MARKOV REGIME-SWITCHING MODELS

by

Lingyun Ye

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
July 2012

DALHOUSIE UNIVERSITY

DEPARTMENT OF MATHEMATICS AND STATISTICS

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "MARKOV REGIME-SWITCHING MODELS" by Lingyun Ye in partial fulfillment of the requirements for the degree of Master of Science.

Dated: July 19, 2012

Co-supervisors: ______________________________

____

Reader: ______________________________

DALHOUSIE UNIVERSITY

DATE: July 19, 2012

AUTHOR: Lingyun Ye

TITLE: MARKOV REGIME-SWITCHING MODELS

DEPARTMENT OR SCHOOL: Department of Mathematics and Statistics

DEGREE: M.Sc. CONVOCATION: October YEAR: 2012

Signature of Author

To the most precious treasure of my life - Laura, Ethan and Elise

Table of Contents

List of Tables

LIST OF FIGURES

ABSTRACT

A regime-switching model is a time-series model in which parameters change values according to the regime at present time. While regime-switching models have been very popular in applied work, there is a lack of literature for simulation studies. New methods based on regime-switching models are often proposed with neither a proof of convergence nor simulations to demonstrate their basic properties. In this thesis, a detailed simulation study of regime-switching models is conducted. A strategy to generate initial search values in the parameter estimation of regime-switching models is proposed. It is shown that this method can dramatically reduce the number of restarts of the optimizer. Even in 3-regime models (with 15 unknown parameters), parameters can be estimated reasonably well with only 5 restarts.

Acknowledgements

I want to thank my supervisor Dr. Yonggan Zhao, for introducing me to the area of financial modeling and regime-switching models. I would also like to thank my co-supervisor Dr. Bruce Smith, for his inspiring guidance, encouragement and great patience. I am grateful to Dr. Joanna Flemming who spent her valuable time serving on my committee. Lastly, I want to express my sincere gratitude to my wife Laura Lin for her love, encouragement and constant prayer, and to my mom who supports me in her own way during the course of this study.

Chapter 1

Introduction

A financial market often changes patterns over time. It can also exhibit dramatic changes due to unexpected events such as natural hazards and financial crisis. A widely accepted fact is that financial markets behave quite differently in different economic situations. Traders often adjust their portfolios according to market trend, which is defined as the long term tendency of a financial market to move in a certain direction. A financial market is traditionally classified into 3 categories: bearish, bullish and neutral. The first two terms describe overall market gain and loss respectively. The term neutral market is used when no strong upward or downward trend is observed.

As an example, we plot the weekly Standard & Poor index from January 2002 to March 2012 in Figure 1.1. The index dropped sharply in 2002 as a consequence of the collapse of internet bubble. We can consider the market over this period as bearish. The stock market recovered slowly since 2003 until the financial crisis hit United States in 2008 when the index experienced dramatic drop in the whole year. One could say the index was in bull market between 2003 and 2008. However, there are clearly patterns in the movement of the S&P index within this time period. The index bounced back sharply between 2003-2004 and increased at a much slower pace from 2005 to 2006. Should the market trend in those two periods be defined as the same or should we consider them differently in the analysis? In general, market categorization is very subjective and complicated. It is common to observe a few days of sharp market drop in a bullish market or a short period of market rally in a bearish market. There is no standard way to clearly define market trends by direct observation of financial data. This makes financial market analysis very challenging. To better understand the changing patterns of financial market, statistical modeling and analyses of financial data are often used to identify the underlying economic stages and examine how the market behaves under different economic regimes.

