturbances u_t, u_{t-1}, ... in terms of linear combinations of the observed time series y_t, y_{t-1}, ... and then substituting into (1.7). It is true that the solution to the minimization problem is also the conditional expectation of $y_{t+\ell}$ relative to the present and past observations y_t, y_{t-1}, ... :

$$\hat{y}_t(\ell) = E[y_{t+\ell} | y_t, y_{t-1}, \ldots].$$

One may recall that projections have also been called wide sense conditional expectations. This optimal forecast and its properties are discussed thoroughly in Doob [1953, Ch. XII].

Since attention is focused on the AR(1) model throughout the dissertation, it is useful to consider its optimal predictor, or forecast. For this special case it is much easier to take the conditional expectation of $y_{t+\ell}$ in its autoregressive form rather than expressing it as a (possibly infinite order) moving average. Thus, for example, the one-step-ahead predictor from time t is

$$\begin{aligned}\hat{y}_t(1) &= E[y_{t+1} | y_t, y_{t-1}, \ldots] \\ &= E[\alpha y_t + u_{t+1} | y_t, y_{t-1}, \ldots] \\ &= \alpha y_t\end{aligned}$$

since the expectation of the future random disturbance u_{t+1} at time t is zero. Upon estimation of α, the estimate of the one-step-ahead predictor for the AR(1) model is given by

$$\hat{y}_t(1) = \hat{\alpha} y_t. \tag{1.8}$$

Since it is assumed that the random disturbances follow a normal distribution with zero mean and some common variance σ^2, then from Box and Jenkins [1976, Ch. V] approximate $1-\varepsilon$ probability limits $\hat{y}^{-}_{t+1}$ and $\hat{y}^{+}_{t+1}$ for the future observation one step ahead, y_{t+1}, are given by

$$\hat{y}^{\pm}_{t+1} = \hat{y}_t(1) \pm U_{\frac{\varepsilon}{2}}\, s,$$

where $U_{\frac{\varepsilon}{2}}$ is the deviate exceeded by a proportion $\frac{\varepsilon}{2}$ of the standard normal distribution and s^2 is an estimate, based on a reasonably sized sample, of the random disturbance variance σ^2.

Again, it should be emphasized that the precision of the forecast ultimately depends on the accuracy of the model, and the optimal property of the forecast in (1.7) depends on the model exactly fitting the time series. In the next section a new class of time series models is motivated. This class has promise for parsimoniously fitting some time series with complicated autocorrelation structures and for forecasting these time series more accurately than is possible with the present models.

1.4 A Random Parameter Model

It is suggested that a more general approach to modeling time series is needed for which the coefficients do not remain fixed but change over time. This approach would include a wide range of possible random parameter models. However, only one special class of random parameter models is investigated in this study. This generalization of

standard time series models is motivated by considering the simple autoregressive model of order one and some of its properties including its characteristic behavior. An example is then given of a time series whose characteristic behavior cannot be modeled well with a fixed coefficient model.

Consider the simple autoregressive model of order one,

$$y_t = \alpha y_{t-1} + u_t, \qquad t \in I \qquad (1.9)$$

where α is a fixed constant such that $|\alpha| < 1$ to ensure stationarity. The u_t's are independent identically normally distributed random disturbances with zero means and some common variance σ^2. The autocorrelation structure of this AR(1) model is of the simple form

$$\rho_k = \alpha^{|k|}, \qquad k = 0, \pm 1, \pm 2, \ldots .$$

Thus the autocorrelation function is completely determined by the value of α.

To illustrate the relationship between α and the corresponding modeled time series behavior, take $\alpha = .9$. This large positive correlation of $\rho_1 = .9$ between successive variables in the time series may be interpreted to mean that "larger than average" time series observations tend strongly to be immediately followed by "larger than average" observations and "smaller than average" time series observations tend to be immediately followed by "smaller than average" observations. With $\rho_2 = (.9)^2 = .81$, large (or small) observations tend to be again followed after two time periods by large (or small) observations. Since

$\rho_k = \alpha^{|k|}$ where $\alpha^{|k|} > 0$ for all k, one has a similar interpretation for the relationship between observations k time periods apart. This interpretation indicates the smooth, slowly changing time series behavior of Figure 1.1.

Now consider $\alpha = -.9$. Now $\rho_1 = -.9$, and successive variables are strongly negatively correlated. Variables two time units apart have strong positive correlations of $\rho_2 = (-.9)^2 = .81$, then $\rho_3 = (-.9)^3 = -.729$ is negative, $\rho_4 = (-.9)^4 = .6561$ is positive, ... Intuitively, this autocorrelation structure calls for "large" observations tending strongly to be immediately followed by "small" observations and then by "large" observations after two time periods, and then by "small" observations after three time periods, ... This interpretation indicates a rapidly changing time series as exemplified by Figure 1.2.

The intuition used above may be pursued to argue that as α ranges from +1 to -1, the behavior of the corresponding AR(1) model changes from that of a slowly changing, smooth time series to that of a rapidly changing time series.

Now consider the observations in Figure 1.3. It illustrates a more complicated time series behavior that appears to "randomly" combine degrees of smooth and then rapidly changing behavior. The intuition developed for the behavior of an AR(1) model in (1.9) indicates its inadequacy for modeling the complicated behavior exhibited in Figure 1.3. Similar intuition for low order MA models would also indicate their inadequacy. It should be repeated that, in theory, stationary time series behavior of the type exhibited in Figure 1.3 can be precisely modeled by a (possibly infinite order) moving average. However, high order

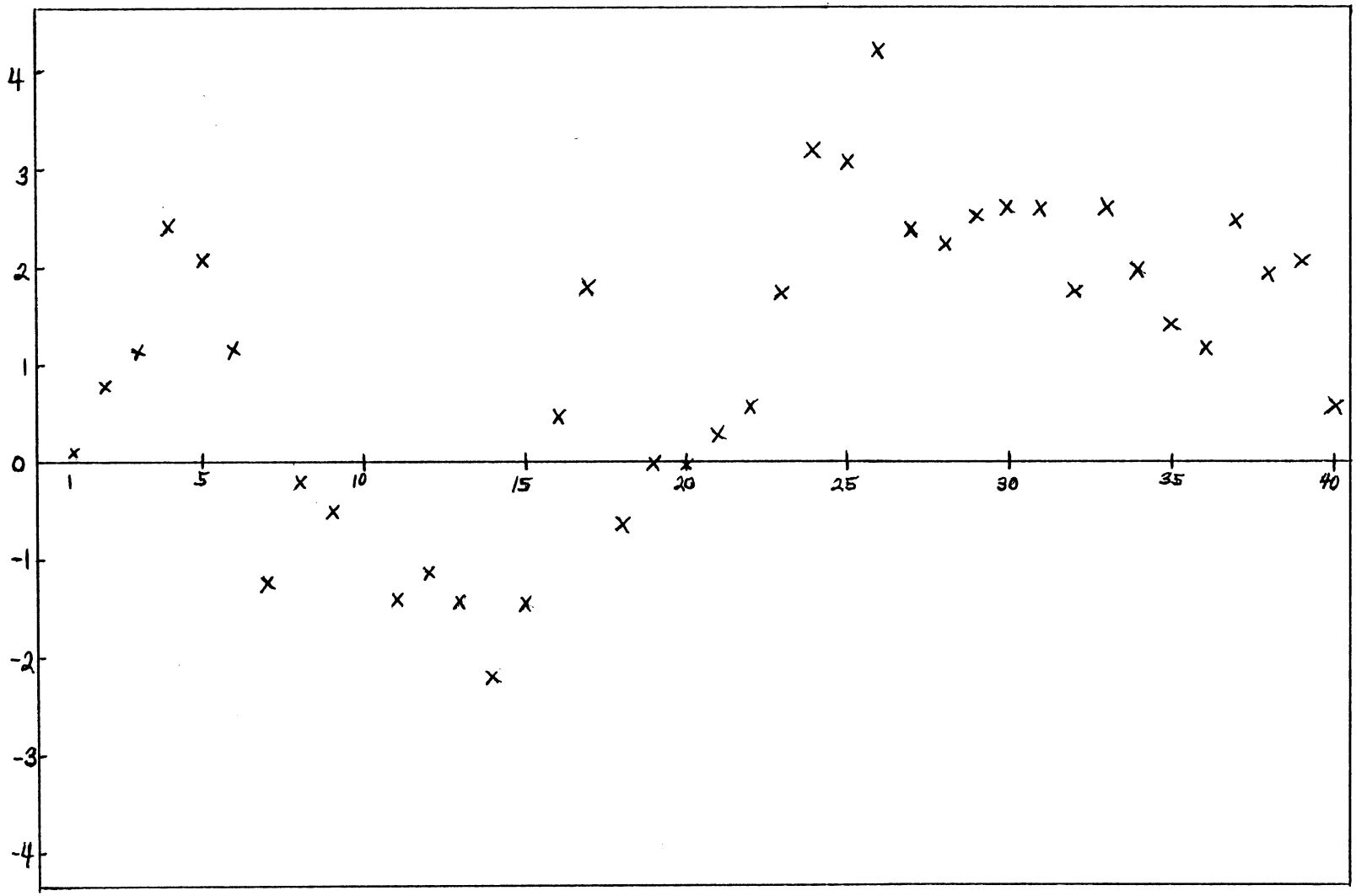

Figure 1.1. Data Generated From AR(1) Model With $\alpha = .9$, $\sigma = 1.0$.

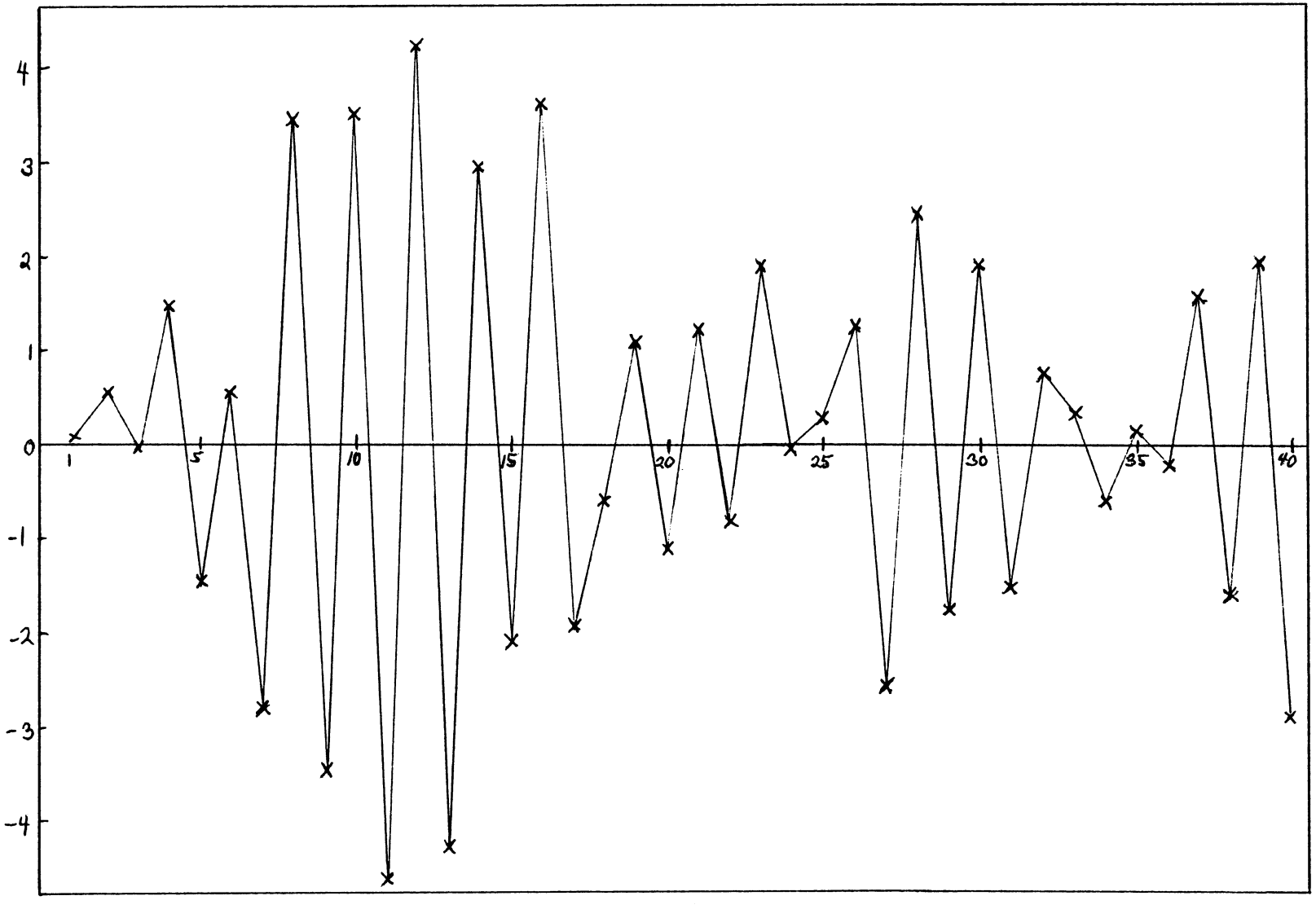

Figure 1.2. Data Generated From AR(1) Model With $\alpha = -.9$, $\sigma = 1.0$.

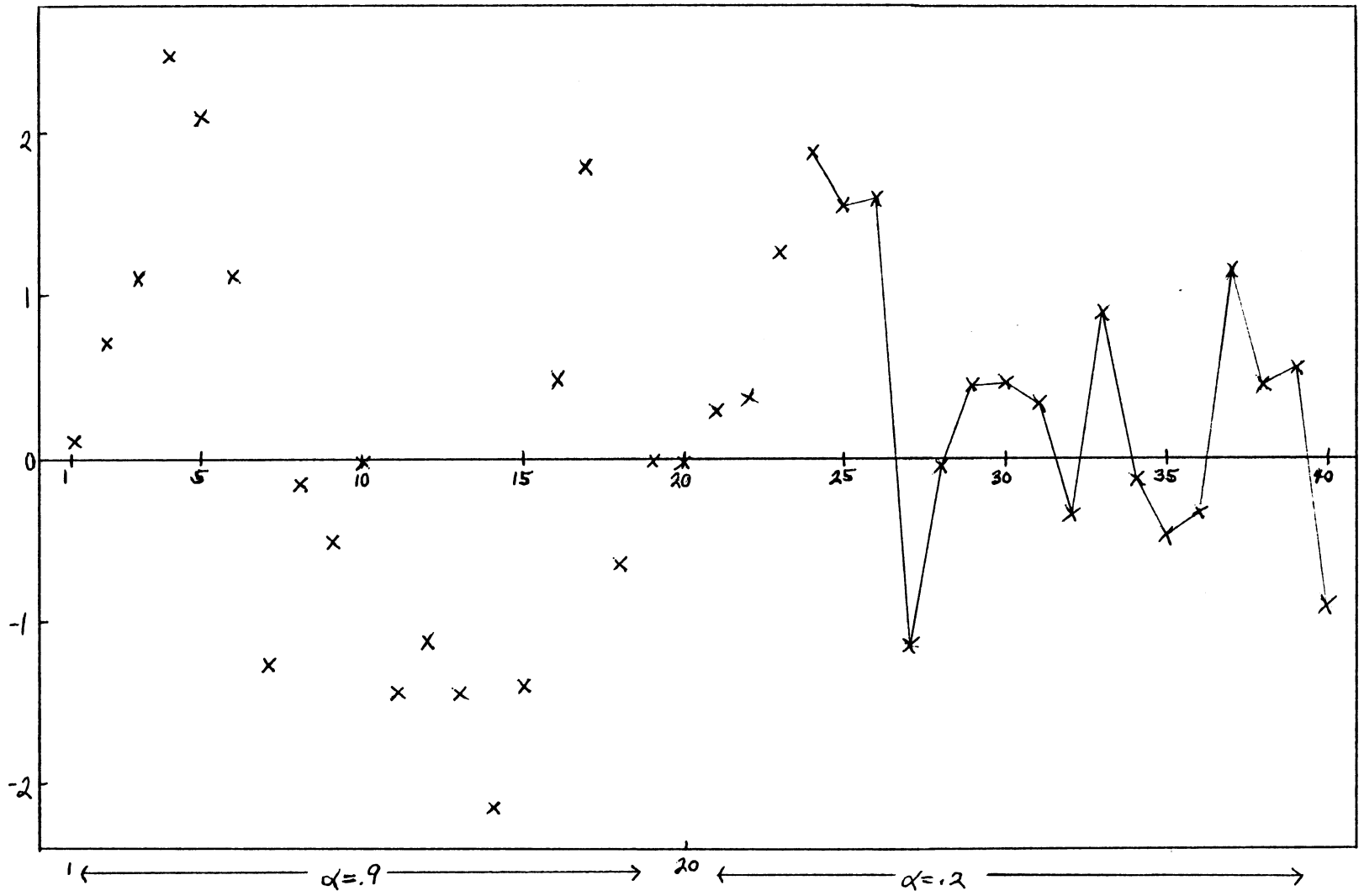

Figure 1.3. Data Generated From AR(1) Model With $\sigma = 1$ and Indicated α's.

models are useless when the parameters must be estimated from a finite number of observations; thus, parsimonious models are to be preferred.

Figures 1.1, 1.2, and 1.3 are now to be used to propose a new, parsimonious generalized autoregressive model of order one for the behavior exhibited in Figure 1.3. Compare Figure 1.1 with the smooth part of the time series in Figure 1.3 between $t = 1$ and $t = 24$. Intuition indicates that, perhaps, an AR(1) model with α near +1 might fit this initial smooth behavior. However, from $t = 25$ to $t = 40$ the time series behavior becomes more and more variable and exhibits the behavior of an AR(1) model as α moves from +1 toward -1 (heuristically speaking). Similarly, behavior along different segments of the time series of Figure 1.3 resembles that of various AR(1) models with differing α values between +1 and 0 in this case.

These heuristic observations motivate a generalization of the AR(1) model where the coefficient α itself changes stochastically with time. The theory for these generalized AR(1) models is developed in Chapter II. To illustrate the theory, a specific model for the random coefficient is chosen in Chapter III. It is shown that the generalized AR(1) model is a weakly stationary stochastic process (with continuous spectrum) under certain conditions, and therefore has a (possibly infinite order) moving average representation. The existence of this representation is all that is needed to construct the optimal forecasts given in (1.7). To the extent that the new model is superior to the existing AR(1) model for describing a given time series, the corresponding forecasts from the new model promise to be superior to forecasts constructed from the AR(1) model.

1.5 Identification

The approach for identifying standard ARMA models is to match the estimated autocorrelation structure of a given time series to the known characteristic structure of a particular model. In order to identify a generalized AR(1) model as opposed to the usual AR(1) model, one might take a similar approach. In the previous section it is indicated that a time series which exhibits "random" changes in autocorrelation structure, yet appears to remain stationary, is characteristic of the new generalized AR(1) model. This behavior would only be recognizable, however, if the "random" changes were fairly slow on the average.

One must realize that the usual AR(1) model with fixed coefficient is a special case of the generalized AR(1) model! In fact, it is of some interest to note that every set of data simulated in this study from the specific generalized AR(1) model to be discussed later would be identified as an AR(1) model by the procedures used in Box and Jenkins [1976]. However, this is not meant to be a conclusive result. It seems reasonable to try to fit a generalized AR(1) model whenever the usual procedures indicate fitting an AR(1) model. The generalized AR(1) model defines a richer class of models from which to choose when fitting a time series. It makes sense to try to obtain the best fitting model from this richer class!

1.6 Literature Review

There are few references in the statistical literature concerning linear stochastic time series models with random parameters. The few references that exist, in general, are more concerned with methods of

tracking values for the randomly varying parameters than with procedures for modeling the underlying behavior including the dependence structure of the random parameters. The proposed generalized AR(1) model with weakly stationary solution y_t investigated in this study appears to be a new model not specifically discussed in the literature. One should note that there is no interest in tracking the values the random parameter takes on in this study. Instead, the underlying probability mechanism generating the random parameter is given a specific structure; and a model is estimated for use in developing an optimal predictor for future time series values.

