School of Mathematics and Statistics

Parameter Estimation of Smooth Threshold Autoregressive Models

Darfiana Nur

This thesis is presented as part of the requirements for the award of the degree of Doctor of Philosophy of the Curtin University of Technology

December 1998

Certification

I certify that the work presented in this thesis is my own work and that all references are duly acknowledged. This work has not been submitted, in whole or in part, in respect of any academic award at Curtin University of Technology or elsewhere.

Darfiana Nur

17 December 1998

Abstract

This thesis is mainly concerned with the estimation of parameters of a first-order Smooth Threshold Autoregressive (STAR) model with delay parameter one. The estimation procedures include classical and Bayesian methods from a parametric and a semiparametric point of view.

As the theoretical importance of stationarity is a primary concern in estimation of time series models, we begin the thesis with a thorough investigation of necessary or sufficient conditions for ergodicity of a first-order STAR process followed by the necessary and sufficient conditions for recurrence and classification for null-recurrence and transience.

The estimation procedure is started by using Bayesian analysis which derives posterior distributions of parameters with a noninformative prior for the STAR models of order p. The predictive performance of the STAR models using the exact one-step-ahead predictions along with an approximation to multi-step-ahead predictive density are considered. The theoretical results are then illustrated by simulated data sets and the well-known Canadian lynx data set.

The parameter estimation obtained by conditional least squares, maximum likelihood, M-estimator and estimating functions are reviewed together with their asymptotic properties and presented under the classical and parametric approaches. These estimators are then used as preliminary estimators for obtaining adaptive estimates in a semiparametric setting. The adaptive estimates

for a first-order STAR model with delay parameter one exist only for the class of symmetric error densities. At the end, the numerical results are presented to compare the parametric and semiparametric estimates of this model.

Acknowledgements

In the last three years or so, I was most fortunate for being given the opportunity to do my PhD studies at Curtin University of Technology. During this time, I have experienced good and bad times and I am indebted to people around me who have always been supportive in many ways in the completion of this thesis.

It is a pleasure for me to express my sincere gratitude to my supervisors, Dr. Nihal D. Yatawara and Dr. M. Gopalan Nair for their guidance and encouragement, depth of knowledge and many valuable criticism and suggestion on the completion of this thesis. Acknowledgement is also made to Professor Louis Caccetta for his assistance and for giving me the opportunity to be a part-time staff in the School.

I am very grateful to Center for Research on the Application and Advancement of Mathematics, Institut Teknologi Bandung, Indonesia for providing me the scholarship during the period of my study, especially to Prof. M. Ansjar and Dr. R. K. Sembiring for their guidance. I would also like to thank Institut Teknologi Bandung for giving me the opportunity to study in Australia.

The most important and sincere gratitude must go to my husband, Andi, and three beautiful children, Ahmad, Musthafa and Zahra who have always been wondering about my enthusiasm to complete my study. My little daughter Zahra needs a special mention as she kept me company during my studies in

Preface

This thesis includes the following paper:

Nair, M.G., Nur, D. and Yatawara, N.D. (1998). Bayesian analysis of Gaussian Smooth Threshold Autoregressive (GSTAR) Models. Technical Report 1/1996, School of Mathematics and Statistics, Curtin University of Technology

This paper is extracted mainly from chapter 3. Apart from the joint work represented by the paper, the rest of the thesis has been done by myself.

Contents

Certification i

Abstract ii

Acknowledgements iv

Preface vi

Contents vii

List of Figures ix

List of Tables x

1 Preliminaries **1**

1.1 Introduction 1

1.2 Notation 6

2 Ergodicity **9**

2.1 Introduction 9

2.2 Recurrence 12

2.3 Ergodicity 23

3 Bayesian Analysis **35**

3.1 Introduction 35

3.2 Model and Distribution 38

3.3 Predictive Performance 43

3.4 Examples 45

3.4.1 Simulation Study 47

3.4.2 Canadian Lynx data 63

4 Some Other Parametric Estimation **68**
4.1 Introduction 68
4.2 Conditional Least Squares 71
4.3 Maximum Likelihood estimators 74
4.4 M-estimators 77
4.5 Estimating Equations 82
5 Adaptive Estimation **84**
5.1 Introduction 84
5.2 Notation and Preliminaries 88
5.3 Local Asymptotic Normality for STAR models 95
5.4 Adaptivity 102
5.5 Efficient estimators 106
5.6 Simulation 109
6 Summary, Conclusion and Future Prospects **116**
Bibliography **120**
Appendix **127**
Appendix A. Ergodicity 127
Appendix B. Parametric Estimation 130
B.1 Conditional Least Squares 130
B.2 Maximum Likelihood estimators 132
B.3 M-estimators 134
B.4 Estimating Equations 136
Appendix C. Adaptivity 138

List of Figures

Figures		Page
Figure 2.1.	Recurrence region	13
Figure 2.2.	Necessary condition for ergodicity	24
Figure 2.3.	Classification of G light-tailed	26
Figure 2.4.	Classification of G thick-tailed	31
Figure 3.1.	Bivariate joint density	47
Figure 3.2.	Bivariate posterior density of lynx data	64
Figure 3.3.	Graph of (r, z_r)	65
Figure 3.4.	RMSE 2-steps prediction	65

List of Tables

Tables	Page
Table 3.1. Bayesian estimates	48
Table 3.2. CLS estimates	49
Table 3.3. RMSE comparison of predictors	50
Table 3.4. Parameter coefficient of five simulation models	51
Table 3.5. Bayesian estimates of model 1	52
Table 3.6. CLS estimates of model 1	53
Table 3.7. Bayesian estimates of model 2	54
Table 3.8. CLS estimates of model 2	55
Table 3.9. Bayesian estimates of model 3	56
Table 3.10. CLS estimates of model 3	57
Table 3.11. Bayesian estimates of model 4	58
Table 3.12. CLS estimates of model 4	59
Table 3.13. Bayesian estimates of model 5	60
Table 3.14. CLS estimates of model 5	61
Table 3.15. Average value of estimates	62
Table 3.16. $P(d\|X)$ of lynx data	64
Table 3.17. RMSE of lynx data	66

Tables   Page

Table 5.1. Average estimates for $a_n = 0.5$	110
Table 5.2. Average estimates for $a_n = 0.6$	111
Table 5.3. Average estimates for $a_n = 0.7$	112
Table 5.4. Average estimates for $a_n = 0.8$	113
Table 5.5. Average estimates for $a_n = 0.9$	114

Chapter 1

Preliminaries

1.1 Introduction

The past twenty years have witnessed some major developments in the field of time series analysis. The twin assumptions of linearity and stationarity which underlie so much of conventional time series have finally been abandoned and the subject has moved in a number of directions. One of these new directions is the study of *nonlinear models* which offer many challenging and unresolved problems.

One class of nonlinear models, called *threshold models*, was introduced by Tong in a long sequence of papers culminating in Tong and Lim (1980); a detailed account of the theory and application of threshold models is given in the monograph by Tong (1983). The class of *Self Exciting Threshold Autoregressive* (SETAR) models of order p introduced by Tong (1983) is typified by the following difference equation:

$$\begin{aligned} X_t &= a_0 + a_1 X_{t-1} + \ldots + a_p X_{t-p} + \\ &\quad (b_0 + b_1 X_{t-1} + \ldots + b_p X_{t-p}) I(X_{t-d} - r) + \varepsilon_t, \end{aligned} \qquad (1.1.1)$$

where d is a delay parameter, $d \in \mathcal{Z}^+$; $p \in \mathcal{Z}^+$; r is a threshold parameter, $r \in \Re$; $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables with ε_t being independent of $\{X_s, s < t\}$; and I is the indicator function given by

$$I(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Essentially, the nonlinear model (1.1.1) is composed of two linear submodels, that is, two regimes with the change point being specified by a (the real valued) threshold parameter r. To remove sudden jumps in the SETAR model (1.1.1), Chan and Tong (1986) replaced the indicator function I by a smooth function G. This leads to a new class of nonlinear time series models called *Smooth Threshold Autoregressive* (STAR) models of which an order p model is defined by

$$\begin{aligned} X_t &= a_0 + a_1 X_{t-1} + \ldots + a_p X_{t-p} + \\ &\quad (b_0 + b_1 X_{t-1} + \ldots b_p X_{t-p}) G\left(\frac{X_{t-d} - r}{z}\right) + \varepsilon_t, \qquad (1.1.2) \end{aligned}$$

where d, p, r, $\{\varepsilon_t\}$ are as defined above, z is a smoothing parameter $z \in \Re^+$ and G is a known distribution function which is assumed to be continuous. The choice of G is left very flexible, the minimal requirement being that it is continuous and nondecreasing.

The study of STAR models (1.1.2) is not as well-developed as the SETAR models (1.1.1) yet. Besides Chan and Tong (1986), there are only a few papers such as Luukkonen, *et al.* (1988) which discussed testing for linearity against STAR models, and the one by Teräsvirta (1994) which considered the specification and estimation of STAR models when G is a logistic or an exponential distribution function. Granger and Teräsvirta (1993) discussed Smooth Transition Regression model as a model for explaining the nonlinearity in economic relationships. Nur (1993) investigated conditions for ergodicity of STAR models when $p = 1 = d$ and applied the theory to some well-known data sets with G assumed to be a Gaussian distribution function. In all these studies the analysis is done from the classical or the frequentist point of view.

In many applications, a smooth threshold model may be more attractive than a threshold model. For instance, macroeconomic time series are most often the results of decisions made by a large number of economic agents. Even if one assumes that the agents make only dichotomous decisions or change their behaviour discretely, it is unlikely that they do this simultaneously. Thus if only an aggregated process is observed, then the regime changes in that process may be more accurately described as being smooth rather than discrete. In animal ecology, it also seems plausible to think that a possible change in population dynamics from a state with a small population to over population and vice versa may be continuous rather than discrete.

A few applications of the smooth threshold autoregressive model, in economics and population biology can be seen in the work of Granger and Teräsvirta (1993) and Tong (1988). Of the two illustrative examples in economics as given in Granger and Teräsvirta (1993), the first one uses international volume of industrial production indices from thirteen countries and the second one considers a possible nonlinear relationship between US gross national product and an index of leading indicators. For industrial production data, they presented testing linearity against logistic or exponential STAR models, and from Table 9.1 of Granger and Teräsvirta (1993), it can be concluded that there are 10 out of 13 countries in which the industrial production data should be modelled by logistic or exponential STAR models. It is seen generally for industrial production data that a superior fit is achieved by the nonlinear STAR models although these are less parsimonious. For the US gross national product data, they also found some evidence of nonlinearity.

Tong (1988) applied the STAR model in population biology. In 1950, the Australian entomologist, A. J. Nicholson conducted a series of experiments with blowflies, *lucilia cuprina.* The bi-daily record of one of Nicholson's experiments extending for two years, in which a caged population of approximately 1000 blowflies was initiated with a reasonably balanced sex ratio. The caged blowflies were fed a limited amount of 500 mg ground liver daily as the only source of protein which is necessary for egg production. Experimental evidence suggests that egg production usually ceases when daily protein intake for the female fly drops below 0.14 mg and levels out at 10 eggs per fly per day when protein supply is plentiful. On the unrealistic assumptions of absolute egalitarianism and sex equality among the fly population, 500 mg of ground liver will maintain 3571 flies above the minimum protein requirement for egg production. It is then transparent that in any reasonable 'balance equation' describing the time evolution of the population size, the number of births, as a function of the population size must peak at a certain critical value as a threshold. The balance equation is a delay-differential equation with the delay corresponding to the biological development time. They first fitted the SETAR model for the data but then evidence suggested that the STAR model improves the statistical goodness of fit.

Having referred the reader to some examples of the applications of the STAR models in real life, we now turn to the motivation for the current investigation and put our objectives in perspective. This thesis is mainly concerned with the estimation of parameters in stationary first-order STAR models both from frequentist and non-frequentiest points of view, and from parametric and semiparametric points of view. The original contributions to the literature of this thesis are contained in Chapters 2, 3, Sections 4.3, 4.4, 4.5 and 5.6. The results in Chapters 4 and 5 are mostly reviewed from the literature and conditions therein are verified for the STAR models as needed.

The thesis is organised as follows. The present chapter closes with a summary of notations and conventions which will be used in subsequent chapters without comment. To begin with, we present an investigation on necessary or sufficient conditions for ergodicity of the first-order STAR processes with delay parameter one as a continuation of earlier work by Nur (1993) in Chapter 2. The sufficient conditions for ergodicity of this process depend on the distribution function G. A necessary condition for ergodicity is presented for any G. These results will then be used as the assumptions needed in Chapters 4 and 5. Furthermore, we give a necessary and sufficient conditions for the process to be recurrent.

In Chapter 3, we discuss the Bayesian analysis of a STAR model of order p. This includes the derivation of the posterior distributions of parameters with a noninformative prior and the marginal distribution of intrinsic parameters, assessing the performance of the exact one-step-ahead and conditional multi-step-ahead predictive densities. Several illustrations using simulated and real data, are also provided.

Chapter 4 investigates the parameter estimation of the STAR models using various methods, namely, Conditional Least Squares, Maximum Likelihood, M-estimation and estimating function methods together with the theoretical properties of the resulting estimates. The conditional least squares and maximum likelihood estimators are reviewed from Tjøstheim (1986) and Tong (1983, 1990) while the M-estimator and the estimating function estimator are constructed based on Koul (1996) and Thavaneswaran and Abraham (1988) respectively.

Chapter 5 contains the semiparametric estimation of the STAR models which is called adaptive estimation. The construction of estimators that are asymptotically efficient in the presence of infinite dimensional nuisance parameters is the main objective of adaptive estimation. The original contributions in this chapter are basically on the adaptation of existing methods to STAR models and the presentation of numerical examples. The theoretical results are mostly verification of conditions for theorems in Koul and Schick (1997), which are not straightforward. The content of this chapter is strongly related to Chapter 2 as the ergodicity assumption is needed for most of the proofs and also to Chapter 4 as the estimator obtained in Chapter 4 is used as the preliminary estimator in adaptive estimation.

Overall conclusions of the research study are presented in Chapter 6, along with some other general comments on the estimation methods and suggestions for possible extensions and future developments.

1.2 Notation

The following notation will be used without comment in the sequel.

$a.s.$	almost surely (i.e. with probability one)
$i.i.d.$	independent and identically distributed
$r.v.$	random variable
$\Re$	set of real numbers
$\Re^+$	set of positive real numbers
$\Re^k$	k – dimensional Euclidean
Θ	an open subset of $\Re^k$
$\mathcal{Z}$	set of integers
$\mathcal{Z}^+$	set of positive integers
$\mathcal{I}$	$\mathcal{Z}$ or $\mathcal{Z}^+$

$\mathcal{B}$	Borel $\sigma-$algebra on $\Re$
$\mathcal{B}_0$	trivial $\sigma-$algebra
$\mathcal{B}_k$	$\sigma-$algebra generated by $\{X_s, 1 \leq s \leq k\}$
$\in$	is a member (belongs to)
$\ni$	such that
$\exists$	there exists
$\forall$	for all
$\wedge$	maximum
$\Leftrightarrow$	if and only if
$\leadsto$	has the same distribution as
$\sim$	approximately equal to
$\equiv$	equivalent to
$:=$	is assigned to be
$O(1)$	a sequence of r.v. that bounded in probability
$o(1)$	a sequence of r.v. that converges to 0 in probability
A^T	transpose of the matrix A
$f^{(1)}$	first derivative of f
CLT	Central Limit Theorem
CLS	Conditional Least Squares
MLE	Maximum Likelihood Estimator
$RMSE$	Root Mean Squared Error

Almost sure convergence, convergence in probability and convergence in distribution are denoted by $\xrightarrow{\text{a.s.}}$, $\xrightarrow{\text{P}}$, $\xrightarrow{\mathcal{D}}$ respectively.

$N(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2.

$\mathcal{N}(\mu, \mathcal{W})$ is the multivariate normal distribution with mean μ and covariance matrix $\mathcal{W}$.

Class of Smooth Threshold Autoregressive (STAR) models of order p defined by

$$X_t = a_0 + \sum_{j=1}^{p} a_j X_{t-j} + (b_0 + \sum_{j=1}^{p} b_j X_{t-j}) G\left(\frac{X_{t-d} - r}{z}\right) + \varepsilon_t,$$

which $G(x)$ is a rewritten expression for $G(\frac{x-r}{z})$, and $G_t = G\left(\frac{X_{t-d}-r}{z}\right)$.

The first-order STAR model with delay parameter $d = 1$ defined by

$$X_t = aX_{t-1} + bX_{t-1} G\left(\frac{X_{t-1} - r}{z}\right) + \varepsilon_t,$$

which can be rewritten as

$$X_t = aX_{t-1} + bX_{t-1} G(X_{t-1}) + \varepsilon_t.$$

Chapter 2

Ergodicity

2.1 Introduction

In this chapter we show the sufficient conditions for ergodicity of a first-order STAR process with delay parameter one under some assumptions on the tail behaviour of the distribution function G and the necessary condition for ergodicity for any G. Also, we derive some conditions under which the model as a Markov chain is recurrent, null recurrent or transient.

In time series analysis, we are generally interested in ergodic models. This is partly because of the theoretical importance of stationarity in estimation. Assuming the process is stationary and has a finite second moment for error allows us to establish large sample properties like consistency and asymptotic normality of estimators.

Unlike the linear models, it is rather hard to obtain necessary and sufficient conditions for ergodicity for a given class of nonlinear time series models. There are no general necessary and sufficient conditions available for the ergodicity of a higher order model, eventhough interesting results exist for some special cases. Consider, for example, the SETAR models as defined in (1.1.1). In relation to these models, Petrucelli and Woolford (1984) derived a necessary and sufficient condition for the ergodicity of a simple first-order threshold process with delay parameter one. In another development, Chan, *et. al.* (1985) obtained necessary and sufficient conditions on the parameters for ergodicity of the multiple first-order threshold processes and showed when the process is transient on a subset of the remainder. Also, they conjectured that the process is null recurrent everywhere else. Subsequently, Guo and Petrucelli (1991) proved the conjecture and under the assumption of finite variance of the error distributions they resolved the remaining questions of transience and null recurrence for this process. Following these, Chen and Tsay (1991) established a necessary and sufficient condition for geometrical ergodicity for the general first-order threshold autoregressive processes with general delay parameter. Their results extended the result of Petrucelli and Woolford (1984). More general results emerged from the recent work of Bhattacharya and Lee (1995) who derived a sufficient condition for geometric ergodicity and a necessary condition for recurrence of nonlinear first-order autoregressive processes.

As a natural extension, Chan and Tong (1986) introduced smoothness into the threshold autoregressive models, leading to a new class encompassing threshold autoregressive models. Consider the first-order STAR model with delay parameter $d = 1$ defined by

$$X_t = aX_{t-1} + bX_{t-1}G\left(\frac{X_{t-1} - r}{z}\right) + \varepsilon_t,$$

where r is a threshold parameter, $r \in \Re$; z is a smoothing parameter, $z \in \Re^+$ and $G(.)$ is a known distribution function. We rewrite the above model as

$$X_t = aX_{t-1} + bX_{t-1}G(X_{t-1}) + \varepsilon_t, \tag{2.1.1}$$

where, hereafter we represent $G\left(\frac{x-r}{z}\right)$ by $G(x)$ for given values of r and z. We assume the following conditions on the error distribution.

(C1) $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables with zero mean and ε_t is independent of X_{t-s}, $s \geq 1$.

(C2) ε_t has an absolutely continuous marginal distribution and positive probability density function over the real line and $E \mid \varepsilon_t \mid < \infty$.

(C3) $E(\varepsilon_t^2) < \infty$.

We note that $\{X_t : t \geq 0\}$ as defined in (2.1.1) is a Markov chain with state space $(\Re, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ- algebra on the real numbers $\Re$. The transition density is given by

$$p(x, y) = f(y - ax - bxG(x)) \tag{2.1.2}$$

where f is a strictly positive density function of ε_t. Using definition in Orey (1971), we note also that $\{X_t : t \geq 0\}$ is ϕ- irreducible and aperiodic, where ϕ is the Lebesgue measure on $\Re$.

The transition law, $\{P(x, .)\}$, for the Markov chain $\{X_t\}$ defined by

$$P(x, A) = \int_A p(x, y)dy, \quad x \in \Re, \quad A \in \mathcal{B}.$$

is strongly continuous if $\forall A \in \mathcal{B}, P(x, .)$ is a continuous function in x. Since we assume that G is continuous everywhere and f is absolutely continuous it follows that $P(x, .)$ is continuous $\forall A \in \mathcal{B}$.

Remark 2.1 *According to Tweedie (1975), the continuity condition on transition function is required to ensure that compact sets are of finite measure (see Lemma 4.1, Tweedie (1975) page 393). However, this condition can be weakened when considering less general models than what Tweedie considers. For example, Chan, et.al. (1985) consider multiple-threshold autoregressive order 1 models and derive Lemma 2.1 of Chan, et.al. (1985) , see Lemma A.1 of Appendix A, which replaces condition (ii) of Theorem 4.2 of Tweedie (1975)(Theorem A.1 of Appendix A) so the transition law $\{P(x, .)\}$ is not necessarily strongly continuous. Moreover, they states that their Lemma 2.1 is true for more general Markov chains than multiple-threshold autoregressive order 1 models. In particular, it applies to STAR models. Using this Lemma*

2.1, (2.1.1) is still ergodic when $\{P(x,.)\}$ *corresponding to the transition density (2.1.2) is not strongly continuous.*

Some notable contributions toward the ergodicity problem of the STAR model are the work of Chan and Tong (1986) who stated the sufficient condition for ergodicity of (2.1.1) and conjectured that it is also an almost necessary condition for ergodicity and the work of Nur (1993) who gave a detailed proof of the sufficient condition for ergodicity of (2.1.1) based on Proposition 2.1 of Chan and Tong (1986) (Proposition A.1, Appendix A). The new results in this chapter will include the necessary and sufficient condition for recurrence of (2.1.1), the necessary condition for ergodicity of (2.1.1) for any distribution function G and the sufficient condition for ergodicity when G is a thick-tailed.

This chapter is organised as follows. In Section 2.2, we prove the necessary and sufficient conditions for recurrence for any G. In Section 2.3, we classify the necessary condition for ergodicity of process (2.1.1) for any distribution function G. Furthermore, we present a sufficient condition for ergodicity when G is light-tailed, thick-tailed or a combination of both. As a consequence of the recurrence and ergodic results, we obtain some conditions for transience and null recurrence. To prove the sufficient conditions for ergodicity and recurrence, we use the theorems due to Tweedie (1975) as stated in Appendix A.

2.2 Recurrence

In this section, we present the necessary and sufficient conditions for recurrence for any distribution function G. The proof of the result is obtained by using Theorem 4.3 of Tweedie (1975) (Theorem A.2 of Appendix A). Mainly there are two conditions to be verified. For the condition (i) we have to show that the functions $L(x)$ or $J(x)$, as mentioned below, are greater or equal zero when x is outside a defined compact set. Condition (ii) will follow easily.

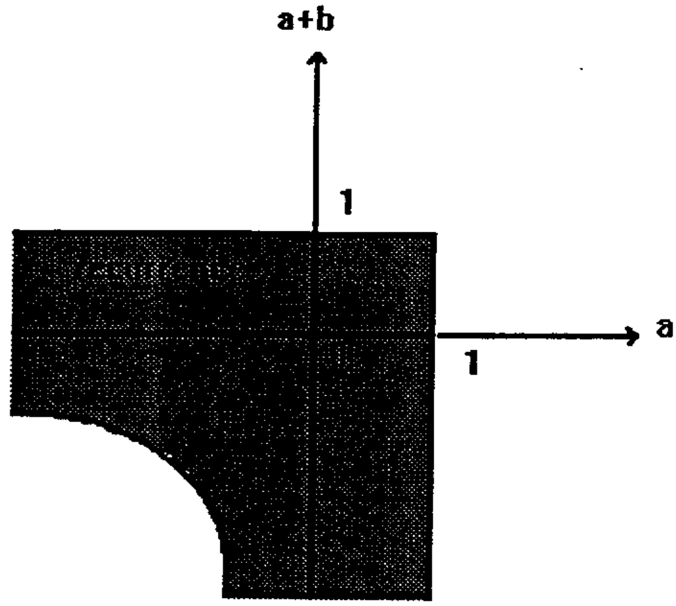

Figure 2.1. Recurrence region

Theorem 2.1 (Necessary and sufficient conditions for recurrence) *The process $\{X_t : t \geq 0\}$ as defined in (2.1.1) is recurrent if and only if*

$$a \leq 1, \quad a+b \leq 1, \text{ and } a(a+b) \leq 1. \tag{2.2.1}$$

Proof.

The sufficient condition

We divide the proof into three cases as follows:

Case (i). $-1 < a \leq 1,\ -1 < a+b \leq 1,$

Case (ii). $a \leq -1,\ -1 < a+b \leq 1,\ a(a+b) \leq 1,$

Case (iii). $-1 < a \leq 1,\ a+b \leq -1,\ a(a+b) \leq 1.$

In each of the cases we use Theorem 4.3 of Tweedie (1975) (Theorem A.2 of

Appendix A).

Proof of case (i).

Let $g(x) = \mid x \mid$, for $x \in \Re$, i.e. $g(x) = x^+ + x^-$, where $x^+ = \max(x, 0)$, $x^- = -\min(x, 0)$.

Let $W = X_{t-1}(a + bG(X_{t-1}))$ so that $X_t = W + \varepsilon_t$. Then

$$E[g(X_t) \mid X_{t-1}] = \int_{\Re} p(x, y) g(y) dy$$

can be written as

$$\begin{aligned} E[g(X_t) \mid X_{t-1}] &= E[g(W + \varepsilon_t) \mid X_{t-1}] \\ &= E[(W + \varepsilon_t)^+ + (W + \varepsilon_t)^- \mid X_{t-1}] \\ &= E[(W + \varepsilon_t) + 2(W + \varepsilon_t)^- \mid X_{t-1}] \\ &= W + 2E[(W + \varepsilon_t)^- \mid X_{t-1}]. \end{aligned} \tag{2.2.2}$$

Equivalently,

$$E[g(X_t) \mid X_{t-1}] = -W + 2E[(W + \varepsilon_t)^+ \mid X_{t-1}]. \tag{2.2.3}$$

For convenience, let $\varepsilon_t \equiv \varepsilon$. For $x > 0$, (2.2.2) becomes,

$$\begin{aligned} E[g(X_t) \mid X_{t-1}] &= x(a + bG(x)) + 2E[x(a + bG(x)) + \varepsilon]^- \\ &= x - x + x(a + bG(x)) + 2E[x(a + bG(x)) + \varepsilon]^- \\ &= g(x) - L(x), \end{aligned}$$

where

$$\begin{aligned} L(x) &= x - x(a + bG(x)) - 2E[x(a + bG(x)) + \varepsilon]^- \\ &= x(1 - (a + b)) + bx(1 - G(x)) \\ &- 2E[x(a + b) - bx(1 - G(x)) + \varepsilon]^-. \end{aligned}$$

According to Theorem 4.3 of Tweedie (1975) (Theorem A.2 of Appendix A) to prove the theorem we have to show that $L(x) \geq 0$ as $x \to \infty$. For notational convenience, let $\delta(x) = 1 - G(x)$. Therefore $L(x)$ becomes

$$L(x) = (1 - (a + b - b\delta(x)))x - 2E[x(a + b - b\delta(x)) + \varepsilon]^-.$$

If $0 < a+b \leq 1$ and $-1 < a \leq 1$ and as $E \mid \varepsilon \mid < \infty$ then

$$E[x(a+b-b\delta(x))+\varepsilon]^- \sim 0, \quad \text{as} \quad x \to \infty, \tag{2.2.4}$$

and if $-1 < a+b < 0$ and $-1 < a \leq 1$ and as $E \mid \varepsilon \mid < \infty$ then

$$E[x(a+b-b\delta(x))+\varepsilon]^- \sim -x(a+b-b\delta(x)), \quad \text{as} \quad x \to \infty$$

Therefore, as $x \to \infty$,

$$L(x) \sim \begin{cases} x(1-(a+b-b\delta(x))), & 0 < a+b \leq 1 \\ x(1+(a+b-b\delta(x))), & -1 < a+b < 0. \end{cases}$$

Also, it is clear that when $a+b=0$ and $-1 < a \leq 1$,

$$\begin{aligned} L(x) &\sim (1+b\delta(x))x - 2E[-bx\delta(x)) + \varepsilon]^- \\ &\geq 0, \quad \text{as } x \to \infty. \end{aligned}$$

Hence for $-1 < a \leq 1$, $-1 < a+b \leq 1$,

$$L(x) \geq 0, \quad \text{as } x \to \infty.$$

When $x < 0$, (2.2.3) becomes

$$\begin{aligned} E[g(X_t) \mid X_{t-1}] &= -x(a+bG(x)) + 2E[x(a+bG(x))+\varepsilon]^+ \\ &= -x + x - x(a+bG(x)) + 2E[x(a+bG(x))+\varepsilon]^+ \\ &= g(x) - J(x), \end{aligned}$$

where

$$\begin{aligned} J(x) &= -x + x(a+bG(x)) - 2E[x(a+bG(x))+\varepsilon]^+ \\ &= -x(1-a) + bxG(x) - 2E[ax + bxG(x) + \varepsilon]^+. \end{aligned}$$

By Theorem 4.3 of Tweedie (1975) (Theorem A.2 of Appendix A) we need to show that $J(x) \geq 0$ as $x \to -\infty$. Now, since $E \mid \varepsilon \mid < \infty$, we have that if $0 < a \leq 1$ and $-1 < a+b \leq 1$

$$E[x(a+bG(x))+\varepsilon]^+ \sim 0, \quad \text{as} \quad x \to -\infty$$

and if $-1 < a < 0$ and $-1 < a + b \leq 1$

$$E[x(a + bG(x)) + \varepsilon]^+ \sim x(a + bG(x)), \quad \text{as} \quad x \to -\infty.$$

Therefore, as $x \to -\infty$,

$$J(x) \sim \begin{cases} -x(1 - (a + bG(x))), & 0 < a \leq 1 \\ -x(1 + (a + bG(x))), & -1 < a < 0. \end{cases}$$

Similarly, it is clear that when $a = 0$ and $-1 < a + b \leq 1$,

$$\begin{aligned} J(x) &\sim -x(1 - bG(x)) - 2E[bxG(x) + \varepsilon]^+ \\ &\geq 0, \quad \text{as} \quad x \to -\infty. \end{aligned}$$

Hence for $-1 < a \leq 1$, $-1 < a + b \leq 1$,

$$J(x) \geq 0, \quad \text{as } x \to -\infty.$$

Now, for sufficiently large B, *choose* $A = [-B, B]$. Then by condition (i) and (ii) of Theorem 4.3 in Tweedie (1975) (Theorem A.2 of Appendix A), $\{X_t\}$ is recurrent. □

Proof of Case (ii).

From this condition, it is possible to choose positive constants c and d such that $c < d$, satisfying

$$-\frac{d}{c} < a \leq -1, \quad -\frac{c}{d} < a + b \leq 1.$$

Suppose we again choose the non-negative measurable function

$$g(x) = \begin{cases} cx, & x > 0 \\ -dx, & x < 0 \end{cases}$$

i.e. $g(x) = cx^+ + dx^-$.

Then as before

$$E[g(X_t) \mid X_{t-1}] = cW + (c + d)E[(W + \varepsilon_t)^- \mid X_{t-1}], \tag{2.2.5}$$

$$E[g(X_t) \mid X_{t-1}] = -dW + (c + d)E[(W + \varepsilon_t)^+ \mid X_{t-1}]. \tag{2.2.6}$$

For $x > 0$ (2.2.5) can be written as

$$\begin{aligned} E[g(X_t) \mid X_{t-1}] &= cx(a+bG(x)) + (c+d)E[x(a+bG(x))+\varepsilon]^- \\ &= cx - cx + cx(a+bG(x)) \\ &+ (c+d)E[x(a+bG(x))+\varepsilon]^- \\ &= g(x) - L(x), \end{aligned}$$

where

$$\begin{aligned} L(x) &= cx - cx(a+bG(x)) - (c+d)E[x(a+bG(x))+\varepsilon]^- \\ &= cx(1-(a+b)) + cbx(1-G(x)) \\ &- (c+d)E[x(a+b) - bx(1-G(x)) + \varepsilon]^-. \end{aligned}$$

Using Theorem 4.3 of Tweedie (1975) (Theorem A.2 of Appendix A) we have to show that $L(x) \geq 0$ as $x \to \infty$. Letting $\delta(x) = 1 - G(x)$, $L(x)$ becomes

$$L(x) = (1 - (a + b - b\delta(x)))cx - (c+d)E[x(a+b-b\delta(x)) + \varepsilon]^-.$$

Since $E \mid \varepsilon \mid < \infty$ yields

$$E[x(a+b-b\delta(x)) + \varepsilon]^- \sim 0, \quad \text{as } x \to \infty,$$

if $0 \leq a+b \leq 1$ and $-\frac{d}{c} < a \leq -1$ and if $-\frac{c}{d} < a+b < 0$ and $-\frac{d}{c} < a \leq -1$

$$E[x(a+b-b\delta(x)) + \varepsilon]^- \sim -x(a+b-b\delta(x)), \quad \text{as } x \to \infty.$$

Therefore, as $x \to \infty$,

$$L(x) \sim \begin{cases} cx(1-(a+b-b\delta(x))), & 0 < a+b \leq 1 \\ x(c+d(a+b-b\delta(x))), & -\frac{c}{d} < a+b < 0. \end{cases}$$

As before, when $a + b = 0$ and $-\frac{d}{c} < a \leq -1$,

$$\begin{aligned} L(x) &\sim (1 + b\delta(x))cx - (c+d)E[-bx\delta(x) + \varepsilon]^- \\ &\geq 0, \quad \text{as } x \to \infty. \end{aligned}$$

Hence, for $a \le -1,\ -1 < a+b \le 1,\ a(a+b) \le 1,$ we have

$$L(x) \ge 0, \quad \text{as } x \to \infty.$$

Consider $x < 0$ so that (2.2.6) becomes

$$\begin{aligned} E[g(X_t) \mid X_{t-1}] &= -dx(a+bG(x)) + (c+d)E[x(a+bG(x))+\varepsilon]^+ \\ &= -dx + dx - dx(a+bG(x)) \\ &+ (c+d)E[x(a+bG(x))+\varepsilon]^+ \\ &= g(x) - J(x), \end{aligned}$$

where

$$\begin{aligned} J(x) &= -dx + dx(a+bG(x)) - (c+d)E[x(a+bG(x))+\varepsilon]^+ \\ &= -dx(1-a) + dbxG(x) \\ &- (c+d)E[ax + bxG(x) + \varepsilon]^+. \end{aligned}$$

Again, we need to show that $J(x) \ge 0$ as $x \to -\infty$. Since $E \mid \varepsilon \mid < \infty$, for $-\frac{d}{c} < a \le -1$ and $-\frac{c}{d} < a+b \le 1$, we have

$$E[x(a+bG(x))+\varepsilon]^+ \sim x(a+bG(x)), \quad \text{as } x \to -\infty.$$

Therefore, as $x \to -\infty$,

$$J(x) \sim -x(d + c(a+bG(x))), \quad -\frac{d}{c} < a \le -1,$$

which implies that, for $a \le -1,\ -1 < a+b \le 1,\ a(a+b) \le 1,$

$$J(x) \ge 0, \quad \text{as } x \to -\infty.$$

As before take, for sufficiently large B, $A = [-B, B]$, then conditions (i) and (ii), Theorem 4.3 of Tweedie (1975) (Theorem A.2 of Appendix A) are satisfied and $\{X_t\}$ is recurrent. □

Proof of Case (iii).

This case becomes the symmetric case of case (ii) by letting $y_t = -x_t$ in (2.1.1). In this case (2.1.1) becomes

$$y_t = y_{t-1}(A_0 + B_0 V(y_{t-1}) + \varepsilon_t \tag{2.2.7}$$

where $A_0 = a + b$, $B_0 = -b$ and $V(y) = 1 - G(-y)$. But (2.2.7) has the same form as (2.1.1) so that by the previous result, (2.2.7) is recurrent if $A_0 = a + b \leq -1$, $-1 < A_0 + B_0 = a \leq 1$, $A_0(A_0 + B_0) = a(a + b) \leq 1$. Combining all cases, the sufficient condition is proved. □

Necessary condition

The necessary part of the theorem follows from Lemma 2.1 and Lemma 2.2 given below. The proof follows the method of Petrucelli and Woolford (1985) given for the threshold AR(1) model. □

Lemma 2.1 *If* $a > 1$ *or* $a + b > 1$ *then the process (2.1.1) is not recurrent.*

Proof.

Consider the case $a + b > 1$. Notice that we have denoted $G\left(\frac{x-r}{z}\right)$ by $G(x)$ where G is a known distribution function.

For $B > 0$ sufficiently large, the model (2.1.1) can be rewritten as

$$X_t = \begin{cases} X_{t-1}(a+b) - bX_{t-1}(1 - G(X_{t-1})) + \varepsilon_t, & X_{t-1} > B \\ X_{t-1}(a + bG(X_{t-1})) + \varepsilon_t, & \mid X_{t-1} \mid \leq B \\ aX_{t-1} + bX_{t-1}G(X_{t-1}) + \varepsilon_t, & X_{t-1} < -B \end{cases}$$

We choose x sufficiently large such that $a + b - b(1 - G(x)) > 1$. Let η be such that $1 < \eta < a + b - b(1 - G(x))$. Then, for $X_{t-1} > B$,

$$\begin{aligned} & P\left(X_t \leq \frac{\eta+1}{2} X_{t-1} \mid X_{t-1}\right) \\ = \; & P\left((a + b - b(1 - G(X_{t-1}))X_{t-1} + \varepsilon_t \leq \frac{\eta+1}{2} X_{t-1} \mid X_{t-1}\right) \\ = \; & P\left(-\varepsilon_t \geq \left(a + b - b(1 - G(X_{t-1})) - \frac{\eta+1}{2}\right) X_{t-1} \mid X_{t-1}\right) \\ \leq \; & P\left(\mid \varepsilon_t \mid \geq \left(a + b - b(1 - G(X_{t-1})) - \frac{\eta+1}{2}\right) X_{t-1} \mid X_{t-1}\right) \\ \leq \; & \frac{E \mid \varepsilon_t \mid}{(a + b - b(1 - G(X_{t-1})) - \frac{\eta+1}{2})X_{t-1}} \quad \text{by Markov's inequality} \\ \leq \; & \frac{2E \mid \varepsilon_t \mid}{(\eta - 1)X_{t-1}}. \end{aligned} \tag{2.2.8}$$

Choose $B > 0$ large enough such that $c = \frac{2E|\varepsilon_t|}{(\eta-1)B} < 1$. Then for $X_{t-1} > B$ we have from (2.2.8)

$$P\left(X_t > \frac{\eta+1}{2}X_{t-1} \mid X_{t-1}\right) \geq 1 - c.$$

Noting that for $X_{t-1} > B$, $\frac{\eta+1}{2}X_{t-1} > X_{t-1}$, by repeating the above argument, we get

$$\begin{aligned}
& P\left(X_{t+1} > \frac{\eta+1}{2}X_t, X_t > \frac{\eta+1}{2}X_{t-1} \mid X_{t-1}\right) \\
= \; & P\left(X_{t+1} > \frac{\eta+1}{2}X_t \mid X_t > \frac{\eta+1}{2}X_{t-1}, X_{t-1}\right) \\
\times \; & P\left(X_t > \frac{\eta+1}{2}X_{t-1} \mid X_{t-1}\right) \\
= \; & P\left(X_{t+1} > \frac{\eta+1}{2}X_t \mid X_t > \frac{\eta+1}{2}X_{t-1}\right) P\left(X_t > \frac{\eta+1}{2}X_{t-1} \mid X_{t-1}\right) \\
\geq \; & \left(1 - \frac{c}{\frac{\eta+1}{2}}\right)(1-c) = (1-c\beta)(1-c)
\end{aligned}$$

where $\beta = \frac{2}{\eta+1} < 1$ for which the first term on the right hand side follows from

$$\begin{aligned}
& P\left(X_{t+1} \leq \frac{\eta+1}{2}X_t \mid X_t > \frac{\eta+1}{2}X_{t-1}\right) \\
= \; & P\left((a+b-b(1-G(X_{t-1})))X_t + \varepsilon_{t+1} \leq \frac{\eta+1}{2}X_t \mid X_t > \frac{\eta+1}{2}X_{t-1}\right) \\
= \; & P\left(-\varepsilon_{t+1} \geq \left(a+b-b(1-G(X_{t-1})) - \frac{\eta+1}{2}\right)X_t \mid X_t > \frac{\eta+1}{2}X_{t-1}\right) \\
\leq \; & P\left(\mid \varepsilon_{t+1} \mid \geq \left(a+b-b(1-G(X_{t-1})) - \frac{\eta+1}{2}\right)X_t \mid X_t > \frac{\eta+1}{2}X_{t-1}\right) \\
\leq \; & P\left(\mid \varepsilon_{t+1} \mid \geq \left(a+b-b(1-G(X_{t-1})) - \frac{\eta+1}{2}\right)\frac{\eta+1}{2}X_{t-1} \mid X_t > \frac{\eta+1}{2}X_{t-1}\right) \\
\leq \; & \frac{E \mid \varepsilon_{t+1} \mid}{(a+b-b(1-G(X_{t-1})) - \frac{\eta+1}{2})\frac{\eta+1}{2}X_{t-1}} \quad \text{(by Markov's inequality)} \\
\leq \; & c\beta.
\end{aligned}$$

Continuing the result, for $X_1 > B$, we have

$$P\left(X_{k+1} > \frac{\eta+1}{2}X_k, k = 1, 2, \ldots, t \mid X_1 > B\right) \geq \prod_{k=1}^{t}\left(1 - c\beta^{k-1}\right) \geq (1-c)^{\frac{1}{1-\beta}}$$

as $t \to \infty$, since

$$\begin{aligned}
\sum_{t=1}^{\infty} \log(1 - c\beta^{t-1}) &= \sum_{t=1}^{\infty}\sum_{k=1}^{\infty}\left(-\frac{c^k\beta^{k(t-1)}}{k}\right) \\
&= \sum_{k=1}^{\infty}\left(-\frac{c^k}{k(1-\beta^k)}\right) \\
&\geq \sum_{k=1}^{\infty}\left(-\frac{c^k}{k(1-\beta)}\right) = \frac{\log(1-c)}{1-\beta}.
\end{aligned}$$

Consequently, for any $X_0 \in \Re$

$$P(X_t \to \infty \mid X_0) \geq (1-c)^{\frac{1}{1-\beta}} P(X_1 > B \mid X_0) > 0$$

since $P(X_1 > B \mid X_0) > 0$ from the assumptions on ε_t. Hence $\{X_t\}$ is not recurrent.