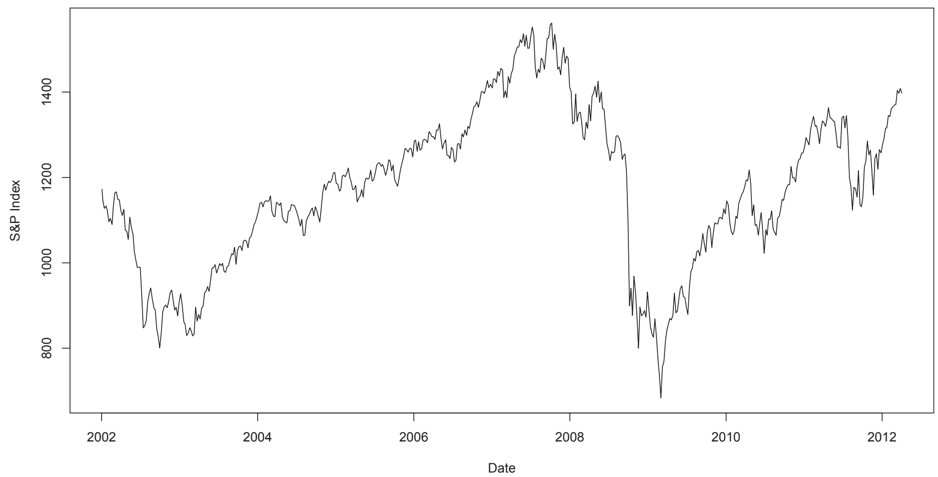


Figure 1.1: Weekly S&P index from January 2002 to March 2012

Among many approaches to this problem, regime-switching models have become very popular in the past decade. A regime-switching model is a time-series model in which parameters change values according to the regime at present time. The changing of regimes is governed by a finite-state stochastic process, which is often unobservable. There are two types of regime-switching models: threshold models and Markov models. In a threshold regime-switching model, regime switching is determined by the value of observed variables and predefined thresholds. A Markov regime-switching model, on the other hand, assumes that the unobservable stochastic process is a Markov chain. In this thesis, we focus our attention on Markov regime-switching models. Therefore the word 'Markov' will be omitted as long as there is no confusion in the context.

Regime-switching model is a special case of a more general framework called hidden Markov Model (see *Zucchini and MacDonald* (2009)). Regime-switching models were first introduced to econometric study by *Hamilton* (1989) and have become very popular particularly in applied works. Applications of regime switching models spread over a broad range of research areas, such as modeling shifts in inflation and interest rates (*Garcia and Perron* (1996); *Ang and Bekaert* (2002)), structural breaks in business cycle (*Kim and Nelson* (1999); *Piger et al.* (2005); *Altug and Bildirici* (2010)), changes in government policy (*Valente* (2003); *Owyang and Ramey* (2004); *Sims and Zha* (2006)) and shifts in exchange rate (*Bekaert and Hodrick* (1992); *Bollen et al.* (2000)). While the basic regime-switching model has been extended in many ways, there is a lack of literature for simulation studies. New methods based on regime-switching models are often proposed with neither a proof of convergence nor simulations to demonstrate their basic properties. The purpose of this thesis is to carry out a detailed simulation study of regime-switching models. We study different ways of estimating parameters in regime-switching models. We show that the convergence of empirical parameter estimates to true values, although it theoretically holds, can be very slow and even impossible to achieve especially on real financial data with high volatility.

The structure of this thesis is as follows: In the reminder of this chapter, we review some statistical prerequisites in order to understand regime-switching models. A very generic regime-switching model, based on which the simulation is carried out, is introduced in detail. Before stepping into the discussion of how to solve this model, we first study two easier regime-switching submodels in Chapter 2, in which some parameters are assumed to be fixed. A full empirical analysis of the generic model is presented in Chapter 3. Finally, a summary of this study and comments regarding future work are presented in Chapter 4.

1.1 A Basic Regime-Switching Model

Regime-switching models rely on a special type of stochastic process called Markov chain, which is slightly more complicated than a sequence of i.i.d random variables. Given all previous states, the present state of a Markov chain depends on and only on the last state. This condition is called the *Markov property*. Formally,

$$P\left(X_t|X_{t-1},\ldots,X_0\right) = P\left(X_t|X_{t-1}\right), \tag{1.1}$$

where X_t denotes the regime at time t. Now we can define *Markov chain* as

Definition 1.1.1. A stochastic process $\{X_t, t = 0\ldots\}$ is said to be a *Markov chain* if the following conditions are satisfied

- X_n only takes on a finite or countable number of values $\{a_i, i = 0\ldots\}$. The set $\{a_i, i = 0\ldots\}$ is called the *state space* of $\{X_t, t = 0\ldots\}$.
- X_n satisfies the Markov property as defined in (1.1).