There is a recent book entitled Dynamic Stochastic Models From Empirical Data by Kashyap and Rao [1976] which briefly discusses some random parameter time series models. In Chapter III types of dynamic stochastic models are described. At the end mention is made of models with time-varying coefficients. In particular an AR(1) model is presented in which the time sequence of values assumed by the coefficient is itself random, obeying a dynamical system as described in Kalman [1963]. Consider the process y_t obeying the first-order time-varying AR model given by

$$y_t = a_0(t) + a_1(t)y_{t-1} + w_t \tag{1.10}$$

where $\{a_0(t),\ t = 1, 2, \ldots\}$ and $\{a_1(t),\ t = 1, 2, \ldots\}$ are sequences obeying the constant coefficient AR models

$$a_i(t) = b_i(t) + c_i,$$

$$b_i(t) = \gamma_i b_i(t-1) + \eta_i(t), \qquad i = 0, 1;\ t = 1, 2, \ldots \tag{1.11}$$

where c_i, γ_i, i = 0, 1 are constants and $\{\eta_i(t), t = 1, 2, \ldots\}$ are the usual zero mean independent and identically distributed random sequences, independent of each other as well as of w_t. Note that the constants γ_i are either less than one, which gives stationarity for $b_i(t)$, or equal to one. Kashyap and Rao state that equations (1.10) and (1.11) can be generalized to handle higher order AR models. They also indicate that "very little work has been done on the analysis of the process" y_t in (1.10) and (1.11) "for properties such as stability and stationarity."

Later in Chapter VI when discussing estimation techniques, Kashyap and Rao propose an estimation procedure for systems with slowly varying coefficients. They extend the earlier AR model to difference equations with time-varying coefficients:

$$y_t = a_1(t) + \sum_{j=2}^{n_1} a_j(t) y_{t-j+1} + \sum_{j=n_1+1}^{n_1+n_2} a_j y_{t-j+1} + w_t, \qquad (1.12)$$

where a_j, $j=n_1+1, \ldots, n_1+n_2$, are constants and $a_j(t)$, $j = 1, \ldots, n_1$, are the time-varying coefficients. The w_t, $t = 1, 2, \ldots$, are the usual zero mean Gaussian disturbances with variance ρ. The time-varying coefficients are modeled as

$$a_j(t) = b_j(t) + c_j, \qquad j = 1, \ldots, n_1 \qquad (1.13)$$

where c_j, $j = 1, \ldots, n_1$, are constants and $\{b_j(t), j = 1, \ldots, n_1\}$ are sequences of zero mean random variables which are independent of w_t for all t and obey

$$b_j(t) = \gamma_j b_j(t-1) + \eta_j(t), \qquad (1.14)$$

where $\{\eta_j(t)\}$ are independent and identically distributed sequences with zero mean and variance q_j. Again, it is mentioned that there has been very little study on the properties of the system given in (1.12), (1.13), and (1.14).

Kashyap and Rao then express the model in a standard form for the application of Kalman theory and give an expression for the likelihood function. However, it is stated that obtaining the maximum likelihood estimates of all of the unknowns using the given observations is a "formidable maximization problem in general", since one must maximize the likelihood function not only with respect to the unknown constants but also the random variables $b_j(t)$ for all j and all t. Two cases are considered. First when γ_j, $j = 1, \ldots, n_1$, and $q_j\rho^{-1}$ are assumed known, an elegant recursive solution is given in terms of the Kalman theory. The $b_j(t)$'s, c_j's, $j = 1, \ldots, n_1$, and a_j's, $j = n_1 + 1, \ldots, n_1+n_2$, are instantaneously estimated or "tracked" in time using a recursive discrete Kalman filter algorithm. In the second case, γ_j, q_j, and ρ are unknown. In general only approximate solutions can be obtained. One can approximate the estimate of ρ by estimating the corresponding quantity in the corresponding constant coefficient system. Again Kalman filtering is used to "track" estimates for the constants c_j's and a_j's and the random variables $b_j(t)$'s in a recursive manner. During the tracking these estimates are assumed to be the actual values and used to "track" estimates of the $\gamma_j(t)$'s and the $q_j(t)$'s at the same time.

Concerning properties of the estimates in the first case described above, one can obtain a sequence that plays the same role as the

residual sequence in constant coefficient systems. Testing to see whether this sequence is random noise is a check on the validity of the model.

From the engineering literature in "Recursive Approaches to Time Series Analysis", Young [1974] surveys recursive least squares analysis in its various forms and discusses recent developments which have particular relevance to the estimation of parameters characterizing discrete and continuous time series models of stochastic dynamic systems. This includes Kalman's work on optimal state estimation and filter theory. Young states that in effect, Kalman utilizes "the principle of orthogonal projection to evolve a more general form of the recursive least squares equations for the case where the unknown parameters are no longer considered constant coefficients ... but are treated as inherently time-variable states described by a general set of linear stochastic state equations." Time varying coefficients are assumed to vary between sampling instants in a manner that can be described by a stochastic difference equation. Hopefully the form of the equation is available from a priori information. Often an AR(1) model is assumed with constant coefficient less than one in absolute value. If not, a random walk model, which is nonstationary but seems to allow for smooth changes in the parameter, is usually used.

In order to apply Kalman's algorithms, the problem is first expressed in a vector-matrix formulation from which recursive equations can be derived. Initial estimates are needed. Then values for the constant coefficients as well as time-varying coefficients are estimated

by "tracking" possible parametric variations over the observation interval. This may require more than one run through the data where at the end of each run initial estimates are updated. The algorithm not only supplies the parameter estimates at each sampling instant, but also provides an indication of the accuracy of these estimates by tracking an error covariance matrix. Often plots are made of the tracked coefficients to check for constancy as well as plots made of the tracked time-varying parameters.

In conclusion, Kalman's algorithms are best applied to estimation problems where there is prior information on the internal structure of the dynamic system, or some structure given to the time-varying parameters, which allows for a formulation of a state-space model of the system. After applying the necessary algorithms which track the time-varying parameters, one has only a sample realization of the time-varying parameters. The most important point to be made is that there is no attempt using Kalman filtering theory to model the underlying probability mechanism that generates the time-varying coefficients using the observed data.

Ledolter [1978] discusses recursive approaches to parameter estimation in some time series models. He presents an algorithm for the recursive estimation of parameters in ARIMA, autoregressive integrated moving average, models. If the data represented by an ARIMA model is differenced a necessary number of times, then the resulting stationary differences of the original data can be represented by an ARMA model. Thus the ARIMA model is nonlinear in the parameters. Ledolter argues that if data is collected sequentially, one would prefer recursive

estimation procedures which update the values of the parameters as each new observation becomes available. Therefore he uses Extended Kalman filtering to develop an algorithm that recursively linearizes the ARIMA model and tracks the stochastically time-varying parameters in order to update parameter estimates. An example is given which uses the procedure for the MA(1) model with constant coefficient. Thus Ledolter makes no attempt to model the underlying probability mechanism of the random coefficient.

In summary, all of the references mentioned above give procedures which use some form of Kalman filtering theory. Instantaneous estimation of the stochastically time-varying parameters seems to be the main concern. No interest is taken in modeling the underlying probability mechanism of the random coefficients from the observed data. Nor is there interest in finding conditions that give weak stationarity for a system of this type. The above procedures seem to assume nonstationarity of the entire process. There seems to be very little concern for the properties of the system.

One more article in the engineering literature merits discussion. Zadeh [1950] gives a theorem which states the relation between the covariance functions of the random disturbance input and the observed output of a linear time-varying-parameter system. He works in the frequency domain approach using spectral theory so all details will be omitted here. Although he works with continuous time, his results can be shown to be true for discrete time. Using spectral representations Zadeh first gets a conditional expectation, conditional on the time-

varying parameters a(t), for the product of the output $y_t y_{t+k}$ for given lag k as follows:

$$E(y_t y_{t+k} | a(t), a(t+k)).$$

By taking expectations again, Zadeh obtains the unconditional covariance structure of the output as a function of the covariance structure of the input. In effect, by conditioning on the time-varying coefficients one is specifying one particular linear stochastic model with fixed coefficients for that instant of time. This is done for each instant of time. Then information from all of these instantaneous models, one for each point in time, is averaged to give the desired result, the unconditional covariance function. Thus Zadeh's linear time-varying-parameter system is a whole set of fixed coefficient models, one for each instant of time. Note that some of the fixed coefficients may be the same, and the model may be identical at several points in time. One should also note that assuming the input is weakly stationary, the covariance function of the output is only a function of the lag k and independent of time. Zadeh's results are general in the sense that they apply to weakly stationary linear stochastic time series models where the coefficients, random or deterministic functions of time, are not given any specific model.

Koopmans [1974, Ch. IV] also mentions the models for which Zadeh's result applies. He gives an example of a linear differential equation with time-dependent coefficients. A number of physical situations can be described by this type of equation. For example, consider a simple

structure subjected to random vibrations. Its characteristics could change due to fatique damage. Thus the response of the system would be described by a differential equation where the time varying coefficients depend on the "stress history" of the structure.

It is evident that very little work has been done in the area of linear stochastic models with random coefficients. Little mention of them is made in the statistical literature. A few informative references from the engineering literature have been given. The reader should note that this is not meant to be an extensive review of the engineering literature.

CHAPTER II

THE EXISTENCE OF AN EXPLICIT SOLUTION

It is informative to consider a more general time series model with random coefficients. Just as the fixed coefficient time series model is a member of a larger class of stochastic linear difference equation models, the time series model with random coefficients is a member of a larger class of stochastic linear difference equation models with random coefficients.

In section 2.1 stochastic linear difference equations are defined on specific domains. What is meant by a solution to a stochastic linear difference equation is discussed; and then given certain conditions, the existence and uniqueness of such a solution is proved. First, solutions are exhibited that depend on the starting point of the time series. An explicit expression for such a solution conditional on a starting point is constructed.

Under certain conditions one can obtain a solution that is not conditioned on a specified starting point. In section 2.2 necessary and sufficient conditions for the existence of an unconditional solution with finite second order moments are presented, and an explicit form for such a solution is given. Finally, sufficient conditions on the random coefficients are given to ensure an unconditional solution that is weakly stationary.

This chapter is based on the theory discussed in Chapters 1 and 4 of Linear Difference Equations by Kenneth S. Miller [1968]. The coefficients in his equations are deterministic functions of time, and constant in the special case. Considered here are equations involving coefficients that are random functions of time, and deterministic in the special case. Since much of the argument here parallels Miller's presentation, similar format and notation is used for the reader's comparison.

2.1 Stochastic Linear Difference Equations

For a sequence $\{y_t \mid t=0, 1, 2, \ldots\}$ the equation

$$y_t + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_n y_{t-n} = w_t, \qquad t = n, n+1, \ldots$$

where the α_i are real constants, $\alpha_n \neq 0$, and w_t a real function of time t is a "linear difference equation of order n" with constant coefficients. By letting the coefficients, the α_i's, be real functions of time t one obtains a more general linear difference equation. If w_t is a random variable, a random function of time t, then one has a "stochastic linear difference equation." These equations are discussed in Miller's book. However, for the purposes of this dissertation, let the coefficients, the α_i's, be random functions of time. The equation

$$y_t + \alpha_1(t) y_{t-1} + \alpha_2(t) y_{t-2} + \ldots + \alpha_n(t) y_{t-n} = w_t, \quad t = n, n+1, \ldots$$

where the $\alpha_i(t)$'s are random functions of t and w_t a random function of t is a stochastic linear difference equation with random coefficients. Attention is restricted to a first order stochastic linear

difference equation with one random coefficient, but the theory developed in this chapter can be generalized in a straightforward manner.

One should make note of the following additional restriction. In a difference equation the argument t varies continously; whereas, in the above equations the argument t takes on only (equally spaced) discrete values, t = ..., -1, 0, 1, ... Thus the theory presented in this chapter is actually for the class of linear recurrence relations which is a special class of linear difference equations.

Let $I = \{..., -1, 0, 1, ...\}$ be the set of all integers. Let $I_a = \{a, a+1, a+2, ...\}$ where $a \in I$. Let $\alpha(t)$ and w_t be defined for all $t \in I_{a+1}$. Let y_t have domain I_a. Consider the stochastic linear recurrence equation

$$y_t = \alpha(t)y_{t-1} + w_t \tag{2.1}$$

where the coefficient $\alpha(t)$ is a random function of time taking on a nonzero real value. Let w_t be a member of a stochastic process $\{w_t \mid t \in I_{a+1}\}$ taking on some real value. If the random variable w_t is not identically zero for $t \in I_{a+1}$, then (2.1) is called a "nonhomogeneous equation." The associated "homogeneous equation" is

$$y_t = \alpha(t)y_{t-1}. \tag{2.2}$$

A "solution of (2.1) for $t \in I_a$" is a stochastic process η_t with the properties that η_a is some well-defined random variable and $\eta_t = \alpha(t)\eta_{t-1} + w_t$ for t = a+1, a+2, ...

Theorem 2.1: Let $a \in I$ and let X be some well-defined random variable. Then there exists one and only one stochastic process $\phi(t)$ defined for $t \in I_a$ such that

$$\phi(a) = X$$

and

$$\phi(t) = \alpha(t)\phi(t-1) + w_t$$

for all $t > a$.

Proof: Define $\phi(a)$ as X and define $\phi(a+1)$ as $\alpha(a+1)X + w_{a+1}$. Then recursively define $\phi(a+n)$ for n a positive integer. Then

$$\phi(a) = X$$

and for $t > a$

$$\phi(a+1) = \alpha(a+1)X + w_{a+1} \tag{2.3}$$

and

$$\phi(t) = \alpha(t)\alpha(t-1) \ldots \alpha(a+1)X + \sum_{r=a+1}^{t-1} \alpha(t)\alpha(t-1) \ldots \alpha(r+1)w_r + w_t, \qquad t = a+2, a+3, \ldots .$$

Thus $\phi(a) = X$ and $\phi(t) = \alpha(t)\phi(t-1) + w_t$ for $t > a$.

Now to prove uniqueness suppose another stochastic process $\psi(t)$, defined for $t \in I_a$, satisfies (2.1) for $t > a$ and has the property that $\psi(a) = X$. Then $\psi(a) = \phi(a)$ and from

$$\psi(t) = \alpha(t)\psi(t-1) + w_t$$

with $t = a+1$ one obtains

$$\psi(a+1) = \alpha(a+1)\psi(a) + w_{a+1}$$

$$= \alpha(a+1)X + w_{a+1}.$$

Thus $\psi(a+1) = \phi(a+1)$. Now, using mathematical induction one can show that $\psi(a+n) = \phi(a+n)$ for all non-negative integers n.

Before going further with the nonhomogeneous equation, it is informative to look at the homogeneous equation and some properties of its solution. Let y'_t be a solution of (2.2) for $t \in I_a$. Then

$$y'_a = X$$

for some well-defined random variable X and

$$y'_t = \alpha(t)y'_{t-1}$$

for $t > a$.

Lemma 2.1: Let the stochastic process y'_t be a solution of (2.2) for $t \in I_a$. Then either $P[y'_t = 0] = 0$ or $P[y'_t = 0] = 1$ for $t \in I_a$.

Proof: By definition $y'_t = \alpha(t)y'_{t-1}$. Thus

$$y'_t = [\alpha(t)\alpha(t-1) \ldots \alpha(s+1)]y'_s,$$

for $t > s \geq a$. But $\alpha(t) \neq 0$ for all $t \in I_{a+1}$.

One can interpret this lemma as either the stochastic process is "never zero" or else is identically zero almost everywhere for $t \in I_a$.

Excluding the trivial solution where the stochastic process $\phi(t)$ is zero almost everywhere for $t \in I_a$, let $\{\phi(a) = X, \phi(t)\}$, where X is some well-defined random variable, be the set of all stochastic processes which are solutions of (2.2) for $t \in I_a$. Let $\phi'(t) = \{1, \phi(t)\}$ be the unique solution of (2.2) for $t \in I_a$ for which X is a variable taking on only the value 1 at $t = a$. Clearly, by Lemma 2.1 $\phi'(t)$ is never zero almost everywhere on I_a. Then the following lemma gives the relationship between any two solutions of (2.2) for $t \in I_a$.

Lemma 2.2: Let the stochastic processes y'_t and z_t be two solutions of (2.2) for $t \in I_a$. Then y'_t and z_t differ by a factor which is some well-defined random variable.

Proof: Let $\phi'(t)$ be the unique solution of (2.2) for $t \in I_a$ introduced above where $\phi(a) \equiv 1$. Then by Theorem 2.1 one can recursively construct

$$y'_t = \phi'(t)y'_a, \qquad t \geq a$$

and

$$z_t = \phi'(t)z_a, \qquad t \geq a.$$

By Lemma 2.1 $P[y'_t = 0] = 0$ and $P[z_t = 0] = 0$, since the trivial solution has been excluded; and, in particular, $P[y'_a = 0] = 0$ and $P[z_a = 0] = 0$. Thus

$$y'_t = z_t(z_a^{-1}y'_a)$$

and

$$z_t = y'_t(y'^{-1}_a z_a).$$

Corollary 2.1: Let η_t be any solution of (2.2) for $t \in I_a$. Let y'_t be a nontrivial solution of (2.2) for $t \in I_a$. Then any solution can be represented by the following:

$$\eta_t = y'_t y'^{-1}_a \eta_a.$$

An explicit expression for a solution to the nonhomogeneous equation (2.1) conditional on the starting point a can be constructed.

Theorem 2.2: Let a be an arbitrary integer in I. Let $\alpha(t)$ be defined for $t \in I_{a+1}$ where the coefficient $\alpha(t)$ is a random function of time taking on a **nonzero** real value. Let the stochastic process w_t be defined for $t \in I_{a+1}$. Let y'_t be a nontrivial solution to the homogeneous equation

$$y_t = \alpha(t) y_{t-1}$$

for $t \in I_a$. Let X be some well-defined random variable. Then the stochastic process ζ_t, where

$$\zeta_a = X$$

$$\zeta_t = \sum_{s=a+1}^{t} y'_t y'^{-1}_s w_s + y'_t y'^{-1}_a X, \qquad t > a, \quad (2.4)$$

is the unique solution to the nonhomogeneous equation

$$y_t = \alpha(t) y_{t-1} + w_t, \qquad t \in I_a,$$

conditional on $\zeta_a = X$.

Proof: Theorem 2.2 can be deduced directly from Theorem 2.1. Since y'_t is a nontrivial solution to the homogeneous equation, one has

$$y'_t = \alpha(t) y'_{t-1}, \qquad t \in I_a,$$

where $P(y'_t = 0) = 0$ for all t. This implies that

$$y'_t y'^{-1}_s = \alpha(t)\alpha(t-1) \ldots \alpha(s+1), \qquad s \in I_a. \quad (2.5)$$

Thus substituting (2.5) into (2.3), one obtains the unique solution ζ_t to the nonhomogeneous equation conditional on $\zeta_a = X$ at the starting point a:

$$\zeta_a = X$$

$$\zeta_t = \sum_{s=a+1}^{t} y'_t y'^{-1}_s w_s + y'_t y'^{-1}_a X, \qquad \text{for } t = a+1, a+2, \ldots,$$

as given in (2.4).

2.2 An Unconditional Solution With Finite Second Order Moments

Before going further it is necessary to impose additional structural assumptions on the model. Consider the stochastic linear recurrence relation

$$y_t = \alpha(t)y_{t-1} + u_t, \qquad t \in I, \quad (2.6)$$

where the coefficient $\alpha(t)$ is a random function of time defined for $t \in I$ which takes on a nonzero real value. The random variable u_t is a member of a stochastic process $\{u_t \mid t \in I\}$. In particular, let u_t, $t \in I$, have mean zero,

$$Eu_t = 0, \qquad \forall t$$

and second order moments,

$$Eu_t u_s = \delta_{ts}\sigma^2, \qquad \forall t,s$$

where $\sigma^2 < \infty$ and where δ_{ts} is defined:

$$\delta_{ts} = \begin{cases} 1 \text{ if } t = s \\ 0 \text{ otherwise.} \end{cases}$$

Equation (2.6) will be called a "generalized autoregressive equation" due to the random coefficient $\alpha(t)$.

One would like to establish conditions under which the generalized autoregressive equation of order one (2.6) determines a stochastic process $\{y_t\}$ for all $t \in I$, independent of the starting point a, such that its first and second order moments are well-defined and finite. The following theorem not only gives necessary and sufficient conditions for the existence of such an unconditional solution, independent of the starting point, but also gives an explicit expression for an unconditional solution for all $t \in I$.