The case $a > 1$ is the symmetric case of $a + b > 1$ when $y_t = -x_t$. The result is obtained by working with the transformed equation

$$y_t = y_{t-1}(A_0 + B_0 V(y_{t-1})) + \varepsilon_t, \tag{2.2.9}$$

where $A_0 = a + b$, $B_0 = -b$ and $V(y) = 1 - G(-y)$. This equation has the same form as (2.1.1) so that by the previous result, (2.2.9) is not recurrent if $A_0 + B_0 = a > 1$. □

Lemma 2.2 *If* $a < 0$ *and* $a(a+b) > 1$ *then the process* $\{X_t\}$ *is not recurrent.*

Proof.

Without loss of generality, consider the case of $a + b < -1$ & $a(a+b) > 1$. Consider the model (2.1.1) as the combination of the last two observations

$$\begin{aligned}
X_t &= (a + bG(X_{t-1}))X_{t-1} + \varepsilon_t \\
&= (a + bG(X_{t-1}))\left((a + bG(X_{t-2}))X_{t-2} + \varepsilon_{t-1}\right) + \varepsilon_t \\
&= (a + bG(X_{t-1}))(a + bG(X_{t-2}))X_{t-2} + (a + bG(X_{t-1}))\varepsilon_{t-1} + \varepsilon_t.
\end{aligned}$$

We rewrite the model as

$$X_t = \begin{cases} (a+bG(X_{t-1}))(a+b-b(1-G(X_{t-2})))X_{t-2} & \\ \quad +(a+bG(X_{t-1}))\varepsilon_{t-1}+\varepsilon_t, & \text{if } X_{t-2} > B, \\ X_t, & \text{otherwise,} \end{cases}$$

or can be rewritten as

$$X_t = \begin{cases} (a+bG(X_{t-1}))(a+b-b(1-G(X_{t-2})))X_{t-2} & \\ \quad +(X_t - E(X_t \mid X_{t-1}, X_{t-2})), & \text{if } X_{t-2} > B, \\ X_t, & \text{otherwise.} \end{cases}$$

Suppose $X_{t-2} > B$ then we can always choose $B > 0$ sufficiently large such that $(a+bG(X_{t-1}))(a+b-b(1-G(X_{t-2}))) \geq a(a+b) > 1$. There exists $1 < \eta < (a+bG(X_{t-1}))(a+b-b(1-G(X_{t-2})))$ such that

$$\begin{aligned}
& P\left(X_t \leq \frac{\eta+1}{2} X_{t-2} \mid X_{t-1}, X_{t-2}\right) \\
= \ & P\big((a+bG(X_{t-1}))(a+b-b(1-G(X_{t-2})))X_{t-2} \\
+ \ & (X_t - E(X_t \mid X_{t-1}, X_{t-2})) \leq \frac{\eta+1}{2} X_{t-2} \mid X_{t-1}, X_{t-2}\big) \\
= \ & P\big(-(X_t - E(X_t \mid X_{t-1}, X_{t-2})) \geq \\
& \quad \left((a+bG(X_{t-1}))(a+b-b(1-G(X_{t-2}))) - \frac{\eta+1}{2}\right) X_{t-2} \mid X_{t-1}, X_{t-2}\big) \\
\leq \ & P\big(\mid X_t - E(X_t \mid X_{t-1}, X_{t-2}) \mid \geq \\
& \quad \left((a+bG(X_{t-1}))(a+b-b(1-G(X_{t-2}))) - \frac{\eta+1}{2}\right) X_{t-2} \mid X_{t-1}, X_{t-2}\big) \\
\leq \ & \frac{E\left(\mid X_t - E(X_t \mid X_{t-1}, X_{t-2}) \mid\mid X_{t-1}, X_{t-2}\right)}{\left((a+bG(X_{t-1}))(a+b-b(1-G(X_{t-2}))) - \frac{\eta+1}{2}\right) X_{t-2}} \\
\leq \ & \frac{2E\left(\mid X_t - E(X_t \mid X_{t-1}, X_{t-2}) \mid \mid X_{t-1}, X_{t-2}\right)}{(\eta-1)X_{t-2}}.
\end{aligned}$$

Whenever $X_{t-2} > B$, $E\left(\mid X_t - E(X_t \mid X_{t-1}, X_{t-2}) \mid\mid X_{t-1}, X_{t-2}\right) \leq \psi < \infty$ for $t \geq 2$. Choose $B > 0$ large enough so that

$$c = \frac{2\psi}{(\eta-1)B} < 1.$$

Then for $X_{t-2} > B$,

$$P\left(X_t > \frac{\eta+1}{2}X_{t-2} \mid X_{t-1}, X_{t-2}\right) \geq 1 - c. \qquad (2.2.10)$$

By a similar argument as in Lemma 2.1, (2.2.10) implies that

$$P\left(X_{t+2} > \frac{\eta+1}{2}X_t, X_t > \frac{\eta+1}{2}X_{t-2} \mid X_{t-1}, X_{t-2}\right) \geq (1 - c\beta)(1 - c)$$

where $\beta = \frac{2}{\eta+1} < 1$.

Continuing in this manner,

$$P\left(X_{2(k+1)} > \frac{\eta+1}{2}X_{2k}, k = 1, 2, ..., t \mid X_2 > B\right)$$

$$\geq \prod_{i=1}^{t}\left(1 - c\beta^{i-1}\right) \geq (1 - c)^{\frac{1}{1-\beta}}.$$

Consequently, for any $X_0 \in \Re$,

$$P(X_{2t} \to \infty \mid X_0) \geq (1 - c)^{\frac{1}{1-\beta}} P(X_2 > B \mid X_0) > 0.$$

Now as in Lemma 2.1, $\{X_{2t}\}$ is not recurrent. Hence $\{X_t\}$ is not recurrent.

2.3 Ergodicity

Chan and Tong (1986) obtained that the process $\{X_t : t \geq 0\}$ as defined in (2.1.1) is ergodic only if (2.3.4), given below, is satisfied (Proposition 2.1 of Chan and Tong (1986), see Proposition A.1 of Appendix A). They pointed out that any sufficiently smooth function G with a rapidly decaying tail will suffice to satisfy the condition. That is, basically they presented a sufficient condition for ergodicity when G is light-tailed distribution function. Nur (1993) presented the proof of the proposition by using general properties of distribution function G. The contribution for ergodicity theory in this thesis will be the necessary conditions for ergodicity of (2.1.1) for any G and the sufficient condition for ergodicity based on the classification of G as a distribution function.

In this section, a necessary condition for ergodicity is given in Theorem 2.2 for any distribution function G. In Theorem 2.3, a sufficient condition for ergodicity is presented when G is a light-tailed distribution function. Theorem 2.4 gives a sufficient condition for ergodicity when G is thick-tailed. Corollary 2.1 and 2.2 present the combination of both. For a special case of a light-tailed distribution function, the necessary condition is also sufficient.

Theorem 2.2 *For any continuous distribution function* G, *if the process* $\{X_t : t \geq 0\}$ *as defined in (2.1.1) is ergodic then*

$$a \leq 1, \quad a+b \leq 1, \text{ and } a(a+b) \leq 1, \quad b \neq 0. \tag{2.3.1}$$

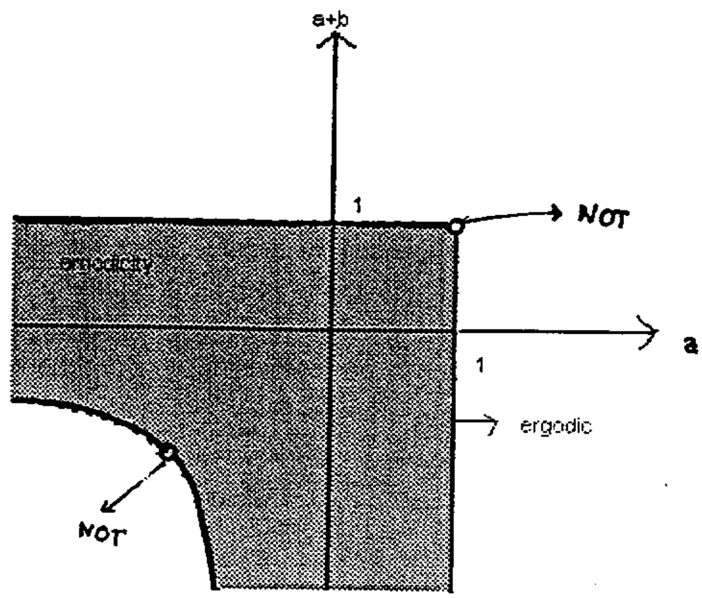

Figure 2.2. Necessary condition for ergodicity for any G

Proof.

By Theorem 2.1, it is enough to show that the process is not ergodic when $a = a + b = 1$, and $a = a + b = -1$. In these cases $b = 0$. Hence in the two cases the model becomes

$$X_t = X_{t-1} + \varepsilon_t$$

and

$$X_t = -X_{t-1} + \varepsilon_t$$

respectively. Clearly these are not ergodic. Hence the proof is complete. □

As mentioned earlier in the proof of sufficient condition for recurrence, to satisfy condition (i) of Theorem 4.3 of Tweedie (1975), the functions $L(x)$ or $J(x)$ were shown to be greater or equal to zero. For the proof of sufficient condition for ergodicity, by using Theorem 4.2 of Tweedie (1975)(Theorem A.1 of Appendix A) and Remark A.1, we have to show that these functions are greater than zero for x outside the defined compact set (condition (i)) and the expectation of the measurable function is finite in the compact set (condition (ii)).

Theorem 2.3 (Light-tailed distribution function.) *Suppose G is a continuous distribution function such that*

$$x(1 - G(x)) \to 0, \quad \text{as } x \to \infty, \tag{2.3.2}$$

$$xG(x) \to 0, \quad \text{as } x \to -\infty. \tag{2.3.3}$$

The process $\{X_t : t \geq 0\}$ as defined in (2.1.1) is ergodic only if

$$a < 1, \quad a + b < 1, \text{ and } a(a + b) < 1. \tag{2.3.4}$$

Remark 2.2 *Distribution functions satisfying (2.3.2) and (2.3.3) includes the class of light-tailed to medium-tailed continuous distribution functions such as Normal, Laplace, Gamma, Lognormal, Beta, Exponential, F, t, and many others.*

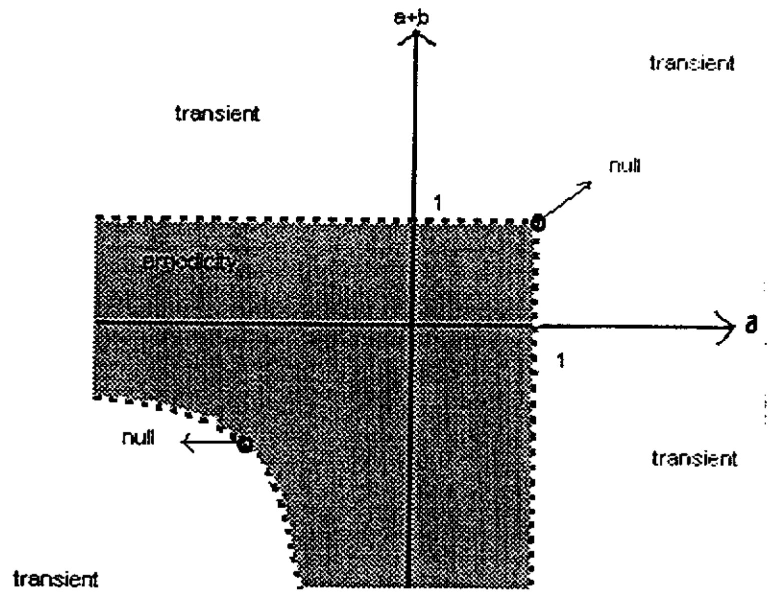

Figure 2.3. Classification for G light-tailed

Proof of Theorem 2.3.

As before, we divide the proof into three cases as follows:

Case (i). $-1 < a < 1,\ -1 < a + b < 1,$

Case (ii). $a \leq -1,\ -1 < a + b < 1,\ a(a + b) < 1,$

Case (iii). $-1 < a < 1,\ a + b \leq -1,\ a(a + b) < 1.$

Proof of Case (i).

As in proof of Theorem 2.1 case (i), let $g(x) = \mid x \mid$ for $x \in \Re$. Using the equation (2.2.2) and (2.2.3) we have, for $x > 0$

$$E[g(X_t) \mid X_{t-1}] = g(x) - L(x), \tag{2.3.5}$$

where

$$L(x) = x(1-(a+b)) + bx(1-G(x)) - 2E[x(a+b) - bx(1-G(x)) + \varepsilon]^-,$$

and $x^- = -\min(x, 0)$.

To prove condition (i), Theorem 4.2 of Tweedie (1975) we will now show that $L(x) > 0$ as $x \to \infty$. By (2.3.2) for large x,

$$L(x) \sim (1-(a+b))x - 2E[x(a+b) + \varepsilon]^-.$$

Also, since $E \mid \varepsilon \mid < \infty$, for large x,

$$\begin{aligned} E[x(a+b)+\varepsilon]^- &\sim 0, \text{ if } 0 < a+b < 1, \\ &= E(\varepsilon^-) \text{ if } a+b = 0, \\ &\sim -x(a+b), \text{ if } -1 < a+b < 0. \end{aligned} \tag{2.3.6}$$

Therefore, for large x, and $-1 < a+b < 1$,

$$L(x) \sim \begin{cases} x(1-(a+b)), & 0 \le a+b < 1 \\ x(1+(a+b)), & -1 < a+b < 0. \end{cases}$$

Hence for a given $a+b$, as $x \to \infty$, $L(x) > 0$.

Consider $x < 0$ so that (2.2.3) becomes

$$E[g(X_t) \mid X_{t-1}] = g(x) - J(x), \tag{2.3.7}$$

where

$$J(x) = -x(1-a) + bxG(x) - 2E[ax + bxG(x) + \varepsilon]^+.$$

We will show that $J(x) > 0$ as $x \to -\infty$. By (2.3.4), as $x \to -\infty$,

$$J(x) \sim -x(1-a) - 2E[ax + \varepsilon]^+.$$

Also, since $E \mid \varepsilon \mid < \infty$, we have,

$$\begin{aligned} E[ax+\varepsilon]^+ &\sim 0, \text{ if } 0 < a < 1, \\ &= E(\varepsilon^+) \text{ if } a = 0, \\ &\sim ax, \text{ if } -1 < a < 0. \end{aligned}$$

Therefore, as $x \to -\infty$,

$$J(x) \sim \begin{cases} -x(1-a), & 0 \le a < 1 \\ -x(1+a), & -1 < a < 0. \end{cases}$$

Hence $J(x) > 0$, as $x \to \infty$.

Now, for sufficiently large B, let

$$A = [-B, B].$$

Then with this A, the condition (i) of Theorem 4.2 of Tweedie (1975) is satisfied. Furthermore, condition (ii) of Theorem 4.2 of Tweedie (1975)(Theorem A.1 of Appendix A) follows, by using (2.3.5) and (2.3.7). Hence $\{X_t\}$ is ergodic. □

Proof of Case (ii).

Under this condition, it is possible to choose positive constants c and d such that $c < d$ satisfying

$$-\frac{d}{c} < a \le -1, \quad -\frac{c}{d} < a + b < 1.$$

Consider the function $g(x)$ as defined in Theorem 2.1 case (ii). Using the equations (2.2.5) and (2.2.6) we have, for $x > 0$

$$E[g(X_t) \mid X_{t-1}] = g(x) - L(x), \tag{2.3.8}$$

where

$$L(x) = cx(1-(a+b)) + cbx(1-G(x)) - (c+d)E[x(a+b) - bx(1-G(x)) + \varepsilon]^-.$$

We will show that $L(x) > 0$ as for large x. As $x \to \infty$, the asymptotic value of $G(x)$ is given in (2.3.2). As in case (i) it can be shown that for large x,

$$L(x) \sim \begin{cases} cx(1-(a+b)), & 0 \le a+b < 1 \\ x(c+d(a+b)), & -\frac{c}{d} < a+b < 0. \end{cases}$$

Hence it follows that there exists B such that

$$L(x) > 0, \quad \text{for } x > B.$$

For $x < 0$ (2.2.6) becomes

$$E[g(X_t) \mid X_{t-1}] = g(x) - J(x), \tag{2.3.9}$$

where

$$J(x) = -dx(1-a) + dbxG(x) - (c+d)E[ax + bxG(x) + \varepsilon]^+.$$

We will show that $J(x) > 0$ as $x \to -\infty$. Again as in case (i) we have

$$J(x) \sim \; -x(d+ac), \quad \text{if} \quad -\frac{d}{c} < a \leq -1$$

Hence we have for large $B > 0$

$$J(x) > 0, \quad \text{for } x < -B.$$

By taking $A = [-B, B]$ it is clear that the condition (i) of Theorem 4.2 of Tweedie (1975) is satisfied. Furthermore condition (ii) of Theorem 4.2 of Tweedie (1975)(Theorem A.1 of Appendix A) follows from (2.3.8) and (2.3.9). Hence $\{X_t\}$ is ergodic. □

Proof of Case (iii).

This case is seen to be the symmetric case of case (ii) if we let the transformation $y_t = -x_t$ in (2.1.1). The transformed process is given by

$$y_t = y_{t-1}(A_0 + B_0 V(y_{t-1})) + \varepsilon_t, \tag{2.3.10}$$

where $A_0 = a + b$, $B_0 = -b$ and $V(y) = 1 - G(-y)$. This has the same form as (2.1.1) so that by the previous result, (2.3.10) is ergodic if $A_0 = a + b \leq -1$, $-1 < A_0 + B_0 = a < 1$, $A_0(A_0 + B_0) = a(a+b) < 1$. □

Combining cases (i) to (iii), Theorem 2.3 is proved. □

Remark 2.3 *For a special case of a light-tailed distribution function G for which*

$$x(1 - G(x)) = 0, \qquad x \; > k_1, \;\; 0 < k_1 < \infty,$$
$$xG(x) = 0, \;\; x \; < -k_1,$$

then the sufficient condition (2.3.4) is also necessary. This follows easily from a similar argument in Theorem 2.2.

Theorem 2.4 (Thick-tailed distribution function.) *Suppose G is a continuous distribution function so that*

$$x(1 - G(x)) \to k_1, \quad k_1 > 0 \quad \text{as} \quad x \to \infty, \tag{2.3.11}$$

$$xG(x) \to k_2, \quad k_2 < 0 \quad \text{as} \quad x \to -\infty. \tag{2.3.12}$$

where k_1 & k_2 are either finite or infinite. The process $\{X_t : t \geq 0\}$ as defined in (2.1.1) is ergodic only if

$$a \leq 1, \quad a + b \leq 1, \quad \text{and} \quad a(a + b) < 1. \tag{2.3.13}$$

Remark 2.4 *Distribution functions satisfying (2.3.11) and (2.3.12) include the class of thick-tailed continuous distribution functions such as Cauchy and Pareto distribution.*

Proof of Theorem 2.4.

By Theorem 2.3 it suffices to prove the theorem for the following cases:

Case (i.a). $-1 < a < 1, \ a + b = 1,$

Case (i.b). $a \leq -1, \ a + b = 1,$

Case (ii.a). $a = 1, \ -1 < a + b < 1.$

Case (ii.b). $a = 1, \ a + b \leq -1.$

Proof of Case (i.a).

As in proof of Theorem 2.1 case (i), let $g(x) = | x |$, for $x \in \Re$. Using equations (2.2.2) and (2.2.3) we have for $x > 0$,

$$E[g(X_t) \mid X_{t-1}] = g(x) - L(x), \tag{2.3.14}$$

where

$$L(x) = bx(1 - G(x)) - 2E[x - bx(1 - G(x)) + \varepsilon]^-.$$

Note that $b > 0$ in this case. Therefore for large x, $\lim_{x \to \infty} E[x - bx(1 - G(x)) + \varepsilon]^- = 0$, so that we have

$$L(x) \sim bx(1 - G(x)), \quad b > 0 \ \text{as} \ x \to \infty.$$

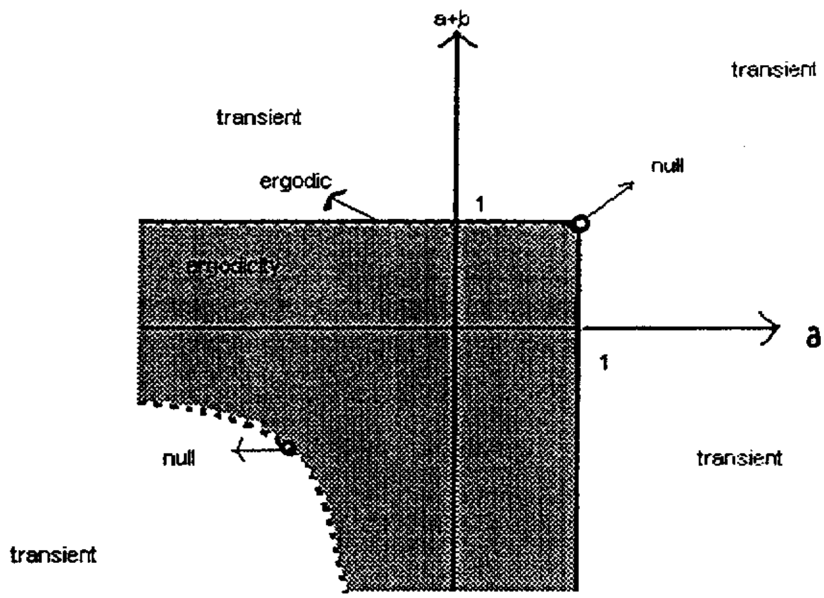

Figure 2.4. Classification for G thick-tailed

Then

$$L(x) > 0, \quad \text{as } x \to \infty.$$

Consider $x < 0$ so that (2.2.3) becomes

$$E[g(X_t) \mid X_{t-1}] = g(x) - J(x), \tag{2.3.15}$$

where

$$J(x) = -x(1-a) + bxG(x) - 2E[ax + bxG(x) + \varepsilon]^+.$$

For $x < 0$, $\mid x \mid$ sufficiently large, $E \mid \varepsilon \mid < \infty$, we have

$$\begin{aligned} E[ax + bk + \varepsilon]^+ &\sim 0, \quad \text{if } 0 \le a < 1, \\ &\sim ax + bxG(x), \quad \text{if } -1 < a < 0. \end{aligned}$$

Therefore, as $x \to -\infty$, by using (2.3.12)

$$J(x) \sim \begin{cases} -x(1-a) + bxG(x), & 0 \leq a < 1 \\ -x(1+a) - bxG(x), & -1 < a < 0. \end{cases}$$

As $x \to -\infty$, the first term on the right hand side goes to infinity faster than the second term. Hence

$$J(x) > 0, \quad \text{as } x \to -\infty.$$

Now by defining, for sufficiently large B, $A = [-B, B]$ the condition (i) of Theorem 4.2 of Tweedie (1975) (Theorem A.1 of Appendix A) is satisfied. Furthermore, condition (ii) of the theorem follows from (2.3.14) and (2.3.15). Hence $\{X_t\}$ is ergodic. □

Proof of Case (i.b).

Under this condition, it is possible to choose positive constants c and d such that $c < d$ satisfying

$$-\frac{d}{c} < a \leq -1, \quad a + b = 1.$$

Consider the function $g(x)$ as defined in Theorem 2.1 case (ii). Using the equation (2.2.5) and (2.2.6), we have, for $x > 0$,

$$E[g(X_t) \mid X_{t-1}] = g(x) - L(x), \tag{2.3.16}$$

where

$$L(x) = cbx(1 - G(x)) - (c + d)E[x - bx(1 - G(x)) + \varepsilon]^{-}.$$

Note also that $b > 0$ in this case. Therefore for large x,

$$\lim_{x \to \infty} E[x - bx(1 - G(x)) + \varepsilon]^{-} = 0.$$

Therefore, as $x \to \infty$,

$$L(x) \sim cbx(1 - G(x)), \quad \text{as } x \to \infty$$

Hence it follows that $L(x) > 0$, as $x \to \infty$.

For $x < 0$ (2.2.6) becomes

$$E[g(X_t) \mid X_{t-1}] = g(x) - J(x), \tag{2.3.17}$$

where

$$J(x) = -dx(1-a) + dbxG(x) - (c+d)E[ax + bxG(x) + \varepsilon]^+.$$

Again we have

$$E[ax + bxG(x) + \varepsilon]^+ \sim ax, \quad \text{if} \quad -\frac{d}{c} < a \leq -1.$$

Therefore, as $x \to -\infty$,

$$J(x) \sim \; -x(d+ac) - cbxG(x), \quad -\frac{d}{c} < a \leq -1.$$

As $b > 0$, $c > 0$ and $k < 0$, we have

$$J(x) > 0, \quad \text{as } x \to -\infty.$$

By taking

$$A = [-B, B]$$

it is clear that the condition (i) of Theorem 4.2 of Tweedie (1975) is satisfied. Furthermore condition (ii) of Theorem 4.2 of Tweedie (1975) follows from (2.3.15) and (2.3.17). Hence $\{X_t\}$ is ergodic. □

Proof of Cases (ii.a) and (ii.b)

These are the symmetric cases of (i.a) and (i.b) respectively, when $y_t = -x_t$ and hence the proof follows as above.

Combining all the cases, Theorem 2.4 is proved. □

From the proof of two theorems the following two corollaries follow.

Corollary 2.1 (Right light and left thick-tailed distribution function.) *Suppose G is a distribution function such that (2.3.3) and (2.3.13) are satisfied. If*

$$a \leq 1, \quad a + b < 1, \text{ and } a(a+b) < 1, \tag{2.3.18}$$

then the process $\{X_t : t \geq 0\}$ as defined in (2.1.1) is ergodic.

Corollary 2.2 (Right thick and left light-tailed distribution function.) *Suppose* G *is a distribution function such that (2.3.4) and (2.3.12) are satisfied. If*

$$a < 1, \quad a + b \leq 1, \text{ and } a(a+b) < 1, \tag{2.3.19}$$

then the process $\{X_t : t \geq 0\}$ *as defined in (2.1.1) is ergodic.*

Finally, from the results on ergodicity and recurrence we have the following simple results on null recurrence and transience of the process,

Corollary 2.3 (Transience) *The process* $\{X_t : t \geq 0\}$ *as defined in (2.1.1) is transient if and only if*

$$a > 1 \text{ or } a + b > 1, \tag{2.3.20}$$

$$a < 0 \text{ and } a(a+b) > 1. \tag{2.3.21}$$

Proof. Follows from Theorem 2.1. □

Corollary 2.4 *For any* G *a continuous distribution function, the process* $\{X_t : t \geq 0\}$ *as defined in (2.1.1) is null recurrent only if*

$$a = 1 \text{ and } a + b = 1, \tag{2.3.22}$$

$$a = -1 \text{ and } a + b = -1. \tag{2.3.23}$$

Proof. Follows from Lemma 2.5 and 2.6. □

For the estimation of parameters of the STAR model, in the subsequent chapters, we assume ergodicity. From the discussion in this chapter it follows that the STAR model is ergodic if $a < 1$, $a + b < 1$, and $a(a+b) < 1$ and if ε_t has a strictly positive density function with respect to the Lebesgue measure.

Chapter 3

Bayesian Analysis

3.1 Introduction

In this chapter, we provide a Bayesian analysis of STAR models of order p including the derivation of the posterior distributions of coefficient parameters with a noninformative prior and the marginal distribution of intrinsic parameters. The performance of the exact one-step-ahead and conditional multi-step-ahead predictive densities are also investigated, with illustrative examples using real and simulated data sets.

Bayesian analysis has enjoyed success in many branches of statistics. Its potential for time series analysis, however, is under developed, because time series analysis often involves highly nonlinear dynamic structures, which causes difficulties in prior specification and in posterior evaluation. The situation is changing, however. There has been a growing interest in Bayesian time series analysis in recent years. The advances in statistical computation makes possible a fully Bayesian analysis in time series.

As early as 1971, Zellner began a systematic study of Bayesian methods in time series analysis. He derived posterior and predictive distributions for first and second order autoregressive processes using a diffuse prior density and gave a complete analysis for regression models with autocorrelated errors and for distributed lag models. Since 1980, Bayesian techniques have also begun to show that they did offer an attractive alternative to the popular Box and Jenkins methodology. This was initiated by Monahan (1983), who used numerical integration to implement a complete time series analysis, including identification, diagnostic checking, estimation and prediction. This was the first Bayesian attempt to perform a comprehensive analysis and was a valuable contribution, in this direction.

Among the nonlinear time series models proposed in the literature, the threshold autoregressive model is perhaps the most popular one. Gowever the threshold autoregressive model has not been widely used in practice due to the difficulty in identifying the threshold variable and in estimating the associated threshold value.

The nonasymptotic Bayesian approach for threshold autoregressive models with two regimes, suggested by Broemeling and Cook (1992), can be considered as an initial attempt to circumvent these difficulties. They developed formulae to obtain the marginal posterior distribution for each of the parameters and the one-step-ahead predictive distributions along with their means and variances. Extending these ideas, Geweke and Terui (1993) provided a Bayesian approach for parameter estimation and related inferences in the threshold autoregressive model. In this thesis, the exact posterior distribution of delay and the threshold parameters are derived as is the multi-step-ahead predictive density. Both Broemeling and Cook (1992) and Geweke and Terui (1993) investigated the two-regime threshold autoregressive model analytically. In contrast, Chen and Lee (1995) proposed a Bayesian analysis for the two-regime model where the desired marginal posterior densities of the threshold value and other parameters were obtained numerically via the Gibbs sampler. Similarly, McCulloch and Tsay (1994) applied the Gibbs sampler in analysing autoregressive processes including random level-shift models.

The smooth threshold autoregressive model is a generalisation of the threshold autoregressive model having an additional intrinsic parameter, a smoothing parameter. The smoothing parameter allows the model to transit between regimes in a continuous manner. In a recent paper, Chan and Tong (1986) presented a method of estimation of parameters of the STAR model using the conditional least squares procedure. They established consistency and normality of the parameters as in Theorem 3.1 of Chan and Tong (1986). To provide tentative identification of threshold and delay parameters, Tong (1990) suggested the use of non-parametric lag regression estimates. More suggestions have been given by some other authors which rely heavily on the use of graphics (Tsay (1989)).

On the contrary, the Bayesian approach allows one to obtain the joint density of the intrinsic parameters in a closed form offering an alternative procedure to estimate the parameters.

This chapter is organized as follows. Section 3.2 contains a description of the STAR model of order p and the derivation of posterior distributions of parameters with a noninformative prior. These posterior distributions are developed using the conditional likelihood function and we show how the marginal distribution of each intrinsic parameter is obtained. In Section 3.3, we consider the predictive performance of the STAR model by obtaining the exact one-step-ahead predictive density and conditional multi-step-ahead predictive density. Moreover, we present an approximation to multi-step-ahead predictive density using Monte Carlo integration. Finally, in section 3.4, we illustrate the results of sections 3.2 and 3.3 by means of simulated data sets and the frequently reported Canadian lynx data.

3.2 Model and Distribution

Consider the class of Gaussian Smooth Threshold Autoregressive (GSTAR) models of order p defined by

$$X_t = a_0 + \sum_{j=1}^{p} a_j X_{t-j} + (b_0 + \sum_{j=1}^{p} b_j X_{t-j}) G\left(\frac{X_{t-d} - r}{z}\right) + \varepsilon_t, \qquad (3.2.1)$$

where G is a Gaussian distribution function with known mean and variance, for convenience; r is a threshold parameter, $r \in \Re$; d is a delay parameter, $d \in \mathcal{Z}^+$; z is a smoothing parameter, $z \in \Re^+$ and $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables, ε_t is $i.i.d.\ \mathcal{N}(0, \sigma^2)$ where $\{X_s,\ s < t\}$ is independent of ε_t. As noted earlier, the minimal requirement of G is being continuous and nondecreasing.

To keep the notation simple, in the Bayesian analysis, we define all parameters in (3.2.1) in a compact form as

$$\theta = (a_0, a_1, ..., a_p, b_0, b_1, ..., b_p)^T, \quad \phi = (r, d, z), \quad \sigma.$$

Conditional on $\mathbf{X}_0 = (X_1, ..., X_m)^T$, $m = \wedge\{d, p\}$ and ϕ, the model (3.2.1) can be written as

$$\mathbf{X} = \mathbf{\Phi}\, \theta + \varepsilon, \qquad (3.2.2)$$

where

$$\mathbf{X} = (X_{m+1}, X_{m+2}, ..., X_N)^T$$

$$\varepsilon = (\varepsilon_{m+1}, \varepsilon_{m+2}, ..., \varepsilon_N)^T$$

$$\mathbf{\Phi} = \begin{pmatrix} 1 & X_m & \dots & X_{m+1-p} & G_{m+1} & X_m G_{m+1} & \dots & X_{m+1-p} G_{m+1} \\ 1 & X_{m+1} & \dots & X_{m+2-p} & G_{m+2} & X_{m+1} G_{m+2} & \dots & X_{m+2-p} G_{m+2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & X_{N-1} & \dots & X_{N-p} & G_N & X_{N-1} G_N & \dots & X_{N-p} G_N \end{pmatrix}$$

and $G_t = G\left(\frac{X_{t-d} - r}{z}\right)$.

Then the conditional likelihood function, based on (3.2.2) and the assumption on ε_t, is

$$L(\mathbf{X} \mid \theta, \sigma, \phi) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{X} - \Phi\theta)^T(\mathbf{X} - \Phi\theta)\right\}, \qquad (3.2.3)$$

where $n = N - m$ is the number of effective data points.

Let $\pi(\phi)$ be a prior density function of ϕ. Assume that conditional on ϕ, independent priors of θ and σ are of standard Jeffreys form. Hence

$$\pi\{\theta, \phi, \sigma\} = \pi_c\{\theta, \sigma \mid \phi\}\pi(\phi), \qquad (3.2.4)$$

where $\pi_c\{\theta, \sigma \mid \phi\} \propto \frac{1}{\sigma}$, for $\theta \in \Re^{2(p+1)}$ and $\sigma \in \Re^+$.

Using Bayes theorem the joint posterior density of θ and σ, conditional on $(\phi, \mathbf{X})$ can be expressed as

$$P(\theta, \sigma \mid \phi, \mathbf{X}) \propto \pi_c(\theta, \sigma \mid \phi) L(\mathbf{X} \mid \theta, \sigma, \phi). \qquad (3.2.5)$$

From (3.2.3),(3.2.4) and (3.2.5) we have

$$P(\theta, \sigma \mid \phi, \mathbf{X}) \propto (2\pi)^{-\frac{n}{2}} \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{X} - \Phi\theta)^T(\mathbf{X} - \Phi\theta)\right\}. \quad (3.2.6)$$

Now,

$$\begin{aligned}
(\mathbf{X} - \Phi\theta)^T(\mathbf{X} - \Phi\theta) &= (\mathbf{X} - \Phi\hat{\theta} + \Phi\hat{\theta} - \Phi\theta)^T(\mathbf{X} - \Phi\hat{\theta} + \Phi\hat{\theta} - \Phi\theta) \\
&= (\mathbf{X} - \Phi\hat{\theta})^T(\mathbf{X} - \Phi\hat{\theta}) + (\mathbf{X} - \Phi\hat{\theta})^T(\Phi\hat{\theta} - \Phi\theta) \\
&+ (\Phi\hat{\theta} - \Phi\theta)^T(\mathbf{X} - \Phi\hat{\theta}) + (\Phi\hat{\theta} - \Phi\theta)^T(\Phi\hat{\theta} - \Phi\theta) \\
&= (\mathbf{X} - \Phi\hat{\theta})^T(\mathbf{X} - \Phi\hat{\theta}) + (\hat{\theta} - \theta)^T\Phi^{\mathbf{T}}\Phi(\hat{\theta} - \theta) \\
&= v\, s^2 + (\hat{\theta} - \theta)^T\Phi^{\mathbf{T}}\Phi(\hat{\theta} - \theta),
\end{aligned}$$

since the cross product term $(\hat{\theta} - \theta)^T\Phi^T(\mathbf{X} - \Phi\hat{\theta}) = 0$. Then (3.2.6) can be reformulated as

$$\begin{aligned}
P(\theta, \sigma \mid \phi, \mathbf{X}) &\propto (2\pi)^{-\frac{n}{2}} \sigma^{-(n+1)} \\
&\times \exp\left\{-\frac{1}{2\sigma^2}(vs^2 + (\hat{\theta} - \theta)^T\Phi^{\mathbf{T}}\Phi(\hat{\theta} - \theta))\right\}, (3.2.7)
\end{aligned}$$

where

$$\hat{\theta} = (\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{X}, \quad s^2 = \frac{(\mathbf{X}-\mathbf{\Phi}\hat{\theta})^T(\mathbf{X}-\mathbf{\Phi}\hat{\theta})}{v} \tag{3.2.8}$$

and $v = N - m - 2(p+1) = n - 2(p+1)$.

We now use (3.2.7) to determine the posterior distributions of θ and σ given $(\phi, \mathbf{X})$ and obtain the marginal posterior distribution of ϕ given $\mathbf{X}$.

From (3.2.7) it is clear that the conditional distribution of θ given $(\sigma, \phi, \mathbf{X})$ is a multivariate normal

$$\theta \mid (\sigma, \phi, \mathbf{X}) \rightsquigarrow \mathcal{N}(\hat{\theta}, \sigma^2(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}).$$

The posterior distribution of σ given $(\phi, \mathbf{X})$ can be obtained by integrating (3.2.7) with respect to θ, resulting in

$$P(\sigma \mid \phi, \mathbf{X}) \propto \frac{1}{\sigma^{v+1}} \exp\left(-\frac{vs^2}{2\sigma^2}\right).$$

This indicates that the density of σ is in the form of an Inverted Gamma density, IG(α, γ), with parameters $\alpha = \frac{v}{2}$ and $\gamma = \frac{2}{vs^2}$.