Theorem 2.3 (The Existence Theorem): Let $\alpha(t)$, $t \in I$, be the coefficient of the generalized autoregressive equation of order one

$$y_t = \alpha(t)y_{t-1} + u_t, \qquad t \in I$$

where $\alpha(t)$ is a random function of time t taking on a nonzero real value for all $t \in I$. Let the stochastic process $\{u_t | t \in I\}$ be a mean zero family of random variables with the covariance function $Eu_t u_s = \delta_{ts}\sigma^2$ where σ^2 is finite. Assume $\alpha(t)$ and u_s are independent for all t and s. Let b be any integer in I such that $b < t$. Then

$$\xi_t = \sum_{r=0}^{n} [\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)]u_{t-r},$$

with $[\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)] \overset{\text{def}}{=} 1$ for the term where r = 0, converges in mean square as n increases to a mean zero stochastic process with finite second order moments if and only if

$$\sum_{b'=1}^{\infty} E[\alpha^2(t)\alpha^2(t-1) \ldots \alpha^2(t-b'+1)] < \infty$$

for all $t \in I$ where $b' = t-b$. The limiting stochastic process, to be denoted by

$$\xi_t = \sum_{r=0}^{\infty} [\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)]u_{t-r}, \qquad (2.7)$$

is a solution of the generalized autoregressive equation

$$y_t = \alpha(t)y_{t-1} + u_t, \qquad t \in I.$$

Proof: First look at the homogeneous equation

$$y'_t = \alpha(t)y'_{t-1}$$

where y'_t is a nontrivial solution. Recursively, it can be shown that

$$y'_t = \alpha(t)y'_{t-1}$$

$$y'_t = \alpha(t)\alpha(t-1)y'_{t-2}$$

$$\vdots$$

$$y'_t = \alpha(t)\alpha(t-1) \ldots \alpha(s+1)y'_s$$

so that $y'_t y'^{-1}_s = \alpha(t)\alpha(t-1) \ldots \alpha(s+1)$, i.e. a product of the random coefficients.

Now looking at the nonhomogeneous equation one can see that

$$y_t = \alpha(t)y_{t-1} + u_t$$

$$y_t = \alpha(t)\alpha(t-1)y_{t-2} + \alpha(t)u_{t-1} + u_t$$

$$y_t = \alpha(t)\alpha(t-1)\alpha(t-2)y_{t-3} + \alpha(t)\alpha(t-1)u_{t-2} + \alpha(t)u_{t-1} + u_t$$

$$\vdots$$

For any $a \in I$, Theorem 2.2 implies that the stochastic process ζ_t defined for $t \in I_a$ as

$$\zeta_a = 0$$

$$\zeta_t = \sum_{s=a+1}^{t} y_t' y_s'^{-1} u_s, \qquad \text{for all } t > a \tag{2.8}$$

is a solution of (2.6) for $t \in I_a$. To emphasize the dependence of ζ_t in (2.8) on a starting point a one can write

$$\xi_t^{(b)} = \sum_{s=b+1}^{t} y_t' y_s'^{-1} u_s, \qquad \text{for all } t > a \geq b, \tag{2.9}$$

which gives a sequence of all solutions conditional on starting points b less than or equal to a where a must be less than t. Substituting for $y_t' y_s'^{-1}$ in (2.9) gives

$$\xi_t^{(b)} = \sum_{s=b+1}^{t} \alpha(t)\alpha(t-1) \ldots \alpha(s+1)u_s, \qquad t > a \geq b.$$

It must be shown that the sequence $\{\xi_t^{(b)} \mid b = a, a-1, \ldots\}$ converges in quadratic mean to some random variable as b decreases. This will follow by the Cauchy criterion upon showing

$$\lim_{b,c \to -\infty} E[\xi_t^{(b)} - \xi_t^{(c)}]^2 = 0 \qquad \text{for all } t \in I.$$

Now with v a negative integer,

$$E[\xi_t^{(b+v)} - \xi_t^{(b)}]^2 = E[\sum_{s=b+v+1}^{t} \alpha(t)\alpha(t-1) \ldots \alpha(s+1)u_s - \sum_{s=b+1}^{t} \alpha(t)\alpha(t-1) \ldots \alpha(s+1)u_s]^2$$

$$= E[\sum_{s=b+v+1}^{b} \alpha(t)\alpha(t-1) \ldots \alpha(s+1)u_s]^2$$

By hypothesis $\alpha(t)$ and u_s are independent for all t and s, and the expectation can be taken inside the summation since the sum is finite giving

$$E[\xi_t^{(b+v)} - \xi_t^{(b)}]^2 = \sum_{s=b+v+1}^{b} E[\alpha^2(t)\alpha^2(t-1) \ldots \alpha^2(s+1)]E[u_s^2]$$

$$= \sigma^2 \sum_{s=b+v+1}^{b} E[\alpha^2(t)\alpha^2(t-1) \ldots \alpha^2(s+1)]. \qquad (2.10)$$

One must show that (2.10) may be made less than any preassigned $\varepsilon < 0$ for -b sufficiently large and for all $v < 0$. By the Cauchy criterion for convergence of series, this follows if and only if

$$\sum_{s=-\infty}^{t-1} E[\alpha^2(t)\alpha^2(t-1) \ldots \alpha^2(s+1)] < \infty. \qquad (2.11)$$

Making the change of variable $b' = t-s$ in (2.11) gives this condition as presented in the statement of the theorem:

$$\sum_{b'=1}^{\infty} E[\alpha^2(t)\alpha^2(t-1) \ldots \alpha^2(t-b'+1)] < \infty.$$

Thus, $\{\xi_t^{(b)}\}$ converges in mean square for all $t \in I$ to the limiting stochastic process ξ_t.

From (2.9) since $Eu_t = 0$, $E\xi_t^{(b)} = 0$. And since

$$[E(\xi_t^{(b)}-\xi_t)]^2 \leq E[\xi_t^{(b)}-\xi_t]^2,$$

where $E[\xi_t^{(b)}-\xi_t]^2 \to 0$ as $b \to -\infty$, one can conclude that ξ_t has mean zero.

One needs to show that ξ_t defined in (2.7) satisfies (2.6). One can write ξ_t as follows:

$$\xi_t = \sum_{s=-\infty}^{t} y_t' y_s'^{-1} u_s.$$

Now,

$$\xi_t - \alpha(t)\xi_{t-1} = \sum_{s=-\infty}^{t} y_t' y_s'^{-1} u_s - \alpha(t) \sum_{s=-\infty}^{t-1} y_{t-1}' y_s'^{-1} u_s$$

$$= y_t' y_t'^{-1} u_t + \sum_{s=-\infty}^{t-1} [y_t' - \alpha(t) y_{t-1}'] \, y_s'^{-1} u_s$$

where $y_t' y_t'^{-1} = 1$. Since y_t' is a solution of the homogeneous equation, $y_t' = \alpha(t) y_{t-1}'$; thus, the above becomes

$$\xi_t - \alpha(t)\xi_{t-1} = u_t + \sum_{s=-\infty}^{t-1} (0) y_s'^{-1} u_s$$

or

$$\xi_t - \alpha(t)\xi_{t-1} = u_t.$$

Therefore, ξ_t does satisfy the generalized autoregressive equation (2.6):

$$y_t = \alpha(t)y_{t-1} + u_t \qquad \text{for all } t \in I.$$

This completes the proof. Henceforth, a solution of this type will be expressed as

$$y_t = \sum_{r=o}^{\infty} [\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)]u_{t-r}, \tag{2.12}$$

with the infinite sum understood to be the mean square limit of finite partial sums.

Theorem 2.3 gives the existence of a solution (2.12) to the generalized autoregressive equation of order one (2.6) for all $t \in I$. This solution is a stochastic process having mean zero and finite second order moments and, in general, with no further restrictions is a nonstationary stochastic process. It is of interest to investigate the special class of solutions that are weakly stationary stochastic processes. Before one can specify this special class of weakly stationary solutions, it is necessary to obtain explicit expressions for the second order moments. Then one can determine sufficient conditions such that these second order moments are independent of time giving weakly stationary solutions.

First, one can express all of the finite second order moments of the solution y_t in (2.12), or the autocovariance structure for any lag k where $k \in I$, as follows:

$$Ey_t y_{t-k} = E\left\{\left[\sum_{r=0}^{\infty} \alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)u_{t-r}\right]\right.$$

$$\left.\cdot \left[\sum_{r=0}^{\infty} \alpha(t-k)\alpha(t-k-1) \ldots \alpha(t-k-r+1)u_{t-k-r}\right]\right\}.$$

For convenience, use the following transformation for the second summation:

Let $s = r+k \Rightarrow r = s-k$ $\qquad u_{t-k-r} = u_{t-k-s+k}$

$r = 0 \Rightarrow s=k$ $\qquad = u_{t-s}$

$r = \infty \Rightarrow s=\infty$

Then one has

$$Ey_t y_{t-k} = E\left\{\left[\sum_{r=0}^{\infty} \alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)u_{t-r}\right]\right.$$

$$\left.\cdot \left[\sum_{s=k}^{\infty} \alpha(t-k)\alpha(t-k-1) \ldots \alpha(t-s+1)u_{t-s}\right]\right\}.$$

This can be partitioned as follows:

$$Ey_t y_{t-k} = E\left\{\left[\sum_{r=0}^{k-1} \alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)u_{t-r}\right]\right.$$

$$\cdot \left[\sum_{s=k}^{\infty} \alpha(t-k)\alpha(t-k-1) \ldots \alpha(t-s+1)u_{t-s}\right]$$

$$+ \left[\sum_{r=k}^{\infty} \alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)u_{t-r}\right]$$

$$\left.\cdot \left[\sum_{s=k}^{\infty} \alpha(t-k)\alpha(t-k-1) \ldots \alpha(t-s+1)u_{t-s}\right]\right\}$$

$$= E\Bigg\{ \sum_{r=0}^{k-1} \sum_{s=k}^{\infty} \alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)\alpha(t-k)$$

$$\cdot \alpha(t-k-1) \ldots \alpha(t-s+1)u_{t-r}u_{t-s}$$

$$+ \sum_{r=k}^{\infty} \alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k)\alpha^2(t-k-1) \ldots$$

$$\cdot \alpha^2(t-r+1)u_{t-r}^2 \qquad \text{(i.e. } r=s\text{)}$$

$$+2 \sum_{r=k}^{\infty} \sum_{s=r+1}^{\infty} \alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k) \ldots \alpha^2(t-r+1)$$

$$r < s$$

$$\cdot \alpha(t-r) \ldots \alpha(t-s+1)u_{t-r}u_{t-s} \Bigg\}; \qquad (2.13)$$

where, in the first summation $\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1) \overset{\text{def}}{=} 1$ for $r=0$ and $\alpha(t-k)\alpha(t-k-1) \ldots \alpha(t-s+1) \overset{\text{def}}{=} 1$ for $s=k$, and in the second and third summations $\alpha^2(t-k)\alpha^2(t-k-1) \ldots \alpha^2(t-r+1) \overset{\text{def}}{=} 1$ for $r=k$. Thus, the second order moments can be partitioned into the expected value of three infinite sums. One needs to impose some conditions in order to interchange expectation and limit for each summation in (2.13). This can be done by applying the following result stated in Rao [1973, p. 111].

Lemma 2.3: If X_i, $i=1, 2, \ldots$ is a sequence of random variables, then

$$E\left(\sum_{i=1}^{\infty} X_i\right) = \sum_{i=1}^{\infty} EX_i$$

provided $\sum_{i=1}^{\infty} E|X_i| < \infty$, which ensures the convergence of $\sum_{i=1}^{\infty} X_i$ with probability one.

Lemma 2.3 implies that the following three conditions, one for each summation in (2.13), respectively, are sufficient for one to interchange expectation and limit for each summation in (2.13). For each lag k, one must require that

(a) $\sum_{s=k}^{\infty} E|\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)\alpha(t-k)\alpha(t-k-1) \ldots \alpha(t-s+1)u_{t-r}u_{t-s}| < \infty$

for fixed r = 0, 1, 2, ..., k-1;

(b) $\sum_{r=k}^{\infty} E|\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k)\alpha^2(t-k-1) \ldots \alpha^2(t-r+1)u_{t-r}^2| < \infty$;

and

(c) $\sum_{r=k}^{\infty} \sum_{s=r+1}^{\infty} E|\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k) \ldots \alpha^2(t-r+1)$

$\cdot \alpha(t-r) \ldots \alpha(t-s+1)u_{t-r}u_{t-s}| < \infty.$

It has been assumed that $\alpha(t)$ and u_s are independent for all t and s, and therefore these conditions can be further simplified. It is known that $E|u_{t-r}^2| = Eu_{t-r}^2 = \sigma^2$ for all t-r where σ^2 is assumed finite. Thus $E|u_{t-r}| = m$ for all t-r where m is some finite constant.

Thus condition (a) becomes

$$\sum_{s=k}^{\infty} E[|\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)\alpha(t-k)\alpha(t-k-1) \ldots \cdot \alpha(t-s+1)||u_{t-r}u_{t-s}|] < \infty$$

and using the independence of α and u and the uncorrelatedness of the u's, one obtains

$$\sum_{s=k}^{\infty} E|\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)\alpha(t-k)\alpha(t-k-1) \ldots \alpha(t-s+1)|[E|u_{t-r}|]^2 < \infty$$

or

$$m^2 \sum_{s=k}^{\infty} E|\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)\alpha(t-k)\alpha(t-k-1) \ldots \alpha(t-s+1)| < \infty.$$

Condition (a) simplifies to a condition just on the random coefficients $\alpha(t)$, $t \in I$:

$$(a)' \quad \sum_{s=k}^{\infty} E|\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)\alpha(t-k)\alpha(t-k-1) \ldots \alpha(t-s+1)| < \infty$$

for fixed r = 0, 1, 2, ..., k-1.

Similarly, condition (b) becomes

$$\sum_{r=k}^{\infty} E|\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k)\alpha^2(t-k-1) \ldots \alpha^2(t-r+1)|E|u^2_{t-r}| < \infty$$

or

$$\sigma^2 \sum_{r=k}^{\infty} E|\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k)\alpha^2(t-k-1) \ldots \alpha^2(t-r+1)| < \infty$$

and $\sigma^2 < \infty$ by assumption. Condition (b) simplifies to a condition just on the random coefficients $\alpha(t)$, $t \in I$:

$$(b)' \quad \sum_{r=k}^{\infty} E|\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k)\alpha^2(t-k-1) \ldots \alpha^2(t-r+1)| < \infty.$$

Similarly, condition (c) becomes

$$\sum_{r=k}^{\infty}\sum_{s=r+1}^{\infty} E\left|\alpha(t)\alpha(t-1)\ \ldots\ \alpha(t-k+1)\alpha^2(t-k)\ \ldots\ \alpha^2(t-r+1)\alpha(t-r)\ \ldots \right.$$

$$\left. \cdot\ \alpha(t-s+1)\right| \left[E\left|u_{t-r}\right|\right]^2 < \infty$$

or

$$m^2 \sum_{r=k}^{\infty}\sum_{s=r+1}^{\infty} E\left|\alpha(t)\alpha(t-1)\ \ldots\ \alpha(t-k+1)\alpha^2(t-k)\ \ldots\ \alpha^2(t-r+1)\alpha(t-r)\ \ldots \right.$$

$$\left. \cdot\ \alpha(t-s+1)\right| < \infty.$$

Condition (c) simplifies to a condition just on the random coefficients $\alpha(t)$, $t \in I$:

$$(c)' \quad \sum_{r=k}^{\infty}\sum_{s=r+1}^{\infty} E\left|\alpha(t)\alpha(t-1)\ \ldots\ \alpha(t-k+1)\alpha^2(t-k)\ \ldots\ \alpha^2(t-r+1)\alpha(t-r)\ \ldots \right.$$

$$\left. \cdot\ \alpha(t-s+1)\right| < \infty.$$

Therefore, if conditions (a)', (b)', and (c)' above are satisfied for the random coefficients $\alpha(t)$, $t \in I$, then one can interchange expectation and limit in (2.13) and obtain explicit expressions for the second order moments of y_t, $t \in I$, for each lag k.

Interchanging expectation and limit in (2.13) one obtains

$$Ey_t y_{t-k} = \sum_{r=0}^{k-1}\sum_{s=k}^{\infty} E[\alpha(t)\alpha(t-1)\ \ldots\ \alpha(t-r+1)\alpha(t-k)\alpha(t-k-1)\ \ldots$$

$$\cdot\ \alpha(t-s+1)u_{t-r}u_{t-s}]$$

$$+ \sum_{r=k}^{\infty} E[\alpha(t)\alpha(t-1)\ \ldots\ \alpha(t-k+1)\alpha^2(t-k)\alpha^2(t-k-1)\ \ldots$$

$$\cdot\ \alpha^2(t-r+1)u_{t-r}^2]$$

$$+ 2 \sum_{\substack{r=k \\ r<s}}^{\infty}\sum_{s=r+1}^{\infty} E[\alpha(t)\alpha(t-1)\ \ldots\ \alpha(t-k+1)\alpha^2(t-k)\ \ldots$$

$$\cdot\ \alpha^2(t-r+1)\alpha(t-r)\ \ldots\ \alpha(t-s+1)u_{t-r}u_{t-s}].$$

Using the independence of α and u and the uncorrelatedness of the u's, one obtains

$$Ey_ty_{t-k} = E[u_{t-r}u_{t-s}] \sum_{r=0}^{k-1} \sum_{s=k}^{\infty} E[\alpha(t)\alpha(t-1)\ \ldots\ \alpha(t-r+1)\alpha(t-k)\alpha(t-k-1)\ldots$$

$$\cdot\ \alpha(t-s+1)]$$

$$+ E[u_{t-r}^2] \sum_{r=k}^{\infty} E[\alpha(t)\alpha(t-1)\ \ldots\ \alpha(t-k+1)\alpha^2(t-k)\,\alpha^2(t-k-1)\ \ldots$$

$$\cdot\ \alpha^2(t-r+1)]$$

$$+ 2E[u_{t-r}u_{t-s}] \sum_{\substack{r=k \\ r<s}}^{\infty} \sum_{s=r+1}^{\infty} E[\alpha(t)\alpha(t-1)\ \ldots\ \alpha(t-k+1)\,\alpha^2(t-k)\ \ldots$$

$$\cdot\ \alpha^2(t-r+1)\alpha(t-r)\ \ldots\ \alpha(t-s+1)]$$

and since $E[u_{t-r}u_{t-s}] = 0$ in the first and third pieces, one is left with

$$Ey_ty_{t-k} = \sigma^2 \sum_{r=k}^{\infty} E[\alpha(t)\alpha(t-1)\ \ldots\ \alpha(t-k+1)\alpha^2(t-k)\alpha^2(t-k-1)\ \ldots$$

$$\cdot\ \alpha^2(t-r+1)]. \qquad (2.14)$$

Given that the conditions of Theorem 2.3, the existence theorem, are satisfied and a solution as expressed in equation (2.12) exists, the following theorem states sufficient conditions for a solution to be a weakly stationary stochastic process.

Theorem 2.4: A solution

$$y_t = \sum_{r=0}^{\infty} [\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)]u_{t-r}$$

of the generalized autoregressive equation of order one

$$y_t = \alpha(t)y_{t-1} + u_t \qquad t \in I$$

is a weakly stationary stochastic process if the set of random coefficients $\{\alpha(i), i \in I\}$ form a strictly stationary stochastic process for which the following conditions are satisfied for each lag k:

$$\text{(a)}' \quad \sum_{s=k}^{\infty} E\left|\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)\alpha(t-k)\alpha(t-k-1) \ldots \alpha(t-s+1)\right| < \infty ,$$

for fixed r = 0, 1, 2, ..., k-1;

$$\text{(b)}' \quad \sum_{r=k}^{\infty} E\left|\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k)\alpha^2(t-k-1) \ldots \alpha^2(t-r+1)\right| < \infty;$$

and

$$\text{(c)}' \quad \sum_{r=k}^{\infty} \sum_{s=r+1}^{\infty} E\left|\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k) \ldots \alpha^2(t-r+1) \cdot \alpha(t-r) \ldots \alpha(t-s+1)\right| < \infty .$$

Proof: Suppose that the set of random coefficients $\{\alpha(i), i \in I\}$ satisfy the three conditions (a)', (b)', and (c)' sufficient to interchange expectation and limit in the general expression (2.13) for the second order moments of y_t, $t \in I$. Then the second order moments of y_t can be expressed as in (2.14):

$$Ey_t y_{t-k} = \sigma^2 \sum_{r=k}^{\infty} E[\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k)\alpha^2(t-k-1) \ldots \cdot \alpha^2(t-r+1)],$$

for each lag k. Suppose also that the set of random coefficients $\{\alpha(i), i \in I\}$ form a strictly stationary stochastic process. Then all moments of any order for $\{\alpha(i), i \in I\}$ would be independent of time t.

Recall that $E(y_t) = 0$ for all t, and thus the mean is the same for all t, i.e. independent of time t. In expression (2.14) for the finite second order moments of y_t, it is assumed that σ^2 is finite and it has been shown that the infinite sum is finite. Each term in the sum is a moment of order 2r-k for $\{\alpha(i), i \in I\}$ and is therefore independent of time t. Since the infinite sum is finite and each term in the sum is independent of time t, then the sum must also be independent of time t. Therefore, solution y_t is a weakly stationary stochastic process.

Although the theory in this chapter applies to the general case where $\alpha(t)$ is any random function of time satisfying the conditions stated in Theorem 2.3 that are required for a solution of the generalized autoregressive equation of order one to exist, it is of interest to consider specific model structures for the random coefficient that may be useful in practical applications. This is done in the next chapter.

CHAPTER III

A SPECIFIC MODEL FOR THE RANDOM COEFFICIENT

To illustrate the theory of the generalized autoregressive equation model of order one and to develop a random parameter approach to modeling and forecasting from this model, one must impose some structure on the random coefficient $\alpha(t)$. Suppose $\alpha(t)$ varies slowly with time but only takes on discrete values. In particular, take the simplest case where $\alpha(t)$ only takes on two different values. Let $\alpha(t)$ be dependent on its previous values. One natural way of modeling this behavior is by using a two-state Markov chain. This random parameter model, the generalized autoregressive equation model of order one with random coefficient $\alpha(t)$ modeled as a two-state Markov chain, is the special class of models investigated in this study. Henceforth, this model is referred to as a two-state Markov chain model.

In section 3.1 a two-state Markov chain model is described. Based on Hoel, et al. [1972], a review of some basic definitions and properties of Markov chains is given. These basic concepts are applied to the model of interest giving properties and necessary assumptions for this random parameter model to be a useful extension of the fixed coefficient autoregressive model of order one.

In section 3.2 it is shown that the two-state Markov chain model satisfies Theorem 2.3 and thus has as a solution a mean zero stochastic process with finite second order moments. Since it is assumed that the two-state Markov chain for $\alpha(t)$ is strictly stationary and since $\alpha(t)$ satisfies the other sufficient conditions given in Theorem 2.4, the solution is a weakly stationary stochastic process. To complete the chapter, a brief discussion is given for the generalization to an n-state Markov chain model.

3.1 A Two-State Markov Chain Model: Its Properties and Assumptions

Consider the following generalized autoregressive equation of order one

$$y_t = \alpha(t)y_{t-1} + u_t, \qquad t \in I$$

where the same assumptions are made as in Chapter II for the random noise variable. Specifically, $\{u_t \mid t \in I\}$ is a mean zero stochastic process with covariance structure given by

$$Eu_t u_s = \delta_{ts}\sigma^2$$

where the variance σ^2 is finite and δ_{ts} is defined as

$$\delta_{ts} = \begin{cases} 1 \text{ for } t = s \\ 0 \text{ for } t \neq s. \end{cases}$$

Thus $\{u_t \mid t \in I\}$ is a weakly stationary stochastic process. It is also assumed that the u_t follow a normal distribution, and therefore the u_t are independent for all t. Since the normal distribution is

completely determined by its first and second order moments, $\{u_t | t \in I\}$ is now strictly stationary.

The random coefficient $\{\alpha(t) | t \in I\}$ is modeled as a two-state Markov chain having state space $S = \{\theta_1, \theta_2\}$ where $0 < \theta_i < 1$. Thus $\alpha(t)$ takes on either the value θ_1 or θ_2 for all t. It is also assumed that $\alpha(t)$ and u_s are independent for all t and s.

Assume the two-state Markov chain for $\alpha(t)$ satisfies the Markov property:

$$P(\alpha(t+1) = \alpha_{t+1} | \alpha(0) = \alpha_0, \ldots, \alpha(t) = \alpha_t) = P(\alpha(t+1) = \alpha_{t+1} | \alpha(t) = \alpha_t) \qquad \forall t$$

where α_i is an element of the state space S. Thus given the present state, the past states have no influence on the future. The conditional probabilities, or transition probabilities,

$$P(\alpha(t+1) = b | \alpha(t) = a)$$

are assumed to be stationary, and thus independent of time t. The transition function is defined as

$$P(a, b) = P(\alpha(t+1) = b | \alpha(t) = a) \qquad t \geq 0$$

such that

$$P(a, b) \geq 0 \qquad a, b \in S$$

$$\sum_b P(a, b) = 1 \qquad a \in S.$$

Let $P(\alpha(t+1) = \theta_2 | \alpha(t) = \theta_1) = p$ $\quad P(\alpha(t+1) = \theta_1 | \alpha(t) = \theta_1) = 1-p$

$P(\alpha(t+1) = \theta_1 | \alpha(t) = \theta_2) = q$ $\quad P(\alpha(t+1) = \theta_2 | \alpha(t) = \theta_2) = 1-q$

where it is assumed that $0 < p < 1$ and $0 < q < 1$. The numbers $P(a, b)$ are the one-step transition probabilities of the Markov chain. Thus the one-step transition matrix is

$$P = \begin{array}{c} \\ \theta_1 \\ \theta_2 \end{array} \begin{array}{c} \begin{array}{cc} \theta_1 & \theta_2 \end{array} \\ \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \end{array} \begin{array}{c} \\ 1 \\ 1 \end{array}$$

where $p + q > 0$.

The m-step transition function $P^m(a, b)$, which gives the probability of going from a to b in m steps, is defined by

$$P^m(a, b) = \sum_{b_1} \cdots \sum_{b_{m-1}} P(a, b_1)P(b_1, b_2) \cdots P(b_{m-2}, b_{m-1})P(b_{m-1}, b) \tag{3.1}$$

for $m \geq 2$, by $P^1(a, b) = P(a, b)$ and by $P^0(a, b) = \begin{cases} 1 & a=b \\ 0 & \text{elsewhere.} \end{cases}$

For Markov chains having a finite number of states, (3.1) allows one to think of P^m as the m^{th} power of the matrix P. In general for the two-state Markov chain

$$P^m = \frac{1}{p+q} \begin{pmatrix} q & p \\ q & p \end{pmatrix} + \frac{(1-p-q)^m}{p+q} \begin{pmatrix} p & -p \\ -q & q \end{pmatrix} \quad \forall m.$$

One can also show that

$$P(\alpha(t+m) = b \mid \alpha(t) = a) = P^m(a, b),$$

and

$$P^{t+m}(a, b) = \sum_c P^t(a, c)P^m(c, b).$$

The function $\pi_0(a)$, $a \in S$, defined by

$$\pi_0(a) = P(\alpha(0) = a) \qquad a \in S$$

is the initial distribution of the Markov chain. It has the following properties:

$$\pi_0(a) \geq 0, \ a \in S$$

$$\sum_a \pi_0(a) = 1.$$

Thus for this two-state Markov chain model

$$\pi_0(\theta_1) + \pi_0(\theta_2) = 1,$$

where

$$\pi_0(\theta_1) = P(\alpha(0) = \theta_1)$$

$$\pi_0(\theta_2) = 1 - \pi_0(\theta_1) = P(\alpha(0) = \theta_2).$$

It is desirable to have the entire distribution of $\alpha(t)$ independent of time t. The distribution of $\alpha(t)$ can be expressed as follows:

$$P(\alpha(t) = \theta_1) = \frac{q}{p+q} + (1-p-q)^t \left(\pi_0(\theta_1) - \frac{q}{p+q}\right)$$

$$P(\alpha(t) = \theta_2) = \frac{p}{p+q} + (1-p-q)^t(\pi_0(\theta_2) - \frac{p}{p+q})$$

One can choose $\pi_0(\theta_1)$ and $\pi_0(\theta_2)$ such that $P(\alpha(t) = \theta_1)$ and $P(\alpha(t) = \theta_2)$ are independent of time t. Suppose one chooses:

$$\pi_0(\theta_1) = \frac{q}{p+q}$$

$$\pi_0(\theta_2) = \frac{p}{p+q} .$$

Then the distribution of $\alpha(t)$ becomes

$$P(\alpha(t) = \theta_1) = \frac{q}{p+q} \qquad \forall t$$

$$P(\alpha(t) = \theta_2) = \frac{p}{p+q} \qquad \forall t,$$

and this is independent of t.

It can be shown that the distribution of $\alpha(t)$ is independent of time t if and only if the initial distribution is a stationary distribution. If $\pi(a)$, $a \in S$, are nonnegative numbers summing to one, and if

$$\sum_a \pi(a)P(a, b) = \pi(b) \qquad b \in S$$

then π is a stationary distribution. It turns out that for this two-state Markov chain model the unique stationary distribution π is determined by

$$\pi(\theta_1) = \frac{q}{p+q} \text{ and } \pi(\theta_2) = \frac{p}{p+q} ,$$

such that

$$\pi(\theta_1) + \pi(\theta_2) = \frac{q}{p+q} + \frac{p}{p+q} = \frac{p+q}{p+q} = 1.$$

Also, one can show for

$$b = \theta_1: \quad \pi(\theta_1)P(\theta_1, \theta_1) + \pi(\theta_2)P(\theta_2, \theta_1) = \frac{q}{p+q}(1-p) + \frac{p}{p+q}q$$

$$= \frac{q-pq+pq}{p+q}$$

$$= \frac{q}{p+q} = \pi(\theta_1)$$

$$b = \theta_2: \quad \pi(\theta_1)P(\theta_1,\theta_2) + \pi(\theta_2)P(\theta_2,\theta_2) = \frac{q}{p+q}p + \frac{p}{p+q}(1-q)$$

$$= \frac{pq+p-pq}{p+q}$$

$$= \frac{p}{p+q} = \pi(\theta_2).$$

It can also be shown that

$$\lim_{t\to\infty} P^t(a, b) = \pi(b), \qquad b \in S.$$

Thus regardless of the initial distribution of the chain, the distribution of $\alpha(t)$ approaches π as $t \to \infty$. In this case, π is a steady state distribution.

The joint distribution of $\alpha(0), \ldots, \alpha(t)$ can be expressed in terms of the transition function and the initial distribution:

$$P(\alpha(0) = \alpha_0, \alpha(1) = \alpha_1, \ldots, \alpha(t) = \alpha_t) = \pi_0(\alpha_0)P(\alpha_0, \alpha_1) \ldots$$

$$\cdot P(\alpha_{t-1}, \alpha_t).$$

Also, with π_0 as the initial distribution,

$$P(\alpha(t) = b) = \sum_a \pi_0(a)P^t(a, b)$$

or

$$P(\alpha(t+1) = b) = \sum_a P(\alpha(t) = a)P(a, b).$$

Let $P_a(\ )$ denote probabilities of various events defined in terms of a Markov chain starting at state a. This notation will be used in some of the discussion that follows.

A state c of a Markov chain is called an absorbing state if $P(c, c) = 1$ or, equivalently, if $P(c, b) = 0$ for $b \neq c$. For the two-state Markov chain for $\alpha(t)$, it is assumed that $p \neq 0$, $q \neq 0$, and thus neither of the two states is absorbing. Otherwise, it would be natural to fit a fixed coefficient AR(1) model to the observations after the absorption had occurred, if this could be determined.

Let T_a, the hitting time of a, be defined by

$$T_a = \begin{cases} \min\ (t > 0:\ \alpha(t) = a) \\ \infty \text{ if } \alpha(t) \neq a \qquad\qquad \forall t > 0. \end{cases}$$

In other words, T_a is the first positive time the Markov chain is in, or hits, state a. Observe that

$$P_a(T_b = t+1) = \sum_{c \neq b} P(a, c)P_c(T_b = t), \qquad t \geq 1.$$

Let $\rho_{ab} = P_a(T_b < \infty)$. Then ρ_{ab} denotes the probability that a Markov chain starting at a will be in state b at some positive time. In particular, ρ_{bb} denotes the probability that a Markov chain

starting at b will ever return to state b. A state b is "recurrent" if $\rho_{bb} = 1$ and "transient" if $\rho_{bb} < 1$. Thus if b is a recurrent state, a Markov chain starting at b returns to b with probability one; but if b is a transient state, a Markov chain starting at b has positive probability $1-\rho_{bb}$ of never returning to b. For the two-state Markov chain for $\alpha(t)$, both states θ_1 and θ_2 will be required to be recurrent. Thus $\rho_{\theta_1\theta_1} = 1$ and $\rho_{\theta_2\theta_2} = 1$.

A nonempty set C of states is closed if no state inside of C leads to any state outside of C, i.e. if

$$\rho_{ab} = 0 \qquad a \in C \text{ and } b \notin C.$$

A closed set C is "irreducible" if a leads to b for all choices of a and b in C. Let C' be a finite irreducible closed set of states. Then every state in C' is recurrent. The two-state Markov chain for $\alpha(t)$ is irreducible with both states recurrent.

Define the following:

$$\ell_b(c) = \begin{cases} 1 & c = b \\ 0 & c \neq b \end{cases}$$

$$E_a(\ell_b(\alpha(t))) = P_a(\alpha(t) = b) = P^t(a, b).$$

Now set

$$N_t(b) = \sum_{m=1}^{t} \ell_b(\alpha(m))$$

where $N_t(b)$ denotes the number of visits of the Markov chain to b during times m = 1, ..., t; and set

$$G_t(a, b) = \sum_{m=1}^{t} P^m(a, b) = E_a(N_t(b))$$

which is the expected number of such visits for the Markov chain starting at a.

One can define $m_b = E_b(T_b)$ as the mean return time to b for a chain starting at b if this return time has finite expectation, and otherwise $m_b = \infty$. Thus once a chain reaches b, it returns to b "on the average every m_b units of time". A recurrent state b is null recurrent if $m_b = \infty$. Requiring $m_b < \infty$ implies that b is positive recurrent. If C is a finite irreducible closed set of states, then every state in C is positive recurrent. Thus the state space $S = \{\theta_1, \theta_2\}$, which is finite, closed and irreducible, implies that both states are positive recurrent; and the chain is positive recurrent.

An irreducible positive recurrent Markov chain has a unique stationary distribution π, given by

$$\pi(a) = \frac{1}{m_a}, \qquad a \in S.$$

It can also be shown that an irreducible Markov chain is positive recurrent if and only if it has a stationary distribution. Let $\alpha(t)$, $t \geq 0$, be an irreducible positive recurrent Markov chain having stationary distribution π. Then with probability one

$$\lim_{t\to\infty} \frac{N_t(a)}{t} = \pi(a), \qquad a \in S.$$

Also, one can show that

$$\lim_{t\to\infty} \frac{1}{t} \sum_{m=1}^{t} P^m(a, b) = \lim_{t\to\infty} \frac{G_t(a, b)}{t} = \pi(b), \qquad a, b \in S.$$

One can define the period d_a by

$$d_a = \text{g.c.d. } \{t \geq 1: \ P^t(a, b) > 0\}.$$

If $P(a, a) > 0$, then $d_a = 1$. It can be shown that the states in an irreducible Markov chain have common period d. For the two-state Markov chain for $\alpha(t)$, $d_a = 1$; and thus the chain is "aperiodic". Let $\alpha(t)$, $t \geq 0$, be an irreducible positive recurrent Markov chain having stationary distribution π. If the chain is aperiodic,

$$\lim_{t\to\infty} P^t(a, b) = \pi(b), \qquad a, b \in S.$$

For details concerning these concepts and definitions and justification of some of the results, the reader is referred to Hoel, et al. [1972].

Since $\alpha(t)$ is modeled as a strictly stationary, two-state Markov chain with the above properties, one can develop the following moment structure for $\alpha(t)$. Consider the distribution of $\alpha(t)$:

$$\alpha(t) = \begin{cases} \theta_1 \text{ with probability } \frac{q}{p+q} \\ \theta_2 \text{ with probability } \frac{p}{p+q} ; \end{cases}$$

where $0 < p < 1$ and $0 < q < 1$. Therefore,

$$E[\alpha(t)] = \frac{q\theta_1 + p\theta_2}{p+q} \qquad \forall t.$$

Also

$$\alpha^2(t) = \begin{cases} \theta_1^2 & \text{with probability } \frac{q}{p+q} \\ \theta_2^2 & \text{with probability } \frac{p}{p+q} \end{cases}.$$

Thus the variance of $\alpha(t)$ can be expressed as:

$$\text{var}[\alpha(t)] = E[\alpha^2(t)] - [E\alpha(t)]^2$$

$$= \frac{q\theta_1^2 + p\theta_2^2}{p+q} - \left[\frac{q\theta_1 + p\theta_2}{p+q}\right]^2$$

$$= \frac{pq(\theta_1-\theta_2)^2}{(p+q)^2} \qquad \forall t.$$

Now consider the joint distribution of $\{\alpha(t), \alpha(t+k)\}$ for $k > 0$:

$$\{\alpha(t)\alpha(t+k)\} = \begin{cases} \theta_1^2 \text{ with probability } \frac{q}{(p+q)^2}[q+p(1-p-q)^k] \\ \theta_1\theta_2 \text{ with probability } \frac{pq}{(p+q)^2}[1-(1-p-q)^k] \\ \theta_2\theta_1 \text{ with probability } \frac{pq}{(p+q)^2}[1-(1-p-q)^k] \\ \theta_2^2 \text{ with probability } \frac{p}{(p+q)^2}[p+q(1-p-q)^k]. \end{cases}$$

Thus the autocovariance structure of $\alpha(t)$ can be expressed as:

$$\text{cov}[\alpha(t), \alpha(t+k)] = E[\alpha(t)\alpha(t+k)] - E[\alpha(t)]E[\alpha(t+k)]$$

$$= \frac{1}{(p+q)^2}[\theta_1^2 q[q+p(1-p-q)^k] + 2\theta_1\theta_2 pq[1-(1-p-q)^k]$$

$$+ \theta_2^2 p[p+q(1-p-q)^k] - (q\theta_1+p\theta_2)^2]$$

$$= \frac{pq(1-p-q)^k(\theta_1-\theta_2)^2}{(p+q)^2} \qquad \forall\, t.$$

Therefore, the autocorrelation structure of $\alpha(t)$ is given by

$$\rho_k = \frac{\text{cov}[\alpha(t),\ \alpha(t+k)]}{\text{var}[\alpha(t)]}$$

$$= \frac{\dfrac{pq(1-p-q)^k(\theta_1-\theta_2)^2}{(p+q)^2}}{\dfrac{pq(\theta_1-\theta_2)^2}{(p+q)^2}}$$

Thus

$$\rho_k = (1-p-q)^k.$$

3.2 An Illustration of the General Theory

It is now shown that the two-state Markov chain model satisfies the conditions of Theorem 2.3 and thus has as a solution a mean zero stochastic process with finite second order moments. From Theorem 2.3, the existence theorem, one can express a solution y_t as follows:

$$y_t = \sum_{r=0}^{\infty} \alpha(t)\alpha(t-1) \ \ldots\ \alpha(t-r+1)u_{t-r}. \qquad (3.2)$$

The condition for a solution with finite second order moments to exist is:

$$\sum_{b=1}^{\infty} E[\alpha^2(t)\alpha^2(t-1) \ \ldots\ \alpha^2(t-b+1)] < \infty.$$

The random coefficient $\alpha(t)$ takes on either the value θ_1 or θ_2 for all t. Recall that $0 < \theta_i < 1$. Thus $\alpha^2(t)$ takes on either the value θ_1^2 or θ_2^2 for all t such that $0 < \theta_1^2 < 1$. Now define $\theta_{max} = \max\{\theta_1, \theta_2\}$. Then $0 < \theta_{max} < 1$. Thus

$$\sum_{b=1}^{\infty} E[\alpha^2(t)\alpha^2(t-1) \ldots \alpha^2(t-b+1)] \leq \sum_{b=1}^{\infty} E[\theta_{max}^{2b}]$$

$$= \sum_{b=1}^{\infty} \theta_{max}^{2b}$$

$$\leq \sum_{b=0}^{\infty} \theta_{max}^{2b}$$

$$= \frac{1}{1-\theta_{max}^2} < \infty.$$

Note that $\theta_{max}^2 \neq 1$ since $\theta_i \neq 1$ for i = 1, 2. Since this condition is satisfied by the two-state Markov chain model, (3.2) has finite second order moments.