Similarly, the posterior distribution of θ conditional on ϕ and $\mathbf{X}$ can be obtained by integrating (3.2.7) over σ, $\sigma > 0$, as

$$P(\theta \mid \phi, \mathbf{X}) \propto \int_0^\infty (2\pi)^{-\frac{n}{2}} \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2}(vs^2 + (\hat{\theta}-\theta)^T\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}(\hat{\theta}-\theta))\right\} d\sigma.$$

Using Gamma function

$$\Gamma(x) = \int_0^\infty t^{x-1} \exp(-t) dt,$$

and letting

$$w = \frac{vs^2 + (\hat{\theta}-\theta)^T\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}(\hat{\theta}-\theta)}{2\sigma^2},$$

$$dw = -\frac{vs^2 + (\hat{\theta}-\theta)^T\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}(\hat{\theta}-\theta)}{\sigma^3} d\sigma,$$

the integral can be simplified as

$$\begin{aligned}
P(\theta \mid \phi, X) &= -(2\pi)^{-\frac{n}{2}} \int_0^\infty \sigma^{-(n+1)} \exp(-w) \frac{-\sigma^3}{2\sigma^2 w} dw, \\
&= (2\pi)^{-\frac{n}{2}} \int_0^\infty (\sigma^2)^{-\frac{n}{2}} \exp(-w) \frac{1}{2w} dw, \\
&= (2\pi)^{-\frac{n}{2}} \int_0^\infty \left\{ \frac{vs^2 + (\hat{\theta} - \theta)^T \mathbf{\Phi}^\mathrm{T} \mathbf{\Phi} (\hat{\theta} - \theta)}{2w} \right\}^{-\frac{n}{2}} \frac{1}{2w} \exp(-w) dw, \\
&= (2\pi)^{-\frac{n}{2}} 2^{\frac{n}{2}-1} \left\{ vs^2 + (\hat{\theta} - \theta)^T \mathbf{\Phi}^\mathrm{T} \mathbf{\Phi} (\hat{\theta} - \theta) \right\}^{-\frac{n}{2}} \\
&\times \int_0^\infty w^{\frac{n}{2}-1} \exp(-w) dw, \\
&\propto (2\pi)^{-\frac{n}{2}} 2^{\frac{n}{2}-1} \Gamma\left(\frac{n}{2}\right) \{ vs^2 + (\hat{\theta} - \theta)^T \mathbf{\Phi}^\mathrm{T} \mathbf{\Phi} (\hat{\theta} - \theta) \}^{-\frac{n}{2}} \qquad (3.2.9)
\end{aligned}$$

which is in the form of a $2(p+1)$ dimensional multivariate t distribution with v degrees of freedom, mean $\hat{\theta}$ and variance-covariance matrix $\frac{v}{v-2} s^2 (\mathbf{\Phi}^\mathrm{T} \mathbf{\Phi})^{-1}$.

In summary, we have the following

$$\begin{aligned}
\theta \mid (\sigma, \phi, \mathbf{X}) &\rightsquigarrow \mathcal{N}(\hat{\theta}, \sigma^2 (\mathbf{\Phi}^\mathrm{T} \mathbf{\Phi})^{-1}), \\
\sigma \mid (\phi, \mathbf{X}) &\rightsquigarrow \mathrm{IG}\left(\frac{v}{2}, \frac{2}{vs^2}\right), \\
\theta \mid (\phi, \mathbf{X}) &\rightsquigarrow \mathrm{T}_{2(p+1)}\left(v, \hat{\theta}, \frac{v}{v-2} s^2 (\mathbf{\Phi}^\mathrm{T} \mathbf{\Phi})^{-1}\right).
\end{aligned}$$

The marginal posterior density of $\phi = (r, d, z)$ can now be obtained by the Bayes theorem,

$$P(\phi \mid \mathbf{X}) = \frac{\pi(\phi) P(\mathbf{X} \mid \phi)}{P(\mathbf{X})},$$

where $P(\mathbf{X} \mid \phi) = \int_{\Re^+} \int_{\Re^{2(p+1)}} \pi_c(\theta, \sigma \mid \phi) L(\mathbf{X} \mid \theta, \sigma, \phi) \, d\theta \, d\sigma$ and $P(\mathbf{X})$ is a normalising constant. Hence, the posterior distribution of ϕ

$$\begin{aligned}
P(\phi \mid \mathbf{X}) &\propto \int_{\Re^{2(p+1)}} \pi(\phi) (2\pi)^{-\frac{n}{2}} 2^{\frac{n}{2}-1} \Gamma(\frac{n}{2}) \left\{ vs^2 + (\hat{\theta} - \theta)^T \mathbf{\Phi}^\mathrm{T} \mathbf{\Phi} (\hat{\theta} - \theta) \right\}^{-\frac{n}{2}} d\theta, \\
&\propto \pi(\phi) (2\pi)^{-\frac{n}{2}} 2^{\frac{n}{2}-1} \Gamma(\frac{n}{2}) \int_{\Re^{2(p+1)}} \left\{ vs^2 + (\hat{\theta} - \theta)' \mathbf{\Phi}^\mathrm{T} \mathbf{\Phi} (\hat{\theta} - \theta) \right\}^{-\frac{n}{2}} d\theta,
\end{aligned}$$

$$\propto \pi(\phi)(2\pi)^{-\frac{n}{2}}2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})$$

$$\times \int_{\Re^{2(p+1)}}(vs^2)^{-\frac{n}{2}}\left\{1+\frac{(\hat{\theta}-\theta)^T\mathbf{\Phi^T\Phi}(\hat{\theta}-\theta)}{vs^2}\right\}^{-\frac{n}{2}} d\theta.$$

Solving the above by using the t-integrals (Box and Tiao (1973), page 145-146)

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left\{1+\frac{(\mathbf{X}-\eta)^T\mathbf{C}^{-1}(\mathbf{X}-\eta)}{v}\right\}^{-\frac{v+n}{2}} dx_1...dx_n$$

$$= \frac{\left(\Gamma(\frac{1}{2})\right)^n\Gamma(\frac{n}{2})}{\Gamma(\frac{v+n}{2})}v^{\frac{n}{2}} \mid \mathbf{C} \mid^{\frac{1}{2}},$$

where $\mathbf{x}=(x_1,...,x_n)^T$, η is an $n\times 1$ vector of constants and $\mathbf{C}$ is an $n\times n$ positive definite symmetric matrix, we have

$$P(\phi \mid \mathbf{X}) \propto \pi(\phi)(2\pi)^{-\frac{n}{2}}2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})$$

$$\times \int_{\Re^{2(p+1)}}(vs^2)^{-\frac{n}{2}}\left\{1+\frac{(\hat{\theta}-\theta)^T\mathbf{\Phi^T\Phi}(\hat{\theta}-\theta)}{vs^2}\right\}^{-\frac{v+2(p+1)}{2}} d\theta,$$

$$\propto \pi(\phi)(2\pi)^{-\frac{n}{2}}2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})(vs^2)^{-\frac{n}{2}}$$

$$\times \int_{\Re^{2(p+1)}}\left\{1+\frac{(\hat{\theta}-\theta)^T\frac{\mathbf{\Phi^T\Phi}}{s^2}(\hat{\theta}-\theta)}{v}\right\}^{-\frac{v+2(p+1)}{2}} d\theta,$$

$$\propto \pi(\phi)(2\pi)^{-\frac{n}{2}}2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})(vs^2)^{-\frac{n}{2}}\frac{\left(\Gamma(\frac{1}{2})\right)^{2(p+1)}\Gamma(\frac{v}{2})}{\Gamma(\frac{n}{2})}(vs^2)^{(p+1)} \mid \mathbf{\Phi^T\Phi} \mid^{-\frac{1}{2}},$$

$$\propto \pi(\phi)2^{-(\frac{v}{2}+1)}\pi^{-\frac{v}{2}}\left(\frac{vs^2}{2}\right)^{-\frac{v}{2}}\Gamma(\frac{v}{2}) \mid \mathbf{\Phi^T\Phi} \mid^{-\frac{1}{2}},$$

from which, we can readily obtain the marginal distributions of each intrinsic parameter r, d, and z.

3.3 Predictive Performance

It is well-known that the joint predictive density function even for the linear autoregressive model of order p does not have a closed form except in the case of one-step-ahead prediction (Zellner (1971)). When cycles exist in the data, Tong (1983) developed a simple method to obtain multi-step-ahead predictions from the threshold model based on the cyclic property. In this section we derive the exact one-step-ahead predictive distribution based on a STAR model of order p conditional on ϕ without utilising the cyclic property and extend this to obtain the exact multi-step-ahead predictive distribution under certain conditions. Monte Carlo approach (Geweke and Terui (1993)) is then used to investigate the unconditional multi-step-ahead predictive performance.

Suppose X_N denotes the last observation and let

$$\Psi = (1, X_N, ..., X_{N-p+1}, G_{N+1}, X_N G_{N+1}, ..., X_{N-p+1} G_{N+1}),$$

so that

$$X_{N+1} = \Psi\ \theta\ +\ \varepsilon_{N+1}, \tag{3.3.1}$$

where θ and G_t are as in Section 3.2. Conditional on ϕ, (3.3.1) is in the form of a usual linear model. Therefore by the usual normal theory, the Bayesian predictive density of X_{N+1} given ϕ can be obtained. Since ε_{N+1} is normally distributed with mean 0 and variance σ^2, the distribution of the next observation conditional on all parameters is given by

$$P(X_{N+1} \mid \theta, \sigma, \phi, \mathbf{X}) \propto \sigma^{-1} \exp\{-\frac{1}{2\sigma^2}(X_{N+1} - \Psi\theta)^2\}. \tag{3.3.2}$$

Combining (3.2.7) and (3.3.2), we find that

$$\begin{aligned} P(X_{N+1}, \theta, \sigma \mid \phi, \mathbf{X}) &\propto P(X_{N+1} \mid \theta, \sigma, \phi, \mathbf{X}) P(\theta, \sigma \mid \phi, \mathbf{X}), \\ &\propto \sigma^{-(n+2)} \exp\left\{-\frac{1}{2\sigma^2}\{(X_{N+1} - \Psi\theta)^2\right. \\ &+ \left. vs^2 + (\hat{\theta} - \theta)^T \mathbf{\Phi}^{\mathbf{T}} \mathbf{\Phi} (\hat{\theta} - \theta)\}\right\}, \end{aligned} \tag{3.3.3}$$

where $\hat{\theta} = (\Phi^T\Phi)^{-1}\Phi^T\mathbf{X}$, $s^2 = \frac{(\mathbf{X}-\Phi\hat{\theta})^T(\mathbf{X}-\Phi\hat{\theta})}{v}$ and $v = N - m - 2(p+1) = n - 2(p+1)$.

Upon simplification of the right-hand side of (3.3.3) and integrating with respect to θ and σ, it is clear that X_{N+1}, conditional on ϕ and $\mathbf{X}$, follows a t distribution with v degrees of freedom, with the conditional predictive mean and variance $\hat{X}_{N+1} = \Psi\hat{\theta}$ and $\Psi \text{ Var}(\theta \mid \phi, \mathbf{X})\Psi^T + \text{E}(\sigma^2 \mid \phi, \mathbf{X})$ respectively. These quantities can be evaluated in a closed form using conditional distributions of θ and σ.

The above exact one-step-ahead predictions, conditional on ϕ, can be easily extended to obtain the exact j-steps-ahead predictions as follows.

Define,

$$\Psi_j = (1, X_{N+j-1}, \ldots, X_{N+j-p}, G_{N+j}, X_{N+j-1}G_{N+j}, \ldots, X_{N+j-p}G_{N+j}),$$

for $1 \leq j \leq h \leq d$, where h is the lead time.

Hence, conditional on ϕ, X_{N+j} can be written as

$$X_{N+j} = \Psi_j\,\theta + \varepsilon_{N+j}, \quad j = 1, 2, \ldots, h, \quad j \leq d. \tag{3.3.4}$$

Therefore, as before, the Bayesian predictive density of X_{N+j}, conditional on ϕ, can be obtained.

An algorithm for obtaining the conditional Bayesian multi-step-ahead predictions based on Monte Carlo integration is outlined below. The procedure is repeated M times, where M is a constant depending upon the convergence rate.

(1). Fix $\phi = (r, d, z)$ at the posterior modal point.

(2). Conditional on ϕ draw σ and θ from the density (3.2.6). This may be done by employing a two step procedure as follows, (a). since $\frac{vs^2}{\sigma^2} \rightsquigarrow \chi^2(v)$ conditional on ϕ and $\mathbf{X}$, draw $w \rightsquigarrow \chi^2(v)$ and set $\sigma^2 = \frac{vs^2}{w}$.

(b). since $\theta \mid \sigma, \phi, \mathbf{X} \rightsquigarrow \mathcal{N}(\hat{\theta}, \sigma^2(\mathbf{\Phi}^{\mathbf{T}}\mathbf{\Phi})^{-1})$, draw θ from this multivariate normal distribution.

(3). Define $\hat{X}_t = X_t$ for $t = 1, 2, ..., T$. Recursively generate

$$\begin{aligned}\hat{X}_{T+j} &= a_0 + \sum_{i=1}^{p} a_i X_{T-i} \\ &+ (b_0 + \sum_{i=1}^{p} b_i X_{T-i}) G\left(\frac{\hat{X}_{T+j-d} - r}{z}\right) + \varepsilon_{T+j},\end{aligned}$$

where $j = 1, 2, ..., h$ and h is the number of steps prediction.

(4). Averaging each of $\hat{X}_{T+j}$, $j = 1, ..., h$ over the M replications, we obtain a numerical approximation to the Bayesian predictor.

In the next section, the Bayesian predictor obtained by the Monte Carlo method is compared with the Conditional Least Squares predictor given in Chan and Tong (1986) using the root-mean-squared- error performance measure defined by

$$\text{RMSE}(h) = \left\{\frac{1}{T^*} \sum_{T'=T}^{T+T^*} (\hat{X}_{T'+h} - X_{T'+h})^2\right\}^{\frac{1}{2}}, \tag{3.3.5}$$

where T^* is the number of samples used in the calculation of this expression for each h, $\hat{X}_{T'+h}$ is the h step ahead predictor from the time origin T'. This measurement was used by Tong (1983) to compare the results of the short, medium and long range forecast errors.

3.4 Examples

This section presents two examples to complement the theoretical results developed in Sections 3.2 and 3.3. Most of the programs to perform the necessary computations were written in Fortran 77, with a few subroutines included from the Nag-library. For illustrative purposes, only the first-order STAR model within the ergodicity region

$$a_1 < 1,\ a_1 + b_1 < 1,\ a_1\,(a_1 + b_1) < 1. \tag{3.4.1}$$

is discussed.

In what follows, a flat prior, $\pi(\phi)$, for $\phi = (r, d, z)$ is employed so that the marginal posterior density of ϕ becomes

$$P(\phi \mid \mathrm{X}) \propto 2^{-(\frac{v}{2}+1)}\pi^{-\frac{v}{2}}\left(\frac{vs^2}{2}\right)^{-\frac{v}{2}}\Gamma(\frac{v}{2}) \mid \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi} \mid^{-\frac{1}{2}}. \tag{3.4.2}$$

Note that an advantage of the Bayesian method compared to Conditional Least Squares is that the marginal posterior density of the implicit parameters (r, d, z) can be obtained in a closed form, using the former.

It is clear that the maximisation of posterior density (3.4.2) depends on the determinant of $\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}$, the estimated variance s^2 and sample size of the data. Specifically, the determinant component in (3.4.2) will tend to infinity when the matrix $\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}$ is almost singular and when this happens this component dominates the maximisation of posterior density causing the bivariate modes of r and z to fall on the boundary of their domain, D_{rz}. To avoid the singularity, the domains of r and z have to be properly defined. As an example, when $p = 1$ and $d = 1$ the matrix $\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}$ is

$$\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi} = \begin{pmatrix} N-1 & \sum_{j=1}^{N-1} X_j & \sum_{j=1}^{N-1} G_{j+1} & \sum_{j=1}^{N-1} X_j G_{j+1} \\ \sum_{j=1}^{N-1} X_j & \sum_{j=1}^{N-1} X_j^2 & \sum_{j=1}^{N-1} X_j G_{j+1} & \sum_{j=1}^{N-1} X_j^2 G_{j+1} \\ \sum_{j=1}^{N-1} G_{j+1} & \sum_{j=1}^{N-1} X_j G_{j+1} & \sum_{j=1}^{N-1} G_{j+1}^2 & \sum_{j=1}^{N-1} X_j G_{j+1}^2 \\ \sum_{j=1}^{N-1} X_j G_{j+1} & \sum_{j=1}^{N-1} X_j^2 G_{j+1} & \sum_{j=1}^{N-1} X_j G_{j+1}^2 & \sum_{j=1}^{N-1} X_j^2 G_{j+1}^2 \end{pmatrix},$$

where $G_t = G\left((X_{t-d} - r)/z\right)$. The singularity is caused by the value of $G(.)$ in the above matrix when r is considerably large with respect to the observation such that the bivariate mode reaches the right boundary of the domain of r, D_r. To reduce the domination of the determinant, one may also need to choose a proper sample size in the variance component of the above posterior density.

3.4.1 Simulation study

The following simulation study was performed to illustrate the theory given in Sections 3.2 and 3.3. With a view to limiting computer time, 50 independent samples of size 200 of

$$X_t = 10 - 2X_{t-1} - 5G\left(\frac{X_{t-1}-5}{0.5}\right) + 1.6X_{t-1}G\left(\frac{X_{t-1}-5}{0.5}\right) + \varepsilon_t,$$

where G and the distribution of ε_t are standard Normal distribution functions, were generated and the computations were carried out with the domain of $(r,\ z)$ restricted to $D_{rz} = \{(r,z) : 0 < r < 10,\ 0.1 < z < 10\}$.

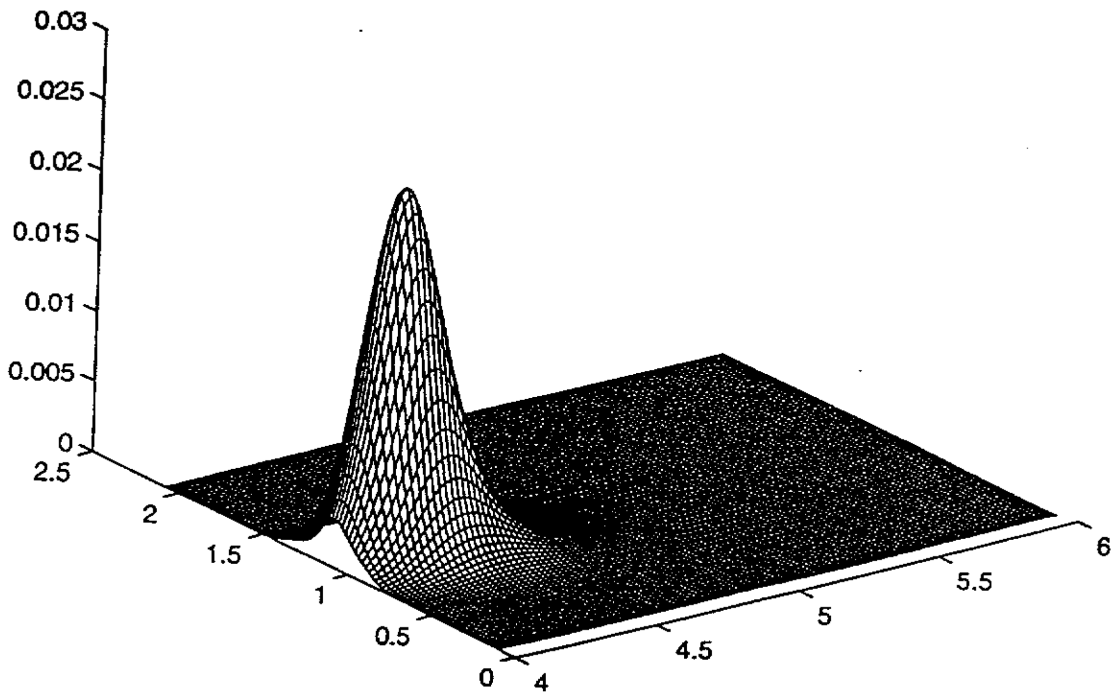

Figure 3.1. Bivariate joint density of $(r, z) = (5, 0.5)$

Table 3.1. Bayesian estimates :

True value of $(a_0, a_1, b_0, b_1, r, z, \sigma^2) = (10, -2, -5, 1.6, 5, 0.5, 1)$

No.	$\hat{a_0}$	$\hat{a_1}$	$\hat{b_0}$	$\hat{b_1}$	$\hat{r}$	$\hat{z}$	$\hat{\sigma^2}$
1	10.1069	−2.0274	−4.1995	1.5065	5.0	0.80	0.9341
2	10.0811	−2.0192	−5.1294	1.6173	5.0	0.40	1.2160
3	10.0676	−2.0330	−3.9349	1.5067	5.0	0.60	1.0806
4	10.0328	−2.0129	−4.4424	1.5564	5.15	0.45	0.9694
5	9.9527	−1.9800	−4.9061	1.5566	4.85	0.65	1.1566
6	10.0344	−1.9866	−5.3471	1.6194	5.10	0.40	1.0574
7	9.9884	−2.0067	−4.8449	1.6152	5.0	0.35	1.0695
8	10.2302	−2.0277	−5.7741	1.6955	4.85	0.50	1.0370
9	9.8725	−1.9112	−5.1845	1.5661	5.10	0.40	0.9280
10	10.1198	−2.0309	−6.1211	1.7437	4.75	0.60	1.0149
11	9.7618	−1.9096	−3.3610	1.3521	5.40	0.60	1.1520
12	10.1162	−1.9595	−5.7680	1.6626	5.15	0.40	0.9653
13	10.0181	−1.9846	−6.0480	1.7450	5.05	0.15	0.8604
14	10.3111	−2.1434	−6.4142	1.8777	3.95	1.25	1.0303
15	10.0058	−2.0507	−4.7598	1.6197	4.9	0.70	1.1672
16	10.0184	−1.9824	−4.8446	1.5501	5.20	0.25	0.8394
17	10.0985	−2.0958	−4.6336	1.6558	4.95	0.60	0.9903
18	9.8583	−1.9843	−5.2753	1.6344	4.75	0.50	1.0956
19	10.1096	−2.0099	−5.4994	1.6667	4.85	0.40	0.8788
20	9.9865	−1.9700	−4.1232	1.4607	5.25	0.40	1.4539
21	9.8541	−1.9199	−5.0267	1.5234	5.0	0.40	0.8704
22	9.8598	−1.9978	−5.5240	1.6891	4.60	0.70	1.0958
23	0.9995	−1.9755	−5.5444	1.6466	4.90	0.40	1.1323
24	10.3172	−2.1241	−5.3343	1.7180	4.85	0.40	0.8426
25	10.1032	−2.0423	−5.9011	1.7325	4.70	0.10	1.3483
26	9.7828	−1.8911	−6.5101	1.6986	5.20	0.25	0.9102
27	9.9583	−2.0567	−4.8873	1.6627	4.95	0.60	0.7897
28	9.8178	−1.9423	−4.8215	1.5294	4.90	0.35	0.9548
29	9.6862	−1.9060	−5.3511	1.5886	5.0	0.60	1.0565
30	10.3105	−2.1705	−5.0536	1.7515	4.85	0.35	1.0410
31	10.2282	−2.0502	−5.4933	1.6857	4.55	0.70	1.0067
32	9.9379	−2.0042	−3.7045	1.4802	5.25	0.65	1.0340
33	9.8834	−1.9363	−5.2186	1.5877	5.15	0.70	0.9109
34	10.1834	−2.0925	−5.3833	1.7086	4.95	0.55	0.9786
35	9.8668	−1.9316	−4.9325	1.5367	4.90	0.25	1.0891
36	9.9843	−1.9647	−4.5851	1.5270	5.15	0.25	0.9561
37	9.6901	−1.8504	−4.5640	1.4245	5.15	0.75	1.1827
38	9.8237	−1.9125	−4.3276	1.4669	5.20	0.70	1.1264
39	9.9964	−1.9656	−5.4179	1.6334	5.15	0.65	1.0392
40	10.2996	−2.1042	−5.2006	1.7111	5.05	0.35	0.9416
41	10.0825	−1.9682	−4.8214	1.5053	5.15	0.35	0.9771
42	10.0598	−2.0665	−3.9949	1.5369	5.0	0.60	1.0171
43	10.3152	−2.1699	−5.4961	1.8054	4.90	0.45	0.9466
44	9.7059	−1.9404	−4.2509	1.4936	5.10	0.55	1.4281
45	9.9424	−1.9274	−4.4041	1.4458	5.05	0.55	1.0564
46	10.0146	−1.9790	−6.5412	1.7815	4.85	0.10	0.8291
47	10.2831	−2.0815	−4.9149	1.6656	5.25	0.45	1.2053
48	9.5228	−1.8266	−5.0730	1.4983	5.0	0.30	1.0039
49	10.1308	−2.0019	−4.6920	1.5535	5.0	0.40	0.8885
50	9.9250	−1.9891	−4.7190	1.5705	5.05	0.35	1.1444

Table 3.2. Conditional Least Squares estimates:

True value of $(a_0, a_1, b_0, b_1, r, z, \sigma^2) = (10, -2, -5, 1.6, 5, 0.5, 1)$

No.	$\hat{a_0}$	$\hat{a_1}$	$\hat{b_0}$	$\hat{b_1}$	$\hat{r}$	$\hat{z}$	$\hat{\sigma^2}$
1	10.0986	−2.0204	−4.2871	1.5091	5.0	0.30	0.9340
2	10.0729	−2.0124	−5.1809	1.6168	5.10	0.40	1.2159
3	10.0595	−2.0259	−4.0371	1.5108	4.95	0.45	1.0805
4	10.0328	−2.0129	−4.4424	1.5564	5.25	0.60	0.9694
5	9.9527	−1.9800	−4.9061	1.5566	5.05	0.30	1.1566
6	10.0344	−1.9866	−5.3471	1.6194	5.0	0.55	1.0574
7	9.9884	−2.0067	−4.8449	1.6152	4.75	0.45	1.0695
8	10.2302	−2.0277	−5.7741	1.6955	5.25	0.60	1.0370
9	9.8725	−1.9112	−5.1845	1.5661	5.05	0.60	0.9279
10	10.1024	−2.0193	−6.0999	1.7315	4.85	0.10	1.0148
11	9.7618	−1.9096	−3.3610	1.3521	2.85	0.10	1.1520
12	10.1162	−1.9595	−5.7680	1.6626	5.10	0.35	0.9653
13	10.0181	−1.9846	−6.0480	1.7450	5.05	0.55	0.8604
14	10.2558	−2.1398	−6.4625	1.8837	5.05	0.60	1.0297
15	10.0058	−2.0507	−4.7598	1.6197	5.0	0.70	1.1672
16	10.0184	−1.9824	−4.8446	1.5501	4.70	0.85	0.8394
17	10.0904	−2.0894	−4.7245	1.6595	4.90	0.10	0.9902
18	9.8583	−1.9843	−5.2753	1.6344	5.05	0.35	1.0956
19	10.0990	−2.0031	−5.5694	1.6690	5.15	0.55	0.8786
20	9.9865	−1.9700	−4.1232	1.4607	4.75	0.50	1.4539
21	9.8541	−1.9199	−5.0267	1.5234	4.85	0.70	0.8704
22	9.8598	−1.9978	−5.5240	1.6891	5.0	0.25	1.0958
23	0.9995	−1.9755	−5.5444	1.6466	4.95	0.80	1.1323
24	10.3172	−2.1241	−5.3343	1.7180	5.25	0.45	0.8426
25	10.1032	−2.0423	−5.9011	1.7325	5.20	0.50	1.3483
26	9.7780	−1.8876	−6.6441	1.7108	5.10	0.65	0.9101
27	9.9583	−2.0567	−4.8873	1.6627	4.95	0.60	0.7897
28	9.8178	−1.9423	−4.8215	1.5294	4.95	0.45	0.9548
29	9.6862	−1.9060	−5.3511	1.5886	5.10	0.30	1.0565
30	10.3105	−2.1705	−5.0536	1.7515	4.95	0.80	1.0410
31	10.2041	−2.0353	−5.5413	1.6787	5.0	1.0	1.0064
32	9.9379	−2.0042	−3.7045	1.4802	5.15	0.45	1.0340
33	9.8372	−1.9030	−6.3404	1.6811	4.80	0.70	0.9101
34	10.1834	−2.0925	−5.3833	1.7086	5.15	0.30	0.9786
35	9.8629	−1.9294	−5.0315	1.5465	5.0	0.45	1.0890
36	9.9843	−1.9647	−4.5851	1.5270	5.10	0.45	0.9561
37	9.6855	−1.8463	−4.6331	1.4282	4.85	0.10	1.1825
38	9.8237	−1.9125	−4.3276	1.4669	5.15	0.30	1.1264
39	9.9964	−1.9656	−5.4179	1.6334	5.05	0.45	1.0392
40	10.2996	−2.1042	−5.2006	1.7111	4.80	0.80	0.9416
41	10.0709	−1.9582	−4.7878	1.4929	4.95	0.40	0.9770
42	10.0598	−2.0665	−3.9949	1.5369	5.0	0.15	1.0171
43	10.3152	−2.1699	−5.4961	1.8054	4.95	0.65	0.9466
44	9.6986	−1.9340	−4.3088	1.4931	5.0	0.65	1.4281
45	9.9424	−1.9274	−4.4041	1.4458	5.0	0.10	1.0564
46	10.0146	−1.9790	−6.5412	1.7815	5.15	0.55	0.8291
47	10.2831	−2.0815	−4.9149	1.6656	5.15	0.75	1.2053
48	9.5320	−1.8312	−5.3598	1.5377	4.75	0.65	1.0038
49	10.1308	−2.0019	−4.6920	1.5535	4.80	0.10	0.8885
50	9.9205	−1.9855	−4.7874	1.5743	5.0	0.60	1.1442

The marginal posterior density of the delay parameter d, calculated for $d = 1, 2, 3, 4$ indicates that the highest posterior density is at $d = 1$ for each of the 50 samples. Hence the computation of the bivariate joint density of (r, z) was restricted to the case $d = 1$, for each sample. As an illustration, the bivariate joint density obtained from one of the samples is given in Figure 3.1. Within the norms of the Bayesian framework, the bivariate posterior modes of these distributions were chosen as the estimators of r and z, for d fixed at 1, and the remaining model parameters θ and σ^2, were estimated using (3.2.8). The estimates based on the Bayesian method are summarised in Table 3.1. For completeness and to make a sample-wise comparison of the Bayesian method with the conditional least square method given in Chan and Tong (1986), we have summarised the corresponding Conditional Least Squares estimates for the same samples in Table 3.2. The results indicate that the Bayesian method performs equally well compared to the Conditional Least Squares method.

Table 3.3. Root Mean Squared Error comparison of predictors

Sample	Average	Sample	Average
1	1.000811	26	1.000384
2	1.004111	27	0.998334
3	1.001733	28	1.001881
4	0.997150	29	0.998914
5	1.000859	30	1.007727
6	0.996887	31	1.008209
7	0.999623	32	1.011499
8	1.010009	33	1.000470
9	0.996768	34	1.001108
10	1.008101	35	0.991953
11	1.002135	36	1.006774
12	0.996713	37	1.008356
13	1.001812	38	0.995705
14	1.002958	39	0.993605
15	0.993719	40	0.990490
16	0.995772	41	0.999583
17	1.006863	42	1.013771
18	0.999396	43	1.002979
19	1.009301	44	1.005271
20	0.9934889	45	0.998269
21	0.996131	46	0.996274
22	0.993405	47	0.992993
23	0.998917	48	0.997356
24	0.999113	49	1.002352
25	1.004841	50	0.999672

To understand the multi-step-ahead predictive performance of the Bayesian approach as compared to the Conditional Least Squares (CLS) method we computed the root mean squared error (RMSE), given in (3.3.5), for up to 20 steps ahead predictions ($h = 1,\ 2,\ \ldots,\ 20$), with $T^* = 5$. The initial estimates of the model were computed using the first 200 observations from each of the samples, of size 300. The data generation model had the same set of parameters as before, and the computations were repeated 50 times. The results showed that the RMSE of Conditional Least Squares (RMSE(C)), and the RMSE of Bayesian (RMSE(B)) are close to each other for all values of $h = 1,\ 2,\ \ldots,\ 20$. That is the ratio of RMSE(B) to RMSE(C) is close to 1. From this we conclude that the predictive performance of the Bayesian method is comparable to that of the CLS method. The average of the ratio of RMSE(B) to RMSE(C), average taken over the values $h = 1,\ 2,\ \ldots,\ 20$, for the 50 independent samples, are given in Table 3.3.

In order to compare the performance of the two methods more thoroughly, further simulation study was performed using different sets of parameter values, $(a_0, a_1, b_0, b_1, r, z, \sigma^2)$ as given in Table 3.4, which are within the ergodic region of (3.4.1).

Table 3.4. Parameter coefficients of five simulation models

Model	a_0	a_1	b_0	b_1	r	z	σ^2
1	1.0	0.8	−2.0	−0.5	0.5	0.3	0.01
2	1.0	−2.0	−0.5	2.4	1.0	0.5	1.00
3	2.0	0.5	−0.5	−1.4	1.0	0.5	1.00
4	5.0	0.8	−0.5	−5.8	1.0	0.5	1.00
5	1.0	−10.0	−5.0	10.5	1.0	0.5	1.00

As in the earlier study, once again fifty independent samples of size 500 were generated for each of the models and both the Bayesian estimates and the conditional least square estimates were computed. The individual

Table 3.5. Bayesian estimates of Model 1: True value of

$(a_0, a_1, b_0, b_1, r, z, \sigma^2) = (1.0, 0.8, -2.0, -0.5, 0.5, 0.3, 0.01)$

No.	$\hat{a}_0$	$\hat{a}_1$	$\hat{b}_0$	$\hat{b}_1$	$\hat{r}$	$\hat{z}$	$\hat{\sigma}^2$
1	1.1096	0.9878	−5.1574	1.6857	0.70	0.40	0.0092
2	0.9960	0.8125	−1.9878	−0.5394	0.50	0.30	0.0099
3	0.9942	0.7549	−1.8207	−0.6557	0.50	0.30	0.0091
4	1.0216	0.8338	−2.1350	−0.4016	0.50	0.30	0.0134
5	1.0570	0.8665	−7.0714	3.4880	0.80	0.40	0.0091
6	0.9902	0.7558	−1.9392	−0.4937	0.50	0.30	0.0104
7	0.9627	0.7117	−3.0465	0.5799	0.60	0.30	0.0093
8	0.9817	0.7194	−1.9620	−0.4217	0.50	0.30	0.0128
9	1.1360	1.0531	−5.3973	1.8138	0.70	0.40	0.0097
10	0.9864	0.7985	−1.9687	−0.5108	0.50	0.30	0.0128
11	1.0043	0.8360	−2.0383	−0.4777	0.50	0.30	0.0107
12	0.9937	0.7923	−2.0280	−0.4470	0.50	0.30	0.0099
13	1.0008	0.7928	−1.9873	−0.4698	0.50	0.30	0.0095
14	1.0074	0.8293	−2.0627	−0.4631	0.50	0.30	0.0099
15	1.0070	0.8492	−1.1692	−1.3483	0.40	0.30	0.0125
16	0.9794	0.7468	−1.9022	−0.5179	0.50	0.30	0.0105
17	1.0166	0.8435	−2.1056	−0.4393	0.50	0.30	0.0136
18	0.9903	0.7993	−1.9896	−0.4916	0.50	0.30	0.0104
19	0.9897	0.8033	−1.9622	−0.5447	0.50	0.30	0.0091
20	1.0060	0.8048	−2.o979	−0.3774	0.50	0.30	0.0126
21	0.9838	0.7772	−1.9245	−0.5390	0.50	0.30	0.0097
22	1.0180	0.8331	−1.2347	−1.2873	0.40	0.30	0.0133
23	1.0310	0.8853	−1.2581	−1.3137	0.40	0.30	0.0104
24	1.0576	0.9119	−7.1917	3.5452	0.80	0.40	0.0098
25	1.0113	0.8188	−2.0570	−0.4720	0.50	0.30	0.0096
26	1.0520	0.9140	−7.1423	3.4880	0.80	0.40	0.0092
27	1.0139	0.8001	−2.0825	−0.4261	0.50	0.30	0.0106
28	0.9899	0.7989	−1.9280	−0.5513	0.50	0.30	0.0111
29	1.0516	0.9016	−7.0799	3.4243	0.80	0.40	0.0124
30	1.0215	0.8312	−2.1124	−0.4274	0.50	0.30	0.0102
31	1.0315	0.8636	−1.2455	−1.3144	0.40	0.30	0.0105
32	1.0147	0.8364	−2.1889	−0.3243	0.50	0.30	0.0128
33	1.0222	0.8032	−6.6358	3.1805	0.80	0.40	0.0127
34	1.0547	0.9146	−7.2325	3.6278	0.80	0.40	0.0098
35	0.9941	0.7865	−1.8985	−0.5956	0.50	0.30	0.0129
36	1.0329	0.8768	−1.2998	−1.2804	0.40	0.30	0.0099
37	1.1242	1.0563	−5.2699	1.7021	0.70	0.40	0.0088
38	1.0009	0.8188	−1.9812	−0.5479	0.50	0.30	0.0089
39	1.1051	0.9869	−5.1442	1.6767	0.70	0.40	0.0129
40	0.9960	0.8021	−2.0202	−0.4858	0.50	0.30	0.0101
41	0.9875	0.7624	−1.9429	−0.5100	5.15	0.35	0.0092
42	0.9727	0.7646	−1.8607	−0.6024	0.50	0.30	0.0105
43	1.0040	0.8044	−2.0539	−0.4450	0.50	0.30	0.0099
44	0.9796	0.7723	−1.8481	−0.6340	0.50	0.30	0.0108
45	0.9941	0.7820	−1.8588	−0.6386	0.50	0.30	0.0108
46	0.9933	0.7933	−2.0051	−0.4769	0.50	0.30	0.0115
47	0.9862	0.7690	−1.9062	−0.5672	0.50	0.30	0.0117
48	1.0060	0.7954	−2.0597	−0.4482	0.50	0.30	0.0105
49	1.0122	0.8254	−2.0157	−0.5379	0.50	0.30	0.0090
50	1.0131	0.8062	−2.0367	−0.4737	0.50	0.30	0.0104

Table 3.6. Conditional Least Squares estimates of Model 1: True value of

$$(a_0, a_1, b_0, b_1, r, z, \sigma^2) = (1.0, 0.8, -2.0, -0.5, 0.5, 0.3, 0.01)$$

No.	$\hat{a}_0$	$\hat{a}_1$	$\hat{b}_0$	$\hat{b}_1$	$\hat{r}$	$\hat{z}$	$\hat{\sigma}^2$
1	1.1096	0.9878	-5.1574	1.6857	0.7000	0.4000	0.0093
2	0.9960	0.8125	-1.9878	-0.5393	0.5000	0.3000	0.0099
3	0.9942	0.7549	-1.8207	-0.6557	0.5000	0.3000	0.0090
4	1.0216	0.8338	-2.1349	-0.4016	0.5000	0.3000	0.0134
5	1.0015	0.7659	-1.9913	-0.4716	0.5000	0.3000	0.0090
6	0.9902	0.7558	-1.9392	-0.4937	0.5000	0.3000	0.0104
7	0.9627	0.7117	-3.0465	0.5799	0.6000	0.3000	0.0093
8	1.0084	0.7803	-1.2496	-1.1844	0.4000	0.3000	0.0128
9	1.1360	1.0531	-5.3973	1.8138	0.7000	0.4000	0.0097
10	0.9864	0.7985	-1.9687	-0.5108	0.5000	0.3000	0.0128
11	1.0043	0.8360	-2.0383	-0.4777	0.5000	0.3000	0.0107
12	0.9937	0.7923	-2.0280	-0.4470	0.5000	0.3000	0.0099
13	1.0008	0.7928	-1.9872	-0.4698	0.5000	0.3000	0.0095
14	1.0074	0.8293	-2.0627	-0.4631	0.5000	0.3000	0.0100
15	1.0070	0.8492	-1.1692	-1.3483	0.4000	0.3000	0.0125
16	0.9794	0.7468	-1.9022	-0.5179	0.5000	0.3000	0.0105
17	1.0166	0.8435	-2.1056	-0.4393	0.5000	0.3000	0.0136
18	0.9903	0.7993	-1.9896	-0.4916	0.5000	0.3000	0.0104
19	0.9897	0.8033	-1.9622	-0.5447	0.5000	0.3000	0.0091
20	1.0060	0.8048	-2.0979	-0.3774	0.5000	0.3000	0.0126
21	0.9838	0.7772	-1.9245	-0.5390	0.5000	0.3000	0.0097
22	1.0179	0.8331	-1.2347	-1.2873	0.4000	0.3000	0.0133
23	1.0310	0.8853	-1.2581	-1.3137	0.4000	0.3000	0.0104
24	1.0575	0.9119	-7.1917	3.5452	0.8000	0.4000	0.0098
25	1.0113	0.8188	-2.0570	-0.4720	0.5000	0.3000	0.0096
26	0.9932	0.8046	-1.9752	-0.5300	0.5000	0.3000	0.0092
27	1.0139	0.8001	-2.0825	-0.4261	0.5000	0.3000	0.0106
28	0.9899	0.7989	-1.9280	-0.5513	0.5000	0.3000	0.0112
29	0.9960	0.8033	-1.9903	-0.5169	0.5000	0.3000	0.0123
30	1.0215	0.8312	-2.1124	-0.4274	0.5000	0.3000	0.0102
31	1.0315	0.8636	-1.2455	-1.3144	0.4000	0.3000	0.0105
32	1.0147	0.8364	-2.1889	-0.3243	0.5000	0.3000	0.0128
33	1.0222	0.8032	-6.6358	3.1805	0.8000	0.4000	0.0127
34	1.0547	0.9146	-7.2325	3.6278	0.8000	0.4000	0.0098
35	0.9941	0.7865	-1.8985	-0.5956	0.5000	0.3000	0.0129
36	1.0329	0.8768	-1.2998	-1.2804	0.4000	0.3000	0.0100
37	1.1242	1.0563	-5.2700	1.7021	0.7000	0.4000	0.0088
38	1.0009	0.8188	-1.9812	-0.5479	0.5000	0.3000	0.0089
39	1.1051	0.9869	-5.1443	1.6767	0.7000	0.4000	0.0129
40	0.9960	0.8021	-2.0202	-0.4858	0.5000	0.3000	0.0101
41	0.9875	0.7624	-1.9429	-0.5100	0.5000	0.3000	0.0092
42	0.9727	0.7646	-1.8607	-0.6024	0.5000	0.3000	0.0105
43	1.0040	0.8044	-2.0539	-0.4450	0.5000	0.3000	0.0099
44	0.9796	0.7723	-1.8481	-0.6340	0.5000	0.3000	0.0108
45	0.9941	0.7820	-1.8588	-0.6386	0.5000	0.3000	0.0108
46	0.9933	0.7933	-2.0051	-0.4769	0.5000	0.3000	0.0115
47	0.9862	0.7690	-1.9062	-0.5671	0.5000	0.3000	0.0117
48	1.0060	0.7954	-2.0596	-0.4482	0.5000	0.3000	0.0105
49	1.0122	0.8254	-2.0157	-0.5379	0.5000	0.3000	0.0090
50	1.0131	0.8062	-2.0367	-0.4737	0.5000	0.3000	0.0104