Theorem 2.3 guarantees that the stochastic process has mean zero; however, this can also be shown directly. One can express the mean of y_t as

$$Ey_t = E\left[\sum_{r=0}^{\infty} \alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)u_{t-r}\right] . \tag{3.3}$$

To interchange expectation and limit, one can use the result stated by Rao and given in Lemma 2.3 upon showing that

$$\sum_{r=0}^{\infty} E\left|\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)u_{t-r}\right| < \infty.$$

Recall that it is assumed that $\alpha(t)$ and u_s are independent for all t and s. This gives

$$\sum_{r=0}^{\infty} E|\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)u_{t-r}|$$

$$= \sum_{r=0}^{\infty} E|\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)|E|u_{t-r}|.$$

From properties of the folded normal distribution

$$E|u_{t-r}| = \sigma\sqrt{\frac{2}{\pi}} \qquad \forall t-r.$$

This gives

$$\sum_{r=0}^{\infty} E|\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)u_{t-r}| = \sigma\sqrt{\frac{2}{\pi}} \sum_{r=0}^{\infty} E|\alpha(t)\alpha(t-1) \ldots \cdot \alpha(t-r+1)|$$

$$\leq \sigma\sqrt{\frac{2}{\pi}} \sum_{r=0}^{\infty} E|\theta_{max}^{r}|$$

$$= \sigma\sqrt{\frac{2}{\pi}} \sum_{r=0}^{\infty} \theta_{max}^{r},$$

since θ_{max} is positive, and

$$= \frac{\sigma\sqrt{\frac{2}{\pi}}}{1-\theta_{max}} < \infty .$$

Note that σ^2 is finite and $\theta_{max} \neq 1$ since $\theta_i \neq 1$. Therefore, interchanging expectation and limit in (3.3) gives

$$Ey_t = \sum_{r=0}^{\infty} E[\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)u_{t-r}]$$

$$= E[u_{t-r}] \sum_{r=0}^{\infty} E[\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)].$$

This summation is finite since

$$\sum_{r=0}^{\infty} E[\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)] \leq \sum_{r=0}^{\infty} E[\theta_{max}^{r}]$$

$$= \frac{1}{1-\theta_{max}}, \text{ where } \theta_{max} \neq 1.$$

Thus since $E[u_{t-r}] = 0$ for all $t-r$ and the summation is finite, $Ey_t = 0$ for all t.

In order to show that the solution to the two-state Markov chain model is a weakly stationary stochastic process, one must show that the sufficient conditions of Theorem 2.4 are satisfied. It is assumed that the two-state Markov chain for the random coefficient $\alpha(t)$ is strictly stationary. Thus one must show that the random coefficients $\{\alpha(t) \mid t \in I\}$ satisfy the following three conditions sufficient to interchange expectation and limit in the general expression (2.13) for the second order moments of y_t, for each lag k:

(a)' $$\sum_{s=k}^{\infty} E|\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)\alpha(t-k)\alpha(t-k-1) \ldots \alpha(t-s+1)| < \infty$$

for fixed $r=0, 1, \ldots, k-1$,

(b)' $$\sum_{r=k}^{\infty} E|\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k)\alpha^2(t-k-1) \ldots \alpha^2(t-r+1)| < \infty,$$

and

$$(c)' \quad \sum_{r=k}^{\infty} \sum_{s=r+1}^{\infty} E\left|\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k) \ldots \alpha^2(t-r+1)\alpha(t-r) \ldots \alpha(t-s+1)\right| < \infty.$$

One should note the following structure for condition (a)':

$$(a)' \quad \sum_{s=k}^{\infty} E\left|\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)\alpha(t-k)\alpha(t-k-1) \ldots \alpha(t-s+1)\right|$$

where there are s-k+r different α's to the first power.

Replacing each $\alpha(t)$ by θ_{max} for all t gives

$$\sum_{s=k}^{\infty} E\left|\alpha(t)\alpha(t-1) \ldots \alpha(t-r+1)\alpha(t-k)\alpha(t-k-1) \ldots \alpha(t-s+1)\right|$$

$$\leq \sum_{s=k}^{\infty} \cdot E\left|\theta_{max}^{s-k+r}\right|$$

$$= \sum_{s=k}^{\infty} \theta_{max}^{s-k+r}$$

since θ_{max} is positive. Make the following transformation:

Let $t = s-k$

then $s = k \Rightarrow t = 0$ and $s-k+r = t+r$

$s = \infty \Rightarrow t = \infty$.

This gives

$$\sum_{s=k}^{\infty} \theta_{max}^{s-k+r} = \sum_{t=0}^{\infty} \theta_{max}^{t+r}$$

$$= \theta_{max}^{r} \sum_{t=0}^{\infty} \theta_{max}^{t}$$

$$= \frac{\theta_{max}^{r}}{1-\theta_{max}} < \infty.$$

Note $\theta_{max} \neq 1$ since $\theta_i \neq 1$. Thus condition (a)' is satisfied.

One should note the following structure for condition (b)':

$$\text{(b)}' \quad \sum_{r=k}^{\infty} E\left|\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k)\alpha^2(t-k-1) \ldots \alpha^2(t-r+1)\right|$$

where there are r total different α's with k α's to the first power and r-k α^2's.

Again, replacing each $\alpha(t)$ by θ_{max} for all t gives

$$\sum_{r=k}^{\infty} E\left|\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k)\alpha^2(t-k-1) \ldots \alpha^2(t-r+1)\right|$$

$$\leq \sum_{r=k}^{\infty} E\left|\theta_{max}^{2(r-k)+k}\right|$$

$$= \sum_{r=k}^{\infty} \theta_{max}^{2(r-k)+k} \qquad \text{(since } \theta_{max} \text{ is positive)}$$

$$= \theta_{max}^{k} \sum_{r=k}^{\infty} \theta_{max}^{2(r-k)}.$$

Again make the following transformation:

Let $t = r-k$

then $r = k \Rightarrow t = 0$

$r = \infty \Rightarrow t = \infty$.

This gives

$$\theta_{max}^{k} \sum_{r=k}^{\infty} \theta_{max}^{2(r-k)} = \theta_{max}^{k} \sum_{t=0}^{\infty} \theta_{max}^{2t}$$

$$= \frac{\theta_{max}^{k}}{1-\theta_{max}^{2}} < \infty.$$

Note $\theta_{max} \neq 1$ since $\theta_i \neq 1$. Thus condition (b)' is satisfied.

One should note the following structure for condition (c)':

$$(c)' \quad \sum_{r=k}^{\infty} \sum_{\substack{s=r+1 \\ r<s}}^{\infty} E\left|\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\, \alpha^2(t-k) \ldots \alpha^2(t-r+1) \cdot \alpha(t-r) \ldots \alpha(t-s+1)\right|$$

where there are s total different α's with $s-r+k$ α's to the first power and $r-k$ α^2's.

Again, replacing each $\alpha(t)$ by θ_{max} gives

$$\sum_{r=k}^{\infty} \sum_{\substack{s=r+1 \\ r<s}}^{\infty} E\left|\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k) \ldots \alpha^2(t-r+1)\alpha(t-r) \ldots \cdot \alpha(t-s+1)\right|$$

$$\leq \sum_{r=k}^{\infty} \sum_{s=r+1}^{\infty} E\left|\theta_{max}^{2(r-k)+s-r+k}\right|$$

$$= \sum_{r=k}^{\infty} \sum_{s=r+1}^{\infty} \theta_{max}^{2(r-k)+s-r+k} \qquad \text{(since } \theta_{max} \text{ is positive)}$$

$$= \theta_{max}^{k} \sum_{r=k}^{\infty} \theta_{max}^{2(r-k)} \sum_{s=r+1}^{\infty} \theta_{max}^{s-r}.$$

Make the following transformation for the second summation:

Let $t = s-r-1$

then $s = r+1 \Rightarrow t = 0$ and $s-r = t+1$

$s = \infty \Rightarrow t = \infty$.

This gives

$$\theta_{max}^{k} \sum_{r=k}^{\infty} \theta_{max}^{2(r-k)} \sum_{s=r+1}^{\infty} \theta_{max}^{s-r} = \theta_{max}^{k} \sum_{r=k}^{\infty} \theta_{max}^{2(r-k)} \sum_{t=0}^{\infty} \theta_{max}^{t+1}$$

$$= \theta_{max}^{k+1} \sum_{r=k}^{\infty} \theta_{max}^{2(r-k)} \sum_{t=0}^{\infty} \theta_{max}^{t}$$

$$= \frac{\theta_{max}^{k+1}}{1-\theta_{max}} \sum_{r=k}^{\infty} \theta_{max}^{2(r-k)}.$$

Make the following transformation:

Let $t = r-k$

then $r = k \Rightarrow t = 0$

$r = \infty \Rightarrow t = \infty$.

This gives

$$\frac{\theta_{max}^{k+1}}{1-\theta_{max}} \sum_{r=k}^{\infty} \theta_{max}^{2(r-k)} = \frac{\theta_{max}^{k+1}}{1-\theta_{max}} \sum_{t=0}^{\infty} \theta_{max}^{2t}$$

$$= \frac{\theta_{max}^{k+1}}{(1-\theta_{max})(1-\theta_{max}^{2})} < \infty.$$

Note that $\theta_{max} \neq 1$ since $\theta_i \neq 1$. Thus condition (c)' is satisfied.

Therefore, the solution y_t is a weakly stationary stochastic process; and the second order moments of y_t are independent of t and a function only of the lag k. An explicit expression for the second order moments of y_t in the general case is given in (2.14):

$$Ey_t E_{t+k} = \sigma^2 \sum_{r=k}^{\infty} E[\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k) \ldots \alpha^2(t-r+1)].$$

In the next chapter expressions for these second order moments of y_t in terms of the parameters of the two-state Markov chain model are developed. This autocovariance structure will then play the major role in developing an estimation procedure for obtaining estimates of the parameters in the two-state Markov chain model as well as developing an optimal predictor for this model.

It is important to note that one could show that Theorem 2.3 is satisfied in the general case where $\alpha(t)$ is modeled as an n-state Markov chain for any positive integer n where $0 < \theta_i < 1$ for i= 1, 2, ..., n. This can be done by defining $\theta_{max} = \max\{\theta_1, \theta_2, \ldots, \theta_n\}$ and using the same techniques used for the two-state Markov chain model. Thus a solution exists as a mean zero stochastic process with finite second order moments. The sufficient conditions for Theorem 2.4 can be shown to be satisfied using the same techniques. Thus assuming that the n-state Markov chain for $\alpha(t)$ is strictly stationary, then Theorem 2.4 is satisfied and the solution is a weakly stationary stochastic process.

Attention is restricted to the two-state Markov chain model in this study.

CHAPTER IV

AN ESTIMATION AND PREDICTION TECHNIQUE FOR THE MARKOV CHAIN MODEL

Let y_t be a weakly stationary solution to the generalized autoregressive model of order one. The solution y_t has autocovariance function given by (2.14) as

$$Ey_t y_{t-k} = \sigma^2 \sum_{r=k}^{\infty} E[\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k)\alpha^2(t-k-1)\ldots\alpha^2(t-r+1)] \qquad t \in I. \tag{4.1}$$

Recall that $\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1) \overset{\text{def}}{=} 1$ for $k = 0$ and $\alpha^2(t-k)\alpha^2(t-k-1) \ldots \alpha^2(t-r+1) \overset{\text{def}}{=} 1$ for $r = k$. In section 4.1 approximate expressions are provided for (4.1) for lags $k = 0, 1, 2$ in the particular case where the random coefficients $\alpha(t)$ are generated by a two-state Markov chain.

These expressions are then used in section 4.2 to obtain estimates of the parameters of this two-state Markov chain model by employing the method of moments. A numerical procedure must be used. A detailed description of the one used in this study is given in Appendix 4A.1 of this chapter.

In section 4.3 a simulation study is used to compare the two-state Markov chain model and the AR(1) model with a fixed coefficient with respect to adequacy of the estimated models and their prediction ability. Some conclusions comparing the estimates of each model and

comparing the optimal predictors for each model are made, but a need for further evaluation and comparison is indicated.

4.1 The Autocovariance Structure of the Two-State Markov Chain Model

One needs to obtain expressions for moments of all orders for $\alpha(t)$ in order to evaluate the infinite sum in expression (4.1) directly for any specified lag k. In practice this cannot be done. It is shown in Chapter II that this infinite sum converges, and thus it seems reasonable to consider a finite truncation of the sum. Since the purpose of this part of the study is to illustrate the theory, and in particular the estimation theory for the two-state Markov chain model; a rather short but reasonable truncation is used. Later it is shown that for lags k = 0, 1, 2 this truncation gives reasonable estimates for the parameters using the method of moments and thus must give reasonable expressions for the autocovariances at lags k = 0, 1, 2. Remember, the goal here is not extreme accuracy. These expressions only give approximations, but one can make the approximations as close as desirable to the true autocovariances by using more terms in the truncated sums.

Consider the following finite truncation for the autocovariance at each lag k, $k \in I$, for the weakly stationary solution y_t of the two-state Markov chain model:

$$E^{*}y_t y_{t-k} = \sigma^2 \sum_{r=k}^{6} E[\alpha(t)\alpha(t-1) \ldots \alpha(t-k+1)\alpha^2(t-k)\alpha^2(t-k-1) \ldots \alpha^2(t-r+1)] \qquad t \in I. \tag{4.2}$$

Evaluating this expression for the different values of k gives explicitly this approximation for the autocovariance structure of y_t. One

should note that using this particular approximation gives $E^*y_t y_{t-k} = 0$ for $k \geq 7$ for all t.

For the values k = 0, 1, and 2 one obtains the approximations for the variance and the first and second order autocovariances of the process y_t listed below. These are used later as part of the estimation procedure. From (4.2) one has for k = 0:

$$\begin{aligned}\text{var}^*(y_t) = E^*y_t^2 = \sigma^2\{1 &+ E[\alpha^2(t)] + E[\alpha^2(t)\alpha^2(t-1)] + E[\alpha^2(t)\alpha^2(t-1) \\ &\cdot \alpha^2(t-2)] + E[\alpha^2(t)\alpha^2(t-1)\alpha^2(t-2)\alpha^2(t-3)] \\ &+ E[\alpha^2(t)\alpha^2(t-1)\alpha^2(t-2)\alpha^2(t-3)\alpha^2(t-4)] \\ &+ E[\alpha^2(t)\alpha^2(t-1)\alpha^2(t-2)\alpha^2(t-3)\alpha^2(t-4)\alpha^2(t-5)]\}.\end{aligned}$$

For k = 1:

$$\begin{aligned}E^*y_t y_{t-1} = \sigma^2\{&E[\alpha(t)] + E[\alpha(t)\alpha^2(t-1)] + E[\alpha(t)\alpha^2(t-1)\alpha^2(t-2)] \\ &+ E[\alpha(t)\alpha^2(t-1)\alpha^2(t-2)\alpha^2(t-3)] + E[\alpha(t)\alpha^2(t-1)\alpha^2(t-2)\alpha^2(t-3) \\ &\cdot \alpha^2(t-4)] + E[\alpha(t)\alpha^2(t-1)\alpha^2(t-2)\alpha^2(t-3)\alpha^2(t-4)\alpha^2(t-5)]\}.\end{aligned}$$

For k = 2:

$$\begin{aligned}E^*y_t y_{t-2} = \sigma^2\{&E[\alpha(t)\alpha(t-1)] + E[\alpha(t)\alpha(t-1)\alpha^2(t-2)] + E[\alpha(t)\alpha(t-1)\alpha^2(t-2) \\ &\cdot \alpha^2(t-3)] + E[\alpha(t)\alpha(t-1)\alpha^2(t-2)\alpha^2(t-3)\alpha^2(t-4)] \\ &+ E[\alpha(t)\alpha(t-1)\alpha^2(t-2)\alpha^2(t-3)\alpha^2(t-4)\alpha^2(t-5)]\}.\end{aligned}$$

Recall that the joint distribution of a set of α's, say $\alpha(s), \ldots, \alpha(t)$ can be expressed in terms of the transition function

$$P(\theta_1, \theta_2) = p \qquad P(\theta_1, \theta_1) = 1-p$$

$$P(\theta_2, \theta_1) = q \qquad P(\theta_2, \theta_2) = 1-q$$

and the initial distribution

$$\pi_0(\theta_1) = \frac{q}{p+q} \text{ and } \pi_0(\theta_2) = \frac{p}{p+q}$$

as

$$P(\alpha(s) = \alpha_s, \alpha(s+1) = \alpha_{s+1}, \ldots, \alpha(t) = \alpha_t) = \pi_0(\alpha_s)P(\alpha_s, \alpha_{s+1}) \ldots \cdot P(\alpha_{t-1}, \alpha_t).$$

Using this probability structure one can evaluate each of the indicated moments in the summations above in terms of the parameters θ_1, θ_2, p and q. This gives the following expressions for the variance and the first and second order autocovariances of the process y_t in terms of the parameters σ^2, θ_1, θ_2, p and q. For k = 0:

$$\begin{aligned}\text{var}^*(y_t) = \frac{\sigma^2}{p+q} \{&p+q + [\theta_1^2 q + \theta_2^2 p] + [\theta_1^4 q(1-p) + \theta_1^2\theta_2^2 \cdot 2pq \\ &+ \theta_2^4 p(1-q)] + [\theta_1^6 q(1-p)^2 + \theta_1^4\theta_2^2 pq(2(1-p)+q) + \theta_1^2\theta_2^4 pq \\ &\cdot (p + 2(1-q)) + \theta_2^6 p(1-q)^2] + [\theta_1^8 q(1-p)^3 + \theta_1^6\theta_2^2 \\ &\cdot 2pq(1-p)(1-p+q) + \theta_1^4\theta_2^4 pq[2(1-p)(1-q) + 2pq + q(1-q) \\ &+ p(1-p)] + \theta_1^2\theta_2^6 \cdot 2pq(1-q)(1+p-q) + \theta_2^8 p(1-q)^3] \\ &+ [\theta_1^{10} q(1-p)^4 + \theta_1^8\theta_2^2 pq(1-p)^2(2-2p+3q) + \theta_1^6\theta_2^4 pq[(1-p)^2\end{aligned}$$

$$\cdot (2+p-2q) + 2q(1-p)(1-q) + pq(4-4p+q)] + \theta_1^4\theta_2^6 pq[(1-q)^2(2-2p+q)$$

$$+ 2p(1-p)(1-q) + pq(4+p-4q)] + \theta_1^2\theta_2^8 pq(1-q)^2(2+3p-2q)$$

$$+ \theta_2^{10} p(1-q)^4] + [\theta_1^{12} q(1-p)^5 + \theta_1^{10}\theta_2^2 \cdot 2pq(1-p)^3(1-p+2q)$$

$$+ \theta_1^8\theta_2^4 pq(1-p)[(1-p)^2(2+p-2q) + 3q(1-p)(1+2p-q) + 3pq^2]$$

$$+ \theta_1^6\theta_2^6 \cdot 2pq\,[(1-p)(1-q)(1-p-q+5pq) + (1-p-q+2pq)[q(1-q)$$

$$+ p(1-p)] + p^2q^2] + \theta_1^4\theta_2^8 pq(1-q)[(1-q)^2(2-2p+q) + 3p(1-q)$$

$$\cdot (1-p+2q) + 3p^2q] + \theta_1^2\theta_2^{10} \cdot 2pq(1-q)^3(1+2p-q) + \theta_2^{12} p(1-q)^5]\}.$$

For k = 1:

$$E\overset{*}{y}_t y_{t-1} = \frac{\sigma^2}{p+q}\{[\theta_1 q + \theta_2 p] + [\theta_1^3 q(1-p) + \theta_1^2\theta_2 pq + \theta_1\theta_2^2 pq + \theta_2^3 p(1-q)]$$

$$+ [\theta_1^5 q(1-p)^2 + \theta_1^4\theta_2 pq(1-p) + \theta_1^2\theta_2^3 pq(1+p-q) + \theta_1^3\theta_2^2 pq$$

$$\cdot(1-p+q) + \theta_2^4\theta_1 pq(1-q) + \theta_2^5 p(1-q)^2] + [\theta_1^7 q(1-p)^3$$

$$+ \theta_1^6\,\theta_2 pq(1-p)^2 + \theta_1^5\theta_2^2 pq(1-p)(1-p+2q) + \theta_1^4\theta_2^3 pq$$

$$\cdot [(1-p)(1+p-q)+pq] + \theta_1^3\theta_2^4 pq[(1-q)(1-p+q)+pq] + \theta_1^2\theta_2^5 pq$$

$$\cdot (1-q)(1+2p-q) + \theta_1\theta_2^6 pq(1-q)^2 + \theta_2^7 p(1-q)^3] + [\theta_1^9 q(1-p)^4$$

$$+ \theta_1^8\theta_2 pq(1-p)^3 + \theta_1^7\theta_2^2 pq(1-p)^2(1-p+3q) + \theta_1^6\theta_2^3 pq(1-p)$$

$$\cdot [(1-p)(1-q) + p(1-p+2q)] + \theta_1^5\theta_2^4 pq[(1-p)(1-q)(1-p+2q)$$

$$+ pq(2-2p+q)] + \theta_1^4\theta_2^5 pq\,[(1-p)(1-q)(1+2p-q) + pq(2+p-2q)]$$

$$+ \theta_1^3\theta_2^6 pq(1-q)[q(1+2p-q) + (1-p)(1-q)] + \theta_1^2\theta_2^7 pq(1-q)^2$$

$$\cdot (1+3p-q) + \theta_1\theta_2^8 pq(1-q)^3 + \theta_2^9 p(1-q)^4] + [\theta_1^{11} q(1-p)^5$$

$$+ \theta_1^{10}\theta_2 pq(1-p)^4 + \theta_1^9\theta_2^2 pq(1-p)^3(1-p+4q) + \theta_1^8\theta_2^3 pq(1-p)^2$$

$$\cdot [(1-p)(1+p-q) + 3pq] + \theta_1^7\theta_2^4 pq(1-p)[(1-p)(1-q)(1-p+3q)$$

$$+ 3pq(1-p+q)] + \theta_1^6\theta_2^5 pq[(1-p)^2(1-q)(1+2p-q) + 2pq(1-p)$$

$$\cdot (2+p-2q) + p^2q^2] + \theta_1^5\theta_2^6 pq[(1-p)(1-q)^2(1-p+2q) + 2pq$$

$$\cdot (1-q)(2-2p+q) + p^2q^2] + \theta_1^4\theta_2^7 pq(1-q)[(1-p)(1-q)(1+3p-q)$$

$$+ 3pq(1+p-q)] + \theta_1^3\theta_2^8 pq(1-q)^2 [(1-q)(1-p+q) + 3pq]$$

$$+ \theta_1^2\theta_2^9 pq(1-q)^3(1+4p-q) + \theta_1\theta_2^{10} pq(1-q)^4 + \theta_2^{11} p(1-q)^5]\}.$$

For k = 2:

$$Ey_t^* y_{t-2} = \frac{\sigma^2}{p+q} \{[\theta_1^2 q(1-p) + \theta_1\theta_2 \cdot 2pq + \theta_2^2 p(1-q)] + [\theta_1^4 q(1-p)^2$$

$$+ \theta_1^3\theta_2 pq(1-p+q) + \theta_1^2\theta_2^2 pq(2-p-q) + \theta_1\theta_2^3 pq(1+p-q)$$

$$+ \theta_2^4 p(1-q)^2] + [\theta_1^6 q(1-p)^3 + \theta_1^5\theta_2 pq(1-p)(1-p+q)$$

$$+ \theta_1^4\theta_2^2 pq(1-p)(2-p) + \theta_1^3\theta_2^3 pq[2pq+q(1-q) + p(1-p)]$$

$$+ \theta_1^2\theta_2^4 pq(1-q)(2-q) + \theta_1\theta_2^5 pq(1-q)(1+p-q) + \theta_2^6 p(1-q)^3]$$

$$+ [\theta_1^8 q(1-p)^4 + \theta_1^7\theta_2 pq(1-p)^2(1-p+q) + \theta_1^6\theta_2^2 pq(1-p)^2$$

$$\cdot (2-p+q) + \theta_1^5\theta_2^3 pq[3pq(1-p)+q(1-p)(1-q) + pq^2 + p(1-p)^2]$$

$$+ \theta_1^4\theta_2^4 pq[2(1-p)(1-q) + pq(2-p-q)] + \theta_1^3\theta_2^5 pq[3pq(1-q)$$

$$+ q(1-q)^2 + p^2q + p(1-p)(1-q)] + \theta_1^2\theta_2^6 pq(1-q)^2(2+p-q)$$

$$+ \theta_1\theta_2^7 pq(1-q)^2(1+p-q) + \theta_2^8 p(1-q)^4] + [\theta_1^{10} q(1-p)^5$$

$$+ \theta_1^9\theta_2 pq(1-p)^3(1-p+q) + \theta_1^8\theta_2^2 pq(1-p)^3(2-p+2q) + \theta_1^7\theta_2^3 pq$$

$$\cdot (1-p)[q(1-p)(1+4p-q) + 2pq^2 + p(1-p)^2] + \theta_1^6\theta_2^4 pq(1-p)$$

$$\cdot [(1-p)(1-q)(2+q) + 2pq(2-p-q) + pq^2] + \theta_1^5\theta_2^5 pq$$

$$\cdot [(1-p-q+3pq)[q(1-q) + p(1-p)] + 2p^2q^2 + 4pq(1-p)(1-q)]$$

$$+ \theta_1^4\theta_2^6 pq(1-q)[(1-p)(1-q)(2+p)+2pq(2-p-q) + p^2q]$$

$$+ \theta_1^3\theta_2^7 pq(1-q)[p(1-q)(1-p+4q)+q(1-q)^2+2p^2q] + \theta_1^2\theta_2^8 pq$$

$$\cdot (1-q)^3(2+2p-q) + \theta_1\theta_2^9 pq(1-q)^3(1+p-q) + \theta_2^{10} p(1-q)^5]\}.$$

In many practical situations, it is reasonable to assume that the random coefficient $\alpha(t)$ jumps between the two states of the Markov chain at the same rate, or with the same probability. Thus $p = q$ for this special case. Since the main purpose of this part of the study is to illustrate the theory associated with the random coefficient models, this special class of two-state Markov chain models with $p = q$ will be studied in detail. With $p = q$ one has the following expressions for the variance and the first and second order autocovariances of the process y_t in terms of the parameters σ^2, θ_1, θ_2, and p. For $k = 0$:

$$\text{Var}^*(y_t) = \frac{\sigma^2}{2}\{2 + \theta_1^2 + \theta_2^2 + \theta_1^4 + \theta_2^4 - p(\theta_1^2 - \theta_2^2)^2 + (1-p)^2$$
$$\cdot (\theta_1^6 + \theta_2^6) + p(2-p)\theta_1^2\theta_2^2(\theta_1^2 + \theta_2^2) + (1-p)^3(\theta_1^8 + \theta_2^8)$$
$$+ 2p(1-p)\theta_1^2\theta_2^2(\theta_1^4 + \theta_2^4) + 2p(1-p+p^2)\theta_1^4\theta_2^4 + (1-p)^4$$
$$\cdot (\theta_1^{10} + \theta_2^{10}) + p(1-p)^2(2+p)\theta_1^2\theta_2^2(\theta_1^6 + \theta_2^6) + p(2-3p+4p^2$$
$$-2p^3)\theta_1^4\theta_2^4(\theta_1^2 + \theta_2^2) + (1-p)^5(\theta_1^{12} + \theta_2^{12}) + 2p(1-p)^3$$
$$\cdot (1+p)\theta_1^2\theta_2^2(\theta_1^8 + \theta_2^8) + p(1-p)[(1-p)^2(2-p) + 3p]$$
$$\cdot \theta_1^4\theta_2^4(\theta_1^4 + \theta_2^4) + 2p[(1-p)(1-p+3p^2-p^3) + p^4]\theta_1^6\theta_2^6\}. \tag{4.3}$$

For k = 1:

$$E^*y_t y_{t-1} = \frac{\sigma^2}{2}\{\theta_1 + \theta_2 + \theta_1^3 + \theta_2^3 - p(\theta_1^3 + \theta_2^3 - \theta_1\theta_2(\theta_1 + \theta_2))$$
$$+ (1-p)^2(\theta_1^5 + \theta_2^5) + p(1-p)\theta_1\theta_2(\theta_1^3 + \theta_2^3) + p\theta_1^2\theta_2^2(\theta_1 + \theta_2)$$
$$+ (1-p)^3(\theta_1^7 + \theta_2^7) + p(1-p)^2\theta_1\theta_2(\theta_1^5 + \theta_2^5) + p(1-p^2)$$
$$\cdot \theta_1^2\theta_2^2(\theta_1^3 + \theta_2^3) + p(1-p+p^2)\theta_1^3\theta_2^3(\theta_1 + \theta_2) + (1-p)^4$$
$$\cdot (\theta_1^9 + \theta_2^9) + p(1-p)^3\theta_1\theta_2(\theta_1^7 + \theta_2^7) + p(1-p)^2(1+2p)$$
$$\cdot \theta_1^2\theta_2^2(\theta_1^5 + \theta_2^5) + p(1-p)(1-p+2p^2)\theta_1^3\theta_2^3(\theta_1^3 + \theta_2^3)$$
$$+ p(1-p+p^2)\theta_1^4\theta_2^4(\theta_1 + \theta_2) + (1-p)^5(\theta_1^{11} + \theta_2^{11}) + p(1-p)^4$$
$$\cdot \theta_1\theta_2(\theta_1^9 + \theta_2^9) + p(1-p)^3(1+3p)\theta_1^2\theta_2^2(\theta_1^7 + \theta_2^7) + p(1-p)^2$$

$$\cdot (1-p+3p^2)\theta_1^3\theta_2^3(\theta_1^5 + \theta_2^5) + p(1-p)(1+2p^3)\theta_1^4\theta_2^4(\theta_1^3 + \theta_2^3)$$

$$+ p[(1-p)^3(1+p) + 2p^2(1-p)(2-p) + p^4]\theta_1^5\theta_2^5(\theta_1 + \theta_2)\}. \quad (4.4)$$

For k = 2:

$$E\overset{*}{y}_t y_{t-2} = \frac{\sigma^2}{2}\{\theta_1^2 + \theta_2^2 - p(\theta_1 - \theta_2)^2 + (1-p)^2(\theta_1^4 + \theta_2^4) + p\theta_1\theta_2$$

$$\cdot (\theta_1^2 + \theta_2^2) + 2p(1-p)\theta_1^2\theta_2^2 + (1-p)^3(\theta_1^6 + \theta_2^6)$$

$$+ p(1-p)\theta_1\theta_2(\theta_1^4 + \theta_2^4) + p(1-p)(2-p)\theta_1^2\theta_2^2(\theta_1^2 + \theta_2^2)$$

$$+ 2p^2\theta_1^3\theta_2^3 + (1-p)^4(\theta_1^8 + \theta_2^8) + p(1-p)^2\theta_1\theta_2(\theta_1^6 + \theta_2^6)$$

$$+ 2p(1-p)^2\theta_1^2\theta_2^2(\theta_1^4 + \theta_2^4) + p^2(2-p)\theta_1^3\theta_2^3(\theta_1^2 + \theta_2^2)$$

$$+ 2p(1-p)(1-p+p^2)\theta_1^4\theta_2^4 + (1-p)^5(\theta_1^{10} + \theta_2^{10}) + p(1-p)^3$$

$$\cdot \theta_1\theta_2(\theta_1^8 + \theta_2^8) + p(1-p)^3(2+p)\theta_1^2\theta_2^2(\theta_1^6 + \theta_2^6) + 2p^2$$

$$\cdot (1-p)\theta_1^3\theta_2^3(\theta_1^4 + \theta_2^4) + p(1-p)[(1-p)^2(2+p) + 4p^2(1-p)$$

$$+ p^3]\theta_1^4\theta_2^4(\theta_1^2 + \theta_2^2) + 2p^2(1-p+p^2)\theta_1^5\theta_2^5\}. \quad (4.5)$$

4.2 Estimation Using the Method of Moments

Since the variance and first and second order autocovariances of the solution y_t are functions of the parameters $\sigma^2, \theta_1, \theta_2$, and p, one can use one of the simplest methods of estimation for the parameters, the method of moments.

In general, suppose one is interested in estimating some parameter θ such that

$$\theta = h(m_1, m_2, \ldots, m_k), \tag{4.6}$$

where h is some known function of the m_j's and m_j is the j^{th}-order moment of the population distribution known to exist for $1 \leq j \leq k$. The method of moments consists in estimating θ by the statistic

$$T(X_1, X_2, \ldots, X_n) = h(\frac{1}{n} \sum_{i=1}^{n} X_i, \frac{1}{n} \sum_{i=1}^{n} X_i^2, \ldots, \frac{1}{n} \sum_{i=1}^{n} X_i^k), \tag{4.7}$$

where $\{X_i | i=1, 2, \ldots, n\}$ is a set of observations. T is a statistic if one assumes that $h:R_k \to R$ is a Borel-measurable function. One can extend the method to the estimation of joint moments by using $\frac{1}{n} \sum_{i=1}^{n} X_i Y_i$ to estimate E(XY), etc. If one estimates parameters of the type θ defined in (4.6) where h is some continuous function, the estimates $T(X_1, X_2, \ldots, X_n)$ defined in (4.7) are consistent for θ. Under some mild conditions on h given in Cramér [1946], the estimate T is also asymptotically normal. Also, to avoid ambiguity one takes the estimate involving the lowest-order sample moments. This discussion of the method of moments is based on a brief description given in Rohatgi [1976, p. 373].

The following estimation procedure uses the method of moments to obtain estimates for the parameters of the two-state Markov chain model with p = q. First, one should note that using expression (4.3) the variance σ^2 of the random noise process can be expressed as an explicit function of the other three parameters: θ_1, θ_2, and p, and the theoretical variance of the y_t process. It is shown later that once one obtains estimates for θ_1, θ_2, and p and an estimate for the variance of

y_t, one can calculate an estimate for σ^2. Therefore, σ^2 is estimated last.

One should also note that if $\theta_1 = \theta_2$ then no matter what value p takes on, the two-state Markov chain model reduces to the usual AR(1) model. Each value of p defines a richer class of models which includes the linear stochastic AR(1) model with a fixed coefficient as a special case. For this study a value of p is chosen and a model is selected from the class defined by p. The selection of p is not addressed in depth; however, the following discussion suggests three different methods for the selection of p.

Recall that p is the probability of going to state θ_j in one step conditional on being in the other state θ_k in the two-state Markov chain with state space $S = \{\theta_1, \theta_2\}$ for the random coefficient $\alpha(t)$. Note that at a given point in time t_0, $\alpha(t_0)$ takes on a specific value, either θ_1 or θ_2. For a fixed coefficient AR(1) model the coefficient α is also the first order autocorrelation ρ_1 of the process. Thus for the random coefficient model one is actually modeling a set of observations where the autocorrelation structure changes randomly between two possible values in a stationary manner. The rate of change in the autocorrelation structure is actually what the value of p indicates. One may have a priori knowledge of this rate of change in autocorrelation structure and select p accordingly. A second approach is to select a value for p based on an examination of the data. The changes in autocorrelation structure can often be detected by observing a plot of the entire set of observations if the changes occur at a fairly slow rate, such as once every 10 observations on the average, once every 20

observations on the average, once every 50 observations on the average, etc. A third approach is to use the method of moments to estimate p along with θ_1 and θ_2. This would require obtaining an expression for $Ey_t y_{t-3}$, the third order autocovariance, in order to have as many equations as unknowns for the method of moments.

Given a value for p, one can obtain estimates for the two states θ_1 and θ_2 in the following manner using the method of moments and a numerical approximation technique. The first and second order autocorrelations

$$\rho_1 = \frac{E^*y_t y_{t-1}}{Var^*(y_t)}$$

$$\rho_2 = \frac{E^*y_t y_{t-2}}{Var^*(y_t)}$$

where $Var^*(y_t)$ is the approximation in (4.3), $E^*y_t y_{t-1}$ is the approximation in (4.4), and $E^*y_t y_{t-2}$ is the approximation in (4.5), are functions of θ_1, θ_2, and p. From the data one can calculate an estimate of the autocorrelation for each lag k:

$$\hat{\rho}_k = \frac{\hat{C}(k)}{\hat{C}(0)}$$

where

$$\hat{C}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y}) \qquad k = 0, 1, 2, \ldots .$$

In particular, one has

$$\hat{C}(0) = \frac{1}{n} \sum_{t=1}^{n} (y_t - \bar{y})(y_t - \bar{y}) = \frac{1}{n} \{ \sum_{t=1}^{n} y_t^2 - \frac{(\sum_{t=1}^{n} y_t)^2}{n} \}$$

$$\hat{C}(1) = \frac{1}{n} \sum_{t=1}^{n-1} (y_t - \bar{y})(y_{t+1} - \bar{y})$$

$$\hat{C}(2) = \frac{1}{n} \sum_{t=1}^{n-2} (y_t - \bar{y})(y_{t+2} - \bar{y})$$

where

$$\bar{y} = \frac{1}{n} \sum_{t=1}^{n} y_t .$$

Thus estimates for the first and second order autocorrelations are given by

$$\hat{\rho}_1 = \frac{\hat{C}(1)}{\hat{C}(0)}$$

and

$$\hat{\rho}_2 = \frac{\hat{C}(2)}{\hat{C}(0)} .$$

Using the method of moments one can equate these sample autocorrelations to the theoretical expressions for the (approximate) autocorrelations

$$\hat{\rho}_1 = \frac{E^* y_t y_{t-1}}{Var^*(y_t)} \tag{4.8}$$

$$\hat{\rho}_2 = \frac{E^* y_t y_{t-2}}{Var^*(y_t)} \tag{4.9}$$

and have two equations in the two unknowns θ_1 and θ_2, since a value for p has been determined. Note that the expression for each autocorrelation is a ratio which does not involve the parameter σ^2. However, the two equations (4.8) and (4.9) are nonlinear and apparently cannot be

solved directly to obtain closed form expressions for the estimates of either state, θ_1 or θ_2. A numerical procedure is therefore used. A detailed description of the procedure is given in Appendix A.1 of this chapter.

Now, using the moment estimates $\hat{\theta}_1$, $\hat{\theta}_2$, and the chosen $\hat{p}$ one can estimate the random noise variance as follows:

$$\hat{\sigma}^2 = 2s_y^2 \{2 + \hat{\theta}_1^2 + \hat{\theta}_2^2 + \hat{\theta}_1^4 + \hat{\theta}_2^4 - \hat{p}(\hat{\theta}_1^2 - \hat{\theta}_2^2)^2 + (1-\hat{p})^2(\hat{\theta}_1^6 + \hat{\theta}_2^6) + \hat{p}(2-\hat{p})\hat{\theta}_1^2\hat{\theta}_2^2(\hat{\theta}_1^2 + \hat{\theta}_2^2) + (1-\hat{p})^3(\hat{\theta}_1^8 + \hat{\theta}_2^8) + 2\hat{p}(1-\hat{p}) \cdot \hat{\theta}_1^2\hat{\theta}_2^2(\hat{\theta}_1^4 + \hat{\theta}_2^4) + 2\hat{p}(1-\hat{p}+\hat{p}^2)\hat{\theta}_1^4\hat{\theta}_2^4 + (1-\hat{p})^4(\hat{\theta}_1^{10} + \hat{\theta}_2^{10}) + \hat{p}(1-\hat{p})^2(2+\hat{p})\hat{\theta}_1^2\hat{\theta}_2^2(\hat{\theta}_1^6 + \hat{\theta}_2^6) + \hat{p}(2-3\hat{p}+4\hat{p}^2-2\hat{p}^3)\hat{\theta}_1^4\hat{\theta}_2^4(\hat{\theta}_1^2+\hat{\theta}_2^2) + (1-\hat{p})^5(\hat{\theta}_1^{12} + \hat{\theta}_2^{12}) + 2\hat{p}(1-\hat{p})^3(1+\hat{p})\hat{\theta}_1^2\hat{\theta}_2^2(\hat{\theta}_1^8 + \hat{\theta}_2^8) + \hat{p}(1-\hat{p})[(1-\hat{p})^2(2-\hat{p}) + 3\hat{p}]\hat{\theta}_1^4\hat{\theta}_2^4 \cdot (\hat{\theta}_1^4 + \hat{\theta}_2^4) + 2\hat{p}[(1-\hat{p}) \cdot (1-\hat{p}+3\hat{p}^2-\hat{p}^3) + \hat{p}^4]\hat{\theta}_1^6\hat{\theta}_2^6\}^{-1}$$

where

$$s_y^2 = \frac{\Sigma y_t^2 - \frac{(\Sigma y_t)^2}{n}}{n}$$

is the estimated variance of y_t. This estimate is the moment estimate.

Once $\hat{\sigma}^2$, $\hat{\theta}_1$, $\hat{\theta}_2$, and p are obtained the estimated model is completely specified for a given set of observations.