Table 3.7. Bayesian estimates of Model 2: True value of $(a_0, a_1, b_0, b_1, r, z, \sigma^2) = (1.0, -2.0, -0.5, 2.4, 1.0, 0.5, 1.0)$

No.	$\hat{a_0}$	$\hat{a_1}$	$\hat{b_0}$	$\hat{b_1}$	$\hat{r}$	$\hat{z}$	$\hat{\sigma^2}$
1	1.0326	-2.0899	-0.4906	2.5122	0.9000	0.5500	0.9510
2	0.9397	-2.3144	-0.5256	2.7289	0.8500	0.5000	1.0582
3	1.0307	-1.9416	-0.3647	2.2759	1.1000	0.3500	1.0048
4	0.8407	-2.3106	0.0234	2.5968	0.9500	0.5500	0.9430
5	1.0484	-1.9914	-0.8172	2.4968	0.8500	0.6000	0.9307
6	1.0936	-1.9283	-0.2989	2.2216	1.0500	0.5500	1.0659
7	1.0103	-2.0342	-0.5726	2.4146	1.0000	0.4500	1.0159
8	0.9365	-1.9273	-0.4562	2.3563	1.1000	0.4500	0.9781
9	0.9610	-2.1849	-0.3497	2.5257	0.9000	0.6000	1.0959
10	1.0133	-1.8818	0.0621	2.1396	1.1000	0.5500	1.0364
11	0.9266	-1.9514	-0.0554	2.1951	1.0000	0.4500	0.8931
12	1.2075	-2.0271	-1.2173	2.5734	1.0500	0.2000	0.9413
13	1.0089	-1.9970	-1.0006	2.5146	1.0000	0.4500	1.0672
14	0.9278	-1.9803	-0.5002	2.3942	0.9500	0.5000	1.0093
15	1.0603	-2.0103	-0.3841	2.3483	1.1000	0.5000	0.9906
16	0.8544	-2.2154	-0.6974	2.7038	0.8000	0.5500	0.9245
17	0.8535	-2.0787	0.7844	2.1801	1.1500	0.8000	1.0245
18	0.9972	-2.0793	-0.1422	2.4112	1.0000	0.7000	0.9385
19	1.0004	-1.8683	-0.1100	2.1387	1.1500	0.3500	1.0555
20	0.7300	-2.3939	0.3231	2.6176	0.9000	0.6500	1.0241
21	1.1059	-1.9841	-0.7340	2.4132	1.0500	0.4000	1.0348
22	0.9429	-2.1432	-0.5889	2.5728	0.9500	0.4000	0.9282
23	1.0016	-1.9614	-0.4141	2.3443	1.0000	0.5500	1.0132
24	0.9617	-2.0570	-0.2541	2.4000	1.1000	0.4500	0.9374
25	0.9835	-2.0278	-0.6912	2.4748	1.1500	0.3500	1.0824
26	1.0468	-1.9347	-0.4629	2.3188	1.0500	0.4000	1.0326
27	0.9485	-2.0009	0.0831	2.2290	0.9500	0.8000	0.9972
28	1.0160	-2.0124	0.0947	2.2500	1.2500	0.5000	0.8954
29	1.1187	-1.8345	-0.7520	2.2713	1.1000	0.2500	1.0372
30	1.0394	-1.9945	-0.6005	2.3834	1.1000	0.4500	1.0807
31	1.0130	-2.1004	-0.7074	2.5373	0.9500	0.5500	0.9958
32	0.9580	-2.0319	-0.6807	2.5334	0.8000	0.6000	0.9968
33	0.9505	-1.9916	-0.1004	2.3134	1.1500	0.6000	0.9284
34	0.9270	-2.1151	-0.4095	2.5043	1.0000	0.5500	0.9962
35	0.9582	-2.0409	0.1348	2.2961	1.1000	0.6500	0.9700
36	1.0204	-1.9135	-0.1197	2.1949	1.1500	0.4500	1.0981
37	1.0214	-1.9639	-0.4977	2.3858	0.9500	0.6500	1.0039
38	1.0495	-1.9567	-0.7144	2.4038	1.0500	0.3500	1.0084
39	0.9699	-1.8754	-0.6349	2.2856	1.0000	0.4500	1.0350
40	1.0919	-1.9214	-0.3872	2.2515	1.1500	0.4000	0.9596
41	0.9888	-2.1409	-0.6156	2.5465	0.9000	0.5500	0.9507
42	1.1657	-1.8446	-0.6332	2.1998	1.1000	0.3500	0.9386
43	0.8761	-2.0993	-0.4267	2.5293	0.8500	0.6500	0.9718
44	0.9467	-2.1816	-0.6545	2.6597	0.9000	0.4500	0.9705
45	1.0798	-2.1143	-0.4843	2.4704	1.0000	0.5000	1.0443
46	0.9967	-2.1700	0.1670	2.4002	1.1000	0.6500	1.0415
47	0.9507	-2.0736	-0.7139	2.5627	0.8500	0.4000	1.0118
48	1.1521	-1.9367	-1.2987	2.5138	1.0500	0.2500	0.8485
49	1.2245	-1.8711	-0.6767	2.2415	1.1000	0.5500	1.0556
50	0.9425	-2.1557	-0.2135	2.5261	0.9000	0.7500	0.9814

Table 3.8. Conditional Least Squares estimates of Model 2: True value of $(a_0, a_1, b_0, b_1, r, z, \sigma^2) = (1.0, -2.0, -0.5, 2.4, 1.0, 0.5, 1.0)$

No.	$\hat{a}_0$	$\hat{a}_1$	$\hat{b}_0$	$\hat{b}_1$	$\hat{r}$	$\hat{z}$	$\hat{\sigma}^2$
1	1.0571	-2.0524	-0.5185	2.4737	0.9500	0.5000	0.9510
2	0.9483	-2.3025	-0.6128	2.7377	0.8500	0.4500	1.0581
3	1.0307	-1.9416	-0.3647	2.2759	1.1000	0.3500	1.0048
4	0.8777	-2.2604	-0.0021	2.5405	1.0000	0.5000	0.9430
5	1.0534	-1.9676	-0.8136	2.4678	0.9000	0.5500	0.9307
6	1.0936	-1.9283	-0.2989	2.2216	1.0500	0.5500	1.0659
7	1.0103	-2.0342	-0.5726	2.4146	1.0000	0.4500	1.0159
8	0.9365	-1.9273	-0.4562	2.3563	1.1000	0.4500	0.9781
9	0.9610	-2.1849	-0.3497	2.5257	0.9000	0.6000	1.0959
10	1.0133	-1.8818	0.0621	2.1396	1.1000	0.5500	1.0364
11	0.9435	-1.9257	-0.1770	2.1996	1.0000	0.4000	0.8931
12	1.2243	-1.9992	-1.1630	2.5247	1.1000	0.2000	0.9413
13	1.0139	-1.9906	-1.0703	2.5244	1.0000	0.4000	1.0672
14	0.9420	-1.9631	-0.5984	2.3994	0.9500	0.4500	1.0091
15	1.0603	-2.0103	-0.3841	2.3483	1.1000	0.5000	0.9906
16	0.8623	-2.2024	-0.7839	2.7114	0.8000	0.5000	0.9244
17	0.8857	-2.0506	0.4400	2.2300	1.1000	0.7500	1.0245
18	1.0199	-2.0540	-0.2699	2.4095	1.0000	0.6500	0.9384
19	1.0004	-1.8683	-0.1100	2.1387	1.1500	0.3500	1.0555
20	0.7300	-2.3939	0.3231	2.6176	0.9000	0.6500	1.0241
21	1.1059	-1.9841	-0.7340	2.4132	1.0500	0.4000	1.0348
22	0.9429	-2.1432	-0.5889	2.5728	0.9500	0.4000	0.9282
23	1.0016	-1.9614	-0.4141	2.3443	1.0000	0.5500	1.0132
24	0.9617	-2.0570	-0.2541	2.4000	1.1000	0.4500	0.9374
25	0.9835	-2.0278	-0.6912	2.4748	1.1500	0.3500	1.0824
26	1.0593	-1.9178	-0.5651	2.3274	1.0500	0.3500	1.0326
27	0.9730	-1.9756	-0.0395	2.2245	0.9500	0.7500	0.9971
28	1.0377	-1.9824	-0.0766	2.2593	1.2500	0.4500	0.8953
29	1.1187	-1.8345	-0.7520	2.2713	1.1000	0.2500	1.0372
30	1.0394	-1.9945	-0.6005	2.3834	1.1000	0.4500	1.0807
31	1.0130	-2.1004	-0.7074	2.5373	0.9500	0.5500	0.9958
32	0.9580	-2.0319	-0.6807	2.5334	0.8000	0.6000	0.9968
33	0.9505	-1.9916	-0.1004	2.3134	1.1500	0.6000	0.9284
34	0.9270	-2.1151	-0.4095	2.5043	1.0000	0.5500	0.9962
35	0.9808	-2.0098	-0.0110	2.2936	1.1000	0.6000	0.9699
36	1.0204	-1.9135	-0.1197	2.1949	1.1500	0.4500	1.0981
37	1.0352	-1.9459	-0.5967	2.3875	0.9500	0.6000	1.0037
38	1.0495	-1.9567	-0.7144	2.4038	1.0500	0.3500	1.0084
39	0.9699	-1.8754	-0.6349	2.2856	1.0000	0.4500	1.0350
40	1.0919	-1.9214	-0.3872	2.2515	1.1500	0.4000	0.9596
41	0.9888	-2.1409	-0.6156	2.5465	0.9000	0.5500	0.9507
42	1.1657	-1.8446	-0.6332	2.1998	1.1000	0.3500	0.9386
43	0.8946	-2.0668	-0.4430	2.4938	0.9000	0.6000	0.9717
44	0.9467	-2.1816	-0.6545	2.6597	0.9000	0.4500	0.9705
45	1.0916	-2.1023	-0.5757	2.4780	1.0000	0.4500	1.0442
46	0.9967	-2.1700	0.1670	2.4002	1.1000	0.6500	1.0415
47	0.9507	-2.0736	-0.7139	2.5627	0.8500	0.4000	1.0118
48	1.1521	-1.9367	-1.2987	2.5138	1.0500	0.2500	0.8485
49	1.2402	-1.8410	-0.6660	2.2011	1.1500	0.5000	1.0554
50	0.9633	-2.1302	-0.3301	2.5219	0.9000	0.7000	0.9813

Table 3.9. Bayesian estimates of Model 3: True value of $(a_0, a_1, b_0, b_1, r, z, \sigma^2) = (2.0, 0.5, -0.5, -1.4, 1.0, 0.5, 1.0)$

No.	$\hat{a}_0$	$\hat{a}_1$	$\hat{b}_0$	$\hat{b}_1$	$\hat{r}$	$\hat{z}$	$\hat{\sigma}^2$
1	1.9863	0.4359	-0.6810	-1.2660	1.1500	0.4500	0.9899
2	2.0895	0.6353	-1.2740	-1.2338	1.0500	0.5500	0.9751
3	2.0175	0.4510	-0.5102	-1.3655	1.0500	0.4000	1.0423
4	1.7172	0.3498	1.0646	-1.7967	0.2500	0.5500	0.9844
5	2.0922	0.6123	-0.6205	-1.5268	0.8500	0.6500	1.0075
6	1.8927	0.4465	-0.3466	-1.4072	1.0500	0.4000	0.9745
7	2.1089	0.5125	-1.2176	-1.2301	1.0500	0.6000	1.0342
8	1.9445	0.5028	-0.5341	-1.3714	1.1000	0.3500	0.9940
9	2.0123	0.5474	-0.4189	-1.4804	0.8000	0.6000	1.0556
10	2.1732	0.5904	-0.4657	-1.5912	0.8000	0.6000	1.1539
11	1.9802	0.4910	-0.0192	-1.4930	0.8500	0.4000	0.9520
12	1.8971	0.4957	-0.5309	-1.3426	1.1000	0.3500	1.0487
13	1.8278	0.3929	-0.0143	-1.3993	1.0000	0.2500	1.0221
14	2.2087	0.7020	-1.2154	-1.4686	0.9000	0.7500	0.9443
15	1.8423	0.5612	-0.3188	-1.4451	1.1000	0.3500	1.0022
16	1.9857	0.5376	-0.6681	-1.4027	1.2000	0.5000	1.0391
17	2.2226	0.5652	-1.9672	-1.1064	1.0500	0.9000	0.9655
18	2.0511	0.5342	-0.9969	-1.2872	1.0500	0.5000	0.9650
19	2.0695	0.4805	-1.4618	-1.0776	1.1500	0.6000	0.9952
20	2.0684	0.5141	-0.8233	-1.3541	1.0500	0.5500	1.0453
21	1.8654	0.3187	-0.0405	-1.3554	1.1000	0.3000	0.9701
22	2.0748	0.5255	-1.4387	-1.1224	1.1000	0.5000	1.0172
23	2.0741	0.5385	-0.4572	-1.4965	0.9000	0.6500	0.9842
24	2.0120	0.3811	-1.2214	-1.0785	1.2000	0.7500	1.1359
25	1.9376	0.5006	-0.4688	-1.4018	1.1000	0.3500	0.9094
26	2.0732	0.5423	-1.5766	-1.1937	1.2000	0.7500	0.9652
27	1.8108	0.4053	-0.9613	-1.0117	1.2000	0.4000	1.0373
28	1.8244	0.4686	0.2950	-1.6285	0.9500	0.4500	0.9543
29	2.2164	0.7104	-0.6460	-1.6554	0.8500	0.4000	1.0265
30	2.2208	0.5880	-1.2556	-1.3395	0.9500	0.7500	1.0015
31	1.8927	0.5575	-0.7147	-1.3353	1.0000	0.5000	1.0053
32	2.2157	0.6053	-0.3504	-1.6550	0.9000	0.4500	0.9640
33	1.7289	0.2920	-0.8041	-0.9899	1.3000	0.4000	1.0427
34	1.8423	0.4882	0.4195	-1.6846	0.9000	0.2000	1.0193
35	2.0927	0.5201	-0.0602	-1.6064	0.9500	0.4000	0.9727
36	1.9755	0.3887	-0.6802	-1.1716	1.0000	0.5500	1.0373
37	1.9720	0.3812	-0.5568	-1.2698	1.0500	0.6500	0.9760
38	1.7766	0.4154	-4.2049	-0.1151	1.7000	0.7500	0.9024
39	1.8546	0.4874	-0.1440	-1.4633	0.8500	0.6500	1.0761
40	2.0096	0.4751	0.0145	-1.5190	0.9500	0.4000	1.0189
41	2.1864	0.5622	-0.3547	-1.5897	0.8500	0.6000	1.0342
42	1.9226	0.6463	0.2727	-1.8014	0.6500	0.5500	1.0087
43	1.9930	0.4538	-0.1366	-1.4705	0.7500	0.4000	0.9112
44	1.8764	0.3915	-0.6818	-1.1702	1.1500	0.4000	1.0249
45	1.8856	0.5207	-1.3032	-1.1415	1.2000	0.5000	1.0622
46	1.9793	0.6217	-0.6781	-1.4292	1.1000	0.4500	0.9984
47	1.7725	0.4059	-0.5152	-1.2610	1.2000	0.2500	0.9895
48	1.9499	0.4464	0.1936	-1.5151	1.0000	0.2000	1.0172
49	2.0036	0.4677	-0.0535	-1.5479	0.9000	0.4500	0.9984
50	3.0865	1.1130	-44.1495	6.5650	3.9500	2.0500	0.9787

Table 3.10. Conditional Least Squares estimates of Model 3: True value of

$(a_0, a_1, b_0, b_1, r, z, \sigma^2) = (2.0, 0.5, -0.5, -1.4, 1.0, 0.5, 1.0)$

No.	$\hat{a}_0$	$\hat{a}_1$	$\hat{b}_0$	$\hat{b}_1$	$\hat{r}$	$\hat{z}$	$\hat{\sigma}^2$
1	1.9651	0.4216	-0.5520	-1.2847	1.1500	0.4000	0.9898
2	2.0895	0.6353	-1.2740	-1.2338	1.0500	0.5500	0.9751
3	2.0175	0.4510	-0.5102	-1.3655	1.0500	0.4000	1.0423
4	1.7043	0.3418	1.0897	-1.7921	0.2000	0.5000	0.9844
5	2.0564	0.5883	-0.5944	-1.4958	0.9000	0.6000	1.0074
6	1.8927	0.4465	-0.3466	-1.4072	1.0500	0.4000	0.9745
7	2.0789	0.4943	-1.0679	-1.2462	1.0500	0.5500	1.0342
8	1.9445	0.5028	-0.5341	-1.3714	1.1000	0.3500	0.9940
9	2.0123	0.5474	-0.4189	-1.4804	0.8000	0.6000	1.0556
10	2.1208	0.5597	-0.3302	-1.5838	0.8500	0.5000	1.1538
11	1.9682	0.4832	0.2214	-1.5642	0.8000	0.2500	0.9516
12	1.8797	0.4830	-0.4059	-1.3647	1.1000	0.3000	1.0486
13	1.8278	0.3929	-0.0143	-1.3993	1.0000	0.2500	1.0221
14	2.1804	0.6866	-0.9339	-1.5322	0.8500	0.7000	0.9441
15	1.8423	0.5612	-0.3188	-1.4451	1.1000	0.3500	1.0022
16	1.9590	0.5194	-0.5117	-1.4232	1.2000	0.4500	1.0389
17	2.1862	0.5456	-1.6539	-1.1635	1.0000	0.8500	0.9653
18	2.0511	0.5342	-0.9969	-1.2872	1.0500	0.5000	0.9650
19	2.0366	0.4583	-1.2932	-1.0946	1.1500	0.5500	0.9951
20	2.0598	0.5103	-0.5810	-1.4229	1.0000	0.5000	1.0451
21	1.8654	0.3187	-0.0405	-1.3554	1.1000	0.3000	0.9701
22	2.0748	0.5255	-1.4387	-1.1224	1.1000	0.5000	1.0172
23	2.0741	0.5385	-0.4572	-1.4965	0.9000	0.6500	0.9842
24	1.9828	0.3632	-1.0883	-1.0857	1.2000	0.7000	1.1358
25	1.9376	0.5006	-0.4688	-1.4018	1.1000	0.3500	0.9094
26	2.0478	0.5284	-1.2552	-1.2620	1.1500	0.7000	0.9651
27	1.8108	0.4053	-0.9613	-1.0117	1.2000	0.4000	1.0373
28	1.8244	0.4686	0.2950	-1.6285	0.9500	0.4500	0.9543
29	2.2164	0.7104	-0.6460	-1.6554	0.8500	0.4000	1.0265
30	2.1851	0.5673	-1.1040	-1.3505	0.9500	0.7000	1.0014
31	1.8695	0.5412	-0.5902	-1.3504	1.0000	0.4500	1.0052
32	2.1949	0.5925	-0.2431	-1.6701	0.9000	0.4000	0.9639
33	1.7289	0.2920	-0.8041	-0.9899	1.3000	0.4000	1.0427
34	1.8423	0.4882	0.4195	-1.6846	0.9000	0.2000	1.0193
35	2.0927	0.5201	-0.0602	-1.6064	0.9500	0.4000	0.9727
36	1.9506	0.3712	-0.5563	-1.1845	1.0000	0.5000	1.0371
37	1.8756	0.3189	0.4733	-1.5142	0.9000	0.2000	0.9746
38	1.7628	0.4081	-3.3712	-0.3377	1.6000	0.7000	0.9021
39	1.8384	0.4779	0.0360	-1.5065	0.8000	0.6000	1.0760
40	2.0096	0.4751	0.0145	-1.5190	0.9500	0.4000	1.0189
41	2.1682	0.5513	-0.2570	-1.6023	0.8500	0.5500	1.0340
42	1.9216	0.6441	0.2090	-1.7764	0.7000	0.5500	1.0087
43	1.9930	0.4538	-0.1366	-1.4705	0.7500	0.4000	0.9112
44	1.8764	0.3915	-0.6818	-1.1702	1.1500	0.4000	1.0249
45	1.8742	0.5152	-0.9846	-1.2371	1.1500	0.4500	1.0620
46	1.9793	0.6217	-0.6781	-1.4292	1.1000	0.4500	0.9984
47	1.7725	0.4059	-0.5152	-1.2610	1.2000	0.2500	0.9895
48	1.9499	0.4464	0.1936	-1.5151	1.0000	0.2000	1.0172
49	2.0036	0.4677	-0.0535	-1.5479	0.9000	0.4500	0.9984
50	2.2511	0.7537	-4.2252	-0.6946	1.5000	1.0500	0.9730

Table 3.11. Bayesian estimates of Model 4: True value of

$(a_0, a_1, b_0, b_1, r, z, \sigma^2) = (5.0, 0.8, -0.5, -5.8, 1.0, 0.5, 1.0)$

No.	$\hat{a}_0$	$\hat{a}_1$	$\hat{b}_0$	$\hat{b}_1$	$\hat{r}$	$\hat{z}$	$\hat{\sigma}^2$
1	4.9731	0.7913	-0.8044	-5.6874	1.1500	0.5000	1.1111
2	4.9270	0.7923	-0.5976	-5.7188	1.0000	0.5000	1.2120
3	5.0267	0.8052	-0.9788	-5.7227	1.0000	0.6000	1.0158
4	4.9607	0.8060	-0.2053	-5.8851	0.9500	0.3000	1.3748
5	4.9623	0.7959	-0.1105	-5.8831	0.9500	0.5000	1.0192
6	4.8292	0.7935	-0.1812	-5.8324	1.0500	0.4000	1.1610
7	5.0154	0.7979	-0.9962	-5.6916	1.1000	0.4000	0.9516
8	4.9717	0.8005	-0.4246	-5.8046	0.9500	0.4000	1.1243
9	4.9676	0.8040	-0.8980	-5.7576	1.0500	0.5000	1.1889
10	5.0770	0.8119	-1.4303	-5.6443	1.0500	0.6000	1.1114
11	4.8893	0.8023	-0.0943	-5.8834	1.1000	0.3500	1.0992
12	5.0863	0.8033	-0.7395	-5.7137	1.0500	0.5000	1.0107
13	5.0422	0.7965	0.0256	-5.9349	0.8500	0.6000	1.0275
14	5.1718	0.8122	-0.1658	-5.9364	0.9000	0.4500	1.0829
15	4.9923	0.7962	-2.3380	-5.3060	1.1000	0.6000	2.0711
16	4.9249	0.7939	-0.7166	-5.7225	1.0500	0.5000	1.0507
17	5.0483	0.8013	-0.6238	-5.7776	1.0000	0.4500	1.0528
18	5.0135	0.7979	-0.6622	-5.7441	1.0000	0.4500	1.0792
19	5.0887	0.8091	-0.7047	-5.7751	1.0500	0.4500	0.9760
20	4.9891	0.8016	-0.4061	-5.7930	0.9500	0.5500	1.0032
21	4.9384	0.8039	-0.8504	-5.7131	1.0000	0.5500	1.2896
22	5.0371	0.7998	-0.6138	-5.7961	1.0000	0.5500	1.0406
23	4.8476	0.7878	0.2361	-5.9211	0.9500	0.5500	0.9925
24	4.8363	0.7837	0.1809	-5.9475	1.0000	0.4500	1.1559
25	4.8876	0.8046	-0.8665	-5.7048	1.1000	0.4500	0.8822
26	4.9970	0.7966	-1.2542	-5.5917	1.0500	0.5000	1.0860
27	4.9282	0.7934	0.0285	-5.8576	1.0000	0.4000	1.5621
28	5.0817	0.7927	-0.5041	-5.7935	1.0500	0.4000	1.4928
29	4.9418	0.8001	-0.9515	-5.6572	1.1500	0.5000	1.6361
30	5.1183	0.8071	-0.0382	-5.9105	0.9500	0.5500	0.9534
31	4.9115	0.7927	-0.1105	-5.8854	1.0000	0.4500	1.0205
32	5.1364	0.8071	-0.8395	-5.7556	0.9500	0.4500	1.5874
33	5.1853	0.8163	-0.6666	-5.8479	1.0000	0.4500	0.9791
34	4.9320	0.7993	-0.3194	-5.8248	1.1000	0.4000	1.0894
35	4.7646	0.7827	-0.0880	-5.8335	1.0500	0.5000	1.8254
36	5.0642	0.8046	-0.8658	-5.6938	1.0000	0.4500	0.9120
37	4.9664	0.8030	-0.4001	-5.8349	1.0500	0.4500	0.9855
38	4.8914	0.7928	-1.0826	-5.6007	1.0500	0.5000	1.3045
39	4.9593	0.7969	-0.0807	-5.8689	0.9000	0.4500	1.0138
40	5.0277	0.8003	-0.7822	-5.7008	1.1500	0.4000	0.9971
41	5.0648	0.7993	-0.9471	-5.7326	1.1000	0.5000	1.1391
42	4.9124	0.7837	-0.2903	-5.8007	1.0000	0.4000	0.8945
43	5.0454	0.8060	-0.3577	-5.8584	0.9500	0.4500	1.0299
44	4.8663	0.7943	-0.6892	-5.7215	1.1000	0.4500	0.9946
45	4.9305	0.7893	-0.9315	-5.6734	1.0500	0.6000	1.0139
46	4.9139	0.7934	-0.6261	-5.7726	1.0500	0.5500	0.9638
47	4.8476	0.7864	-0.4566	-5.7405	1.0500	0.4500	1.0610
48	5.0223	0.8010	-0.5920	-5.8064	1.0000	0.4500	1.0025
49	5.1032	0.8016	-0.6979	-5.7692	0.9000	0.5000	0.9639
50	5.0033	0.8080	-0.5631	-5.7926	1.0000	0.5500	1.0846

Table 3.12. Conditional Least Squares estimates of Model 4: True value of

$(a_0, a_1, b_0, b_1, r, z, \sigma^2) = (5.0, 0.8, -0.5, -5.8, 1.0, 0.5, 1.0)$

No.	$\hat{a_0}$	$\hat{a_1}$	$\hat{b_0}$	$\hat{b_1}$	$\hat{r}$	$\hat{z}$	$\hat{\sigma^2}$
1	4.9731	0.7913	-0.8044	-5.6874	1.1500	0.5000	1.1111
2	4.9270	0.7923	-0.5976	-5.7188	1.0000	0.5000	1.2120
3	5.0267	0.8052	-0.9788	-5.7227	1.0000	0.6000	1.0158
4	4.9607	0.8060	-0.2053	-5.8851	0.9500	0.3000	1.3748
5	4.9623	0.7959	-0.1105	-5.8831	0.9500	0.5000	1.0192
6	4.8292	0.7935	-0.1812	-5.8324	1.0500	0.4000	1.1610
7	5.0154	0.7979	-0.9962	-5.6916	1.1000	0.4000	0.9516
8	4.9717	0.8005	-0.4246	-5.8046	0.9500	0.4000	1.1243
9	4.9676	0.8040	-0.8980	-5.7576	1.0500	0.5000	1.1889
10	5.0770	0.8119	-1.4303	-5.6443	1.0500	0.6000	1.1114
11	4.8893	0.8023	-0.0943	-5.8834	1.1000	0.3500	1.0992
12	5.0863	0.8033	-0.7395	-5.7137	1.0500	0.5000	1.0107
13	5.0422	0.7965	0.0256	-5.9349	0.8500	0.6000	1.0275
14	5.1718	0.8122	-0.1658	-5.9364	0.9000	0.4500	1.0829
15	4.9923	0.7962	-2.3380	-5.3060	1.1000	0.6000	2.0711
16	4.9249	0.7939	-0.7166	-5.7225	1.0500	0.5000	1.0507
17	5.0483	0.8013	-0.6238	-5.7776	1.0000	0.4500	1.0528
18	5.0135	0.7979	-0.6622	-5.7441	1.0000	0.4500	1.0792
19	5.0887	0.8091	-0.7047	-5.7751	1.0500	0.4500	0.9760
20	4.9891	0.8016	-0.4061	-5.7930	0.9500	0.5500	1.0032
21	4.9384	0.8039	-0.8504	-5.7131	1.0000	0.5500	1.2896
22	5.0371	0.7998	-0.6138	-5.7961	1.0000	0.5500	1.0406
23	4.8476	0.7878	0.2361	-5.9211	0.9500	0.5500	0.9925
24	4.8363	0.7837	0.1809	-5.9475	1.0000	0.4500	1.1559
25	4.8876	0.8046	-0.8665	-5.7048	1.1000	0.4500	0.8822
26	4.9970	0.7966	-1.2542	-5.5917	1.0500	0.5000	1.0860
27	4.9282	0.7934	0.0285	-5.8576	1.0000	0.4000	1.5621
28	5.0817	0.7927	-0.5041	-5.7935	1.0500	0.4000	1.4928
29	4.9619	0.8015	-0.7189	-5.7135	1.1000	0.5000	1.6360
30	5.1183	0.8071	-0.0382	-5.9105	0.9500	0.5500	0.9534
31	4.9115	0.7927	-0.1105	-5.8854	1.0000	0.4500	1.0205
32	5.1364	0.8071	-0.8395	-5.7556	0.9500	0.4500	1.5874
33	5.1853	0.8163	-0.6666	-5.8479	1.0000	0.4500	0.9791
34	4.9320	0.7993	-0.3194	-5.8248	1.1000	0.4000	1.0894
35	4.7646	0.7827	-0.0880	-5.8335	1.0500	0.5000	1.8254
36	5.0642	0.8046	-0.8658	-5.6938	1.0000	0.4500	0.9120
37	4.9664	0.8030	-0.4001	-5.8349	1.0500	0.4500	0.9855
38	4.8914	0.7928	-1.0826	-5.6007	1.0500	0.5000	1.3045
39	4.9593	0.7969	-0.0807	-5.8689	0.9000	0.4500	1.0138
40	5.0277	0.8003	-0.7822	-5.7008	1.1500	0.4000	0.9971
41	5.0648	0.7993	-0.9471	-5.7326	1.1000	0.5000	1.1391
42	4.9124	0.7837	-0.2903	-5.8007	1.0000	0.4000	0.8945
43	5.0454	0.8060	-0.3577	-5.8584	0.9500	0.4500	1.0299
44	4.8663	0.7943	-0.6892	-5.7215	1.1000	0.4500	0.9946
45	4.9305	0.7893	-0.9315	-5.6734	1.0500	0.6000	1.0139
46	4.9139	0.7934	-0.6261	-5.7726	1.0500	0.5500	0.9638
47	4.8656	0.7877	-0.2563	-5.7914	1.0000	0.4500	1.0610
48	5.0223	0.8010	-0.5920	-5.8064	1.0000	0.4500	1.0025
49	5.1032	0.8016	-0.6979	-5.7692	0.9000	0.5000	0.9639
50	5.0033	0.8080	-0.5631	-5.7926	1.0000	0.5500	1.0846

Table 3.13. Bayesian estimates of Model 5: True value of the parameter

$(a_0, a_1, b_0, b_1, r, z, \sigma^2) = (1.0, -10.0, -5.0, 10.5, 1.0, 0.5, 1.0)$

No.	$\hat{a}_0$	$\hat{a}_1$	$\hat{b}_0$	$\hat{b}_1$	$\hat{r}$	$\hat{z}$	$\hat{\sigma}^2$
1	0.8838	-10.0647	-4.8344	10.5593	0.9000	0.5500	1.0079
2	0.8298	-10.0179	-4.7827	10.5150	1.1000	0.4000	1.0787
3	1.1938	-9.9866	-5.1797	10.4898	1.0500	0.5000	0.9403
4	1.5600	-9.8830	-5.5791	10.3809	0.6500	0.8500	3.6639
5	1.0175	-9.9700	-4.9473	10.4673	1.0000	0.5000	1.0365
6	1.2300	-9.9315	-5.1929	10.4305	1.0500	0.4000	0.9890
7	0.8411	-10.0432	-4.9024	10.5449	1.0500	0.4500	1.0114
8	0.9298	-10.0168	-4.8876	10.5168	1.0000	0.4500	1.0866
9	1.1342	-10.0037	-5.0504	10.5019	1.0000	0.5500	1.0729
10	0.8989	-10.0362	-4.9844	10.5374	1.0000	0.5000	1.0905
11	0.7969	-10.0206	-4.7502	10.5210	0.9500	0.5500	0.8875
12	1.1859	-10.0123	-5.2402	10.5124	0.8500	0.5000	0.9987
13	1.2622	-9.7790	-5.2649	10.2771	1.0500	0.5000	6.8745
14	1.0924	-9.9783	-4.8731	10.4720	1.0000	0.6000	1.0469
15	0.9602	-9.9958	-4.8291	10.4929	1.0500	0.5000	1.5699
16	0.9459	-10.0152	-4.9260	10.5175	1.0000	0.5500	0.9765
17	0.8707	-10.0514	-4.9045	10.5485	1.0000	0.4500	0.9583
18	1.2071	-9.7902	-5.1455	10.2896	1.0500	0.5000	6.6542
19	1.0030	-9.9470	-5.0121	10.4499	1.0500	0.4000	1.0572
20	0.8468	-10.0039	-4.7698	10.5030	1.1000	0.3500	1.4497
21	0.9149	-10.0253	-4.9582	10.5273	1.0000	0.5000	1.1582
22	0.9260	-10.0129	-4.8328	10.5115	0.9500	0.6000	0.9992
23	1.0091	-9.9929	-5.0184	10.4954	1.0500	0.4500	1.0824
24	1.1123	-9.9795	-5.2432	10.4815	0.9500	0.4500	1.0039
25	0.8806	-10.0257	-5.0181	10.5316	1.0000	0.4500	1.1270
26	1.0452	-9.9596	-4.9788	10.4598	1.0000	0.4000	0.9409
27	1.0314	-9.9453	-5.0911	10.4461	0.8500	0.5500	1.0694
28	0.8491	-10.0102	-4.8804	10.5121	0.9500	0.5000	1.8243
29	0.9265	-10.0560	-4.7674	10.5522	1.0000	0.5500	1.0338
30	0.9787	-10.0318	-5.1042	10.5325	0.9000	0.5000	0.9114
31	1.2043	-9.9375	-5.0361	10.4386	1.1000	0.5000	9.8254
32	0.7717	-10.0561	-4.7114	10.5532	1.0500	0.4500	1.0202
33	0.8003	-10.0471	-4.6422	10.5415	0.9500	0.5500	0.9527
34	0.9791	-10.0072	-5.1286	10.5106	0.9500	0.5000	0.8854
35	1.0663	-9.9751	-5.0908	10.4746	0.9000	0.5500	0.9709
36	0.9902	-10.0230	-4.7684	10.5186	1.0000	0.5000	4.0711
37	0.9693	-10.0210	-4.9549	10.5211	1.0000	0.5000	1.3374
38	0.9020	-10.0127	-4.8284	10.5049	1.0000	0.4500	0.9259
39	0.7840	-10.0276	-4.7190	10.5249	1.0500	0.4500	1.0128
40	1.1086	-10.0063	-5.0733	10.5058	1.0000	0.4500	0.8480
41	0.9792	-9.9987	-4.8357	10.4912	1.0000	0.5500	0.8841
42	0.8683	-10.0590	-4.4878	10.5496	1.1000	0.4500	7.0901
43	0.9726	-9.9867	-5.0231	10.4897	0.9000	0.6500	1.0268
44	0.8018	-10.0488	-4.5564	10.5412	1.0500	0.5000	2.7167
45	0.9488	-9.9963	-4.3359	10.4814	1.0000	0.5500	8.9981
46	0.8602	-10.0253	-4.8710	10.5256	1.0000	0.4500	0.8479
47	0.8885	-10.0222	-4.9626	10.5261	1.0500	0.4000	0.8951
48	1.1853	-9.9749	-5.1074	10.4733	1.0500	0.4500	1.0086
49	0.9086	-9.9790	-4.8107	10.4762	1.0500	0.4500	1.0032
50	0.9178	-10.0100	-5.0255	10.5110	1.0500	0.4500	1.0424

Table 3.14. Conditional Least Squares estimates of Model 5: True value of $(a_0, a_1, b_0, b_1, r, z, \sigma^2) = (1.0, -10.0, -5.0, 10.5, 1.0, 0.5, 1.0)$

No.	$\hat{a}_0$	$\hat{a}_1$	$\hat{b}_0$	$\hat{b}_1$	$\hat{r}$	$\hat{z}$	$\hat{\sigma}^2$
1	0.8838	-10.0647	-4.8344	10.5593	0.9000	0.5500	1.0079
2	0.8298	-10.0179	-4.7827	10.5150	1.1000	0.4000	1.0787
3	1.1938	-9.9866	-5.1797	10.4898	1.0500	0.5000	0.9403
4	1.5600	-9.8830	-5.5791	10.3809	0.6500	0.8500	3.6639
5	1.0175	-9.9700	-4.9473	10.4673	1.0000	0.5000	1.0365
6	1.2300	-9.9315	-5.1929	10.4305	1.0500	0.4000	0.9890
7	0.8411	-10.0432	-4.9024	10.5449	1.0500	0.4500	1.0114
8	0.9298	-10.0168	-4.8876	10.5168	1.0000	0.4500	1.0866
9	1.1342	-10.0037	-5.0504	10.5019	1.0000	0.5500	1.0729
10	0.8989	-10.0362	-4.9844	10.5374	1.0000	0.5000	1.0905
11	0.7969	-10.0206	-4.7502	10.5210	0.9500	0.5500	0.8875
12	1.1859	-10.0123	-5.2402	10.5124	0.8500	0.5000	0.9987
13	1.2622	-9.7790	-5.2649	10.2771	1.0500	0.5000	6.8745
14	1.0924	-9.9783	-4.8731	10.4720	1.0000	0.6000	1.0469
15	0.9602	-9.9958	-4.8291	10.4929	1.0500	0.5000	1.5699
16	0.9459	-10.0152	-4.9260	10.5175	1.0000	0.5500	0.9765
17	0.8707	-10.0514	-4.9045	10.5485	1.0000	0.4500	0.9583
18	1.2146	-9.7876	-5.1416	10.2867	1.1000	0.4500	6.6541
19	1.0030	-9.9470	-5.0121	10.4499	1.0500	0.4000	1.0572
20	0.8468	-10.0039	-4.7698	10.5030	1.1000	0.3500	1.4497
21	0.9149	-10.0253	-4.9582	10.5273	1.0000	0.5000	1.1582
22	0.9260	-10.0129	-4.8328	10.5115	0.9500	0.6000	0.9992
23	1.0091	-9.9929	-5.0184	10.4954	1.0500	0.4500	1.0824
24	1.1123	-9.9795	-5.2432	10.4815	0.9500	0.4500	1.0039
25	0.8746	-10.0274	-5.0398	10.5341	1.0000	0.4000	1.1270
26	1.0452	-9.9596	-4.9788	10.4598	1.0000	0.4000	0.9409
27	1.0314	-9.9453	-5.0911	10.4461	0.8500	0.5500	1.0694
28	0.8491	-10.0102	-4.8804	10.5121	0.9500	0.5000	1.8243
29	0.9265	-10.0560	-4.7674	10.5522	1.0000	0.5500	1.0338
30	0.9787	-10.0318	-5.1042	10.5325	0.9000	0.5000	0.9114
31	1.1909	-9.9411	-5.0493	10.4430	1.1000	0.4500	9.8253
32	0.7717	-10.0561	-4.7114	10.5532	1.0500	0.4500	1.0202
33	0.8003	-10.0471	-4.6422	10.5415	0.9500	0.5500	0.9527
34	0.9791	-10.0072	-5.1286	10.5106	0.9500	0.5000	0.8854
35	1.0663	-9.9751	-5.0908	10.4746	0.9000	0.5500	0.9709
36	0.9902	-10.0230	-4.7684	10.5186	1.0000	0.5000	4.0711
37	0.9693	-10.0210	-4.9549	10.5211	1.0000	0.5000	1.3374
38	0.9020	-10.0127	-4.8284	10.5049	1.0000	0.4500	0.9259
39	0.7840	-10.0276	-4.7190	10.5249	1.0500	0.4500	1.0128
40	1.1086	-10.0063	-5.0733	10.5058	1.0000	0.4500	0.8480
41	0.9792	-9.9987	-4.8357	10.4912	1.0000	0.5500	0.8841
42	0.8683	-10.0590	-4.4878	10.5496	1.1000	0.4500	7.0901
43	0.9726	-9.9867	-5.0231	10.4897	0.9000	0.6500	1.0268
44	0.8018	-10.0488	-4.5564	10.5412	1.0500	0.5000	2.7167
45	0.9488	-9.9963	-4.3359	10.4814	1.0000	0.5500	8.9981
46	0.8602	-10.0253	-4.8710	10.5256	1.0000	0.4500	0.8479
47	0.8885	-10.0222	-4.9626	10.5261	1.0500	0.4000	0.8951
48	1.1853	-9.9749	-5.1074	10.4733	1.0500	0.4500	1.0086
49	0.9086	-9.9790	-4.8107	10.4762	1.0500	0.4500	1.0032
50	0.9178	-10.0100	-5.0255	10.5110	1.0500	0.4500	1.0424

Table 3.15. Average values of estimates in Model 1 to 5
with the standard error of estimates of
B=Bayesian, C=Conditional Least Squares

		$\hat{a}_0$	$\hat{a}_1$	$\hat{b}_0$	$\hat{b}_1$	$\hat{r}$	$\hat{z}$	$\hat{\sigma}^2$
1	B	1.0157	0.8297	-2.8069	0.0943	0.5440	0.3200	0.0107
		(0.0379)	(0.0737)	(1.8492)	(1.4345)	(0.1163)	(0.0404)	(0.0010)
	C	1.0129	0.8247	-2.4859	-0.1594	0.5240	0.3140	0.0107
		(0.0366)	(0.0709)	(1.5045)	(1.1602)	(0.0981)	(0.0351)	(0.0010)
2	B	0.9984	-2.0331	-0.4155	2.4072	1.0130	0.5040	0.9959
		(0.0918)	(0.1231)	(0.3820)	(0.1519)	(0.1068)	(0.1324)	(0.0564)
	C	1.0055	-2.0233	-0.4512	2.4042	1.0180	0.4850	0.9959
		(0.0903)	(0.1209)	(0.3494)	(0.1488)	(0.1034)	(0.1226)	(0.0564)
3	B	2.0063	0.5115	-1.5056	-1.2020	1.0700	0.5280	1.0046
		(0.2053)	(0.1261)	(6.2016)	(1.1516)	(0.4615)	(0.2686)	(0.0484)
	C	1.9769	0.4965	-0.6000	-1.3679	1.0110	0.4740	1.0044
		(0.1344)	(0.0989)	(0.8567)	(0.2598)	(0.2081)	(0.1682)	(0.0484)
4	B	4.9824	0.7988	-0.5814	-5.7724	1.0200	0.4780	1.1335
		(0.0919)	(0.0075)	(0.4534)	(0.1101)	(0.0678)	(0.0679)	(0.2413)
	C	4.9831	0.7989	-0.5728	-5.7746	1.0180	0.4780	1.1335
		(0.0913)	(0.0075)	(0.4527)	(0.1091)	(0.0661)	(0.0679)	(0.2413)
5	B	0.9677	-10.0008	-4.9520	10.4997	0.9960	0.5270	0.9856
		(0.2119)	(0.0649)	(0.2726)	(0.0657)	(0.8970)	(0.1170)	(0.1018)
	C	0.9692	-10.0003	-4.9531	10.4992	1.0000	0.5240	0.9856
		(0.2113)	(0.0644)	(0.2725)	(0.0651)	(0.0904)	(0.1179)	(0.1018)

simulations are shown in Tables 3.5-3.14, and the average of the estimates of each parameter, averaged over the 50 samples, are presented in Table 3.15. The standard deviations are in parantheses. It is clear from the simulations that the the Bayesian estimates are once again comparable to the CLS estimates.