4.3 A Simulation Study: Comparison to the AR(1) Model

It is desirable to evaluate the performance of a two-state Markov chain model with respect to adequacy of fit and prediction ability for a set of observed data as compared to existing models, and in particular the AR(1) model. The following simulation study is done using 25 independent sets of 1000 observations generated from the two-state Markov chain model with the specific set of parameters:

$$\sigma^2 = 1.2 \qquad \theta_1 = .2 \qquad \theta_2 = .8 \qquad p = .1.$$

Thus, conditionally on θ, the generated weakly stationary process y_t jumps randomly back and forth between observations having a relatively low positive autocorrelation of .2 to observations having a high positive autocorrelation of .8 at a rate of once every ten observations on the average. Each of the 25 sets of observations is modeled with a two-state Markov chain model using the method of moments estimation procedure described in section 4.2; and each set of observations is modeled with an AR(1) model having fixed coefficient α using the techniques exemplified in Box and Jenkins [1976], which give conditional maximum likelihood estimates. Table 4.1 lists both sets of estimates for each of the 25 sets of observations.

From Table 4.1 one can make the following observations concerning the estimates. First consider the random disturbance variance σ^2. The value used for generating the data is $\sigma^2 = 1.2$. In all 25 cases Box and Jenkins overestimate the random disturbance variance. Their estimates range from about 1.27 to 1.51. However, the simple method of moments estimation procedure for the two-state Markov chain model

Table 4.1. Simulation Study: Estimation.

Data generated from:		$\sigma^2 = 1.200$	$\theta_1 = .200$	$\theta_2 = .800$	$p = .1$
Run #	Estimates:				
1	AR(1)	$\hat{\sigma}^2 = 1.395$	$\hat{\alpha} = .564$		
	MC	$\hat{\sigma}^2 = 1.167$	$\hat{\theta}_1 = .099$	$\hat{\theta}_2 = .807$	$\hat{p} = .1$
2	AR(1)	$\hat{\sigma}^2 = 1.270$	$\hat{\alpha} = .612$		
	MC	$\hat{\sigma}^2 = 1.146$	$\hat{\theta}_1 = .288$	$\hat{\theta}_2 = .798$	$\hat{p} = .1$
3	AR(1)	$\hat{\sigma}^2 = 1.420$	$\hat{\alpha} = .630$		
	MC	$\hat{\sigma}^2 = 1.315$	$\hat{\theta}_1 = .379$	$\hat{\theta}_2 = .787$	$\hat{p} = .1$
4	AR(1)	$\hat{\sigma}^2 = 1.420$	$\hat{\alpha} = .660$		
	MC	$\hat{\sigma}^2 = 1.075$	$\hat{\theta}_1 = .123$	$\hat{\theta}_2 = .908$	$\hat{p} = .1$
5	AR(1)	$\hat{\sigma}^2 = 1.297$	$\hat{\alpha} = .623$		
	MC	$\hat{\sigma}^2 = 1.168$	$\hat{\theta}_1 = .300$	$\hat{\theta}_2 = .807$	$\hat{p} = .1$
6	AR(1)	$\hat{\sigma}^2 = 1.507$	$\hat{\alpha} = .661$		
	MC	$\hat{\sigma}^2 = 1.077$	$\hat{\theta}_1 = .062$	$\hat{\theta}_2 = .927$	$\hat{p} = .1$
7	AR(1)	$\hat{\sigma}^2 = 1.425$	$\hat{\alpha} = .532$		
	MC	$\hat{\sigma}^2 = 1.272$	$\hat{\theta}_1 = .169$	$\hat{\theta}_2 = .745$	$\hat{p} = .1$
8	AR(1)	$\hat{\sigma}^2 = 1.359$	$\hat{\alpha} = .557$		
	MC	$\hat{\sigma}^2 = 1.294$	$\hat{\theta}_1 = .347$	$\hat{\theta}_2 = .701$	$\hat{p} = .1$
9	AR(1)	$\hat{\sigma}^2 = 1.483$	$\hat{\alpha} = .567$		
	MC	$\hat{\sigma}^2 = 1.233$	$\hat{\theta}_1 = .100$	$\hat{\theta}_2 = .810$	$\hat{p} = .1$
10	AR(1)	$\hat{\sigma}^2 = 1.442$	$\hat{\alpha} = .574$		
	MC	$\hat{\sigma}^2 = 1.277$	$\hat{\theta}_1 = .210$	$\hat{\theta}_2 = .780$	$\hat{p} = .1$
11	AR(1)	$\hat{\sigma}^2 = 1.446$	$\hat{\alpha} = .566$		
	MC	$\hat{\sigma}^2 = 1.242$	$\hat{\theta}_1 = .146$	$\hat{\theta}_2 = .793$	$\hat{p} = .1$
12	AR(1)	$\hat{\sigma}^2 = 1.280$	$\hat{\alpha} = .607$		
	MC	$\hat{\sigma}^2 = 1.117$	$\hat{\theta}_1 = .229$	$\hat{\theta}_2 = .813$	$\hat{p} = .1$

Table 4.1. Continued.

Data generated from:		$\sigma^2 = 1.200$	$\theta_1 = .200$	$\theta_2 = .800$	$p = .1$
Run #	Estimates:				
13	AR(1)	$\hat{\sigma}^2 = 1.382$	$\hat{\alpha} = .630$		
	MC	$\hat{\sigma}^2 = 1.294$	$\hat{\theta}_1 = .389$	$\hat{\theta}_2 = .782$	$\hat{p} = .1$
14	AR(1)	$\hat{\sigma}^2 = 1.449$	$\hat{\alpha} = .597$		
	MC	$\hat{\sigma}^2 = 1.169$	$\hat{\theta}_1 = .099$	$\hat{\theta}_2 = .843$	$\hat{p} = .1$
15	AR(1)	$\hat{\sigma}^2 = 1.409$	$\hat{\alpha} = .629$		
	MC	$\hat{\sigma}^2 = 1.297$	$\hat{\theta}_1 = .342$	$\hat{\theta}_2 = .800$	$\hat{p} = .1$
16	AR(1)	$\hat{\sigma}^2 = 1.424$	$\hat{\alpha} = .661$		
	MC	$\hat{\sigma}^2 = 1.300$	$\hat{\theta}_1 = .390$	$\hat{\theta}_2 = .824$	$\hat{p} = .1$
17	AR(1)	$\hat{\sigma}^2 = 1.374$	$\hat{\alpha} = .633$		
	MC	$\hat{\sigma}^2 = 1.198$	$\hat{\theta}_1 = .266$	$\hat{\theta}_2 = .832$	$\hat{p} = .1$
18	AR(1)	$\hat{\sigma}^2 = 1.455$	$\hat{\alpha} = .649$		
	MC	$\hat{\sigma}^2 = 1.314$	$\hat{\theta}_1 = .351$	$\hat{\theta}_2 = .823$	$\hat{p} = .1$
19	AR(1)	$\hat{\sigma}^2 = 1.302$	$\hat{\alpha} = .539$		
	MC	$\hat{\sigma}^2 = 1.172$	$\hat{\theta}_1 = .196$	$\hat{\theta}_2 = .744$	$\hat{p} = .1$
*20	AR(1)	$\hat{\sigma}^2 = 1.275$	$\hat{\alpha} = .557$		
	MC	$\hat{\sigma}^2 = .980$	$*\hat{\theta}_1 = -.010$	$\hat{\theta}_2 = .835$	$\hat{p} = .1$
21	AR(1)	$\hat{\sigma}^2 = 1.393$	$\hat{\alpha} = .551$		
	MC	$\hat{\sigma}^2 = 1.138$	$\hat{\theta}_1 = .058$	$\hat{\theta}_2 = .807$	$\hat{p} = .1$
22	AR(1)	$\hat{\sigma}^2 = 1.271$	$\hat{\alpha} = .605$		
	MC	$\hat{\sigma}^2 = 1.020$	$\hat{\theta}_1 = .101$	$\hat{\theta}_2 = .851$	$\hat{p} = .1$
23	AR(1)	$\hat{\sigma}^2 = 1.501$	$\hat{\alpha} = .607$		
	MC	$\hat{\sigma}^2 = 1.191$	$\hat{\theta}_1 = .095$	$\hat{\theta}_2 = .855$	$\hat{p} = .1$
*24	AR(1)	$\hat{\sigma}^2 = 1.464$	$\hat{\alpha} = .599$		
	MC	$\hat{\sigma}^2 = 1.034$	$*\hat{\theta}_1 = -.051$	$\hat{\theta}_2 = .891$	$\hat{p} = .1$

Table 4.1. Continued.

Data generated from:					
		$\sigma^2 = 1.200$	$\theta_1 = .200$	$\theta_2 = .800$	$p = .1$
Run #	**Estimates:**				
25	AR(1)	$\hat{\sigma}^2 = 1.351$	$\hat{\alpha} = .628$		
	MC	$\hat{\sigma}^2 = 1.037$	$\hat{\theta}_1 = .084$	$\hat{\theta}_2 = .882$	$\hat{p} = .1$

*The initial estimates of θ_1 were on the border of the initial grid of values used and the grid had to be extended to include negative values for further estimation.

underestimates the random disturbance variance in 15 cases, or 60% of the time, and overestimates it in 10 cases, or 40% of the time. The estimates range from about .98 to 1.31. It is important that in 21 cases, or 84% of the time, the new model's estimates are closer to the generation value for σ^2 than Box and Jenkins' estimates and in all 25 cases are smaller than Box and Jenkins' estimates. Thus whenever the Box and Jenkins estimate is closer to the generating value of 1.2, the new model's estimation procedure underestimates σ^2. Some summary statistics for the estimates of σ^2 are given in Table 4.2. One can observe that on the average the estimates for the two-state Markov chain model are closer to the generating value of 1.2 but are more variable. The increased variability is due to the fact that the estimation procedure is based on the method of moments whereas Box and Jenkins use maximum likelihood estimation.

Table 4.2.

		Mean	Variance
AR(1)	$\hat{\sigma}^2$	1.39176	.0054692
MC	$\hat{\sigma}^2$	1.18108	.0104829

Consider the other estimated parameters. One should note that upon observation in most cases the method of moments gives fairly close estimates for the generating values of θ_1 and θ_2. The estimates are more often too low for θ_1 but more often too high for θ_2. Table 4.3 gives some summary statistics for the estimates of α, θ_1, and θ_2. Note that on the average $\hat{\theta}_1$ and $\hat{\theta}_2$ closely approximate the generating values

of $\theta_1 = .2$ and $\theta_2 = .8$. Note again the increased variability of the estimates for the two-state Markov chain model.

Table 4.3.

		Mean	Variance
AR(1)	$\hat{\alpha}$	.601496	.0015534
MC	$\hat{\theta}_1$	.190480	.0170844
MC	$\hat{\theta}_2$	.817800	.0026262

One should note that these observations concerning the estimates of the parameters are based on only the 25 sets of data in this simulation study. It would be desirable to run a more extensive simulation study as well as investigate other tendencies or characteristics of the parameters. It would be of interest to study different pairs of $\{\theta_1, \theta_2\}$ as well as a range of values for σ^2 given a fixed pair $\{\theta_1, \theta_2\}$. These topics are left for further research.

Given the estimated models, how can one effectively compare the adequacy of fit of the two-state Markov chain model with the adequacy of fit of the AR(1) model? Usual checks for adequacy of the model involve looking at the residuals. However, residuals cannot be calculated for the two-state Markov chain model since the random coefficients $\{\alpha(t)\}$ are not known at each time of observation. Recall that one of our major goals is to forecast future time series observations. Therefore, a better method of comparison of the adequacy of the two models is to compare them with respect to their prediction ability.

Consider comparing the one-step-ahead predictions from the AR(1) model with those from the two-state Markov chain model. From Chapter I

recall that the optimal predictor for a weakly stationary time series is just the conditional expectation of the future value to be predicted, conditional on the given observations. In particular, the one-step-ahead predictor for the AR(1) model is given by

$$\hat{y}_t(1) = E[y_{t+1}|y_t, y_{t-1}, \ldots]$$

$$= \alpha y_t.$$

Upon estimation of the fixed coefficient α, one has as an estimate of the one-step-ahead predictor

$$\hat{y}_t(1) = \hat{\alpha} y_t \tag{4.10}$$

as given in (1.8). Now, the one-step-ahead predictor for the new model, the two-state Markov chain model, is again the conditional expectation of y_{t+1} given the observed data and is given by the following Lemma.

Lemma 4.1: The optimal linear one-step-ahead predictor for the two-state Markov chain model is given by

$$\hat{y}_t(1) = y_t E[\alpha(t+1)|y_t, y_{t-1}, \ldots] \tag{4.11}$$

Proof: Recall that the optimal linear one-step-ahead predictor is the conditional expectation of y_{t+1} given the observed data:

$$\hat{y}_t(1) = E[y_{t+1}|y_t, y_{t-1}, \ldots]$$

$$= E[\alpha(t+1)y_t + u_{t+1}|y_t, y_{t-1}, \ldots]$$

$$= E[\alpha(t+1)y_t|y_t, y_{t-1}, \ldots]$$

since the expectation of the future random disturbance u_{t+1} at time t is zero. Since y_t is given, one has

$$\hat{y}_t(1) = y_t E[\alpha(t+1) | y_t, y_{t-1}, \ldots].$$

It is important to note that the one-step-ahead prediction coefficient for the AR(1) model is a fixed constant whereas the one-step-ahead prediction coefficient for the two-state Markov chain model may be a function of time that depends on the observed time series up through time t. Therefore, it is evident that the two models give different predictions for the future observation one step ahead, neglecting zero probabilities.

One could use expression (4.11) directly to obtain the optimal linear predictor $\hat{y}_t(1)$ if the conditional expectation $E[\alpha(t+1) | y_t, y_{t-1}, \ldots]$ were known. However, since $E[\alpha(t+1) | y_t, y_{t-1}, \ldots]$ is unknown for this model, the following indirect procedure for approximating $\hat{y}_t(1)$ is to be used. In practice, for this model one can express $\hat{y}_t(1)$ as a linear combination of the observed time series and get the following estimate of the optimal linear predictor:

$$\hat{y}_t(1) = \sum_{k=0}^{t-1} \hat{d}_k^{(1)} y_{t-k} \tag{4.12}$$

where the $d_k^{(1)}$ are obtained from the function

$$D_1(\lambda) = \sum_{k=0}^{t-1} \hat{d}_k^{(1)} e^{-i\lambda k}$$

defined in the frequency (λ) domain approach to time series analysis. Thus the predictor can be viewed as a time series in t. Expression

(4.12) is obtained from linear prediction theory which is developed in detail in Koopmans [1974, Ch. VII]. An outline of the linear prediction theory and the numerical techniques used in this study is given in Appendix A.2 of this chapter.

In the simulation study estimates of the prediction coefficient $\hat{\alpha}$ for the AR(1) model and the $\{\hat{d}_k^{(1)} | k = 0, 1, \ldots, 20\}$ for the two-state Markov chain model are obtained and then used to predict the future observation one-step-ahead, y_{1001}, from expressions (4.10) and (4.12), respectively. Information concerning the predictions from both models for each of the 25 sets of observations in the simulation study is summarized in Table 4.4. For each model a prediction coefficient is given along with the last observation, y_{1000}. The product of these gives the one-step-ahead prediction $\hat{y}_{1000}(1)$. Then the generated value y_{1001} is given and the difference $y_{1001} - \hat{y}_{1000}(1)$ is given. This difference indicates the error in the prediction.

One should note that in 22 cases for the two-state Markov chain model $\hat{d}_0^{(1)}$ is listed as the prediction coefficient since all of the other coefficients $\{\hat{d}_k^{(1)} | k = 1, 2, \ldots, 20\}$ have negligible values. However for Run #5 the optimal predictor is expressed as a linear combination of most of the past twenty observations and Runs #8 and 22 have optimal predictors based on y_t and y_{t-10}. These predictors are listed at the end of the table.

How do the two models compare with respect to prediction ability? For each case in Table 4.4 where the prediction from the two-state Markov chain model is closer, a star is given in the right margin. One should note that the two-state Markov chain model gives a closer

Table 4.4. Simulation Study: Prediction Information.

Run #		Prediction Coefficient	y_{1000}	One-Step-Ahead Prediction $\hat{y}_{1000}(1)$	y_{1001}	$y_{1001}-\hat{y}_{1000}(1)$ Difference	
1	AR(1)	.564	-2.279	-1.285	.191	1.476	*
	MC	.508		-1.158		1.349	
2	AR(1)	.612	-1.415	- .866	-2.796	-1.930	
	MC	.583		- .824		1.972	
3	AR(1)	.630	.122	.077	.844	.767	
	MC	.618		.075		.769	
4	AR(1)	.660	- .725	- .479	.147	.625	*
	MC	.614		- .445		.592	
5	AR(1)	.623	-2.490	-1.550	-1.523	.027	
	MC	**		-2.469		.946	
6	AR(1)	.661	-3.154	-2.084	-2.865	- .781	
	MC	.600		-1.892		- .973	
7	AR(1)	.532	-1.871	- .995	- .319	.676	*
	MC	.494		- .925		.606	
8	AR(1)	.557	.491	.273	.237	- .036	
	MC	***		.277		- .040	
9	AR(1)	.567	-1.503	- .853	-1.249	- .396	
	MC	.517		- .777		- .472	

Table 4.4. Continued.

Run #		Prediction Coefficient	y_{1000}	One-Step-Ahead Prediction $\hat{y}_{1000}(1)$	y_{1001}	$y_{1001}-\hat{y}_{1000}(1)$ Difference	
10	AR(1)	.574	.046	.026	- .756	- .782	*
	MC	.535		.024		- .780	
11	AR(1)	.566	- .550	- .311	.253	.564	*
	MC	.516		- .284		.537	
12	AR(1)	.607	.811	.493	1.035	.542	
	MC	.571		.463		.572	
13	AR(1)	.630	1.382	.870	1.764	.894	
	MC	.605		.836		.928	
14	AR(1)	.597	-1.224	- .731	.866	1.597	*
	MC	.540		- .661		1.527	
15	AR(1)	.629	3.096	1.948	3.245	1.297	
	MC	.606		1.875		1.370	
16	AR(1)	.661	.800	.529	- .219	- .747	*
	MC	.642		.514		- .733	
17	AR(1)	.633	1.811	1.147	- .219	-1.366	*
	MC	.595		1.077		-1.296	
18	AR(1)	.649	-1.716	-1.114	-1.912	- .797	
	MC	.619		-1.062		- .849	

Table 4.4. Continued.

Run #		Prediction Coefficient	y_{1000}	One-Step-Ahead Prediction $\hat{y}_{1001}(1)$	y_{1001}	$y_{1001}-\hat{y}_{1000}(1)$ Difference	
19	AR(1)	.539	- .030	- .016	2.186	2.202	*
	MC	.500		- .015		2.201	
20	AR(1)	.557	.840	.468	1.236	.768	
	MC	.480		.403		.833	
21	AR(1)	.551	1.490	.822	1.517	.695	
	MC	.488		.727		.790	
22	AR(1)	.605	- .888	- .537	.678	1.215	*
	MC	****		- .466		1.144	
23	AR(1)	.607	-2.012	-1.220	-3.200	-1.980	
	MC	.546		-1.098		-2.103	
24	AR(1)	.599	1.222	.732	.936	.204	
	MC	.504		.616		.320	
25	AR(1)	.628	.448	.282	- .169	- .450	*
	MC	.566		.254		- .423	

Run #

** 5 $\hat{y}_{1000}(1) = .999y_{999} + .002y_{998} - .001y_{995} - .005y_{994} + .013y_{993} - .002y_{992} - .003y_{991} - .007y_{990} - .044y_{989} + .091y_{988} - .005y_{987} - .011y_{986} - .028y_{985} - .371y_{984} + .621y_{983} + .004y_{982} - .001y_{980}$

*** 8 $\hat{y}_{1000}(1) = .539y_{1000} - .009y_{990}$

**** 22 $\hat{y}_{1000}(1) = .532y_{1000} - .007y_{990}$

prediction in 11 cases, or 44% of the time; and the AR(1) model gives a closer prediction in 14 cases, or 56% of the time. One criterion for comparison is mean square prediction error, which is given by

$$\text{MSPE} = E[y_{t+1} - \hat{y}_t(1)]^2$$

for the one-step-ahead predictor. The optimal predictor should minimize the mean square prediction error. For the simulation study the estimates of the mean square prediction error for the two models are

AR(1) $\hat{\text{MSPE}} = 1.1597018$

MC $\hat{\text{MSPE}} = 1.2195594$

Note that the two estimates are fairly close. It is apparent that no conclusive results can be drawn concerning the superiority of either model with respect to prediction ability for data of the type generated for this one simulation study. Both models appear to do equally well. A more extensive simulation study is needed.

Appendix 4A.1. A Numerical Procedure for Estimation by the Method of Moments.

The two estimating equations

$$\hat{\rho}_1 = \frac{E^{*}y_t y_{t-1}}{Var^{*}(y_t)} \tag{4A.1.1}$$

and

$$\hat{\rho}_2 = \frac{E^{*}y_t y_{t-2}}{Var^{*}(y_t)} \tag{4A.1.2}$$

are functions of the two unknown states θ_1 and θ_2. Recall that in this study θ_1 and θ_2 may take on values between 0 and 1. For the numerical estimation procedure a grid of values for θ_1 and θ_2 are used to calculate the righthand sides of equations (4A.1.1) and (4A.1.2), and these values are compared to the sample autocorrelations $\hat{\rho}_1$ and $\hat{\rho}_2$ on the lefthand side of equations (4A.1.1) and (4A.1.2). Due to the natural symmetry of the Markov chain's structure and therefore the symmetry of the righthand sides of equations (4A.1.1) and (4A.1.2), it is assumed that $\theta_1 \leq \theta_2$. Since it is required that $\theta_i \neq 0$ and $\theta_i \neq 1$ for $i = 1, 2$; in the grid each state θ_i is allowed to take on values between .02 and .98 at intervals of .01. The first criterion that must be met in this grid procedure is that the absolute differences in the two sides of each equation must be simultaneously smaller than some specified amount. The value .007 is specified in this study. For the subset of pairs $\{\theta_1, \theta_2\}$ where the above indicated absolute differences satisfy this criterion,

the values of θ_1 and θ_2 are listed along with the two absolute differences, each absolute difference as a percentage of its corresponding sample autocorrelation (the lefthand side of the equation), and a statistic which gives the total of these two percentages. The pair of estimates $\{\theta_1^*, \theta_2^*\}$ are chosen as the pair $\{\theta_1, \theta_2\}$ that gives the minimum total percent difference, or total percent error, for the two absolute differences. A second finer grid with intervals of .001 for each θ_1, i = 1, 2 is used for the neighborhood of pairs about $\{\theta_1^*, \theta_2^*\}$ to obtain even more accurate estimates. The same procedure is used with the specified amount as the criterion to be satisfied by the absolute differences generally smaller. In this study the value specified is .005. In most cases two iterations are adequate. Again, the final estimates $\{\hat{\theta}_1, \hat{\theta}_2\}$ are chosen as the pair that gives the minimum total percent error for the two absolute differences. A Fortran computer program is used to facilitate calculations.

Appendix 4A.2. Outline of Prediction Theory and Numerical Techniques Used.

In many cases it is possible to express the γ-step-ahead predictor $\hat{Y}_t(\gamma)$ as an explicit linear combination of the random variables $Y(s)$, $s \leq t$:

$$\hat{Y}_t(\gamma) = \sum_{k=0}^{\infty} d_k^{(\gamma)} Y(t-k),$$

where the $d_k^{(\gamma)}$ are obtained from the function

$$D\gamma(\lambda) = \sum_{k=0}^{\infty} d_k^{(\gamma)} e^{-i\lambda k} \qquad (4A.2.1)$$

defined in the frequency (λ) domain approach to time series analysis.

A general outline of the methodology needed to obtain this optimal linear predictor is given first. Then a numerical procedure is outlined for calculating an approximation for the predictor based on a finite set of observations. Finally details specific to the numerical procedure used to approximate the one-step-ahead predictor for the two-state Markov chain model are given.

For a zero mean, weakly stationary time series $\{Y_t | t = 0, \pm 1, \pm 2, \ldots\}$, one has autocovariance structure given by $\{C(k) | -\infty < k < \infty\}$. In the frequency domain approach to time series analysis one can define the spectral density $f(\lambda)$ as

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} C(k) e^{-i\lambda k} \qquad \text{for } -\pi < \lambda \leq \pi \qquad (4A.2.2)$$

for a discrete time series. If the process Y_t has a continuous spectrum and if

$$\int_{-\pi}^{\pi} \log f(\lambda)d\lambda > -\infty,$$

then Y_t has an infinite one-sided moving average representation given by

$$y_t = \sum_{k=0}^{\infty} a_k u_{t-k} \qquad \text{with } a_0 = 1 \tag{4A.2.3}$$

where $\sigma_u^2 > 0$ and where $\sum_{k=0}^{\infty} a_k^2 < \infty$.

Now the spectral density $f(\lambda)$ given in (4A.2.2) can also be expressed in the following factored form:

$$f(z) = \frac{\sigma^2}{2\pi} |B(\lambda)|^2$$

where

$$B(\lambda) = \sum_{k=0}^{\infty} b_k e^{-i\lambda k}, \; b_0 = 1, \text{ and } \sum_{k=0}^{\infty} b_k^2 < \infty . \tag{4A.2.4}$$

Usually a class of functions satisfy these conditions. Consider the sequence $\{a_k | k \geq 0\}$ given by the function

$$A(z) = \sum_{k=0}^{\infty} a_k z^k, \qquad |z| < 1 \tag{4A.2.5}$$

where $A(z)$ is analytic in the region $D = \{z:|z| < 1\}$ for z complex valued. If $A(\lambda) = A(e^{-i\lambda})$ satisfies the conditions of 4A.2.4 such that $A(z)$ is analytic in the region $D = \{z:|z| \leq 1\}$, i.e. $A(z)$ can be extended to the boundary of D, then one can obtain the one-sided Fourier series expansion of $A(\lambda)$ and obtain the necessary coefficients for the one-sided moving average representation of y_t given in 4A.2.3. Remember that the

existence of this representation is all that is needed to obtain the optimal linear predictor. This will be shown later. First one must obtain the a_k's and construct $A(\lambda)$.

One can construct $A(z)$ in the following manner and then extend it to the boundary of D in such a way that (4A.2.5) remains valid. The constants $\{a_k\}$ satisfy and are uniquely determined by the conditions

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{k=0}^{\infty} a_k e^{-i\lambda k} \right|^2$$

and those listed in (4A.2.4) and can be calculated using

$$\log f(\lambda) \sim \sum_{k=-\infty}^{\infty} c_k e^{i\lambda k}$$

where

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda) e^{-i\lambda k} d\lambda$$

and the fact that

$$f(z) = e^{c_0} A(z)A(z^{-1}) = e^{\sum_{k=-\infty}^{\infty} c_k z^k},$$

which gives

$$A(z) = \sum_{k=0}^{\infty} a_k z^k = e^{\sum_{k=1}^{\infty} c_k z^k} \qquad \text{for } |z| < 1. \qquad (4A.2.6)$$

Now it is necessary to extend $A(z)$ to the boundary of D so that $A(z)$ is analytic for $|z| \leq 1$ and z may take on the value $e^{-i\lambda}$ on the unit circle. Consider (4A.2.6) and make the following transformation. Let

$$y = 1.1z \quad \rightarrow \quad z = \frac{y}{1.1} \qquad \text{for } |y| < 1.1.$$

Substituting gives

$$A\left(\frac{y}{1.1}\right) = \sum_{k=0}^{\infty} a_k \left(\frac{y}{1.1}\right)^k = e^{\sum_{k=1}^{\infty} c_k \left(\frac{y}{1.1}\right)^k}$$

or

$$A(y) = \sum_{k=0}^{\infty} \frac{a_k}{(1.1)^k} y^k = e^{\sum_{k=1}^{\infty} \frac{c_k}{(1.1)^k} y^k} \qquad \text{for } |y| < 1.1.$$

Note that the region $\{y:|y| < 1.1\}$ covers the unit circle. Now substituting $y = e^{-i\lambda}$ one has

$$A(\lambda) = A(e^{-i\lambda}) = e^{\sum_{k=1}^{\infty} \frac{c_k}{(1.1)^k} e^{-i\lambda k}}.$$

One can calculate the Fourier coefficients c_k as follows:

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda) e^{-i\lambda k} d\lambda$$

and then calculate

$$c_k^* = \frac{c_k}{(1.1)^k} \qquad \forall k$$

in order to obtain $A(\lambda)$. To get the coefficients a_k, one can use the expansion

$$A(\lambda) = \sum_{k=0}^{\infty} \frac{a_k}{(1.1)^k} e^{-i\lambda k}.$$

Thus, $\frac{a_k}{(1.1)^k}$ are Fourier coefficients and can be obtained by

$$a_k^* = \frac{a_k}{(1.1)^k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda) e^{i\lambda k} d\lambda.$$

Then

$$a_k = (1.1)^k a_k^* \qquad \forall k$$

with $a_0 = 1$. Thus one has the necessary coefficients for the one-sided moving average representation given in (4A.2.3).

Now the function $D_\gamma(\lambda)$ can be expressed as

$$D_\gamma(\lambda) = \frac{A_\gamma(\lambda)}{A(\lambda)}$$

where

$$A_\gamma(\lambda) = e^{i\lambda\gamma} \sum_{k=\gamma}^{\infty} a_k e^{-i\lambda k}$$

and

$$A(\lambda) = \sum_{k=0}^{\infty} a_k e^{-i\lambda k}.$$

Thus given the a_k's one can obtain the function $D_\gamma(\lambda)$, and from the Fourier series expansion given in (4A.2.1) one can calculate the necessary prediction coefficients $d_k^{(\gamma)}$ from

$$d_k^{(\gamma)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} D_\gamma(\lambda) e^{-i\lambda k} d\lambda.$$

In practice one only has a finite set of observations and therefore only a truncated set of autocovariances $\{C(k) \mid -M < k < M\}$ for M finite. One should choose M so that $\{C(k) \mid |k| > M\}$ are negligibly small. Note that $C(-k) = C(k)$ and $C(k)$ where $k \geq 0$ is theoretically a monotonically decreasing function. One can calculate an approximation for the optimal linear predictor using the following numerical procedure. First express the spectral density as a real function. Express the complex function $e^{-i\lambda k}$ as the sum

$$e^{-i\lambda k} = \cos(-\lambda k) + i \sin(-\lambda k).$$

One can then represent the spectral density as a sum of a real and an imaginary part. The imaginary part integrates to zero leaving the real function

$$\hat{f}(\lambda) = \frac{1}{2\pi} [C(0) + 2 \sum_{k=1}^{M} C(k) \cos(\lambda k)].$$

Then using standard integration procedures one can obtain the Fourier coefficients $\{\hat{c}_j\}$ by

$$\hat{c}_j = \frac{1}{\pi} \int_0^{\pi} \log [\frac{1}{2\pi}[C(0) + 2 \sum_{k=1}^{M} C(k) \cos (\lambda k)]] \cos (\lambda j) d\lambda$$

and calculate

$$\hat{c}_j^* = \frac{c_j}{(1.1)^j} \qquad \forall j.$$

One can then express $A(\lambda)$ as a real function and integrate to obtain the coefficients $\{\hat{a}_k^*\}$ by

$$\hat{a}_k^* = \frac{1}{\pi} \int_0^{\pi} e^{\sum_{j=1}^{M} c_j^* \cos(\lambda j)} \cos(\lambda k - \sum_{j=1}^{M} c_j^* \sin(\lambda j)) d\lambda$$

and calculate

$$\hat{a}_k = (1.1)^k \hat{a}_k^* \qquad \forall k$$

where again $a_0 = 1$. One can then express $D\gamma(\lambda)$ as a real function and integrate to obtain the $\{\hat{d}_k^{(\gamma)}\}$. In particular, upon expressing $D_1(\lambda)$ as a real function and integrating one obtains the following expression for the Fourier coefficients $\{\hat{d}_k^{(1)}\}$:

$$\hat{d}_j^{(1)} = \frac{1}{\pi}\int_0^{\pi} \left\{ \cos(\lambda(j-1)) - \frac{\cos(\lambda(j-1))\left(1+\sum_{k=1}^{M} a_k\cos(\lambda k)\right) + \sin(\lambda(j-1))\sum_{k=1}^{M} a_k\sin(\lambda k)}{\left[1+\sum_{k=1}^{M} a_k\cos(\lambda k)\right]^2 + \left[\sum_{k=1}^{M} a_k\sin(\lambda k)\right]^2} \right\} d\lambda.$$

Thus an approximation for the optimal linear one-step-ahead predictor is given by

$$\hat{y}_t(1) = \sum_{k=0}^{M} \hat{d}_k^{(1)}\, y(t-k).$$

For the simulation study for the two-state Markov chain model M is chosen to be 20 since some experimentation showed that a larger value gave little change in the coefficients. A Fortran program is used to facilitate the calculations. The program uses an IMSL (International Mathematical & Statistical Libraries) function called DCADRE in order to numerically evaluate the necessary integrals. The function DCADRE integrates a function f(x) from a to b using cautious adaptive Romberg extrapolation.

One should also note that since there is no closed form expression for the autocovariance at each lag k for k = 0, 1, ..., 20, a large sample of 100,000 observations is generated from the estimated model in each case and the autocovariances estimated empirically.

CHAPTER V

CONCLUSIONS AND FURTHER RESEARCH

5.1 Summary and Conclusions

It should be emphasized that the purpose of this study is to investigate a new approach to modeling and forecasting time series. Although this study concentrates on developing the new approach for the AR(1) model, the results look promising for extension to other linear stochastic models from the class of ARMA(p,q) models. In this study a richer class of models, the generalized AR(1) model, is developed which includes the possibility of coefficients slowly varying over time in a stochastic manner, thus not remaining constant over the observation interval for a given time series. Necessary and sufficient conditions are given to obtain a solution for the model, and sufficient conditions are given to ensure weak stationarity of the solution.

To illustrate the theory of this study the random coefficient $\alpha(t)$ is specifically modeled as a two-state Markov chain, although it is shown that the results can be extended to any n-state Markov chain for n finite. The solution for this specific generalized AR(1) model, the two-state Markov chain model, is shown to be weakly stationary. Properties of the weakly stationary solution as well as its autocovariance structure are developed in detail. Then using the autocovariance structure, an estimation procedure based on the method of moments is given

for estimating a model. This model is used to develop optimal forecasts for future observations.

One simulation study is run to compare the adequacy of the new two-state Markov chain model and the AR(1) model with respect to prediction ability. However, due to the limited size of the study, no conclusive results can be drawn concerning the superiority of either model with respect to prediction ability for data of the type generated for this simulation study.

One must conclude that further research is needed to investigate the promise of the new approach to modeling and forecasting time series for better fitting some time series with complicated dependence structures and for forecasting these time series more accurately than is possible with the currently used linear stochastic models.

5.2 Further Research

There are many directions one could take for further research. One direction is to further investigate the specific generalized AR(1) model considered in this study. A more extensive simulation study using more sets of observations may lead to more conclusive results. It is also desirable to look at different pairs of $\{\theta_1, \theta_2\}$ for the two states with different distances $|\theta_1-\theta_2|$ between the states. It is of interest to see what effect changing the random disturbance variance σ^2 has on the properties of the model. As mentioned earlier in the study, it is desirable to investigate procedures for selecting p, or perhaps estimating p along with the other parameters.

Recall from the motivation of the new generalized AR(1) model that the random coefficient $\alpha(t)$ intuitively represents the randomly changing first order autocorrelation of an AR(1) model. Thus it is of interest to extend the range of possible values for the states θ_1 and θ_2 from 0 to 1 to the range -1 to 1 and allow for negative autocorrelations in the randomly changing process. Recall that in this study the special case where $\alpha(t)$ is modeled as a two-state Markov chain with equal transition probabilities $p = q$ is investigated. One could relax this assumption and let the probabilities p and q take on different values. However, it is desirable to ensure that $p + q < 1$ so that the autocorrelation function for the two-state Markov chain given by

$$\rho_k = (1 - p - q)^k \qquad \forall k$$

remains positive. One main concern of this entirely new approach to modeling and forecasting is to allow for slowly varying random coefficients so that the variation can be detected and modeled in a suitable manner.

It would also be of some interest to get forecasts of future observations more than one step ahead to see how fast the prediction limits spread out as compared to the AR(1) model.

One important area of further research for this particular two-state Markov chain model is a better estimation procedure. It would be desirable to develop the likelihood function for observations having this structure and use maximizing procedures to get the maximum likelihood estimates of the parameters. It is of interest to develop the

distributional properties of the weakly stationary solution y_t represented by the two-state Markov chain model. Asymptotic properties should also be investigated.

Further research could be done in the area of modeling the random coefficient $\alpha(t)$. The Markov chain used in this study is only one application of the general theory. Other models that satisfy the general theory could be developed for $\alpha(t)$. Perhaps a model employing the continuous Beta distribution with range from 0 to 1 and yet allowing for the dependence of the random coefficients could be used.

Finally, the general theory of this new approach to modeling and forecasting could be extended to higher order AR models as well as other linear stochastic models from the class of ARMA models, or perhaps ARIMA models which represent some forms of nonstationary behavior. Other extensions of the theory might include relaxing some of the assumptions made in this study, such as the independence of the random coefficient $\alpha(t)$ and the random disturbance. It may be true that the random noise input has a part in randomly changing the correlation structure of the AR(1) model. Another random parameter model may consider the variance of the random disturbances as changing stochastically over time. This leads to the topic of nonstationary models with random parameters which will not be pursued here.

In conclusion, it appears that more research should be done concerning the new approach to modeling and forecasting that employs models with slowly varying random parameters.

BIBLIOGRAPHY

1. Box, G. E. P., and Jenkins, G. M. (1976), Time Series Analysis: Forecasting and Control, Holden-Day, Inc., San Francisco.

2. Cramer, H. (1946), Mathematical Methods of Statistics, Princeton University Press, Princeton.

3. Doob, J. L. (1953), Stochastic Processes, John Wiley & Sons, Inc., New York.

4. Hoel, P. G., Port, S. C., and Stone, C. J. (1972), Introduction to Stochastic Processes, Houghton Mifflin Company, Boston.

5. Kalman, R. E. (1963), "New Methods in Wiener Filtering Theory", Proc. Symp. Eng. Appl. Random Functions, (J. Bogdanov and F. Kozin, eds.), John Wiley & Sons, Inc., New York.

6. Kashyap, R. L., and Rao, A. R. (1976), Dynamic Stochastic Models From Empirical Data, Academic Press, Inc., New York.

7. Koopmans, L. H. (1974), The Spectral Analysis of Time Series, Academic Press, Inc., New York.

8. Ledolter, J. (1978), "A Recursive Approach to Parameter Estimation in Regression and Time Series Model", Paper presented at the Special Topics Meeting on Time Series Analysis, the 162nd meeting of the Institute of Mathematical Statistics, Ames, Iowa, May 1-3.

9. Miller, K. S. (1968), Linear Difference Equations, W. A. Benjamin, Inc., New York.

10. Rao, C. R. (1973), Linear Statistical Inference and Its Applications, John Wiley & Sons, Inc., New York.

11. Rohatgi, V. K. (1976), An Introduction to Probability Theory and Mathematical Statistics, John Wiley & Sons, Inc., New York.

12. Young, P. C. (1974), "Recursive Approaches to Time Series Analysis", Bulletin of Institute of Mathematics and Its Applications, 10:209-244.

13. Zadeh, L. A. (1950), "Correlation Functions and Power Spectra in Variable Networks", Proceedings of the I.R.E., 38:1342-1345.

The vita has been removed from the scanned document

A RANDOM PARAMETER APPROACH TO MODELING AND FORECASTING TIME SERIES

by

Deborah A. Guyton

(ABSTRACT)

The dependence structure of a stationary time series can be described by its autocorrelation function ρ^k. Consider the simple autoregressive model of order 1: $y_t = \alpha y_{t-1} + u_t$ where $\alpha \in (-1, 1)$ is a fixed constant and the u_t's are i.i.d. $N(0,\sigma^2)$. Here $\rho^k = \alpha^{|k|}$, $k = 0, \pm 1, \pm 2, \ldots$. It can be argued that as α ranges from 1 to -1, the behavior of the corresponding AR(1) model changes from that of a slowly changing, smooth time series to that of a rapidly changing time series. This motivates a generalized AR(1) model where the coefficient itself changes stochastically with time: $y_t = \alpha(t)y_{t-1} + u_t$ where $\alpha(t)$ is a random function of time. This dissertation gives necessary and sufficient conditions for the existence of a mean zero stochastic process with finite second order moments which is a solution to the generalized AR(1) model and gives sufficient conditions for the existence of a weakly stationary solution. The theory is illustrated with a specific model structure imposed on the random coefficient $\alpha(t)$; $\alpha(t)$ is modeled as a strictly stationary, two-state Markov chain with states taking on values between 0 and 1. The resulting generalized

AR(1) process is shown to be weakly stationary. Techniques are provided for estimating the parameters of this specific model and for obtaining the optimal predictor from the estimated model.