The above demonstrates that the Bayesian method is possible to implement and its performance is comparable to the conventional CLS method. However, one of the main advantages of the Bayesian method is the relative ease in which one can obtain the marginal posterior density of the implicit parameters (r, d, z) in a closed form. Also, in the Bayesian method as mentioned by Geweke and Terui (1993) the computation of multi-step-ahead predictions is no more difficult than the one-step-ahead prediction.

3.4.2 Canadian Lynx data

The second study is based on the well-known Canadian lynx data which consists of the annual record of the numbers of the lynx trapped in MacKenzie River, Canada for the period 1821 to 1934 inclusively. This data was previously analysed by Nur (1993) using a STAR model of order 1. The analysis included initial data analysis using graphical methods, testing for linearity, model selection using Akaike Information Criterion, estimation and diagnostics checking. In the present illustration, we fit a STAR model for the first 100 observations of the series and compare the resulting forecasts and parameter estimates using Conditional Least Squares and the Bayesian methods.

Let X_t denote $\log_{10}$(number recorded as trapped in year 1820+t, $t = 1, ..., 100$). Using Conditional Least Squares, we first fitted the STAR order 1 model

$$\begin{aligned} X_t &= -0.0348 + 1.1539X_{t-1} - 6.9209G\left(\frac{X_{t-2} - 3.83}{0.721}\right) \\ &+ 1.3073X_{t-1}G\left(\frac{X_{t-2} - 3.83}{0.721}\right) + \varepsilon_t, \end{aligned}$$

where G is the cumulative distribution function of a Gaussian distribution and Var(ε_t)= 0.0474 to obtain forecasts. The Conditional Least Squares estimates of r and z, $\hat{r} = 3.83$ and $\hat{z} = 0.721$, the delay parameter d, $\hat{d} = 2$ used in the above were taken from Tong (1983, 1990) which use non-parametric lag regression for a tentative identification of delay parameter.

Following the results derived in the previous section we then fitted a STAR order 1 model by the Bayesian method. Assuming that the autoregressive order $p = 1$, for $d \in \{1, 2, 3, 4\}$, $D_r = (0, 4.0)$ and $D_z = (0.5, 4.0)$, the marginal posterior density of d, obtained using the first 100 observations of the lynx data, are given in Table 3.16.

Table 3.16. $P(d \mid \mathbf{X})$ *of the lynx data*

d	1	2	3	4
$P(d \mid \mathbf{X})$	6.38×10^{-16}	9.99×10^{-1}	9.14×10^{-5}	3.19×10^{-7}

The Bayesian analysis seems to indicate that the posterior mode of the delay parameter d is at $\hat{d} = 2$. Similar results have also been obtained by Tong (1983,1990) from the marginal histogram of the data and the non-parametric sample estimates of $E(X_t \mid X_{t-j} = x)$.

The joint posterior density of (r, z), conditional on $d = 2$ obtained by taking $D_r = (0, 7)$ and $D_z = (0.5, 7.5)$, is shown in Figure 3.2. From the graph, it is clear that the bivariate mode is not well-defined, as it is on the right boundary of D_r. This apparent behaviour is caused by the dominant

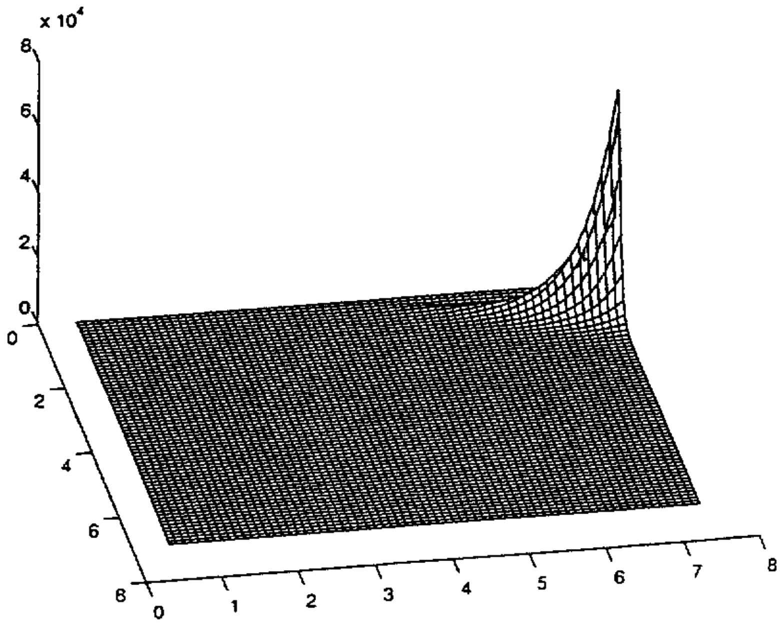

Figure 3.2. The bivariate posterior density of (r, z) *of the lynx data conditional on* $d = 2$.

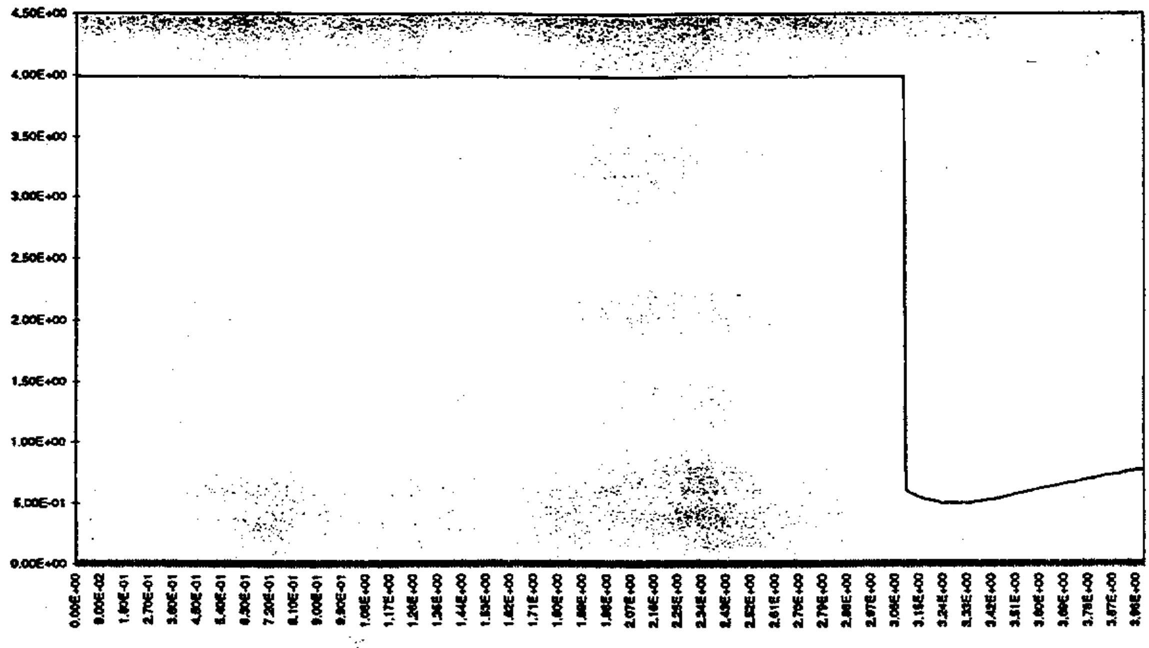

Figure 3.3. Graph of (r, z_r) : conditional modes of $P(z \mid r)$.

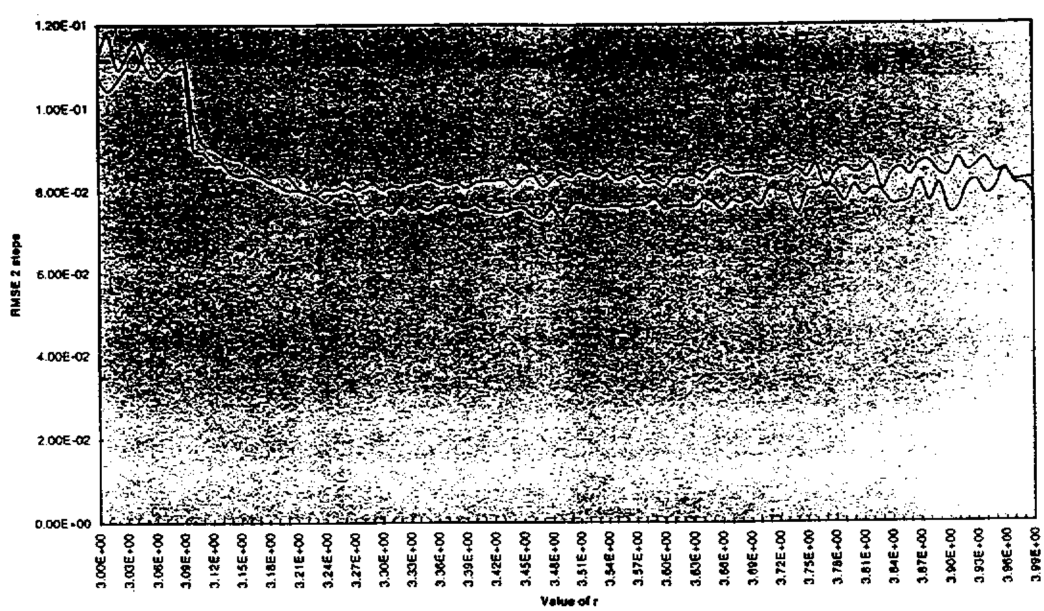

Figure 3.4. Root mean squared error of 2 steps ahead predictions for different conditional modes

determinant in the posterior density which tends to zero as r becomes large, a fact which we encountered for small sample sizes, in our simulation study. Therefore as given in Figure 3.3, we restricted the defined region to $D_r = (0, 4)$ and $D_z = (0.5, 4)$ and obtained a conditional mode, r_z, i.e. mode of $P(r \mid z)$ for all $z \in D_z$. To decide on reasonable estimators for (r, z), we then compared the root mean squared error for 2-steps ahead predictions based on these different conditional modes. As shown in Figure 3.4 the conditional mode $(r, z) = (3.49, 0.56)$ gives the minimum root mean squared error of prediction, leading to the STAR model of order 1, given by

$$\begin{aligned} X_t &= -0.0986 + 1.1715X_{t-1} - 4.1607G\left(\frac{X_{t-2} - 3.49}{0.56}\right) \\ &+ 0.7534X_{t-1}G\left(\frac{X_{t-2} - 3.49}{0.56}\right) + \varepsilon_t, \end{aligned}$$

where G is the cumulative distribution function of a Gaussian distribution with Var(ε_t)= 0.0477.

Table 3.17. Root Mean Squared Error comparison of predictors : lynx data.

Step	RMSE(CB)	RMSE(CLS)	CB/CLS
1	0.083822	0.082491	1.016134
2	0.075814	0.078407	0.966933
3	0.093368	0.159040	0.587072
4	0.079930	0.120405	0.663846
5	0.126805	0.126492	1.002471
6	0.129563	0.123758	1.046910
7	0.132121	0.240750	0.548788
8	0.130433	0.251386	0.518856
9	0.113613	0.091031	1.248067
10	0.099158	0.081716	1.213441
Average			0.881252

The multi-step-ahead predictive performance of this model was then compared to that obtained by the Conditional Least Squares method, through the root mean squared error for 10 steps ahead predictions, $h = 10$, with $T^* = 5$. The results are presented in Table 3.17, where CB/CLS represents the ratio RMSE(B)/RMSE(C) and the average of the comparison is given in the last row. It shows that, on the average, the root mean squared error of the Bayesian predictor is about 88% less than its Least Squares counterpart, suggesting that the Bayesian predictor outperforms the Conditional Least Squares for the data set considered.

It should be noted that the presentation of lynx data in this section illustrates the predictive performance only and the complete model-building strategy needs to be done using Bayesian approach, prior to this.

Chapter 4

Some Other Parametric Estimation Methods

4.1 Introduction

In this chapter, we present a few other parameter estimation techniques for the smooth threshold autoregressive model. These methods include Conditional Least Squares, Maximum Likelihood, M-estimator and an estimating function estimator. We also discuss the theoretical properties of the resulting estimates. The conditional least squares and maximum likelihood estimators are reviewed from Chan and Tong (1986), Tjøstheim (1986) and Tong (1983, 1990) while the M-estimator and the estimating function estimator are constructed respectively from the work of Koul (1996) and Thavaneswaran and Abraham (1988). These various estimators will be used as initial estimates for obtaining an adaptive estimator for parameters of the smooth threshold autoregressive model in the next chapter.

An estimation procedure for stochastic processes based on the minimisation of a sum of squared deviations about conditional expectation, was developed by Klimko and Nelson (1978), which they called conditional least squares and showed that strong consistency, asymptotic joint normality and an iterated logarithm rate of convergence are satisfied by these estimators under a variety of conditions. Later, Tjøstheim (1986) developed a more general framework for analysing estimates in nonlinear time series giving general conditions for strong consistency and asymptotic normality, both for conditional least squares and maximum likelihood type estimators. A comprehensive discussion of the estimation theory of nonlinear time series models based on the maximum likelihood and conditional least squares procedures are given in Tong (1990), showing asymptotic properties of estimates obtained from the both methods. Furthermore, Tong (1990) also discussed estimation equation approach as an alternative for recursive estimation.

The class of M-estimators has played a prominent role in recent research on robust estimation. M-estimators minimise functions of the deviations of the observations from the estimates that are more general than the sum of squared deviations or the sum of absolute deviations. In this way, the class of M-estimators generalises least square estimation, and also, in another way, generalises the idea of the maximum likelihood estimation for the location parameter in a specified distribution (Hoaglin, *et.al.*(1983)). The consistency of M-estimates was given in terms of two theorems by Huber (1981). The first one was concerned with estimates defined through a minimum property and the second one with estimates defined through a system of implicit equations. Furthermore, he presented results showing the asymptotic normality of M-estimates under some assumptions. Koul (1996) established the asymptotic uniform linearity of M-score in a family of nonlinear time series and regression models which he used to obtain the asymptotic normality of certain classes of M-estimators. Later, Koul and Schick (1996) proved the asymptotic normality of generalized M-estimators of the parameters of random coefficient autoregressive models which includes the least squares and least absolute deviations estimators.

The theory of estimating function was originally proposed by Godambe in 1960 for *i.i.d.* observations and recently extended to discrete-time stochastic processes (Godambe (1985)). The estimating equation method is a merger between maximum likelihood and the least squares methods, it has the strengths of both methods and the weaknesses of neither (Godambe (1991)). The discrete-time stochastic processes (Godambe (1985)) was extended to nonlinear time series models by Thavaneswaran and Abraham (1988). The superiority of the estimating function approach over the conditional least squares had been demonstrated through a random coefficient autoregressive example.

For the smooth threshold autoregressive models, the estimation of parameters using conditional least squares method can be found in Chan and Tong (1986) or Tong (1990). One can obtain the maximum likelihood estimators of parameters of this model using the methods developed for nonlinear autoregressive models as mentioned in Tong (1990). For both methods, properties like consistency and asymptotic normality have been well-proven. The M-estimator and estimating functions estimator for the STAR model will be developed in this chapter based on previous work on other models in the literature. The contributions to this chapter includes Theorem 4.2, the presentation of least absolute deviation and Huber's estimator and estimating equations.

This chapter is arranged as follows. In Section 4.2, we present the conditional least squares estimates of the STAR models as mentioned in Chan and Tong (1986). Similarly, in Section 4.3, we discuss the maximum likelihood estimator for the STAR model which is obtained as a special case of Tjøstheim (1986). Based on Koul (1996), the sampling properties of M-estimator for the STAR model is presented in Section 4.4 and, finally, in Section 4.5, we discuss the sampling properties of the estimator obtained by the estimating function approach.

4.2 Conditional Least Squares

In this section, the consistency and asymptotic normality properties of the Conditional Least Squares estimator for general STAR models of order p are presented. This is essentially due to Chan and Tong (1986). The consistency and asymptotic normality properties are obtained based on Theorems 2.1 and 2.2 of Klimko and Nelson (1978) (Theorems B.1 and B.2 of Appendix B.1) for general stochastic processes. Tjøstheim (1986) derived general conditions for strong consistency and asymptotic normality for conditional least squares estimates for both ergodic strictly stationary processes and certain nonstationary processes. The parameter estimates when $p = 1$ are presented at the end of this section.

In the ensuing discussion, it is assumed that $\{X_t\}$ is ergodic and the stationary distribution of X_t has a finite second moment unless otherwise stated. Let $X_1, X_2, ..., X_n$ be a sample record from a STAR model of order p, with delay parameter d known as defined in (1.1.2) and assume that not all b_j's are zeros for identifiability of r and z. The minimal requirement of G being that it is continuous and nondecreasing.

When

$$\theta = (a_0, a_1, ..., a_p, b_0, b_1, ..., b_p, r, z)^T,$$

the natural space for the parameter vector θ is $\Theta = \cup_{i=1}^{p+1} \Theta_i$, where

$$\Theta_i = \underbrace{\Re\times\ldots\times\Re\times\Re\times\ldots\times\Re}_{p+1} \underbrace{\times\Re\times\ldots\times \overset{(p+1+i)-th}{(\Re\backslash\{0\})} \times\ldots\times\Re}_{p+1} \times\Re\times\Re^+.$$

Let θ_0 be the true parameter. Given the set of observations X_t, $t = 1, 2, ..., n$, θ_0 is estimated by minimizing the conditional sum of squares

$$Q_n(\theta) = \sum_{t=m+1}^{n} [X_t - E_\theta(X_t \mid \mathcal{B}_{t-1})]^2 \tag{4.2.1}$$

where $m = \wedge(d, p)$, and $\mathcal{B}_t$ is the σ-algebra generated by $\{X_1, X_2, ..., X_t\}$.

Let $\hat{\theta}$ be such that $Q_n(\hat{\theta}) = \min_\theta Q_n(\theta)$. The variance of ε_t, σ^2, is estimated by

$$\hat{\sigma}^2 = \frac{Q_n(\hat{\theta})}{n}.$$

The consistency and asymptotic normality of the parameters of the STAR model are proved in the following theorem of Chan and Tong (1986).

Theorem 4.1 (Theorem 3.1 of Chan and Tong (1986)) *Let* $V = E_{\theta_0}\left(\frac{\partial E_\theta(X_t|\mathcal{B}_m)}{\partial \theta_i} \cdot \frac{\partial E_\theta(X_t|\mathcal{B}_m)}{\partial \theta_j}\right)$. *Then* V *is positive definite and it follows that*

(i) there exists a sequence of estimators $\hat{\theta}_n$ *such that* $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$,

(ii) for $\varepsilon > 0$ *there exists an event* E *with* $P(E) > 1 - \varepsilon$ *and an* n_0 *such that on* E, *for* $n > n_0$, Q_n *attains a relative minimum at* $\hat{\theta}_n$,

(iii) $n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2 V^{-1})$, and $\hat{\sigma}_n^2 \xrightarrow{a.s.} \sigma^2$,

(iv) $\forall \mathbf{c} \in \Re^{2p+4}$ *then*

$$\limsup_{n \to \infty} \frac{n^{\frac{1}{2}} \mathbf{c}^T (\hat{\theta} - \theta_0)}{(2\nu^2 \ln \ln n\nu^2)^{\frac{1}{2}}} = 1, \quad \text{a.s.}$$

where $\nu^2 = \sigma^2 \mathbf{c}^T V^{-1} \mathbf{c}$

The detailed proofs of the results are given in Chan and Tong (1986) and Tong (1990).

As an illustration of the above theorem, consider the first-order STAR model without intercepts with delay parameter $d = 1$, $m = \wedge(d, p) = 1$ defined by

$$X_t = aX_{t-1} + bX_{t-1} G\left(\frac{X_{t-1} - r}{z}\right) + \varepsilon_t, \tag{4.2.2}$$

which can be rewritten as

$$X_t = aX_{t-1} + bX_{t-1} G(X_{t-1}) + \varepsilon_t, \tag{4.2.3}$$

where $G\left(\frac{x-r}{z}\right)$ rewritten as $G(x)$, G a known distribution function. The stationary region for (4.2.2) is

$$a < 1,\ a + b < 1 \ \text{ and } \ a(a+b) < 1. \tag{4.2.4}$$

Take $\hat{\theta} = (a, b, r, z)^T$ then (4.2.1) is satisfied and

$$V = \begin{pmatrix} E_{\theta_0} X_1^2 & I_1 & -\frac{b}{z} I_2 & \frac{rb}{z^2} I_2 \\ I_1 & I_3 & -\frac{b}{z} I_4 & \frac{rb}{z^2} I_4) \\ -\frac{b}{z} I_2 & -\frac{b}{z} I_4 & \frac{b^2}{z^2} I_5 & -\frac{rb^2}{z^3} I_5 \\ \frac{rb}{z^2} I_2 & \frac{rb}{z^2} I_4 & -\frac{rb^2}{z^3} I_5 & \frac{r^2 b^2}{z^4} I_5 \end{pmatrix},$$

where $I_1 = E_{\theta_0}(X_1^2 G(X_1))$, $I_2 = E_{\theta_0}(X_1^2 g(X_1)), I_3 = E_{\theta_0}(X_1^2 G^2(X_1)), I_4 = E_{\theta_0}(X_1^2 G(X_1) g(X_1))$, $I_5 = E_{\theta_0}(X_1^2 g^2(X_1))$ and $g(.)$ is a derivative of $G(.)$ such that $n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2 V^{-1})$.

As (r, z) cannot be solved explicitly from the conditional least square equation, Chan and Tong (1986) presented an algorithm to implement the CLS method. The algorithm consists of two steps : fixing (r, z) and minimise the conditional sum of squares function with respect to coefficient parameters and then varying (r, z) to minimise the conditional sum of squares function as follows.

Given (r, z), let $\theta_1 = (a, b)^T$ so that

$$\hat{\theta}_1 = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \mathrm{C}^{-1} \begin{pmatrix} \sum_{t=2}^{n} X_t X_{t-1} \\ \sum_{t=2}^{n} X_t X_{t-1} G(X_{t-1}) \end{pmatrix},$$

where

$$\mathrm{C} = \begin{pmatrix} \sum_{t=2}^{n} X_{t-1}^2 & \sum_{t=2}^{n} X_{t-1}^2 G(X_{t-1}) \\ \sum_{t=2}^{n} X_{t-1}^2 G(X_{t-1}) & \sum_{t=2}^{n} X_{t-1}^2 G^2(X_{t-1}) \end{pmatrix}.$$

After obtaining the estimates of θ_1, then one can varies the values of (r, z) to obtain their estimates which minimising (4.2.1). Chan and Tong (1986) also have given some examples on the performance of the conditional least squares estimator for the STAR model. The examples include a simulation study based on a first order STAR model and an application based on Nicholson's blowfy data. In earlier work, Nur (1993) applied the conditional least squares method to analyse the Canadian lynx and the sunspot data.

4.3 Maximum likelihood estimators

In this section, we present Theorem 4.2 which discusses the consistency and asymptotic normality of maximum likelihood estimators of STAR models of order 1. This result is based on Theorems 5.1 and 5.2 of Tjøstheim (1986) (Theorem B.3 and B.4 of Appendix B.2).

Let $\{X_t,\ t \in I\}$ be a discrete time stochastic process taking values in $\Re^k$ and defined on a probability space $(\Omega, \mathcal{B}, P)$. The index set I is either the set $\mathcal{Z}$ or the set $\mathcal{Z}^+$. We assume that observations $(X_1, ..., X_N)$ are available. The parameter vector $\theta = (\theta_1, ..., \theta_p)^T$ will be assumed to be lying in some open set Θ of Euclidean p-space with the true value denoted by θ_0. The notation $\tilde{X}_{t|t-1} = \tilde{X}_{t|t-1}(\theta)$ is used for the conditional expectation $E_\theta(X_t \mid \mathcal{B}_{t-1})$. The $k \times k$ conditional prediction error matrix of $\{X_t\}$ is denoted by

$$Z_{t|t-1} = E\{(X_t - \tilde{X}_{t-1})(X_t - \tilde{X}_{t-1})^T \mid \mathcal{B}_{t-1})\}.$$

If $X_t - \tilde{X}_{t-1}$ independent of $\mathcal{B}_{t-1}$ then

$$Z_{t|t-1} = E\{(X_t - \tilde{X}_{t-1})(X_t - \tilde{X}_{t-1})^T\}.$$

Let L_n be the likelihood type penalty function

$$L_n = \sum_{t=m+1}^{n} [\ln\{det(Z_{t|t-1})\} + (X_t - \tilde{X}_{t|t-1})^T Z_{t|t-1}^{-1}(X_t - \tilde{X}_{t|t-1})] = \sum_{t=m+1}^{n} \phi_t \qquad (4.3.1)$$

where m is the order of the nonlinear autoregressive process such that $t-m \leq s \leq t-1$, s denotes the number of components of the parameter vector θ appearing in $L_n(\theta)$. Due to the presence of $Z_{t|t-1}$ in L_n, in general $s > r$ with r being the number of parameters in the model.

For the first-order STAR model (4.2.2) with the stationary region (4.2.4), the following theorem shows the consistency and asymptotic normality of the maximum likelihood estimators.

Theorem 4.2 *Assume that (r, z) are known, let $\{X_t\}$ be defined by (4.2.2). Assume that (4.2.4) holds, and that ε_t has a density function with infinite support such that $E(\varepsilon_t^2) < \infty$. Then there exists a sequence of estimators $\{\hat{\theta}_n\}$ minimizing the penalty function L_n of (4.3.1), such that $\hat{\theta}_n \xrightarrow{a.s.} \theta$. Also,*

$$n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, (U^T)^{-1}),$$

where

$$U^T = \frac{1}{\sigma^2} \begin{pmatrix} E_\theta X_{t-1}^2 & E_\theta(X_{t-1}^2 G(X_{t-1})) \\ E_\theta(X_{t-1}^2 G(X_{t-1})) & E_\theta(X_{t-1}^2 G^2(X_{t-1})) \end{pmatrix}$$

Proof.

Let $\theta = (a, b)^T$. It follows that

$$\tilde{X}_{t|t-1} = E_\theta(X_t \mid \mathcal{B}_{t-1}) = aX_{t-1} + bX_{t-1}G(X_{t-1}) \qquad (4.3.2)$$

and

$$Z_{t|t-1} = E(\varepsilon_t^2) = \sigma^2.$$

The likelihood function is

$$L_n = \sum_{t=2}^{n}[\ln \sigma^2 + \frac{1}{\sigma^2}(X_t - (aX_{t-1} + bX_{t-1}G(X_{t-1}))^2] = \sum_{t=2}^{n} \phi_t,$$

where

$$\phi_t = \ln \sigma^2 + \frac{1}{\sigma^2}(X_t - (aX_{t-1} + bX_{t-1}G(X_{t-1}))^2.$$

To apply Theorems 5.1 and 5.2 of Tjøstheim (1986) (Theorem B.3 and B.4 of Appendix B.2), we verify conditions (i)-(iii) of Theorem 5.1 and conditions in Theorem 5.2 of Tjøstheim (1986) as follows. We have

$$\frac{\partial \phi_t}{\partial a} = -\frac{2}{\sigma^2}\varepsilon_t X_{t-1}, \quad \frac{\partial \phi_t}{\partial b} = -\frac{2}{\sigma^2}\varepsilon_t X_{t-1} G(X_{t-1}),$$

$$\frac{\partial^2 \phi_t}{\partial a \partial b} = \frac{2}{\sigma^2} X_{t-1}^2 G(X_{t-1}), \quad \frac{\partial^2 \phi_t}{\partial a^2} = \frac{2}{\sigma^2} X_{t-1}^2, \quad \frac{\partial^2 \phi_t}{\partial b^2} = \frac{2}{\sigma^2} X_{t-1}^2 G^2(X_{t-1}).$$

From the independence of $\{\varepsilon_t\}$ and $\{X_s : s < t\}$ and as $\{X_t\}$ is ergodic so that the condition $E(\varepsilon_t^2) < \infty$ implies $E(X_t^2) < \infty$ we have

$$E\left|\frac{\partial \phi_t}{\partial a}\right| < \infty, \quad E\left|\frac{\partial \phi_t}{\partial b}\right| < \infty.$$

Using Theorem 5.6 of Karlin and Taylor (1975) (Theorem A.4 of Appendix A), we have

$$E\left|\frac{\partial^2 \phi_t}{\partial a \partial b}\right| < \infty, \quad E\left|\frac{\partial^2 \phi_t}{\partial a^2}\right| < \infty, \quad E\left|\frac{\partial^2 \phi_t}{\partial b^2}\right| < \infty.$$

Hence condition (i) is satisfied.

As the third derivative of ϕ_t is zero then condition (iii) follows.

Furthermore, to see condition (ii), we have that

$$\frac{\partial \tilde{X}_{t|t-1}}{\partial a} = X_{t-1}, \quad \frac{\partial \tilde{X}_{t|t-1}}{\partial b} = X_{t-1} G(X_{t-1}).$$

As $Z_{t|t-1}$ does not depend on the parameters then the second term of condition (ii) is zero. Thus if two arbitrary real numbers c_1, c_2, exist such that

$$E\left(\left|\sigma\left(c_1 \frac{\partial \tilde{X}_{t|t-1}}{\partial a} + c_2 \frac{\partial \tilde{X}_{t|t-1}}{\partial b}\right)\right|^2\right) = 0$$

implying

$$c_1 X_{t-1}\sigma + c_2 X_{t-1} G(X_{t-1})\sigma = 0 \quad \text{a.s.},$$

then $c1 = c2 = 0$ since $E(X_t^2) \geq E(\varepsilon_t^2) = \sigma^2 > 0$. Hence condition (ii) holds. As $Z_{t|t-1}$ does not depend on the parameters $S = 0$ in Theorem 5.2 of Tjøstheim (1986) so that the condition is obviously satisfied.

Using Theorems 5.1 and 5.2 of Tjøstheim (1986) (Theorem B.3 and B.4 of Appendix B.2), the proof is complete. □

Remark 4.1 *On the above discussion, MLE is presented conditional on fixed* (r, z, d). *In the case of estimating* (r, z, d), *Akaike Information Criterion can be applied as in Tong (1983).*

4.4 M-estimators

This section presents asymptotic properties of Huber(k) and least absolute deviation estimator of a first-order STAR model with delay parameter one. These estimators are well-known examples of M- estimators. As the estimators are $\sqrt{n}$-consistent, the estimators can be used as preliminary estimates for adaptive estimation in Chapter 5.

The asymptotic results of Huber(k) and least absolute deviation estimators follow from theorems and corollaries given in Koul (1996). We use the same notation and terminology as in Koul (1996).

Let m, n and p be integers, $m \wedge p \geq 1$, $n \geq m$, Θ be an open subset of the m-dimensional Euclidean space $\Re^m, \Re = \Re^1$, $\mathbf{t} \in \Re^p$ and $\| \mathbf{t} \|$ its Euclidean norm. Let $\{\varepsilon_i, i = 1, 2, ...\}$ be $i.i.d.$ r.v's with distribution function F; $\mathbf{Y}_0 := (X_0, ..., X_{1-p})^T$ be an observable random variable, independent of $\{\varepsilon_i, i = 1, 2, ...\}$; B_{n1} be the σ field generated by $\{\mathbf{Y_0}\}$ and B_{ni} be the σ- field generated by $\{\mathbf{Y_0}; \varepsilon_j, 1 \leq j < i\}$, $2 \leq i \leq n$.

In a p-th order nonlinear time series model considered here, one observes an array of the process $\{X_{ni}, i = 1, 2, ..., n\}$ satisfying the relation

$$X_{ni} = h(\theta, \mathbf{Y}_{n,i-1}) + \varepsilon_i, \quad i = 1, 2, ..., n, \tag{4.4.1}$$

for some $\theta \in \Theta$, where $\mathbf{Y}_{n0} := \mathbf{Y}_0$, $\mathbf{Y}_{n,i-1} := (X_{n,i-1}, ..., X_{n,i-p})^T$, and h is a known function from $\Theta \times \Re^p$ to $\Re$ that is measurable in the last p coordinates so that $h_{ni}(\theta) := h(\theta, \mathbf{Y}_{n,i-1})$ is $\mathcal{B}_{ni}$-measurable, $\theta \in \Theta, 1 \leq i \leq n$.

Fix a $\theta \in \Theta$ and let P_θ^n denote the probability distribution of $(\mathbf{Y}_0, X_{n1}, ..., X_{nn})$ under (4.4.1) when θ is the true parameter value. Suppose there exists a vector of functions $\dot{\mathbf{h}}_n$ from $\Theta \times \Re^p$ to $\Re^m$ as mentioned in Appendix B.3.

Assuming (r, z) are known, consider the first-order STAR model of (4.2.2) and let

$$\theta = (a, b)^T, \quad h_{ni}(\mathbf{t}) \equiv \mathbf{t}^T \mathbf{W}_i,$$

where $\mathbf{W}_i \equiv (X_{i-1}, X_{i-1} G(X_{i-1}))^T$.
Then it is easy to see that

$$\dot{\mathbf{h}}_{ni}(\mathbf{t}) \equiv \mathbf{W}_i.$$

Let $\mathcal{F} = \mathcal{F}_0^+$, the set of all positive and uniformly continuous Lebesgue densities with zero- mean and finite variances ; i.e. $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) < \infty$. Take

$$\Theta = \{\theta = (a, b) \in \Re^2 : a < 1, a + b < 1, a(a + b) < 1\}.$$

For each $(\theta, \phi) \in \Theta \times \mathcal{F}$, the above model is ergodic by Theorem 2.3 and there is a unique stationary process satisfying the above model (Proposition 2.1, Chan and Tong (1986), see Proposition A.1 of Appendix A). Furthermore, the existence of the s-th absolute moment of ε_t implies that of the stationary distribution, that is, $E \mid \varepsilon_t \mid^s < \infty$ implies $E \mid X_0 \mid^s < \infty$. In the following, we present the propositions on asymptotic properties of Huber and least absolute deviation estimators respectively.

Proposition 4.1 *(Huber estimator) Let* $\dot{\mathbf{h}}_{ni} = \mathbf{W}_i$, *where* $\mathbf{W}_i \equiv (X_{i-1}, X_{i-1} G(X_{i-1}))^T$ *and* $\psi(x) = x I[\mid x \mid \leq c] +$

$c\,\mathrm{sgn}(x)I[|\,x\,|> c]$, $x \in \Re$, *where* c *is a known positive constant. Assume that* F *is symmetric around zero,* F *is continuous and* $[F(c)-F(-c)] > 0$. *Then*

$$n^{1/2}(\hat{\theta}_M - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}_2(\mathbf{0}, \textstyle\sum_\theta^{-1} \nu(\psi, F)),$$

where $\hat{\theta}_M$ *is the Huber estimator and*

$$\sum\nolimits_\theta = \begin{pmatrix} E_\theta X_0^2 & E_\theta(X_0^2 G(X_0)) \\ E_\theta(X_0^2 G(X_0)) & E_\theta(X_0^2 G^2(X_0)) \end{pmatrix}. \tag{4.4.2}$$

with

$$\nu(\psi, F) = \frac{\sigma^2 - E(\varepsilon_i^2 I[|\,\varepsilon_i\,|> c]) + c^2(1 - F(c) + F(-c))}{[F(c) - F(-c)]^2}.$$

Proof.

First we need to verify assumptions (h1) to (h6), (F) and (M1) of Appendix B.3 in order to apply Corollary 1.1 of Koul (1996) (Corollary B.1 of Appendix B.3). The verification proceeds as follows.

(i) Since $\mathcal{F} = \mathcal{F}_0^+$ then assumption (F) is automatically satisfied.

(ii) Since $\dot{\mathbf{h}}_i(\mathbf{t}) \equiv \mathbf{W}_i$ does not depend on parameter θ, then assumptions (h1), (h4), (h5) and (h6) hold.

(iii) To prove assumptions (h2) and (h3), we use Remark 1.1 of Koul (1996) for stationary and ergodic processes that the condition (h_s1) in Remark 1.1 of Koul (1996) (see Appendix B.3) implies condition (h2) and (h3). We have to show that

$$E_\theta \parallel \dot{\mathbf{h}}_1(\theta) \parallel^2 < \infty,$$

where E_θ denotes the expectation under the stationary distribution. Clearly

$$E_\theta \parallel \dot{h}_1(\theta) \parallel^2 = E_\theta(X_0^2 + X_0^2 G^2(X_0)) = E_\theta X_0^2 + E_\theta(X_0^2 G^2(X_0)).$$

And also, $E_\theta X_0^2 < \infty$. Using Theorem 5.6 of Karlin and Taylor (1975) (Theorem A.4 of Appendix A) then yields $E_\theta(X_0^2 G^2(X_0)) < \infty$.

Hence $E_\theta \parallel \dot{\mathbf{h}}_1(\theta) \parallel^2 < \infty$, so that assumptions (h2) and (h3) hold. The existence of a positive -definite matrix $\sum_\theta$ in (h2) is guaranteed by the ergodicity of the processes and finite second moments using Theorem 5.5 and 5.6 of Karlin and Taylor (1975) (Theorem A.3 and A.4 of Appendix A) where $\sum_\theta$ is as defined in (4.4.2).

(iv) To prove condition (M1), we have

$$\mathbf{e}^T\mathbf{M}(\theta + n^{-1/2}r\mathbf{e}) = n^{-1/2}\sum_{i=1}^{n}\mathbf{e}^T\mathbf{W}_i\ \psi(X_i - h_{ni}(\theta + n^{-1/2}r\mathbf{e}))$$

$$= n^{-1/2}\sum_{i=1}^{n}\mathbf{e}^T\mathbf{W}_i\ \psi(X_i - (\theta + n^{-1/2}r\mathbf{e})^T\mathbf{W}_i)$$

$$= n^{-1/2}\sum_{i=1}^{n}\mathbf{e}^T\mathbf{W}_i\ \psi(\varepsilon_i - rn^{-1/2}\mathbf{e}^T\mathbf{W}_i).$$

We have to show that the function is monotonic. Consider a function $\sum_{i=1}^n x_i\psi(y_i - rx_i)$ where $x_i, y_i, r \in \Re$. Let $r_1 > r_2,\ r_1,\ r_2 \in \Re$. If $x_i > 0$ then

$$\psi(y_i - r_1x_i) \leq \psi(y_i - r_2x_i)$$

so that

$$x_i\psi(y_i - r_1x_i) \leq x_i\psi(y_i - r_2x_i).$$

If $x_i < 0$ then, since ψ is nondecreasing,

$$\psi(y_i - r_1x_i) \geq \psi(y_i - r_2x_i)$$

so that

$$x_i\psi(y_i - r_1x_i) \leq x_i\psi(y_i - r_2x_i).$$

Hence it follows that $x_i\psi(y_i - rx_i)$ is monotonic, which implies that for any $\mathbf{e}^T\mathbf{W}_i \in \Re$, $\mathbf{e}^T\mathbf{M}(\theta + n^{-1/2}r\mathbf{e})$ is monotonic. Hence (M1) holds.

The assumption in Corollary 1.1 of Koul (1996) (Corollary B.1 of Appendix B.3) states that

$$\int f d\psi > 0.$$

From the assumption in the proposition,

$$\int f d\psi = F(c) - F(-c) > 0.$$

As the assumptions (F),(h1) to (h6) and (M1) of Appendix B.3 hold then one can apply Corollary 1.1 of Koul (1996) (Corollary B.1 of Appendix B.3). Hence the proposition is proved. □

Proposition 4.2 *(Least absolute deviation estimator) Let* $\dot{\mathbf{h}}_i = \mathbf{W}_i$, *where* $\mathbf{W}_i \equiv (X_{i-1}, X_{i-1}G(X_{i-1}))^T$ *and* $\psi(x) = sgn(x)$, $x \in \Re$. *Assume that* F *has a continuous, bounded and even density* f *and* $f(0) > 0$. *Then*

$$n^{1/2}(\hat{\theta}_{lad} - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}_2(\mathbf{0}, {\textstyle\sum_\theta^{-1}}/4f^2(0))$$

where $\hat{\theta}_{lad}$ *is the least absolute deviation estimator and* $\sum_\theta$ *is defined in (4.4.2)*

Proof. Assumptions (h1) to (h6) of Appendix B.3 hold as mentioned in Proposition 4.1 above. With $\psi(x) = \text{sgn}(x)$, ψ is nondecreasing, then assumption (M1) holds. Also the assumptions in the Corollary 1.2 are satisfied. By applying Corollary 1.2 of Koul (1996) (Corollary B.2 of Appendix B.3) the proposition follows. □

4.5 Estimating Functions

In this section, we recall Godambe's (1985) theorem on stochastic processes and apply it to obtain optimal estimates for STAR models.

Let $X_1, X_2, ..., X_n$ be a sample record from a STAR model of order p :

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + (\sum_{j=1}^{p} b_j X_{t-j}) G\left(\frac{X_{t-d} - r}{z}\right) + \varepsilon_t. \qquad (4.5.1)$$

Assuming (r, z, d) known, let

$$\theta = (a_1, ..., a_p, b_1, ..., b_p)^T.$$

$$\mathbf{X} = (X_{t-1}, ..., X_{t-p}, X_{t-1}G_t, ..., X_{t-p}G_t),$$

where $G_t = G\left(\frac{X_{t-d}-r}{z}\right)$, $t = m+1, ..., n$, and $m = \wedge(d,p)$.

Then the above model becomes

$$X_t = \mathbf{X}\theta + \varepsilon_t.$$

Now, let

$$h_t = X_t - E[X_t \mid \mathcal{B}_{t-1}] = X_t - \mathbf{X}\theta.$$

The optimal estimating equation for θ_i, $i = 1, ..., 2p$ is given by

$$g_i^* = \sum_{t=m+1}^{n} h_t a_{i,t-1}^*,$$

where

$$a_{i,t-1}^* = \frac{E\left[\frac{\partial h_t}{\partial \theta} \mid \mathcal{B}_{t-1}\right]}{E\left[h_t^2 \mid \mathcal{B}_{t-1}\right]}.$$

From the model,

$$E\left[h_t^2 \mid \mathcal{B}_{t-1}\right] = E[\varepsilon_t^2 \mid \mathcal{B}_{t-1}] = E[\varepsilon_t^2] = \sigma^2.$$

The optimal estimate for θ can be obtained by solving the equation

$$\sum_{t=m+1}^{n} h_t a_{i,t-1} = 0$$

which yields,

$$\hat{\theta} = \left(\sum_{t=m+1}^{n} \mathbf{X}^T\mathbf{X}\right)^{-1} \sum_{t=1}^{n} X_t \mathbf{X}^T.$$

For the special case when $p = 1, d = 1$, assuming (r, z) are known

$$X_t = aX_{t-1} + bX_{t-1}G(X_{t-1}) + \varepsilon_t,$$

where $G(X_{t-1}) = G\left(\frac{X_{t-1}-r}{z}\right)$.

Let $\mathbf{X} = (X_{t-1} \;\; X_{t-1}G(X_{t-1}))$, $\theta = (a \;\; b)^T$ so that the above model becomes

$$X_t = \mathbf{X}\theta + \varepsilon_t.$$

Taking $h_t = X_t - E[X_t \mid \mathcal{B}_{t-1}] = X_t - \{aX_{t-1} + bX_{t-1}G(X_{t-1})\} = X_t - \mathbf{X}\theta$ and

$$E\left[\frac{\partial h_t}{\partial \theta} \mid \mathcal{B}_{t-1}\right] = \begin{pmatrix} -X_{t-1} \\ -X_{t-1}G(X_{t-1}) \end{pmatrix}.$$

The optimal estimating equation for θ is

$$\sum_{t=2}^{n} (X_t - \{aX_{t-1} + bX_{t-1}G(X_{t-1})\}) \frac{1}{\sigma^2} \begin{pmatrix} -X_{t-1} \\ -X_{t-1}G(X_{t-1}) \end{pmatrix} = 0,$$

$$\begin{pmatrix} \sum_{t=2}^{n} X_t X_{t-1} \\ \sum_{t=2}^{n} X_t X_{t-1} G(X_{t-1}) \end{pmatrix} = \sum_{t=2}^{n} \begin{pmatrix} X_{t-1} \\ X_{t-1}G(X_{t-1}) \end{pmatrix} (X_{t-1} \;\; X_{t-1}G(X_{t-1})) \begin{pmatrix} a \\ b \end{pmatrix}.$$

Hence

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} \sum_{t=2}^{n} X_{t-1}^2 & \sum_{t=2}^{n} X_{t-1}^2 G(X_{t-1}) \\ \sum_{t=2}^{n} X_{t-1}^2 G(X_{t-1}) & \sum_{t=2}^{n} X_{t-1}^2 G(X_{t-1})^2 \end{pmatrix}^{-1} \times \begin{pmatrix} \sum_{t=2}^{n} X_t X_{t-1} \\ \sum_{t=2}^{n} X_t X_{t-1} G(X_{t-1}) \end{pmatrix}.$$

Remark 4.2 *Similar to MLE, M-estimator and Estimating Function are presented conditional on fixed (r, z, d). For estimating (r, z, d), one can combine the method with CLS or Akaike Information Criterion.*

One should note that the above estimates are the same as conditional least squares estimates obtained in section 2. Hence by asymptotic properties of the conditional least squares estimator it follows that these estimates are consistent and asymptotically normally distributed.

Chapter 5

Adaptive Estimation

5.1 Introduction

This chapter presents a study on adaptive estimation for the class of STAR models and closely follows the theory developed in Koul and Schick (1997). The main contributions to this chapter includes the presentation of numerical examples and the verification of conditions of theorems of Koul and Schick (1997) for the STAR models. The content of this chapter is strongly related to Chapters 2 and 4 as the ergodicity assumptions are needed for the proofs and also the estimator obtained in Chapter 4 is used as the preliminary estimator in adaptive estimation.

The construction of estimators that are asymptotically efficient in the presence of infinite dimensional nuisance parameters has been the focus of numerous researchers in the last three decades. Consider estimation of an Euclidean parameter θ and an infinite dimensional nuisance parameter g. To study what is best possible asymptotically, one needs a bound on the asymptotic performance of estimators of θ. Hàjek (1970) established a lower bound for the local asymptotic minimax risk of a sequence of estimates under local asymptotic normal condition and showed that in the one dimensional case, a condition close to regularity (see Fabian and Hannan (1982)) is necessary for an estimate to be locally asymptotically minimax, that is, for the variance of an estimate to attain the lower bound. A vast majority of models are Locally Asymptotically Normal (LAN) and when this holds the Hajek-Le Cam convolution theorem yields an appropriate lower bound. On an ad hoc basis, it is often possible to find estimators of θ that have the right rate of consistency. Typically, such estimators may be used to construct efficient estimators, which attain the bound of the convolution theorem. If this bound is the same as in the parametric model with g known, then such estimators are called *adaptive*. From the general asymptotic theory for adaptive estimation in locally asymptotically normal (LAN) families it follows that adaptive estimation is not always possible (Koul and Schick (1997)). Necessary conditions for adaptive estimation for these families are given by Fabian and Hannan (1982).

For the independent identically distributed case, a comprehensive account on the present theory along these lines is given in Bickel, Klaassen, Ritov and Wellner (1993). Survey papers obtaining adaptive estimation in linear time series models was probably started by Beran (1976) who constructed adaptive estimates for the parameters of a stationary autoregressive process and obtained their relative asymptotic efficiency with respect to the least squares estimates. The nature of the adaptive estimates encourages stable behaviour for moderate sample sizes. Akritas and Johnson (1982) obtained the asymptotic power for tests of hypotheses concerning the autoregressive process when the error distribution is nonnormal, by employing the concept of contiguity and this leads directly to an expression for the Pitman efficiency tests as well as estimators. The numerical values of the efficiencies suggested a lack of robustness for the normal theory least square estimators when the error distribution is thick tailed. This work was then extended by Kreiss (1987a) to obtain adaptive estimates in stationary autoregressive and moving average processes in a wide class of symmetric density errors, and the estimates turn out to be asymptotically optimal. Moreover, Kreiss (1987b) generalised his earlier work to obtain adaptive estimation in autoregressive models when the error density is asymmetric. For linear regression model, Koul and Susarla (1983) constructed adaptive estimator of the regression parameter vector in the linear model and studied its asymptotic properties. Zwanzig (1994) obtained adaptive estimates for linear regression model for a general error density, and for nonlinear regression with symmetric density errors. A review of results on some basic elements of large sample theory in a restricted structural framework, as described in detail in LeCam and Yang (1990), was comprehensively given in Jeganathan (1995). He illustrated the asymptotic inference problems associated with many linear time series regression models.

For nonlinear time series models, Linton (1993) constructed efficient estimators of the identifiable parameters in a regression model when the errors follow a stationary ARCH(p) process under the assumption of symmetric density error and showed that the identifiable parameters of this process are adaptively estimable. Koul and Schick (1996) proved the existence of adaptive estimation in a stationary and ergodic random coefficient autoregressive model if the distributions of the innovations and the errors in the models are symmetric around zero. Drost and Klaassen (1997) constructed adaptive and hence efficient estimates for semiparametric GARCH in mean-type models with a symmetric error density. The GARCH models include integrated GARCH models which are often used to model financial data sets. Drost, Klaassen and Werker (1997) constructed adaptive estimators in time series models especially with time varying location and scale and showed how a sample splitting technique can be used to construct adaptive estimates for these models and applied the construction to ARMA, Threshold Autoregressive and ARCH models. Koul and Schick (1997) also discussed efficient estimation for a class of nonlinear time series models with unknown error densities. They gave several methods for constructing efficient estimates and these results were then applied for SETAR, EXPAR and ARMA models. They found that adaptation is not possible in SETAR models with asymmetric error densities. Recently, Koul and Schick (1997) considered the construction of adaptive estimators in time series models with time varying location, which are less general than models considered by Drost, *et.al* (1997). But, the former obtained efficient estimates in general error models which will be automatically adaptive under the symmetry conditions.

Consider the first-order STAR model with delay parameter one defined by

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-1} G\left(\frac{X_{t-1} - r}{z}\right) + \varepsilon_t,$$

which can be rewritten as

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-1} G(X_{t-1}) + \varepsilon_t, \tag{5.1.1}$$

where r is a threshold parameter, $r \in \Re$; z is a smoothing parameter, $z \in \Re^+$, $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables with ε_t independent of X_s, $s < t$ and $G\left(\frac{x-r}{z}\right)$, which is redefined as $G(x)$, is a distribution function. Throughout the chapter it is assumed that $G(.)$, r and z are known. The problem of interest is the construction of adaptive estimator of $\theta = (\theta_1, \theta_2)^T$ in the presence of the nuisance parameter f, where f denotes the unknown density of the innovations in the above model. If $X_t \to \infty$, (5.1.1) becomes essentially linear with coefficient $\theta_1 + \theta_2$ and when $X_t \to -\infty$, (5.1.1) becomes linear with coefficient θ_1.

This chapter is organised as follows. In Section 5.2, for convenience, the assumptions, definitions, conditions and theorems from Koul and Schick (1997) are briefly summarised. Section 5.3 discusses local asymptotic normality of the above STAR semiparametric model whereas Section 5.4 addresses the question of adaptive estimation of θ, where the necessary condition for adaptive estimation is verified for the model when f is symmetric about zero. The algorithms for construction of adaptive and efficient estimates are presented in Section 5.5 in which the adaptive estimate is constructed without splitting the sample whereas the efficient estimate is constructed with splitting the sample. Finally in Section 5.6, we present simulation results to compare the conditional least squares estimate with the adaptive and efficient estimates in STAR models.

5.2 Notation and Preliminaries

The following notations of Koul and Schick (1997) are being used in this chapter. Let m and p be positive integers, $\mathcal{F}$ be a class of Lebesgue densities, Θ be an open subset of $\Re^m$, and $\mathcal{P} = \{P_{\vartheta,\phi} : (\vartheta, \phi) \in \Theta \times \mathcal{F}\}$ be a family of probability measures. Let $X_{1-p}, ..., X_0, X_1, X_2, ...$ be random variables and, for each $j = 1, 2, ...,$ let h_j be a measurable map from $\Re^{p+j-1} \times \Theta$ into $\Re$. Define $\mathbf{X}_j = (X_{1-p}, ..., X_j)^T$, $j = 0, 1, ...,$ and

$$H_j(\vartheta) = h_j(\mathbf{X}_{j-1}, \vartheta), \quad \vartheta \in \Theta, \ j = 1, 2, ...$$

Under each $P_{\vartheta,\phi} \in \mathcal{P}$, the following is assumed for the time series $\{X_j : j \geq 1 - p\}$. The random vector $\mathbf{X}_0$ has a Lebesgue density $g_{\vartheta,\phi}$, and the random variables

$$\varepsilon_j(\vartheta) = X_j - H_j(\vartheta), \quad j = 1, 2, ...,$$

are independent with common density ϕ and are independent of $\mathbf{X}_0$.

Let θ and f be the true parameters, F denote the distribution function corresponding to nuisance parameter density f. The expectation under $P_{\vartheta,\phi}$ is denoted by $E_{\vartheta,\phi}$ where $(\vartheta, \phi) \in \Theta \times \mathcal{F}$. For convenience, $P_{\vartheta,\phi}$

and $E_{\vartheta,\phi}$ are abbreviated by P_ϑ and E_ϑ.

Moreover, let $\mathcal{F}_0$ denote the set of all Lebesgue densities with zero-mean, finite variances and finite Fisher information for location, and $\mathcal{F}_0^+$ consist of all positive densities in $\mathcal{F}_0$.

As the statistical information contained in the sample regarding the parameter of interest is described in terms of the likelihood ratios of the sample, satisfactory answers to obtain an optimal estimator when the sample size is large will involve the study of asymptotic behaviour of likelihood ratios. Indeed, one of the central objectives of asymptotic theory, in its simplest form, is to provide methods of recovering the likelihood ratio's of the sample, at least approximately when the sample size is large, by means of a suitable estimator of the parameter. It is assumed that the approximation to the likelihood ratios is quadratic in parameter, and it is shown, in LeCam (1960) that this approximation simply reduces the inference procedures and related problems to those of a Gaussian distribution. The likelihood ratios satisfying the quadratic approximations are called locally asymptotically normal (LAN) likelihood ratios (Jeganathan (1995)).

Various LAN results have been proven in special cases by many authors such as Akritas and Johnson (1982), Kreiss (1987), Drost, *et al.* (1994) and Jeganathan (1995). In the first two of the above papers the error density f is not parametrized, and the latter two prove the LAN property, which is uniform in θ but the error density f is fixed. Koul and Schick (1997) prove LAN in both the parameter of interest and the nuisance parameter, with uniformity in the former.

The following give the assumptions and definitions required to prove LAN result for nonlinear time series models.

Assumption 5.1 (Assumption 2.1 of Koul and Schick (1997)) *The density f has finite Fisher information for location, i.e. f is absolutely continuous with a.e-derivative $f^{(1)}$ and*

$$J = \int l^2 dF \ < \ \infty, \quad \text{where } l = -\frac{f^{(1)}}{f}. \tag{5.2.1}$$

Moreover,

$$\int \mid g_{\vartheta,f}(\mathbf{x}) - g_{\theta,f}(\mathbf{x}) \mid d\mathbf{x} \ \to \ 0, \quad \text{as } \vartheta \ \to \ \theta, \tag{5.2.2}$$

where $g_{\vartheta,f}$ is a Lebesgue density of $\mathbf{X}_0$ under $P_{\vartheta,f} \in \mathcal{P}$.

Assumption 5.2 (Assumption 2.2 of Koul and Schick (1997)) *There exists a $\nu \in \Re^m$, a positive definite $m \times m$ matrix M and measurable functions $\dot{h}_j$ from $\Re^{p+j-1} \times \Theta$ to $\Re^m$, $j = 1, 2, \ldots$ such that for all local sequences $< \vartheta_n >$ and $< \theta_n >$*

$$\sum_{j=1}^{n} \mid H_j(\vartheta_n) - H_j(\theta_n) - (\vartheta_n - \theta_n)^T \dot{H}_j(\theta_n) \mid^2 = o_{\theta_n}(1), \tag{5.2.3}$$

$$\max_{1 \le j \le n} \frac{1}{\sqrt{n}} \parallel \dot{H}_j(\theta_n) \parallel = o_{\theta_n}(1), \tag{5.2.4}$$

$$\frac{1}{n} \sum_{j=1}^{n} \dot{H}_j(\theta_n) = \nu + o_{\theta_n}(1), \tag{5.2.5}$$

$$\frac{1}{n} \sum_{j=1}^{n} \dot{H}_j(\theta_n) \dot{H}_j^T(\theta_n) = M + o_{\theta_n}(1), \tag{5.2.6}$$

where $\dot{H}_j(\vartheta) = \dot{h}_j(X_j, \vartheta)$ for $j = 1, 2, \ldots$ and $\vartheta \in \Theta$.

Definition 5.1 (Definition 2.3 of Koul and Schick (1997)) *By an s-dimensional path we mean a map $\eta \mapsto f_\eta$ from a neighbourhood Δ of the origin in $\Re^s$ into $\mathcal{F}$ such that $f_0 = f$. The path $\eta \mapsto f_\eta$ is said to be ξ-smooth if ξ is a measurable function from $\Re$ to $\Re^s$ such that $\int \parallel \xi \parallel^2 dF \ < \ \infty$, $\int \xi \xi^T dF$ is nonsingular, and*

$$\int \left(\sqrt{f_\eta(x)} - \sqrt{f(x)} - \frac{1}{2} \eta^T \xi(x) \sqrt{f(x)} \right)^2 dx = o(\parallel \eta \parallel^2). \tag{5.2.7}$$

The path $\eta \mapsto f_\eta$ *is said to be* ξ*-regular if it is* ξ*-smooth and if*

$$\int | g_{\vartheta,f_\eta}(\mathrm{x}) - g_{\theta,f}(\mathrm{x}) | \, d\mathrm{x} \;\rightarrow\; 0, \quad \text{as } \vartheta \;\rightarrow\; \theta \text{ and } \eta \;\rightarrow 0. \qquad (5.2.8)$$

Remark 5.1 *Let* $\mathcal{F} = \mathcal{F}_0$. *Then it can be shown that for each measurable function* ξ *such that* $\int \xi(x)dF(x) = 0$, $\int x\xi(x)dF(x) = 0$ *and* $0 < \int \xi^2(x)f(x)dx < \infty$, *there exists an one-dimensional path that is* ξ *smooth.*

Fix an s-dimensional ξ-regular path $\eta \rightarrow f_\eta$. Define $(m+s)$-dimensional random vectors S_j, for $j = 1, 2, ...,$ by

$$S_j(\vartheta, \xi) = \begin{pmatrix} \dot{H}_j(\vartheta) l(\varepsilon_j(\vartheta)) \\ \xi(\varepsilon_j(\vartheta)) \end{pmatrix}$$

and an $(m+s) \times (m+s)$ matrix V by

$$V(\xi) = \begin{pmatrix} JM & \nu \int l\xi^T dF \\ \int l\xi dF \nu^T & \int \xi\xi^T dF \end{pmatrix}.$$

Let $P^n_{\vartheta,\eta}$ be the restriction of P_{ϑ,f_η} to the σ-field generated by $\mathbf{X}_n$. For $\vartheta_1, \vartheta_2 \in \Theta$ and $\eta \in \Delta$, let $\Lambda_n(\vartheta_1, \vartheta_2, \eta)$ denote the log- likelihood ratio of $P^n_{\vartheta_2,\eta}$ to $P^n_{\vartheta_1,0}$:

$$\Lambda_n(\vartheta_1, \vartheta_2, \eta) = \sum_{j=1}^{n} \log \frac{f_\eta(X_j - H_j(\vartheta_2))}{f(X_j - H_j(\vartheta_1))} + \log \frac{g_{\vartheta_2,f_\eta}(X_0)}{g_{\vartheta_1,f}(X_0)}.$$

If Assumption 5.1 holds and Assumption 5.2 equation (5.2.3), (5.2.4) and (5.2.6) are met with $\theta_n = \theta$, then it implies that $\mathcal{L}(\mathbf{X}_n \mid P_{\vartheta_n})$ and $\mathcal{L}(\mathbf{X}_n \mid P_\theta)$ are mutually contiguous for each local sequence $< \vartheta_n >$. That is, in terms of likelihoods formed by probability measures P_{ϑ_n} and P_θ, $\Lambda_n(\vartheta_1, \vartheta_2, \eta) = O(1)$ under both P_{ϑ_n} and P_θ.

The following gives the LAN result for the models described above.

Theorem 5.1 (Theorem 2.4 of Koul and Schick (1997)) *Suppose Assumptions 5.1 and 5.2 hold and the path* $\eta \rightarrow f_\eta$ *is* ξ*-regular and* $V(\xi)$ *is positive definite. Let* $< \theta_n >$ *be a local sequence and* $< \vartheta_n > = << t_n, u_n >>$

be a bounded sequence in $\Re^m \times \Re^s$. *Then*

$$\Lambda_n(\theta_n, \theta_n + \frac{t_n}{\sqrt{n}}, \frac{u_n}{\sqrt{n}}) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \vartheta_n^T S_j(\theta_n, \xi) - \frac{1}{2} \vartheta_n^T V(\xi) \vartheta_n + o_{\theta_n}(1)$$

and

$$\mathcal{L}\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n} S_j(\theta_n, \xi) \mid P_{\theta_n} \right) \Rightarrow \mathcal{N}(0, V(\xi)).$$

Consequently,

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} (S_j(\theta_n, \xi) - S_j(\theta, \xi)) + \begin{bmatrix} JM \\ \int l\xi dF \vartheta^T \end{bmatrix} \sqrt{n}(\theta_n - \theta) = o_\theta(1).$$

Under the semiparametric version of LAN as mentioned above, one can characterize efficient estimates and explain Stein's (1956) necessary condition for adaptation.

Condition 5.1 (Necessary condition for adaptation, Stein (1956)) *Let* $\mathcal{Q}$ *be the family of all regular paths. The necessary condition for adaptation is*

$$\nu \int l\xi_q dF = 0, \quad \text{for each} \quad q \in \mathcal{Q}, \tag{5.2.9}$$

where ν *as defined in Assumption 5.2,* l *as defined in Assumption 5.1 and* ξ_q *is the* ξ *given in Definition 5.1 for the path* q.

Furthermore, under the following additional assumption, Condition 5.2, the construction of adaptive estimates is possible without splitting the sample for symmetric error models as given in Theorem 5.2. The estimates obtained without splitting the sample should give better estimates for moderate sample sizes (Koul and Schick (1997)).

Let $< a_n >$ and $< b_n >$, be sequences of positive numbers converging to 0, $< c_n >$, be a sequence of positive numbers tending to infinity and $< d_n >$ be a sequence of positive integers such that $d_n/n \to 0$.

Condition 5.2 (Condition 5.1 of Koul and Schick (1997)) *For every local sequence* $< \theta_n >$ *for* θ *and every sequence* $< c_n >$ *tending to infinity,*

$$\frac{1}{n}\sum_{j=1}^{n} \|\dot{H}_j(\theta_n)\|^2 I\left[\|\dot{H}_j(\theta_n)\| > c_n\right] = o_{\theta_n}(1).$$

Define the estimate

$$\hat{\theta}_n = \tilde{\theta}_n + (\hat{J}_n M_n)^{-1}\frac{1}{N_n}\sum_{j=1}^{n} \dot{H}_{n,j}\hat{l}_n(\epsilon_{n,j}), \qquad (5.2.10)$$

where $\hat{\theta}_n$ is an estimate for θ, $\tilde{\theta}_n$ be a preliminary estimate which is a discretized $\sqrt{n}$- consistent estimator of θ. The functions $\dot{H}_{n,j}$, $\hat{l}_n$, $\hat{J}_n$, M_n are the estimates of $\dot{H}_j$, l, J, M respectively as defined in the algorithm of the construction of adaptive estimates in Section 5.4 (refer Koul and Schick (1997) for more details).

In classical parametric models the maximum likelihood estimator is typically asymptotically efficient. In semiparametric models such an estimation principle yielding efficient estimator does not exist. However, there exist methods to upgrade $\sqrt{n}$-consistent estimators to be efficient by a Newton-Raphson techniques, provided it is possible to estimate the relevant score or influence functions sufficiently accurate. In Klaassen (1987), such a method based on 'sample splitting' is described. Schick (1986) uses both 'sample splitting' and Le Cam's 'discretization' in *i.i.d.* models. Suppose $< \tilde{\theta}_n >$ is a $\sqrt{n}$-consistent estimator of θ. Then $< \tilde{\theta}_n >$ is called a discrete sequence of estimators when $\tilde{\theta}_n$ is given by one of the vertices of $\{\theta : \theta = n^{-1/2}(i_1, ..., i_{p+q}), i_j \in Z\}$ nearest to θ_n. For example, if $\tilde{\theta}_n = (\tilde{\theta}_{n1}, \tilde{\theta}_{n2}) \in \Re^2$, then $\tilde{\theta}_n$ is discretized by changing its value in $\Re \times \Re$ into (one of) the nearest point(s) in the grid $(c\sqrt{n})(\mathcal{Z} \times \mathcal{Z})$.

Theorem 5.2 (Theorem 5.2 of Koul and Schick (1997)) *Let Assumptions 5.1 and 5.2 and Condition 5.2 hold, and let* f *be symmetric about zero. Suppose* $< \tilde{\theta}_n >$ *is a discretized* $\sqrt{n}$*-consistent estimator of* θ *and the* *sequences* $< a_n >, < b_n >$ *and* $< c_n >$ *satisfy in addition*

$$n^{-1}a_n^{-3}b_n^{-1}c_n^2 \to 0.$$

Then $< \hat{\theta}_n >$, *given in (5.2.10), satisfies*

$$\sqrt{n}(\hat{\theta}_n - \theta) - \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(JM)^{-1}\dot{H}_j(\theta)l(\varepsilon_j(\theta)) = o_\theta(1) \tag{5.2.11}$$

and is adaptive.

For the set Q of all regular paths the above theorem states that if $\mathcal{F}$ contains only symmetric densities around zero, this estimate given in (5.2.10) is adaptive for the LAN subproblems generated by every $q \in Q$. Candidates for $\sqrt{n}$-consistent estimators in the above theorem are Conditional Least Squares, Maximum likelihood, M-estimator or estimating function estimator. These estimators are discussed in Chapter 4.

To see the efficiency of the estimates, the following theorems can be applied under the additional assumption, Assumption 5.3, given below. Let T_Q denote the closed linear span generated by $\cup_{q \in Q} T_q$, where $T_q = \{a^T \xi_q : a \in \Re^{s_q}\}$, s_q denote the dimension of Q and ξ_q is the smoothness parameter. Let l_* denotes the projection of l onto T_Q.

Assumption 5.3 (Assumption 3.1 of Koul and Schick (1997)) *The score function* l *does not belong to* T_Q. *There exists a path* $q_* \in Q$ *such that* $l_* \in T_{q_*}$.

Define the efficient information for estimating θ by

$$I_* = JM - \nu\nu^T \int l_*^2 dF,$$

where J is defined in Assumption 5.1, ν, M are defined in Assumption 5.2 and l_* as mentioned in Assumption 5.3. The efficiency for general estimates is given in Theorem 5.3.

Theorem 5.3 (Theorem 3.2 of Koul and Schick (1997)) *Let Assumptions 5.1, 5.2 and 5.3 hold. Let* $< Z_n >$ *be an estimate satisfying*

$$\sqrt{n}(Z_n - \theta) - \frac{1}{\sqrt{n}}\sum_{j=1}^{n} I_*^{-1}\left(\dot{H}_j(\theta)l(\varepsilon_j(\theta)) - \nu l_*(\varepsilon_j(\theta))\right) = o_\theta(1). \quad (5.2.12)$$

Then

$$\mathcal{L}\left(\sqrt{n}(Z_n - \theta_n) \mid P_{\theta_n, q(u_n/\sqrt{n})}\right) \Longrightarrow \mathcal{N}\left(0, I_*^{-1}\right)$$

for every local sequence $< \theta_n >$, *every* $q \in \mathcal{Q}$ *and every bounded sequence* u_n *in* $\Re^s$. *Consequently,* $< Z_n >$ *is efficient.*

The following theorem gives a method of obtaining the efficient estimates with splitting the sample. Drost, *et.al.* (1997) adopted the principle of splitting the sample because it yields a relatively easy way to obtain efficient estimators under minimal conditions.

Define the estimate $< \hat{\theta}_n >$ by

$$\hat{\theta}_n = \tilde{\theta}_n + \left(\frac{1}{n}\sum_{j=d_n}^{n} \hat{\psi}_{n,j}\hat{\psi}_{n,j}^T\right)^{-1} \frac{1}{n}\sum_{j=d_n}^{n} \hat{\psi}_{n,j}, \quad (5.2.13)$$

where the construction of $\psi(.)$, which contains functions L_n and $L_{*,n}$, is explained in Section 5 (refer Koul and Schick (1997) for more details).

Theorem 5.4 (Theorem 4.1 of Koul and Schick (1997)) *Let Assumptions 5.1, 5.2 and 5.3 hold. Suppose that* $< \tilde{\theta}_n >$ *is a discretized* $\sqrt{n}$*-consistent estimate of* θ *and that the functions* L_n *and* $L_{*,n}$ *are such that for independent random variables* $Y_1, ..., Y_n$ *with density* f

$$\int (L_n(x, Y_1, ..., Y_n) - l(x))^2 f(x)dx \to 0 \quad \text{in probability},$$

$$\int (L_{*,n}(x, Y_1, ..., Y_n) - l_*(x))^2 f(x)dx \to 0 \quad \text{in probability},$$

$$\sqrt{n}\int (L_n(x, Y_1, ..., Y_n) - L_{*,n}(x, Y_1, ..., Y_n))f(x)dx \to 0 \quad \text{in probability}.$$

Then $< \hat{\theta}_n >$ *satisfies (5.2.12) and hence is efficient.*

5.3 Local Asymptotic Normality for STAR models

In this section, it is shown that model (5.1.1) is LAN using the sufficient conditions for LAN given in Theorem 5.1. To apply the theorem, one needs to show that Assumptions 5.1, 5.2 and Definition 5.1 are satisfied. For this model, at first stage, Assumption 5.1 is verified, followed by the verification of Assumption 5.2 and then the existence of a smooth and regular path, as given in Definition 5.1, is proved.

Consider the first-order STAR model of (5.1.1). Let $\mathcal{F} = \mathcal{F}_0^+$, the set of all positive Lebesgue densities with zero-mean, finite variances and finite Fisher information for location. Take the sufficient condition for ergodicity for any distribution function G as mentioned in Chapter 2,

$$\Theta = \{\theta \in \Re^2 : \theta_1 < 1, \theta_1 + \theta_2 < 1, \theta_1(\theta_1 + \theta_2) < 1\}. \qquad (5.3.1)$$

By Theorem 2.2, for each $(\vartheta, \phi) \in \Theta \times \mathcal{F}$, the above model is ergodic and there is a unique stationary process satisfying the above model (see Proposition 2.1 of Chan and Tong (1986) (Proposition A.1 of Appendix A)).

For STAR model of order 1 given in (5.1.1), let

$$H_j(\theta_n) = \theta_{n1} X_{j-1} + \theta_{n2} X_{j-1} G(X_{j-1}), \qquad (5.3.2)$$

and

$$\dot{H}_j(\theta_n) = \left(\frac{\partial}{\partial \theta_{ni}} H_j(\theta_n)\right) = \begin{pmatrix} X_{j-1} \\ X_{j-1} G(X_{j-1}) \end{pmatrix}. \qquad (5.3.3)$$

On Assumption 5.1

If we take $\mathcal{F} = \mathcal{F}_0$, that is, the set of all Lebesgue densities with zero-mean, finite variances and finite Fisher information for location, we only need to verify condition (5.2.2) for the Assumption 5.1 to hold. It is clear that condition (5.2.2) together with condition (5.2.8) of Definition 5.1 reflect that the initial distribution has negligible effect which is guaranteed by the L_1-continuity of the map $(\vartheta, \phi) \rightarrow g_{\vartheta,\phi}$ at (θ, f). The sufficient condition for the L_1-continuity in stationary and ergodic NLAR(1) models was given by Koul and Schick (1997) in their appendix. The verification of condition (5.2.2) is implied by the verification of condition (5.2.8), as the latter is for any (θ, f) and the former is for any θ but fixed f.

We prove in the following proposition that Assumption 5.2 is satisfied for the STAR model (5.1.1).

Proposition 5.1 *For the STAR model (5.1.1) with* Θ *given in (5.3.1) and* $\mathcal{F} = \mathcal{F}_0^+$, *the Assumption 5.2 is satisfied.*

Proof

By Remark 2.6 of Koul and Schick (1997), under Assumption 5.1, it is enough to verify condition (5.2.3)-(5.2.6) hold with $\theta_n = \theta$ and that

$$\frac{1}{n}\sum_{j=1}^{n} \| \dot{H}_j(\theta_n) - \dot{H}_j(\theta) \|^2 = o_\theta(1).$$

But, as $\dot{H}_j(\theta)$ given in (5.3.3) does not depend on the parameter θ, the above last condition is obviously satisfied.

To verify condition (5.2.3), it should be proved that

$$P_{\theta_n}\left[\sum_{j=1}^{n} | H_j(\vartheta_n) - H_j(\theta_n) - (\vartheta_n - \theta_n)^T \dot{H}_j(\theta_n) |^2 > \varepsilon\right] \rightarrow 0.$$

Since $\theta_n = \theta$ and $\vartheta_n = \theta + n^{-1/2}t_n$ and by using (5.3.2) and (5.3.3), we have

$$\begin{aligned}
\textstyle\sum_{j=1}^{n} \quad & | H_j(\vartheta_n) - H_j(\theta_n) - (\vartheta_n - \theta_n)^T \dot{H}_j(\theta_n) |^2 \\
= \quad & \sum_{j=1}^{n} | \theta_1 X_{j-1} + n^{-1/2}t_n X_{j-1} + \theta_2 X_{j-1}G(X_{j-1}) + n^{-1/2}t_n X_{j-1}G(X_{j-1}) \\
- \quad & \theta_1 X_{j-1} - \theta_2 X_{j-1}G(X_{j-1}) - n^{-1/2}t_n X_{j-1} - n^{-1/2}t_n X_{j-1}G(X_{j-1}) |^2 \\
= \quad & 0,
\end{aligned}$$

which proves (5.2.3).

To prove (5.2.4), one needs to show that

$$P_{\theta_n}\left[\max_{1\le j\le n}\frac{1}{\sqrt{n}}\sqrt{X_{j-1}^2(1+G^2(X_{j-1}))}\ >\ \varepsilon\right]\ \to 0.$$

Let $\mid Y_{j-1}\mid=\mid X_{j-1}\sqrt{(1+G^2(X_{j-1}))}\mid$. As $G(x)\ \le\ 1$ for any $x\ \in\ \Re$, then it follows that $\mid Y_{j-1}\mid\le 2\mid X_{j-1}\mid$, $\forall\, j\ \ge\ 1$. Under the assumption that $\{X_j\}$ is stationary and ergodic, and that $EX_{j-1}^2\ <\ \infty$, we have $EY_{j-1}^2\ <\ \infty$. Then by using conditional Chebyshev inequality, the stationarity of the process $\{X_j\}$ and the square integrability of X_0 we have

$$\begin{aligned}
&P_{\theta_n}\left(\max_{1\le j\le n}\frac{\mid Y_{j-1}\mid}{\sqrt{n}}>\varepsilon\right)\le P_{\theta_n}\left(\max_{1\le j\le n}\frac{\mid X_{j-1}\mid}{\sqrt{n}}>\frac{\varepsilon}{2}\right)\\
\le\ &\frac{4}{\varepsilon^2 n}\sum_{j=1}^{n}EX_{j-1}^2I\left(\mid X_{j-1}\mid>\frac{\varepsilon\sqrt{n}}{2}\right)\\
=\ &\frac{4}{\varepsilon^2}EX_0^2I\left(\mid X_0\mid>\frac{\varepsilon\sqrt{n}}{2}\right)\to 0.
\end{aligned}$$

Hence (5.2.4) follows.

To prove condition (5.2.5) and (5.2.6), we need to use the Theorems 5.5 and 5.6 of Karlin and Taylor (1975) (Theorem A.3 and A.4, Appendix A). We need to prove that:

$$\begin{pmatrix}\frac{1}{n}\sum_{j=1}^{n}X_{j-1}\\ \frac{1}{n}\sum_{j=1}^{n}X_{j-1}G(X_{j-1})\end{pmatrix}\xrightarrow{P}\nu,$$

where

$$\nu=\begin{pmatrix}E_\theta X_0\\ E_\theta(X_0G(X_0))\end{pmatrix},$$

and

$$\begin{pmatrix}\frac{1}{n}\sum_{j=1}^{n}X_{j-1}^2 & \frac{1}{n}\sum_{j=1}^{n}X_{j-1}^2G(X_{j-1})\\ \frac{1}{n}\sum_{j=1}^{n}X_{j-1}^2G(X_{j-1}) & \frac{1}{n}\sum_{j=1}^{n}X_{j-1}^2G^2(X_{j-1})\end{pmatrix}\xrightarrow{P}M,$$

where

$$M = \begin{pmatrix} E_\theta X_0^2 & E_\theta(X_0^2 G(X_0)) \\ E_\theta(X_0^2 G(X_0)) & E_\theta(X_0^2 G^2(X_0)) \end{pmatrix}.$$

Under the assumption that $\{X_j\}$ is ergodic it follows, by using Theorem 5.5 of Karlin and Taylor (1975) (Theorem A.3, Appendix A), that

$$\frac{1}{n}\sum_{j=1}^{n} X_{j-1} \xrightarrow{a.s.} E_\theta X_0.$$

Moreover, using Theorem 5.6 of Karlin and Taylor (1975) (Theorem A.4, Appendix A), it follows that

$$\begin{aligned}
&\tfrac{1}{n} \sum_{j=1}^{n} X_{j-1} G(X_{j-1}) \xrightarrow{a.s.} E_\theta(X_0 G(X_0)), \\
&\tfrac{1}{n} \sum_{j=1}^{n} X_{j-1}^2 \xrightarrow{a.s.} E_\theta X_0^2, \\
&\tfrac{1}{n} \sum_{j=1}^{n} X_{j-1}^2 G(X_{j-1}) \xrightarrow{a.s.} E_\theta(X_0^2 G(X_0)), \\
&\tfrac{1}{n} \sum_{j=1}^{n} (X_{j-1}^2 G^2(X_{j-1}))a. \xrightarrow{a.s.} E_\theta X_0^2 G^2(X_0)
\end{aligned}$$

From the finiteness of second moments and as almost sure convergence implies convergences in probability, it follows that the STAR process satisfy the conditions (5.2.5) and (5.2.6).

Combining all, the proof of Proposition 5.1 is completed. □

Now we show the existence of a regular path for the STAR model. That is, we show that there exists a path and a measurable function ξ which satisfy (5.2.7) and (5.2.8). If we take $\xi(x) = x\frac{f^{(1)}(x)}{f(x)} + 1$ then the following lemma shows that there exists a ξ-smooth path.

Lemma 5.1 *Let* $\mathcal{F} = \mathcal{F}_0^+$, *and* $\xi(x) = x\frac{f^{(1)}(x)}{f(x)} + 1$. *Then there exists a one-dimensional* ξ*-smooth path.*

Proof. Note that $\int f(x)dx = 1$, $\int xf(x)dx = 0$, $\int xl(x)f(x)dx = 1$. Then

$$\int \xi(x)dF(x) = \int \left(x\frac{f^{(1)}(x)}{f(x)} + 1\right) dF(x)$$

$$= \int xf^{(1)}(x)dx + \int f(x)dx = 0,$$

which follows from integration by parts.

Moreover,

$$\int x\xi(x)dF(x) = \int x\left(x\frac{f^{(1)}(x)}{f(x)}+1\right)dF(x)$$
$$= \int x^2 f^{(1)}(x)dx + \int xf(x)dx = 0,$$

which also follows from integration by parts.

Since $\int \xi(x)dF(x) = 0$

$$\int (\xi(x))^2 dF(x) = Var(\xi(\varepsilon_t)) > 0.$$

Hence by Remark 5.1 it follows that there exists a one-dimensional path that is ξ-smooth. □

In order to show that the STAR model is LAN, we need to show that there exists a smooth path which satisfy Definition 5.1 equation (5.2.8). By Remark 2.7 of Koul and Schick (1997) for a stationary and ergodic NLAR(1) process so that $H_j(\vartheta) = h(X_{j-1}, \vartheta)$, it follows that to show (5.2.8) for given $\xi(x)$, it is enough to show that there exists positive constant A and a measurable non-negative function ψ such that

$$| h(x,\vartheta) | \leq A\psi(x), \quad x \in \Re, \tag{5.3.4}$$

$$| h(x,\vartheta) - h(x,\theta) | \leq \| \vartheta - \theta \| A\psi(x), \quad x \in \Re, \tag{5.3.5}$$

for all ϑ close to θ, and for a ξ-smooth path $\eta \mapsto f_\eta$ that satisfies $\limsup_{\eta\to 0} \int | x | f_\eta(x)dx < \infty$ so that

$$\limsup_{\vartheta\to\theta\; \eta\to 0} E_{\vartheta,f_\eta}\psi(X_0) < \infty. \tag{5.3.6}$$

As condition (5.3.6) for the STAR models is not in a closed form, we verify this condition using the sufficient condition given in Remark 2.7 of Koul and Schick (1997)(Remark C.1 of Appendix C). The following verifies the above conditions.

Proposition 5.2 *Let* $\mathcal{F} = \mathcal{F}_0^+$ *and* $\{f_\eta\}_{\eta>0}$ *be a* ξ*-smooth path that satisfies* $\limsup_{\eta\to 0} \int |x| f_\eta(x)dx < \infty$.
Let, for $\theta = (\theta_1, \theta_2) \in \Theta$, *given in (5.3.1),*

$$h(x,\theta) = \theta_1 x + \theta_2 x G(x), \quad x \in \Re.$$

Then there exists a measurable non-negative function ψ *and a constant* A *such that (5.3.4), (5.3.5) and (5.3.6) are satisfied.*

Proof.

In order to do this, we use the function ψ introduced in Chapter 2 for proving the ergodicity of the models. In the following, we summarise the proof to relate it to the condition (5.3.6) (see Example 2.8 of Koul and Schick (1997) for $SETAR(2;1,1)$ model). Let $\mathcal{F} \subset \mathcal{F}_0^+$ and Θ as defined in (5.3.1). From the model (5.1.1), we have

$$h(x,\theta) = \theta_1 x + \theta_2 x G(x), \quad x \in \Re,$$

is ergodic for each $(\theta, \phi) \in \Theta \times \mathcal{F}$. From Chapter 2, the ergodicity proof is divided into three cases. For convenience, we combine the cases as follows. From the condition in (5.3.1), it is possible to choose positive constants c, d, e with $e < 1$, satisfying

$$-\frac{d}{c} < \theta_1 < e, \quad -\frac{c}{d} < \theta_1 + \theta_2 < e,$$

and define the required non-negative measurable function as

$$\psi(x) = \begin{cases} cx, & x > 0 \\ -dx, & x \le 0. \end{cases}$$

To prove (5.3.4), we have

$$| h(x,\vartheta) | = | \vartheta_1 x + \vartheta_2 x G(x) | = | \vartheta_1 + \vartheta_2 G(x) | \, | x | \le A\psi(x),$$

for some positive constant A. Similarly, to prove (5.3.5), we have

$$| \quad h(x,\vartheta) - h(x,\theta) | = | \vartheta_1 x + \vartheta_2 x G(x) - \theta_1 x - \theta_2 x G(x) |$$

$$
\begin{aligned}
&= \mid (\vartheta_1 - \theta_1)x + (\vartheta_2 - \theta_2)xG(x) \mid \\
&= \mid (\vartheta_1 - \theta_1 \quad \vartheta_2 - \theta_2)(x \quad xG(x))^T \mid \\
&\leq \parallel \vartheta - \theta \parallel A\psi(x), \quad x \in \Re.
\end{aligned}
$$

Hence (5.3.4) and (5.3.5) hold. Now, we show that (5.3.6) holds as follows. As in the proof of ergodicity, we have

$$
\begin{aligned}
E(\psi(X_t) \mid X_{t-1} = x) &= \begin{cases} cx(\theta_1 + \theta_2 G(x)) + (c+d)E[h(x,\theta) + \varepsilon_t]^-, & x > 0 \\ -dx(\theta_1 + \theta_2 G(x)) + (c+d)E[h(x,\theta) + \varepsilon_t]^+, & x \leq 0 \end{cases} \\
&\leq \psi(x)(\theta_1 + \theta_2 G(x)) \\
&+ (c+d)E \mid h(x,\theta) + \varepsilon_t \mid, \quad x \in \Re \\
&\leq e\psi(x) + (c+d)E \mid h(x,\theta) + \varepsilon_t \mid, \quad x \in \Re,
\end{aligned}
$$

as $\theta_1 + \theta_2 G(x) < e$ from the above condition.

From condition (5.3.4) and $E(\varepsilon_t^2) < \infty$, it is clear that

$$
E \mid h(x,\theta) + \varepsilon_t \mid \leq \mid h(x,\theta) \mid + E \mid \varepsilon_t \mid \leq A\psi(x) + E \mid \varepsilon_t \mid < \infty, \; x \in \Re.
$$

Therefore (2.24) in Remark 2.7 of Koul and Schick (1997) (equation (C.0.1), Remark C.1 of Appendix C) is satisfied with $E \mid h(x,\theta) + \varepsilon_t \mid \leq C, \; x \in \Re$ and $\delta = \frac{1-e}{2}$. □

From the above proposition it follows that (5.2.2) holds and every smooth path that satisfy the conditions in Proposition 5.2 is regular. Hence we have shown that when $\mathcal{F} = \mathcal{F}_0^+$ the STAR model given in (5.1.1) is LAN which is stated in the following theorem.

Theorem 5.5 *Let $\mathcal{F} = \mathcal{F}_0^+$. Then the STAR model (5.1.1) is LAN.*

Remark 5.2 *It should be noted that the verification of condition (5.2.8) of Assumption 5.2 implies the verification of condition (5.2.2) of Assumption 5.1.*

5.4 Adaptivity

This section discusses the adaptivity of the STAR model. In order to show that STAR model is adaptive, we need to verify Condition 5.2. The following gives a construction of an adaptive estimator given by Koul and Schick (1997). Starting with a discretized $\sqrt{n}$-consistent estimator, we prove in the next theorem, that the estimator is adaptive.

(1) Choose $< a_n >$ and $< b_n >$, sequences of positive numbers converging to 0 and $< c_n >$, a sequence of positive numbers converging to infinity such that

$$n^{-1}a_n^{-3}b_n^{-1}c_n^2 \to 0.$$

(2) Let $< d_n >$ be a sequence of positive integers such that $d_n/n \to 0$. Set $N_n = n - d_n + 1$.

(3) Let $\tilde{\theta}_n$ be a preliminary estimate of θ which is a discretized and $\sqrt{n}$-consistent estimate.

(4) Let χ_n denote the map from $\Re^m$ into $\Re^m$ defined by

$$\chi_n(\mathbf{x}) = \mathbf{x}I\left[\|\mathbf{x}\| \le c_n\right] + c_n\frac{\mathbf{x}}{\|\mathbf{x}\|}I\left[\|\mathbf{x}\| > c_n\right], \quad \mathbf{x} \in \Re^m.$$

For the model, calculate

$$\begin{aligned}
\dot{H}_{n,j} &= \chi_{N_n}(\dot{H}_j(\tilde{\theta}_n)) \\
&= \begin{pmatrix} X_{j-1} \\ X_{j-1}G(X_{j-1}) \end{pmatrix} \Big\{ I\left[\sqrt{X_{j-1}^2 + X_{j-1}^2G^2(X_{j-1})} \le c_n\right] \\
&+ \frac{c_n}{\sqrt{X_{j-1}^2 + X_{j-1}^2G^2(X_{j-1})}} I\left[\sqrt{X_{j-1}^2 + X_{j-1}^2G^2(X_{j-1})} > c_n\right] \Big\}
\end{aligned} \tag{5.4.1}$$

(5) Set $\varepsilon_{n,j} = \varepsilon_j(\tilde{\theta}_n) = X_j - \left(\tilde{\theta}_{n1}X_{j-1} + \tilde{\theta}_{n2}X_{j-1}G(X_{j-1})\right).$

(6) Estimate the score function l by

$$\hat{l}_n(\varepsilon_{n,i}) = L_{N_n}(\varepsilon_{n,i}, \varepsilon_{n,d_n}, ..., \varepsilon_{n,n}), \ x \ \in \ \Re,$$

where L_n is defined as

$$L_{N_n} \quad (\varepsilon_{n,i}, \varepsilon_{n,d_n}, ..., \varepsilon_{n,n}) = \frac{f_n^{(1)}(\varepsilon_{n,i}, \varepsilon_{n,d_n}, ..., \varepsilon_{n,n}) - f_n^{(1)}(-\varepsilon_{n,i}, \varepsilon_{n,d_n}, ..., \varepsilon_{n,n})}{b_n + f_n(\varepsilon_{n,i}, \varepsilon_{n,d_n}, ..., \varepsilon_{n,n}) + f_n(-\varepsilon_{n,i}, \varepsilon_{n,d_n}, ...\varepsilon_{n,n})}$$

and

$$f_n(\varepsilon_{n,i}, \varepsilon_{n,d_n}, ..., \varepsilon_{n,n}) = \frac{1}{na_n}\sum_{j=1}^{n} k\left(\frac{\varepsilon_{n,i} - \varepsilon_{n,j}}{a_n}\right), \tag{5.4.2}$$

$$f_n^{(1)}(\varepsilon_{n,i}, \varepsilon_{n,d_n}, ..., \varepsilon_{n,n}) = \frac{1}{na_n^2}\sum_{j=1}^{n} k^{(1)}\left(\frac{\varepsilon_{n,i} - \varepsilon_{n,j}}{a_n}\right) \tag{5.4.3}$$

for $i = 1, ..., n$, and k be the logistic density,

$$k(x) = \exp(-x)(1 + \exp(-x))^{-2}, \quad x \ \in \ \Re.$$

(7) Calculate

$$\hat{J}_n = \frac{1}{n}\sum_{j=d_n}^{n} \hat{l}_n^2(\varepsilon_{n,j}) \ \text{ and } \ M_n = \frac{1}{n}\sum_{j=d_n}^{n} \dot{H}_{n,j}\dot{H}_{n,j}^T.$$

(8) Finally, define the estimate (5.2.10)

$$\hat{\theta}_n = \tilde{\theta}_n + (\hat{J}_n M_n)^{-1}\frac{1}{N_n}\sum_{j=1}^{n} \dot{H}_{n,j}\hat{l}_n(\varepsilon_{n,j}). \tag{5.4.4}$$

Theorem 5.6 *Let $\mathcal{F} = \mathcal{F}_S^+$, the set of all positive Lebesgue densities which are symmetric, having zero means, finite variances and finite Fisher information for location. Then $\hat{\theta}_n$ given in (5.4.4) is adaptive.*

Proof.

If $\mathcal{F} = \mathcal{F}_S^+$ then l is odd and hence (5.2.9) holds. Therefore Condition 5.1 is satisfied. Hence to prove the theorem using Theorem 5.2, one needs to verify Condition 5.2 along with Assumptions 5.1 and 5.2. From the previous section, it is shown that Assumptions 5.1 and 5.2 hold when $\mathcal{F} = \mathcal{F}_0^+$. For a stationary and ergodic NLAR(1) models, Koul and Schick (1997) noted in Remark 5.4 (Remark C.2 of Appendix C) that one can provide simple sufficient condition for Condition 5.2 as follows. A sufficient condition for Condition 5.2 to hold is that there exists a function ψ such that

(i) $E_\theta \psi(X_0) < \infty$,

(ii) $\psi(x) \geq \sup_{\|\vartheta-\theta\|<\delta} \| \dot{h}(x,\vartheta) \|^2$, for all $x \in \Re$ and some $\delta > 0$.

This condition holds for the stationary and ergodic STAR models given in (5.1.1) by taking $\psi(x) = Ax^2$, $A > 2$.

- For condition (i), $E_\theta A X_0^2 < \infty$ which follows from $E_\theta X_0^2 < \infty$.
- For condition (ii),

$$\sup_{\|\vartheta-\theta\|<\delta} \| \dot{h}(x,\vartheta) \|^2 = \sup_{\|\vartheta-\theta\|<\delta} x^2(1 + G^2(x)) \leq Ax^2$$

  for $x \in \Re$ and for some $\delta > 0$, $A > 2$.

Therefore by Theorem 5.2, each regular path generates a LAN-subproblem and the estimator $< \hat{\theta}_n >$ given in (5.2.11) is adaptive for every class of LAN-subproblems that satisfies Stein's condition (5.2.9). Hence the theorem is proved. □

Remark 5.3 *Note that the Stein's necessary condition (5.2.9) given in Condition 5.1, is satisfied if either*

$$\nu = 0 \tag{5.4.5}$$

or

$$\int l\xi_q dF = 0, \quad \forall\, q \in \mathcal{Q}. \tag{5.4.6}$$

For STAR order 1 model,

$$\nu = \begin{pmatrix} E_\theta X_0 \\ E_\theta(X_0 G(X_0)) \end{pmatrix} \neq 0.$$

If $\mathcal{F}$ includes asymmetric densities, then (5.4.6) fails to hold and Stein's condition is equivalent to (5.4.5). In this case, adaptive estimation is ruled out if $\nu \neq 0$ which is the case for STAR models. Therefore in Theorem 5.6 we have taken $\mathcal{F} = \mathcal{F}_S^+$, the set of all positive Lebesgue densities which are symmetric, having zero means, finite variances and finite Fisher information for location in the above theorem.

5.5 Efficient Estimator

In this section we adopt the construction of efficient estimator given in Koul and Schick (1997) to STAR models. We give the construction of efficient estimates under sample splitting techniques. First we show that Assumption 5.3 is satisfied and use Theorem 5.3 to infer that the constructed estimator is efficient. As given in Example 3.5 of Koul and Schick (1997), we let $\mathcal{F} = \mathcal{F}_0^+$, $\mathcal{Q}$ is the set of all regular paths. Then

$$T_\mathcal{Q} = \{a \in L_2(F) : \int a(x)f(x)dx = 0 \text{ and } \int xa(x)f(x)dx = 0\}.$$

As $\int xl(x)dF(x) = 1$, the score function l does not belong to $T_\mathcal{Q}$.
Let

$$l_*(x) = l(x) - \frac{x}{\sigma^2}, \quad x \in \Re,$$

where σ^2 denotes the variance of f. Clearly

$$\begin{aligned} \int l_*(x)dF(x) &= 0, \\ \int xl_*(X)dF(x) &= 0, \\ 0 < \int l_*^2(x)dF(x) &< \infty. \end{aligned}$$

Hence by Definition 5.1, there exists a one dimensional path q_* which is l_*-smooth. Also the l_*-smooth path q_*, is regular since

$$\lim_{\vartheta \to 0} \sup_{\eta \to 0} E_{\vartheta, f_\eta} \psi(X_0) < \infty$$

by Proposition 5.2. Therefore, Assumption 5.3 holds.

Hence for the STAR models we have shown the following.

Theorem 5.7 *Let* $\mathcal{F} = \mathcal{F}_0^+$. *Then* Z_n *satisfying (5.2.12) is efficient.*

For STAR model, if $\mathcal{F} = \mathcal{F}_S^+$, then Stein's condition for adaptivity is satisfied. Hence the efficient estimator is also adaptive. Whereas, if $\mathcal{F} = \mathcal{F}_0^+$ then Stein's condition is not satisfied. Hence Z_n is efficient but not adaptive.

The following gives a method of construction of estimator which satisfy (5.2.14). By choosing $\mathcal{F} = \mathcal{F}_0^+$, and using Theorem 5.3, the estimator given is efficient by the above theorem.

Sample Splitting method

The method is to split the residuals $\hat{\varepsilon}_{n1}, ..., \hat{\varepsilon}_{nn}$ into two samples which may be viewed as independent. As shown in (5.5.1), the first group of residuals is used to estimate the function $\psi(.)$ of the second group errors, whereas the second group of residuals is used to estimate the function $\psi(.)$ of the first group errors. Following gives an efficient estimator when the error density is in $\mathcal{F}_0^+$.

(1) Let $< \tilde{\theta}_n >$ be a preliminary estimate of θ which is discretized and $\sqrt{n}$-consistent.

(2) Set $\varepsilon_{n,j} = \varepsilon_j(\tilde{\theta}_n)$, $j = 1, ..., n$.

(3) Let $< d_n >$ and $< m_n >$ be sequences of positive integers such that $d_n \leq m_n \leq n$, $d_n/n \to 0$ and $m_n/n \to 1/2$ and choose $< a_n >$ and

$< b_n >$ to be sequences of positive numbers converging to 0 such that

$$n^{-1}a_n^{-3}b_n^{-1} \to 0.$$

(4) Set $N_n' = m_n - d_n + 1$ and $N_n'' = n - m_n$.

(5) Calculate

$$\begin{aligned} \hat{\nu}_{1,n} &= \frac{1}{N_n'} \sum_{j=d_n}^{m_n} \dot{H}_j(\tilde{\theta}_n), \\ &= \frac{1}{N_n'} \begin{pmatrix} \sum_{j=d_n}^{m_n} X_{j-1} \\ \sum_{j=d_n}^{m_n} X_{j-1} G(X_{j-1}) \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \hat{\nu}_{2,n} &= \frac{1}{N_n''} \sum_{j=m_n+1}^{n} \dot{H}_j(\tilde{\theta}_n), \\ &= \frac{1}{N_n''} \begin{pmatrix} \sum_{j=m_n+1}^{n} X_{j-1} \\ \sum_{j=m_n+1}^{n} X_{j-1} G(X_{j-1}) \end{pmatrix} \end{aligned}$$

(6) Calculate

$$L_{N_n'}(\varepsilon_{n,i}, \varepsilon_{n,d_n}, \ldots, \varepsilon_{n,m_n}) = -\frac{f_{N_n'}^{(1)}(\varepsilon_{n,i}, \varepsilon_{n,d_n}, \ldots, \varepsilon_{n,m_n})a}{b_n + f_{N_n'}(\varepsilon_{n,i}, \varepsilon_{n,d_n}, \ldots, \varepsilon_{n,m_n})}$$

for $i = m_n + 1, \ldots, n$ where f_n and $f_n^{(1)}$ as in (5.4.4) and (5.4.5).

(7) Similarly, calculate

$$L_{N_n''}(\varepsilon_{n,i}, \varepsilon_{n,m_n+1}, \ldots, \varepsilon_{n,n}) = -\frac{f_{N_n''}^{(1)}(\varepsilon_{n,i}, \varepsilon_{n,m_n+1}, \ldots, \varepsilon_{n,n})}{b_n + f_{N_n''}(\varepsilon_{n,i}, \varepsilon_{n,m_n}, \ldots, \varepsilon_{n,n})}$$

for $i = d_n, \ldots, m_n$.

(8) Calculate

$$L_{*,N_n'}(\varepsilon_{n,i}, \varepsilon_{n,d_n}, \ldots, \varepsilon_{n,m_n}) = L_{N_n'}(\varepsilon_{n,i}, \varepsilon_{n,d_n}, \ldots, \varepsilon_{n,m_n}) - \frac{\varepsilon_{n,i}}{\frac{1}{n}\sum_{j=d_n}^{m_n} \varepsilon_{n,j}^2},$$

for $i = m_n + 1, \ldots, n$.

(9) Similarly, calculate

$$L_{*,N_n''}(\varepsilon_{n,i}, \varepsilon_{n,m_n+1}, \ldots, \varepsilon_{n,n}) = L_{N_n''}(\varepsilon_{n,i}, \varepsilon_{n,m_n+1}, \ldots, \varepsilon_{n,n}) - \frac{\varepsilon_{n,i}}{\frac{1}{n}\sum_{j=m_n+1}^{n} \varepsilon_{n,j}^2},$$

for $i = d_n, \ldots, m_n$.

(10) Let $\mathbf{e}_{n,2} = (\varepsilon_{n,m_n+1}, ..., \varepsilon_{n,n})$ and $\mathbf{e}_{n,1} = (\varepsilon_{n,d_n}, ..., \varepsilon_{n,m_n})$ and calculate

$$\hat{\psi}_{n,j} = \begin{cases} \dot{H}_j(\tilde{\theta}_n) L_{N_n''}(\varepsilon_{n,j}, \mathbf{e}_{n,2}) - \hat{\nu}_{2,n} L_{*,N_n''}(\varepsilon_{n,j}, \mathbf{e}_{n,2}) & \text{if } j = d_n, ...m_n, \\ \dot{H}_j(\tilde{\theta}_n) L_{N_n'}(\varepsilon_{n,j}, \mathbf{e}_{n,1}) - \hat{\nu}_{2,n} L_{*,N_n'}(\varepsilon_{n,j}, \mathbf{e}_{n,1}) & j = m_n + 1, ...n. \end{cases} \tag{5.5.1}$$

(11) Finally, the estimate $< \hat{\theta}_n >$ using (5.2.13)

$$\hat{\theta}_n = \tilde{\theta}_n + \left(\frac{1}{n} \sum_{j=d_n}^{n} \hat{\psi}_{n,j} \hat{\psi}_{n,j}^T \right)^{-1} \frac{1}{n} \sum_{j=d_n}^{n} \hat{\psi}_{n,j}. \tag{5.5.2}$$

Simulation results for this estimator is given in the next section.

By Theorem 5.3, the above estimator satisfies (5.2.12) and hence by Theorem 5.7 we have the following result for STAR models.

Theorem 5.8 *Let $\mathcal{F} = \mathcal{F}_0^+$, the set of all positive Lebesgue densities having zero means, finite variances and finite Fisher information for location. Then $\hat{\theta}_n$ given in (5.5.2) is efficient.*

5.6 Simulation

To demonstrate the applicability the theoretical results of the previous sections, a simulation experiment is presented below. The data comes from a first-order STAR model with delay parameter one;

$$X_t = -2.0 X_{t-1} + 1.6 X_{t-1} G\left(\frac{X_{t-1} - 5.0}{0.5} \right) + \varepsilon_t,$$

where $G(.)$ is a standard Normal distribution function. In the simulation study, 2500 independent replications were generated each with sample size 1000. The following five densities of the errors ε_t were chosen:

$$f_1(x) = (0.5/\sqrt{2\pi}) \exp(-(x-3)^2/50) + (0.5/\sqrt{2\pi}) \exp(-(x+3)^2/2),$$

$$f_2(x) = (0.05/\sqrt{50\pi})\exp(-x^2/50) + (0.95/\sqrt{2\pi})\exp(-x^2/2),$$

$$f_3(x) \leadsto t_5, \quad f_4(x) \leadsto t_7, \quad f_5(x) \leadsto t_9.$$

For each series, conditional on $(r, z) = (5.0, 0.5)$, we estimated the parameter coefficients $\theta = (\theta_1, \theta_2)$ using three methods, viz, conditional least squares, adaptive and efficient estimation. The preliminary estimate $\tilde{\theta}$, used as the initial value for these estimators, is a discretized estimate obtained by the conditional least squares method. Note that $\tilde{\theta}$ is a $\sqrt{n}$-consistent estimator by the results given in Section 4.2.

Table 5.1. Average values of estimates and their sample mean squared error for $a_n = 0.5$.

Estimate	MN1	MN 2	t_5	t_7	t_9
$\hat{\theta}_1(C)$	-2.00006	-1.99997	-1.99996	-1.99994	-1.99956
	(0.000044)	(0.000075)	(0.000124)	(0.000114)	(0.000101)
$\hat{\theta}_2(C)$	1.60007	1.60017	1.60018	1.59992	1.59970
	(0.000057)	(0.000095)	(0.000156)	(0.000140)	(0.000130)
$\hat{\theta}_1(Ad1)$	-2.00006	-1.99997	-1.99989	-2.00002	-1.99955
	(0.000044)	(0.000075)	(0.000101)	(0.000103)	(0.000095)
$\tilde{\theta}_2(Ad1)$	1.59995	1.60007	1.60000	1.60006	1.59974
	(0.000062)	(0.000104)	(0.000139)	(0.000136)	(0.000130)
$\hat{\theta}_1(Ad2)$	-2.00006	-1.99991	-2.00010	-2.00013	-1.99979
	(0.000044)	(0.000078)	(0.000102)	(0.000100)	(0.000097)
$\tilde{\theta}_2(Ad2)$	1.60009	1.59989	1.60017	1.60043	1.59972
	(0.000058)	(0.000103)	(0.000135)	(0.000125)	(0.000123)
$\tilde{\theta}_1(Ef)$	-1.99997	-2.00002	-2.00000	-1.99983	-1.99962
	(0.000050)	(0.000082)	(0.000125)	(0.000117)	(0.000106)
$\tilde{\theta}_2(Ef)$	1.60007	1.60025	1.60022	1.59998	1.59965
	(0.000066)	(0.000106)	(0.000168)	(0.000148)	(0.000136)

The adaptive estimates are constructed by two methods using (5.2.10). The first adaptive estimate, $\hat{\theta}_{Ad1}$, denotes the adaptive estimate with $c_n < \infty$, that is, satisfying Condition 5.2. The second adaptive estimate, $\hat{\theta}_{Ad2}$, is constructed with $c_n = \infty$. The efficient estimate $\hat{\theta}_{Ef}$, as defined in (5.2.13), is constructed using sample splitting technique. The value $(\hat{\theta}_1(C), \hat{\theta}_2(C))$ given in tables is the usual conditional least squares estimator. For all estimates, we used standardised logistic kernels with a bandwith in the interval $0.5 \leq a_n \leq 0.9$.

Table 5.2. Average values of estimates and their sample mean squared error for $a_n = 0.6$.

Estimate	MN1	MN 2	t_5	t_7	t_9
$\hat{\theta}_1(C)$	-1.99988	-2.00003	-1.99981	-1.99994	-2.00002
	(0.000044)	(0.000078)	(0.000129)	(0.000105)	(0.000102)
$\hat{\theta}_2(C)$	1.59992	1.60008	1.59996	1.60023	1.60014
	(0.000057)	(0.000098)	(0.000165)	(0.000135)	(0.000131)
$\hat{\theta}_1(Ad1)$	-1.99990	-2.00003	-2.00000	-2.00001	-1.99992
	(0.000045)	(0.000079)	(0.000109)	(0.000096)	(0.000098)
$\tilde{\theta}_2(Ad1)$	1.59994	1.60008	1.59996	1.60015	1.60002
	(0.000070)	(0.000116)	(0.000162)	(0.000141)	(0.000139)
$\hat{\theta}_1(Ad2)$	-1.99996	-1.99966	-2.00005	-1.99958	-1.99956
	(0.000044)	(0.000076)	(0.000107)	(0.000096)	(0.000095)
$\tilde{\theta}_2(Ad2)$	1.59996	1.59971	1.60028	1.59962	1.59965
	(0.000058)	(0.000098)	(0.000134)	(0.000127)	(0.000118)
$\tilde{\theta}_1(Ef)$	-1.99979	-2.00005	-1.99964	-1.99999	-2.00008
	(0.000058)	(0.000091)	(0.000134)	(0.000111)	(0.000111)
$\tilde{\theta}_2(Ef)$	1.59993	1.59999	1.59992	1.60008	1.60022
	(0.000080)	(0.000116)	(0.000186)	(0.000153)	(0.000150)

The simulation results are presented in Tables 5.1 to 5.5. In each table, for each bandwith, the average values of estimates from 2500 independent replications are given for different error densities. The sample mean squared error of estimates are given in the parantheses.

In Table 5.1 for $a_n = 0.5$, the adaptive estimates with $c_n = \infty$ give the smallest sample mean squared error compared to other estimates when the error densities are t_5, t_7 and t_9 followed by the adaptive estimates with $c_n < \infty$, the CLS and the efficient estimates.

Table 5.3. Average values of estimates and their sample mean squared error for $a_n = 0.7$.

Estimate	MN1	MN 2	t_5	t_7	t_9
$\hat{\theta}_1(C)$	-1.99989	-1.99973	-2.00010	-2.00023	-2.00035
	(0.000041)	(0.000079)	(0.000122)	(0.000108)	(0.000094)
$\hat{\theta}_2(C)$	1.59984	1.59969	1.60032	1.60036	1.60057
	(0.000052)	(0.000102)	(0.000155)	(0.000137)	(0.000125)
$\hat{\theta}_1(Ad1)$	-1.99991	-1.99978	-1.99986	-2.00025	-2.00032
	(0.000043)	(0.000080)	(0.000107)	(0.000103)	(0.000091)
$\tilde{\theta}_2(Ad1)$	1.59981	1.59980	1.59972	1.60038	1.60042
	(0.000080)	(0.000134)	(0.000177)	(0.000160)	(0.000149)
$\hat{\theta}_1(Ad2)$	-2.00014	-2.00012	-1.99979	-1.99965	-1.99990
	(0.000043)	(0.000081)	(0.000108)	(0.000098)	(0.000099)
$\tilde{\theta}_2(Ad2)$	1.60006	1.60021	1.59979	1.59986	1.59987
	(0.000055)	(0.000107)	(0.000142)	(0.000126)	(0.000129)
$\tilde{\theta}_1(Ef)$	-1.99998	-1.99969	-2.00012	-2.00012	-2.00033
	(0.000061)	(0.000100)	(0.000131)	(0.000124)	(0.000111)
$\tilde{\theta}_2(Ef)$	1.59984	1.59984	1.60044	1.60048	1.60041
	(0.000061)	(0.000134)	(0.000183)	(0.000166)	(0.000159)

For mixed normal densities, both adaptive estimates are as good as CLS for estimating θ_1, and in general, CLS gives the smallest sample mean squared error. It seems that, for all densities, the efficient estimate gives greater sample mean squared error compare with others.

In Table 5.2, for $a_n = 0.6$, the adaptive estimates with $c_n = \infty$ give the smallest sample mean squared error compared to others for all density error types. The CLS estimate is better than the adaptive estimate with $c_n < \infty$ for mixture normal densities whereas the latter is better than the former for t-densities. Once again, the efficient estimates with splitting the sample performs worse than others.

Table 5.4. Average values of estimates and their sample mean squared error for $a_n = 0.8$.

Estimate	MN1	MN 2	t_5	t_7	t_9
$\hat{\theta}_1(C)$	-1.99974	-1.99981	-2.00034	-1.99965	-1.99936
	(0.000043)	(0.000078)	(0.000126)	(0.000110)	(0.000102)
$\hat{\theta}_2(C)$	1.59975	1.59982	1.60035	1.59981	1.59954
	(0.000056)	(0.000099)	(0.000161)	(0.000141)	(0.000129)
$\hat{\theta}_1(Ad1)$	-1.99970	-1.99979	-2.00017	-1.99951	-1.99948
	(0.000046)	(0.000082)	(0.000118)	(0.000107)	(0.000104)
$\tilde{\theta}_2(Ad1)$	1.60003	1.59990	1.60023	1.59935	1.59943
	(0.000110)	(0.000153)	(0.000230)	(0.000196)	(0.000186)
$\hat{\theta}_1(Ad2)$	-1.99979	-2.00029	-2.00038	-2.00026	-1.99991
	(0.000043)	(0.000083)	(0.000118)	(0.000104)	(0.000095)
$\tilde{\theta}_2(Ad2)$	1.59983	1.60019	1.60037	1.60031	1.59994
	(0.000056)	(0.000105)	(0.000152)	(0.000130)	(0.000124)
$\tilde{\theta}_1(Ef)$	-1.99969	-1.99979	-2.00018	-1.99943	-1.99937
	(0.000079)	(0.000114)	(0.000141)	(0.000137)	(0.000130)
$\tilde{\theta}_2(Ef)$	1.59987	1.60001	1.60063	1.59991	1.59981
	(0.000113)	(0.000157)	(0.000217)	(0.000190)	(0.000178)

Table 5.5. Average values of estimates and their sample mean squared error for $a_n = 0.9$.

Estimate	MN1	MN 2	t_5	t_7	t_9
$\hat{\theta}_1(C)$	-1.99990	-2.00002	-1.99987	-1.99972	-2.00032
	(0.000042)	(0.000078)	(0.000129)	(0.000111)	(0.000094)
$\hat{\theta}_2(C)$	1.59999	1.60019	1.60002	1.59987	1.60033
	(0.000054)	(0.000100)	(0.000163)	(0.000149)	(0.000121)
$\hat{\theta}_1(Ad1)$	-1.99996	-2.00008	-2.00008	-1.99966	-2.00044
	(0.000045)	(0.000082)	(0.000144)	(0.000113)	(0.000103)
$\tilde{\theta}_2(Ad1)$	1.60027	1.60007	1.60001	1.59947	1.60058
	(0.000141)	(0.000179)	(0.000298)	(0.000237)	(0.000204)
$\hat{\theta}_1(Ad2)$	-1.99982	-1.99995	-2.00001	-1.99997	-1.99990
	(0.000045)	(0.000080)	(0.000132)	(0.000106)	(0.000101)
$\tilde{\theta}_2(Ad2)$	1.59974	1.60001	1.60020	1.59976	1.59996
	(0.000059)	(0.000103)	(0.000164)	(0.000142)	(0.000127)
$\tilde{\theta}_1(Ef)$	-1.99969	-1.99992	-1.99985	-1.99982	-2.00034
	(0.000103)	(0.000131)	(0.000122)	(0.000152)	(0.000142)
$\tilde{\theta}_2(Ef)$	1.59993	1.60012	1.60021	1.59991	1.60054
	(0.000150)	(0.000180)	(0.000245)	(0.000229)	(0.000194)

In Table 5.3 for $a_n = 0.7$, it is shown that when error densities are t_5 and t_7, the best performing estimator, in term of minimum sample mean squared error, is the adaptive estimator with $c_n = \infty$. For the other densities, the CLS estimate gives minimum sample mean squared error.

In Table 5.4 for $a_n = 0.8$, the adaptive estimate with $c_n = \infty$ give the smallest sample mean squared error when the error densities are the first type of mixture normal, t_5 and t_7.

In Table 5.5 for $a_n = 0.9$, the adaptive estimates with $c_n = \infty$ give smaller sample mean squared error when the error density is t_7. For the other densities, the CLS estimates perform better than the other estimators.

The simulation study leads to the conclusion that, in general, the adaptive estimates with $c_n = \infty$ perform slightly better than the CLS for the t-densities. The CLS estimates, however, performs better for mixture normals. The efficient estimates with sample splitting exhibit greater sample mean squared error compared to others.

Chapter 6

Summary, Conclusions and Future Prospects

After seven decades of domination by linear Gaussian models, the time is certainly ripe for a serious study of ways of removing the many limitations of these models. Once we decide to incorporate features in addition to the autocovariances, the class of models would have to be greatly enlarged to include those besides the Gaussian ARMA models. We may either retain the general ARMA framework and allow the white noise to be non-Gaussian, or we may completely abandon the linearity assumption.

We have been witnessing almost an exponential growth in the applications of nonlinear time series models, ranging from solar sciences to earth sciences, from biosciences to economics and finances. It can safely be predicted that this growth will continue for a considerable time to come.

In this thesis, we have principally looked at the estimation of parameters of a first-order Smooth Threshold Autoregressive model with delay parameter one either in a parametric or a semiparametric setting. From the parametric point of view, we considered estimation using both classical (conditional least squares, maximum likelihood, M-estimator, estimating function) and Bayesian approaches. On the other hand, from a semiparametric point of view, the construction of estimators that are asymptotically efficient in the presence of infinite dimensional nuisance parameters is presented. In all our analysis we have considered STAR(1) model. An extension of the work to the STAR(p) model, where $p > 1$, can be accomplished by using ideas developed for STAR(1) model. However, this is a subject for possible future investigations.

Given in Chapter 2 of this thesis, is a necessary and sufficient condition for ergodicity of a first-order STAR process with delay parameter one under various assumptions on the tail behaviour of the smoothing distribution function in the model. The necessary and sufficient condition for recurrence in terms of the parameters, holds for any smoothing distribution function. The sufficient condition for ergodicity in terms of the parameters, depends on the behaviour of the smoothing distribution function. For light-tailed smoothing distribution functions, the sufficient condition for ergodicity is within the region, whereas for thick-tailed distribution functions, the sufficient condition is in terms of a larger parameter set. Combination of both type of smoothing distribution functions resulted in a general ergodicity region. Based on the results of necessary or sufficient conditions for ergodicity, recurrence and transience respectively, we can identify some sufficient conditions for null recurrence for any smoothing distribution function. In all these cases the error density function is assumed to be nonnegative in $\Re$.

It will be nice to get necessary and sufficient conditions for recurrence and ergodicity of the STAR model for any smoothing distribution function G. In this thesis we answered only a part of this problem. A necessary and sufficient condition for ergodicity has been proved for SETAR model by Petrucelli and Woolford (1985). Also, recently, Chen and Tsay (1991) established a necessary and sufficient condition for geometric ergodicity for the first-order threshold autoregressive process with general delay parameter d. By following this approach, we believe that it is possible to obtain a necessary and sufficient condition for ergodicity for the first-order STAR model with general delay parameter d. Moreover, Chan and Tong (1986) also presented the sufficient condition for ergodicity of STAR(p) processes with general delay parameter d without proof in Proposition 2.1(ii) (see Appendix A). This condition might be weakened and proving the condition would also be a good future research topic connected to this model.

In Chapter 3, we presented a Bayesian analysis to estimate the parameters of a STAR(p) model. Posterior distributions for the autoregressive coefficients, the threshold, smoothing, delay parameters and the error variance were developed in closed form. The posterior distributions of the intrinsic parameters (threshold, smoothing and delay) were calculated numerically and it was sighted as a substantial attraction of the Bayesian approach. Furthermore, the exact one-step-ahead predictive density was presented along with the Monte Carlo approach for evaluating the multi-step-ahead predictive density. Examples based on simulation results and well-known Canadian lynx data were presented in the last section of this chapter. It turned out that both the Bayesian approach and the Conditional Least Squares method give similar parameter estimates and produce equally good predictions. However the Bayesian approach yielded the posterior density of parameters as an added benefit.

Future work on Bayesian analysis of STAR(p) model could include extensions by varying the autoregressive order p of the processes or by choosing a different prior distribution such as a proper distribution to express prior information about the parameters. Also, in our Bayesian study we have taken G as the standard normal distribution function. It could be interesting to study the effect of varying G on the Bayes estimates.

In Chapter 4, we presented the parameter estimation of STAR models using Conditional Least Squares, Maximum Likelihood, M-estimator and Estimating Function methods. These estimators can be used as initial estimates for obtaining adaptive estimates given in Chapter 5. Under some conditions, all these methods yielded estimates which are strongly consistent and asymptotically normally distributed.

In Chapter 5, the adaptive estimator for a first-order STAR model with delay parameter one was presented. The adaptive estimator of this model only exists for a class of symmetric error densities. The construction of efficient and adaptive estimates were also reviewed in this chapter followed by a thorough investigation using simulations in the last section. The simulation results were presented to compare conditional least squares, adaptive and efficient estimators, employing sample splitting method. The error densities chosen were mixture normals, t_5, t_7 and t_9. In most cases, for mixture normals, conditional least squares estimator outperformed adaptive and efficient estimators. For t-densities, adaptive estimator outperformed the others.

In summary we have completed an extensive study of the STAR(1) model with respect to aspects of classification and estimation.

Bibliography

[1] Akritas, M.G. and Johnson, R.A. (1980). Efficiencies of tests and estimators of p-order autoregressive processes when the error distribution is nonnormal. *Ann. Inst. Statist. Math.*, **34**, 579-589.

[2] Beran, R. (1976). Adaptive estimates for autoregressive processes. *Ann. Inst. Statist. Math.*, **28**, 77-89.

[3] Bhattacharya, R. N. and Lee, C. (1995). Ergodicity of nonlinear first order AR models. *J. Theo. Prob.* **8(1)**, 207-219.

[4] Bickel, P.J. (1982). On Adaptive Estimation. *Ann. Statist.*, **10**, 647-671.

[5] Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Baltimore, MD; John Hopkins Univ. Press.

[6] Box, G.E.P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading, Mass.

[7] Broemeling, L.D. (1985). *The Bayesian Analysis of Linear Models.* Marcel Dekker, New York.

[8] Broemeling, L.D. and Cook, P. (1992). Bayesian Analysis of Threshold Autoregression. *Commun.Statist-Theory Meth.* **21**, 2459-2482.

[9] Chan, K. S., Petrucelli, J. D., Tong, H. and Woolford, S. W. (1985). A multiple threshold AR(1) model. *J. Appl. Prob,* **22,** 267-279.

[10] Chan, K.S. and Tong, H. (1986). On Estimating Threshold in Autoregressive models . *Journ. of Time Series Anal.* **7**, 179-190.

[11] Chen, C.W.S. and Lee, J.C (1995). Bayesian Inference of Threshold Autoregressive Models. *Journ.Time Series Anal.* **16**, 483-492.

[12] Chen, R. and Tsay, R.S. (1991). On the ergodicity of TAR(1) processes. *Annals Appl. Prob.,* **1,** 613-634.

[13] Drost, F.C., Klaassen, C.A.J. and Werker, B.J.M. (1993). Adaptiveness in time series models. In P.Mandl and M.Huskova (eds), *Asymptotic Statistics, Proceedings of the Fifth Prague Symposium,* p.203-211. Heidelberg: Physica-Verlag.

[14] Drost, F.C., Klaassen, C.A.J. and Werker, B.J.M. (1994). Adaptive estimation in time series models. Discussion Paper Series, No. 9488, CentER, Tilburg University, The Netherlands.

[15] Drost, F.C., Klaassen, C.A.J. and Werker, B.J.M. (1997). Adaptive estimation in time series models. *Annals of Statistics,* **25,** 786-817.

[16] Drost, F.C. and Klaassen, C.A.J. (1997). Efficient estimation in semiparametric GARCH models. *Journal of Econometrics*, **81**, 193-221.

[17] Fabian, V. and Hannan, J. (1982). On estimation and adaptive estimation for locally asymptotically normal families. *Z. Wahrsvheinlichkeitstheorie Verw. Geb.,* **59**, 459-478.

[18] Fabian, V. and Hannan, J. (1987). Local asymptotic behaviour of densities. *Statist. Decisions*, **5**, 105-138. Correction:**6**,(1988), 195.

[19] Geweke, J. and Terui, N. (1993). Bayesian Threshold Autoregressive models for Nonlinear Time Series. *Journ. Time Series Anal.*, **14**,441-454.

[20] Godambe, V.P. (1960). An optimum property of regular maximum likelihood equation. *Ann. Math. Stat.,* **31,** 1208-1211.

[21] Godambe, V.P. (1985). The foundations of finite sample estimation in stochastic processes. *Biometrika,* **72,** 319-328.

[22] Godambe, V.P. (1991). *Estimating Functions.* Oxford Univ. Press., New York.

[23] Granger, C.W.J. and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships.* Oxford Univ. Press, New York.

[24] Guo, M. and Petrucelli, J. D. (1991). On the null recurrence and transience of a first-order SETAR model. *J. Appl. Prob.* **28,** 584-592.

[25] Hájek, J. (1970). A characterisation of limiting distributions of regular estimates. *Zeitschrift für Wahrscheinlich keitstheorie und Verwandte Gebiete,* **14,** 323-330.

[26] Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and Applications.* New York: Academic Press.

[27] Hoaglin, D.C., Mosteller, F. and Tukey, J.W. (1983). *Understanding Robust and Exploratory Data Analysis.* John Wiley & Sons Inc., Canada.

[28] Huber, P.J. (1981). *Robust Statistics.* John Wiley & Sons, New York.

[29] Ibragimov, I. A. and Hasminskii (1981). *Statistical Estimation : Asymptotic Theory.* New York: Springer.

[30] Jeganathan, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory*, **11**; 818-887.

[31] Johnson, N. L. and Kotz, S. (1970). *Continuous Univariate Distributions, Part 1 and 2.* Houghton Mifflin Company, Boston.

[32] Karlin, S. and Taylor,H.M. (1975). *A First Course in Stochastic Processes* 2nd edition. Academic Press, New York.

[33] Klaassen, C.A.J. (1987). Consistent estimation of the influence function of locally asymptotically linear estimators. *Annals of Statistics,* **15**, 1548-1562.

[34] Klimko, L.A. and Nelson, P.I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statis.,* **6**, 629-642

[35] Koul, H.L. (1996). Asymptotic of some estimators and sequential residual empiricals in nonlinear time series. *Ann. Statist.*, **24**, 380-404.

[36] Koul, H.L. and Schick, A. (1996). Adaptive estimation in a random coefficient autoregressive model. *Ann. Statist.*, **24**, 1025-1052.

[37] Koul, H.L. and Schick, A. (1997). Efficient estimation in nonlinear autoregressive time-series models. *Bernoulli*, **3**, 247-277.

[38] Koul, H.L. and Susarla, V. (1983). Adaptive estimation in linear regression. *Statist. Decisions*, **1**, 379-400.

[39] Kreiss, J.P. (1987a). On adaptive estimation in stationary ARMA processes. *Ann. Statist.*, **15**, 112-133.

[40] Kreiss, J.P. (1987b). On adaptive estimation in autoregressive models when there are nuisance functions. *Statist. Decisions.* **5**, 59-76.

[41] LeCam, L. (1960). Locally asymptotically normal families of distributions. *Univ. California Publ. Statist.,* **3**, 37-98.

[42] LeCam, L. and Yang, G.L. (1990). *Asymptotics in Statistics: Some Basic Concepts.* Springer, New York.

[43] Linton, O. (1993). Adaptive estimation in ARCH models. *Econometric Theory,* **9**, 539-569.

[44] Luukkonen, R.,Saikkonen, P. and Terãsvirta, T. (1988). Testing Linearity against Smooth Transition Autoregressive models. *Biometrika* **75**,491-499.

[45] McCulloch, R.E. and Tsay, R.S. (1994). Bayesian analysis of autoregressive time series via Gibbs sampler. *J. Time Ser. Anal.* **15**, 235-250.

[46] Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chain and Stochastic Stability.* Springer, London.

[47] Monahan, J.F. (1983). Fully Bayesian Analysis of ARMA Time Series Models. *Journ. of Econometrics,* **21**, 307-331.

[48] Nur, D. (1993). *Smooth Threshold Autoregressive(1) models* . Unpublished Master's thesis, Univ. of Western Australia.

[49] Orey, S. (1971). *Lecture Notes on Limit Theorems for Markov Chain Transition Probabilities.* Van Nostrand, London.

[50] Petrucelli, J. D. and Woolford, S. W. (1984). A threshold AR(1) model. *J.Appl.Prob.* **21**, 270-286.

[51] Schick, A. (1986). On asymptotically efficient estimation in semiparametric models. *Ann. Statist..* **14**, 1486-1521.

[52] Schick, A. (1993). On efficient estimation in regression models. *Ann. Statist..* **21**, 1139-1151.

[53] Stein, C. (1956). Efficient nonparametric estimation and testing. In J. Neyman (ed.). *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability,* Vol.1, 187-195. Berkeley, University of California Press.

[54] Teräsvirta,T. (1994). Specification,Estimation and Evaluation of Smooth Transition Autoregressive models. *Journ. Amer. Statist. Assoc.* **89**, 208-218.

[55] Thavaneswaran, A. and Abraham, B. (1988). Estimation for non-linear time series models using estimating equations. *J. Time Ser. Anal.* **9**, 99-108.

[56] Tjøstheim, D. (1986). Estimation in nonlinear time series models. *Stoch. Proc. Appl.*, **21**, 251-273.

[57] Tjøstheim, D. (1994). Nonlinear time series: a selective review. *Scand. Journ. Statist.*, **21**, 97-130.

[58] Tong, H.(1983). *Threshold Models in Nonlinear Time Series Analysis.* New York : Springer-Verlaag.

[59] Tong, H.(1988). Nonlinear time series modelling in population biology : a preliminary case study. *Journ of Animal Ecology.* , 75-87.

[60] Tong, H. (1990). *Nonlinear Time Series Analysis: a Dynamical System Approach.* London : Oxford Univ.Press.

[61] Tong, H. and Lim,K.S. (1980). Threshold Autoregression, Limit Cycles and Cyclical Data (with discussion). *Journ.Royal Statis.Soc B* **42**, 245-292.

[62] Tsay, R. S. (1989). Testing and modelling threshold autoregressive processes. *Journ. Amer. Stat. Assoc.,* **84**, 231-240.

[63] Tweedie, R. L. (1975). Sufficient conditions for ergodicity and recurrence of Markov chain on a general state space. *Stoch.Proc. and their Appl.* **3**, 385-403.

[64] Tweedie, R. L. (1976). Criteria for classifying Markov chains. *Adv. Appl. Prob.*, **8**, 737-771.

[65] Tweedie, R. L. (1983). The existence of moments for stationary Markov chains. *Journ. Appl. Prob.*, **20**, 191-196.

[66] Tweedie, R.L. (1974). R-theory for Markov chains on a general state space I : solidarity properties and R-recurrent chains. *Ann. Probab.* **2**, 840-864.

[67] Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics.* New York : Wiley and Sons

[68] Zwanzig, S. (1994). On adaptive estimation in nonlinear regression. *Kybernetika.* **30**, 359-367.

Appendix A

Ergodicity

Suppose $\{X_t\}$ is a Markov chain taking values in some arbitrary measurable space $(\chi, \mathcal{F})$ where $\mathcal{F}$ is a σ-field on χ. Let μ denote a fixed subinvariant measure for $\{X_t\}$. In most applications of Markov chains, the state space will be equipped with a topology τ.

Theorem A.1 (Theorem 4.2 of Tweedie (1975)(Sufficient condition for ergodicity)) *Let $\{X_t\}$ be a ϕ-irreducible Markov chain on a topological space (χ, τ). If $P(x, y)$, the transition function, is strongly continuous, a sufficient condition for $\{X_t\}$ to be ergodic is the existence of a compact set $K \in \tau$ and a non-negative measurable function g on χ such that*

(i)

$$\int_{\chi} P(x, dy) g(y) \leq g(x) - 1, \quad x \notin K, \tag{A.0.1}$$

(ii) and, for some fixed $B > 0$,

$$\int_{\chi} P(x, dy) g(y) = \lambda(x) \leq B < \infty, \quad x \in K. \tag{A.0.2}$$

Remark A.1 *Theorem 9.1 of Tweedie (1976) states that the chain $\{X_t\}$ is positive if there exists $\epsilon > 0$, $\theta < \infty$ and a status set A for $\{X_t\}$ such that*

(i)

$$\int_X P(x, dy)g(y) \leq g(x) - \epsilon, \quad x \notin A, \tag{A.0.3}$$

(ii) and, for some fixed $B > 0$,

$$\int_{A^c} P(x, dy)g(y) \leq \theta, \quad x \in A. \tag{A.0.4}$$

Thus condition (i) of Theorem 4.2 of Tweedie (1975) can be weakened for $\epsilon > 0$.

Theorem A.2 (Theorem 4.3 of Tweedie (1975)(Sufficient condition for recurrence)) *Under the same strong continuity condition as in Theorem 4.2, a sufficient condition for* $\{X_t\}$ *to be recurrent is the existence of a compact set* $K \in \tau$ *and a non-negative measurable function* g *on* X *such that*

(i) $\int_X P(x, dy)g(y) \leq g(x), \quad x \notin K$;

(ii) g *is strictly unbounded, in the sense that there exists, for every sufficiently large* M, *a set* $K_M \in \tau$ *with*

$$g(x) \geq M, \quad x \notin K_M, \tag{A.0.5}$$

Lemma A.1 (Lemma 2.1 of Chan, et.al. (1985)) *Let* $\{P(x, .)\}$ *be the transition law corresponding to the transition density (2.2.1). Then if* $\mathcal{K}$ *is the set of compact sets in* $\mathcal{B}$ *having positive Lebesgue measure, then* $0 < \pi(K) < \infty, \forall K \in \mathcal{K}$, *where* $\pi(.)$ *is a subinvariant measure for* $\{X_t\}$.

Proposition A.1 (Proposition 2.1 of Chan and Tong (1986))

Consider the STAR model of order p. *If either*

(i) $p = 1,\ d = 1,\ a_1 < 1,\ a_1 + b_1 < 1$ and $a_1(a_1 + b_1) < 1$ *or*

(ii) $\sup_{0 \leq \theta \leq 1} (\sum_{i=1}^{p} |a_i + \theta b_i|) < 1$

holds, then $\{X_t\}$ *as defined for STAR model is ergodic and there is a unique stationary process satisfying the model. Furthermore, the existence of the* s*th absolute moment of* ε_t *implies that of the stationary distribution.*

Theorem A.3 (Theorem 5.5 of Karlin and Taylor (1975)) *Let* $\{X_n\}$ *be an ergodic stationary processes having a finite mean* m*. Then with probability one,*

$$\lim_{n\to\infty} \frac{1}{n}(X_1 + X_2 + \dots + X_n) = m, \quad \text{or } \frac{\sum_{i=1}^{n} X_i}{n} \to m, \ a.s.$$

Theorem A.4 *(Theorem 5.6 of Karlin and Taylor (1975)) Let* $\{X_n\}$ *be a stationary process. The following conditions are equivalent:*

1. $\{X_n\}$ *is ergodic*
2. *For every* k *and every function* ϕ *of* $(k+1)$*- variables*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \phi(X_j, \dots, X_{j+k}) = E[\phi(X_0, \dots, X_k)]$$

provided the expectation exists.

Appendix B

Parametric Estimation

B.1 Conditional Least Squares

Let $X^N = (X_1, X_2, ..., X_N)$ denote a sample of N consecutive observations from a time series $\{X_t : t \in \mathcal{Z}\}$. Assume that X^N has a probability density $P_N(x_1, ..., x_N)$, which depends on $\theta \in \Theta$, an open subset of $\Re^p$. Suppose that $E \mid X_t \mid < \infty$, for $t \geq 1$. We may estimate θ by minimizing the 'residual sum of squares' (or equivalently the sum of squares of the innovations)

$$Q_N(\theta) = \sum_{j=1}^{N} [X_j - E_\theta(X_j \mid \mathcal{B}_{j-1})]^2 \tag{B.1.1}$$

with respect to θ, with $\mathcal{B}_k$ being the σ-algebra generated by $\{X_s, 1 \leq s \leq k\}$ and $\mathcal{B}_0$ being the trivial σ-algebra. The estimate, $\hat{\theta}_N$, is given by the solutions of the system of algebraic equations

$$\frac{\partial Q_N(\theta)}{\partial \theta} = 0, \quad i = 1, ..., p, \tag{B.1.2}$$

where the partial derivatives are assumed to exist. In fact, we assume that $E_\theta(X_j \mid \mathcal{B}_{j-1})$ is almost surely twice differentiable with respect to θ in some neighbourhood S of θ_0, the true parameter. We may consider a Taylor series expansion in a neighbourhood of θ_0 within S. For $\delta > 0$,

$\| \theta - \theta_0 \| < \delta$, for some $\theta^*, 0 < \| \theta_0 - \theta^* \| < \delta$ (henceforth, θ^* denotes an appropriate intermediate point not necessarily the same from line to line). Let $D_\theta \mathcal{Q}_N(\theta_0)$ denote the column vector of first partial derivatives of $\mathcal{Q}_N(\theta)$ evaluated at θ_0. Let $D^2_\theta \mathcal{Q}_N(\theta^*)$ denote the matrix of second partial derivatives evaluated at θ^* :

$$\begin{aligned}
\mathcal{Q}_N(\theta) &= \mathcal{Q}_N(\theta_0) + (\theta - \theta_0)^T D_\theta \mathcal{Q}_N(\theta_0) + \frac{1}{2}(\theta - \theta_0)^T D^2_\theta \mathcal{Q}_N(\theta^*)(\theta - \theta_0) \\
&= \mathcal{Q}_N(\theta_0) + (\theta - \theta_0)^T D_\theta \mathcal{Q}_N(\theta_0) + \frac{1}{2}(\theta - \theta_0)^T V_N(\theta_0)(\theta - \theta_0) \\
&+ \frac{1}{2}(\theta - \theta_0)^T T_N(\theta^*)(\theta^* - \theta_0), \qquad \text{(B.1.3)}
\end{aligned}$$

where $V_N(\theta_0)$ is the $p \times p$ matrix of second partial derivatives of $\mathcal{Q}_N(\theta)$ evaluated at θ_0 and

$$T_N(\theta^*) = D^2_\theta \mathcal{Q}_N(\theta^*) - V_N(\theta_0). \qquad \text{(B.1.4)}$$

Specifically, for $1 \le i \le p,\ 1 \le j \le p$,

$$\begin{aligned}
\left(\frac{1}{2N} V_N\right)_{ij} = \sum_{k=1}^{N} \Bigg[\frac{1}{N} \left(\frac{\partial}{\partial \theta_i} E_\theta(X_k \mid \mathcal{B}_{k-1}) \right) \left(\frac{\partial}{\partial \theta_j} E_\theta(X_k \mid \mathcal{B}_{k-1}) \right) \\
- \frac{1}{N} \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} E_\theta(X_k \mid \mathcal{B}_{k-1}) \right) (X_k - E_\theta(X_k \mid \mathcal{B}_{k-1})) \Bigg]. \qquad \text{(B.1.5)}
\end{aligned}$$

Theorem B.1 (Theorem 2.1 of Klimko and Nelson(1978)) *Assume that*

(i) $\lim_{N\to\infty} \sup_{\delta\to 0}(| T_N(\theta^*)_{ij} | / N\delta) < \infty$ *a.s.*, $1 \le i \le p,\ 1 \le j \le p$;

(ii) $(2N)^{-1} V_N \xrightarrow{a.s.} V$, *where* V *is a positive definite symmetric* $p \times p$ *matrix of constants;*

(iii) $N^{-1} D_\theta \mathcal{Q}_N(\theta_0) \xrightarrow{a.s.} 0$ *component-wise.*

Then, there exists a sequence of estimators $\{\hat{\theta}_N\}$ *such that* $\hat{\theta}_N \xrightarrow{a.s.} \theta_0$, *and for any* $\varepsilon > 0$ *there exists an event* E *with* $P(E) > 1 - \varepsilon$ *and an* N_0 *such that on* E, *for* $N > N_0$, $\hat{\theta}_N$ *satisfies conditional least squares equation (4.2.1) and* $\mathcal{Q}_N$ *attains a relative minimum at* $\hat{\theta}_N$.

Theorem B.2 (Theorem 2.2 of Klimko and Nelson(1978)) *Suppose that the conditions of Theorem 4.1 hold and, in addition,* $\frac{1}{2}N^{-\frac{1}{2}}D_\theta Q(\theta_0)$ *converges to multivariate normal,* $\mathcal{N}(\mathbf{0}, W)$ *in distribution as* $N \to \infty$, *where* W *is a* $p \times p$ *positive definite matrix. Then* $N^{\frac{1}{2}}(\hat{\theta}_N - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, V^{-1}WV^{-1})$ *as* $N \to \infty$.

Also, under quite mild conditions, for any non-zero vector $\mathbf{c}$ *of constants*

$$\limsup_{N\to\infty} \frac{N^{\frac{1}{2}}\mathbf{c}^T(\hat{\theta}_N - \theta_0)}{(2\sigma^2 \ln\ln N)^{\frac{1}{2}}} = 1, \quad \text{a.s.,} \tag{B.1.6}$$

where

$$\sigma^2 = \mathbf{c}^T V^{-1} W V^{-1} \mathbf{c}.$$

B.2 Maximum likelihood estimators

Theorem B.3 (Theorem 5.1 of Tjøstheim (1986)). *Assume that* $\{X_t\}$ *is a* d- *dimensional strictly stationary and ergodic process with* $E(|X_t|^2) < \infty$, *and that* $\tilde{X}_{t|t-1}(\theta)$ *and* $Z_{t|t-1}(\theta)$ *are almost surely three times continuously differentiable in an open set* S *containing* θ_0. *Moreover, if* $\phi(t)$ *is defined by (4.3.1), assume that*

(i) $E\left(\left|\frac{\partial\phi(t)}{\partial\theta_i}(\theta_0)\right|\right) < \infty$ and $E\left(\left|\frac{\partial^2\phi(t)}{\partial\theta_i\partial\theta_j}(\theta_0)\right|\right) < \infty$ for $i, j = 1, 2, ..., s$.

(ii) For arbitrary real numbers $a_1, ..., a_s$ *such that, for* $\theta = \theta_0$,

$$E\left(\left| Z_{t|t-1}^{1/2} \sum_{i=1}^{s} a_i \frac{\partial \tilde{X}_{t|t-1}}{\partial\theta_i}\right|^2\right) + E\left[\left| Z_{t|t-1}^{-1/2} \bigotimes Z_{t|t-1}^{-1/2} \sum_{i=1}^{s} a_i \frac{\partial}{\partial\theta_i}\{vec(Z_{t|t-1})\}\right|^2\right] = 0,$$

then we have $a_1 = a_2 = ... = a_s = 0$,

(iii) For $\theta \in \Theta$, *there exist functions* $H_t^{ijk}(X_1, ..., X_t)$ *such that*

$$\left|\frac{\partial^3\phi_t}{\partial\theta_i\partial\theta_j\partial\theta_k}(\theta)\right| \leq H_t^{ijk}, \text{ and } E(H_t^{ijk}) < \infty$$

for $i, j, k = 1, 2, ..., s,\ t = m + 1, ..., N$

Then there exists a sequence of estimators $\{\hat{\theta}_N\}$ *minimising* L_N *of (4.3.1) such that* $\hat{\theta}_N \xrightarrow{a.s.} \theta_0$, *and for* $\varepsilon > 0$ *there exists an event* E *with* $P(E) > 1 - \varepsilon$ *and an* n_0 *such that on* E, *for* $N > n_0$, Q_N *attains a relative minimum at* $\hat{\theta}_N$.

Theorem B.4 (Theorem 5.2 of Tjøstheim (1986)) *Assume that the conditions of Theorem 5.1 of Tjøtheim (1986) are fulfilled and that, for* $\theta = \theta_0$ *and* $i, j = 1, ..., s$,

$$\begin{aligned} S_{ij} &= \frac{1}{4}E\left\{\frac{1}{Z_{t|t-1}^4}\left(\frac{\partial Z_{t|t-1}}{\partial \theta_i}\frac{\partial Z_{t|t-1}}{\partial \theta_j}[E\{(X_t - X_{t|t-1})^4 \mid \mathcal{B}_{t-1}\} - 3Z_{t|t-1}^2]\right.\right. \\ &+ \ 2E\{(X_t - X_{t|t-1})^3 \mid \mathcal{B}_{t-1}\}Z_{t|t-1} \\ \text{x} & \quad \left.\left.\left(\frac{\partial \tilde{X}_{t|t-1}}{\partial \theta_i}\frac{\partial Z_{t|t-1}}{\partial \theta_j} + \frac{\partial \tilde{X}_{t|t-1}}{\partial \theta_j}\frac{\partial Z_{t|t-1}}{\partial \theta_i}\right)\right)\right\} < \infty. \end{aligned}$$

Let $S = (S_{ij})$, *and let* $\{\hat{\theta}_N\}$ *be the estimators obtained in Theorem 5.1 . Then we have*

$$S_{ij} = \frac{1}{4}E\left(\frac{\partial \phi_t}{\partial \theta_i}\frac{\partial \phi_t}{\partial \theta_j}\right) - U'_{ij},$$

where

$$U_{ij}^T = E\left\{\frac{1}{Z_{t|t-1}^2}\left(Z_{t|t-1}\frac{\partial \tilde{X}_{t|t-1}}{\partial \theta_i}\frac{\partial \tilde{X}_{t|t-1}}{\partial \theta_j} + \frac{1}{2}\frac{\partial Z_{t|t-1}}{\partial \theta_i}\frac{\partial Z_{t|t-1}}{\partial \theta_j}\right)\right\}.$$

Furthermore,

$$N^{1/2}(\hat{\theta}_N - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, (U')^{-1} + (U')^{-1}S(U')^{-1}).$$

In the case where $Z_{t|t-1}$ does not depend on θ we have $S = 0$. In addition, if $X_t - \tilde{X}_{t|t-1}$ is independent of $\mathcal{B}_{t-1}$ then

$$U_{ij}^T = E\left[\frac{1}{Z_{t|t-1}^2}\left(Z_{t|t-1}\frac{\partial \tilde{X}_{t|t-1}}{\partial \theta_i}\frac{\partial \tilde{X}_{t|t-1}}{\partial \theta_j}\right)\right],$$

and estimation using likelihood penalty function or sum of squares function essentially gives identical results.

B.3 M-estimators

A class of submodels this chapter shall focus on in some detail is the family of nonlinear autoregressive models where $X_{ni} \equiv X_i$ satisfying

$$X_i = H(\theta, \mathbf{Y}_{i-1}) + \varepsilon_i, \quad i = 0, \pm 1, \pm 2, ..., \tag{B.3.7}$$

where $\{\varepsilon_i, i = 1, 2, ...\}$ are *i.i.d.* with distribution function F and $h(\theta, \mathbf{y}) \equiv H(\theta, \mathbf{y})$.

Fix a $\theta \in \Theta$ and let P_θ^n denote the probability distribution of $(\mathbf{Y}_0, X_{n1}, ..., X_{nn})$ under (4.4.1) when θ is the true parameter value. Suppose there exists a vector of functions $\dot{\mathbf{h}}_n$ from $\Theta \times \Re^p$ to $\Re^m$ as mentioned below.

Define

$$\mathbf{M}(\mathbf{t}) := n^{-1/2} \sum_{i=1}^{n} \dot{\mathbf{h}}_{ni}(\mathbf{t}) \psi(X_i - h_{ni}(\mathbf{t})), \tag{B.3.8}$$

where ψ is a bounded nondecreasing real valued function on $\Re$. The M-estimator of θ to be considered is

$$\hat{\theta}_M := argmin \parallel \mathbf{M}(\mathbf{t}) \parallel .$$

About h we shall assume the following:

(h1) There exists a vector of functions $\dot{\mathbf{h}}_n$ from $\Theta \times \Re^p$ to $\Re^m$ such that $\dot{\mathbf{h}}_{ni} := \dot{\mathbf{h}}_n(\mathbf{t}, \mathbf{Y}_{n,i-1})$ is $\mathcal{B}_{ni}$-measurable, $\mathbf{t} \in \Theta, 1 \leq i \leq n$, and satisfies the following : $\forall \alpha > 0,\ k < \infty,\ \mathbf{s} \in \Theta$,

$$\limsup_n P_\theta^n \left(\sup_{1 \leq i \leq n, \|\mathbf{t}-\mathbf{s}\| \leq k n^{-1/2}} \frac{| h_{ni}(\mathbf{t}) - h_{ni}(\mathbf{s}) - (\mathbf{t}-\mathbf{s})^T \dot{\mathbf{h}}_{ni}(\mathbf{s}) |}{\| \mathbf{t} - \mathbf{s} \|} > \alpha \right) = 0.$$

(F) F has a uniformly continuous density f which is positive *a.s.*

(h2) $n^{-1}\sum_i \dot{\mathbf{h}}_{ni}(\theta)\dot{\mathbf{h}}_{ni}^T(\theta) = \sum_\theta + o_p(1)$, where $\sum_\theta$ is a positive-definite matrix.

(h3) $n^{-1/2}\max_i \| \dot{\mathbf{h}}_{ni}(\theta) \| = o_p(1)$.

(h4) $n^{-1}\sum_i E_\theta^n \| \dot{\mathbf{h}}_{ni}(\theta + n^{-1/2}\mathbf{t}) - \dot{\mathbf{h}}_{ni}(\theta) \|^2 = o(1)$, $\mathbf{t} \in \Theta$.

(h5) $n^{-1/2}\sum_i \| \dot{\mathbf{h}}_{ni}(\theta + n^{-1/2}\mathbf{t}) - \dot{\mathbf{h}}_{ni}(\theta) \| = O_p(1)$, $\mathbf{t} \in \Theta$.

(h6) For every $\alpha > 0$, there exists a $\delta > 0$ and an $N < \infty$, such that

$$P_\theta^n\left(\sup_{\|\mathbf{t}-\mathbf{s}\|\le\delta} n^{-1/2}\sum_i \| \dot{\mathbf{h}}_{ni}(\theta + n^{-1/2}\mathbf{t}) - \dot{\mathbf{h}}_{ni}(\theta + n^{-1/2}\mathbf{s}) \| \le \alpha\right) \ge 1-\alpha,$$

for all $n > N$.

As noted in Remark 1.1 of Koul (1996), if the underlying process and h are such that $\{\dot{\mathbf{h}}_{ni}(\theta)\}$ does not depend on n and the process is stationary and ergodic, then the distribution of $\mathbf{Y}_0$ will depend on θ and the following hold:

(h_s1) $E_\theta \| \dot{\mathbf{h}}_1(\theta) \|^2 < \infty$ implies assumptions (h2) and (h3),

(h_s2) $E_\theta \| \dot{\mathbf{h}}_1(\theta + n^{-1/2}\mathbf{t}) - \dot{\mathbf{h}}_1(\theta) \|^2 = o(1)$, $\mathbf{t} \in \Theta$ is equivalent to assumption (h4),

(h_s3) $n^{1/2}E_\theta \| \dot{\mathbf{h}}_1(\theta + n^{-1/2}\mathbf{t}) - \dot{\mathbf{h}}_1(\theta) \| = O(1)$, $\mathbf{t} \in \Theta$ implies assumption (h5),

where E_θ denotes the expectation under the stationary distribution.

To state the result of the asymptotic uniform linearity of the M -scores, one needs

$$\Psi := \Big\{\psi : \Re \to \Re, \text{ nondecreasing right continuous}, \int \psi dF = 0,\ \psi(\infty) - \psi(-\infty) \le 1\Big\}.$$

Theorem B.5 (Theorem 1.1 of Koul (1996)) *In addition to (4.4.1), assume that (F) and (h1)-(h6) hold. Then, for every* $0 < b < \infty$,

$$\sup_{\psi \in \Psi, \mathbf{t} \in N_b} \left\| \mathbf{M}(\theta + n^{-1/2}\mathbf{t}) - \mathbf{M}(\theta) - \textstyle\sum_{\theta} \mathbf{t} \int f d\psi \right\| = o_p(1). \tag{B.3.9}$$

Corollary B.1 (Corollary 1.1 of Koul (1996)) *In addition to the assumptions of Theorem B.5, assume that* $\int f d\psi > 0$ *and*

(M1) $\mathbf{e}'\mathbf{M}(\theta + n^{-1/2}r\mathbf{e})$ *is monotonic in* $r \in \Re$, $\forall\, \mathbf{e} \in \Re^m$, $\| \mathbf{e} \| = 1, n \geq 1, a.s.$

Then, $\forall \psi \in \Psi$,

$$n^{1/2}(\hat{\theta}_M - \theta) = \left\{ \int \textstyle\sum_{\theta} d\psi \right\}^{-1} \mathbf{M}(\theta) + o_p(1) \tag{B.3.10}$$

and

$$n^{1/2}(\hat{\theta}_M - \theta) \rightsquigarrow \mathcal{N}_q(0, \textstyle\sum_{\theta}^{-1} \nu(\psi, F)) \tag{B.3.11}$$

where $\nu(\psi, F) := \int \psi^2 dF / (\int f d\psi)^2$.

Corollary B.2 (Corollary 1.2 of Koul (1996)) *Assume (4.4.1) and (h1)-(h6) hold. In addition, if (M1) with* $\psi(x) \equiv sgn(x)$ *holds and if the density function* F *has density* f *in an open neighbourhood of* 0 *such that* f *is positive and continuous at* 0, *then* $n^{1/2}(\hat{\theta}_{lad} - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}, \sum_{\theta}^{-1}/4f^2(0))$ *where* p *is the order of the autoregressive model.*

B.4 Estimating Equations

Let $\{X_t : t \in Z^+\}$ be a discrete-time stochastic process taking values in $\Re$ and defined on a probability space (Θ, A, F). We assume that observations $(X_1, X_2, ..., X_n)$ are available and that the parameter $\theta \in \Theta$, a compact subset of $\Re$. Let $\mathcal{F}$ be a class of distributions and $\mathcal{B}_t^x$ be the σ-field generated by X up to time t.

Following Godambe (1985), any real function g of the variates $X_1, X_2, ..., X_n$ and the parameter θ, satisfying certain regularity conditions, is called a regular unbiased estimating function if,

$$E_F[g(X_1, X_2, ..., X_n : \theta(F))] = 0, \quad F \in \mathcal{F},$$

E_F denoting the expectation under F.

Let L be the class of estimating functions g of the form

$$g = \sum_{t=1}^{n} h_t a_{t-1}$$

where the function h_t is such that $E[h_t \mid \mathcal{B}_{t-1}^x] = 0$, $t = 1, 2, ..., n$ and a_{t-1} is a function of $X_1, X_2, ..., X_{t-1}$ and θ, for $t = 1, 2, ..., n$.

The efficiency in estimating θ, through the equation $g(X_1, X_2, ..., X_n) = 0$, is defined by $\lambda(g, F)$, where

$$\lambda(g, F) = \frac{\{E_F\left(\frac{\partial g}{\partial \theta}\right)\}^2}{E_F(g^2)},$$

or, if $\partial g / \partial \theta$ does not exist,

$$\lambda(g, F) = \lim_{\varepsilon \to 0} E_F[\{g(x, \theta + \varepsilon) - g(x, \theta)\}/\varepsilon]^2 / \{E_F(g^2)\}.$$

The estimating function $g^* \in L$ is said to be most efficient or optimum if

$$\lambda(g^*, F) \geq \lambda(g, F), \; F \in \mathcal{F}, \; g \in L.$$

Theorem B.6 (Theorem 1 of Godambe (1985)) *In the class L of unbiased estimating functions g, the optimum estimating function g^* is the one which minimizes*

$$\frac{E(g^2)}{E\left(\frac{\partial g}{\partial \theta}\right)^2}$$

and this is given by

$$g^* = \sum_{t=1}^{n} h_t a_{t-1}^* \quad \text{where} \quad a_{t-1}^* = \frac{E\left[\frac{\partial h_t}{\partial \theta} \mid \mathcal{B}_{t-1}^y\right]}{E\left[h_t^2 \mid \mathcal{B}_{t-1}^y\right]}$$

Appendix C

Adaptivity

Remark C.1 *The following result can be used to verify the condition (5.3.6) if no closed forms for* $E_{\vartheta,\phi}\psi(X_0)$ *is available. Suppose the densities in* $\mathcal{F}$ *are positive,* $\psi(x) \to \infty$ *as* $|x| \to \infty$ *and*

$$\int \psi(y + h(x,\vartheta))\phi(y)dy \le C + (1-2\delta)\psi(x), \quad x \in \mathbb{R}$$

for positive constants C *and* δ*. Then for all sufficiently large* K,

$$E_{\vartheta,\phi}\psi(X_0) \le \frac{1}{\delta}(C + (1-2\delta)\sup_{|x|\le K}\psi(x)). \tag{C.0.1}$$

Remark C.2 *In stationary and ergodic NLAR(1) models one can provide simple sufficient conditions for Condition 5.2 (Condition 5.1 of Koul and Schick (1997)). Suppose there is a map* ψ *such that*

$$E_\theta\psi(X_0) < \infty \text{ and } \psi(x) \ge \sup_{\|\vartheta-\theta\|<\delta} \|\dot{h}(x,\vartheta)\|^2, \quad x \in \Re, \text{ for some } \delta > 0.$$

Then Condition 5.2 (Condition 5.1 of Koul and Schick (1997)) holds. To see this, fix a local sequence $<\theta_n>$ *and a sequence* $<c_n>$ *tending to infinity. From ergodic theorem that if* $\{X_t\}$ *is stationary process then it is ergodic and implies that*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\phi(X_j,...,X_{j+k}) = E\left[\phi(X_0,...,X_k)\right]$$

provided the expectation exists.

Therefore

$$\limsup_n \frac{1}{n}\sum_{j=1}^{n} \|\dot{H}_j(\theta_n)\|^2 I\left[\|\dot{H}_j(\theta_n)\| > c_n\right] \leq E_{\theta_0}\psi(X_0) I\left[\psi(X_0) > c\right]$$

almost surely P_θ *for every* $c > 0$.

As $\lim_{c\to\infty} E_\theta\psi(X_0)I\left[\psi(X_0) > c\right] = 0$, *we find*

$$\frac{1}{n}\sum_{j=1}^{n} \|\dot{H}_j(\theta_n)\|^2 I\left[\|\dot{H}_j(\theta_n)\| > c_n\right] = o_\theta(1).$$

This and contiguity argument yield Condition 5.2 (Condition 5.1 of Koul and Schick (1997)).