If the state space $\{a_i, i = 0 \ldots\}$ is a finite set, we say $\{X_t, t = 0 \ldots\}$ is a *finite Markov chain*. A Markov chain is said to be *time-homogeneous* if the transition probability distribution $P(X_t | X_{t-1})$ is independent of time t. In this thesis, we will only consider finite state time-homogeneous Markov chains, which will simply be called Markov chains when there is no confusion in the context.

Associated with each Markov chain is a transition probability matrix $\boldsymbol{P}$, where entry $p_{i,j}$ is the probability of switching from regime i to regime j. In general, the states of a Markov chain are directly observable so that the only parameter is $\boldsymbol{P}$. In certain situations, the states are not visible but can emit observable outputs. In those cases, true states as well as transition matrix $\boldsymbol{P}$ should be inferred from observations. The underlying statistical models are called hidden Markov models, which include regime-switching models as a special case.

In what follows, we will describe the regime-switching model for our experimentation. Suppose we are interested in modeling and predicting a stationary time series $\{y_t\}$, which is generated by the following AR(1) model:

$$y_t = c_{s_t} + \phi_{s_t} y_{t-1} + \epsilon_t \tag{1.2}$$

where

$$\epsilon_t \sim N\left(0, \sigma_{s_t}^2\right).$$

Here s_t represents the present regime at time t. s_t can shift among N possible values $1, \ldots, N$ according to a Markov chain. The parameters $(c_{s_t}, \phi_{s_t}, \sigma_{s_t})$ all depend on the value of s_t. The regime-shifting feature of this model enables $\{y_t\}$ to exhibit different behaviors over time, which could be helpful in better describing the dynamics of financial markets.

1.2 The Hamilton Filter

In this section, we present a well-known algorithm called the Hamilton Filter, which is used to make inference about current regime s_t in model (1.2). First, we need to introduce some notations:

- p_{ij}: transition probability between regime i and j.

$$P(s_t = j | s_{t-1} = i, \ldots, s_1 = k, y_{t-1}, \ldots, y_0) = P(s_t = j | s_{t-1} = i) = p_{ij}$$

- $\boldsymbol{\theta}$: the vector of parameters:

$$\boldsymbol{\theta} = (\phi_1, \ldots, \phi_N, c_1, \ldots, c_N, \sigma_1, \ldots, \sigma_N, p_{11}, \ldots, p_{N,N-1})$$

- $\boldsymbol{Y}_t$: the vector of historical observations up to time t

$$\boldsymbol{Y}_t = (y_t, \ldots, y_0)$$

- η_{jt}: conditional probability density function of y_t given current regime s_t and past observations up to time $t-1$

$$\begin{aligned} &\eta_{jt} \\ = \ & f(y_t | s_t = j, \boldsymbol{Y}_{t-1}; \boldsymbol{\theta}) \\ = \ & f(y_t | s_t = j, y_{t-1}; \boldsymbol{\theta}) \\ = \ & \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left[-\frac{(y_t - c_j - \phi_j y_{t-1})^2}{2\sigma_j^2}\right] \end{aligned} \tag{1.3}$$

- ξ_{jt}: conditional regime probability distribution given all observations up to time t

$$\xi_{jt} = P(s_t = j | \boldsymbol{Y}_t; \boldsymbol{\theta})$$

In general, ξ_{jt} is the quantity of primary interest and we will show next how to calculate it. Using Bayes rule, ξ_{jt} can be represented as

$$\begin{aligned} &\xi_{jt} \\ = \ & P(s_t = j | y_t, \boldsymbol{Y}_{t-1}; \boldsymbol{\theta}) \\ = \ & \frac{f(y_t | s_t = j, \boldsymbol{Y}_{t-1}; \boldsymbol{\theta}) P(s_t = j | \boldsymbol{Y}_{t-1}; \boldsymbol{\theta})}{f(y_t | \boldsymbol{Y}_{t-1}; \boldsymbol{\theta})} \\ = \ & \frac{\eta_{jt} \cdot P(s_t = j | \boldsymbol{Y}_{t-1}; \boldsymbol{\theta})}{f(y_t | \boldsymbol{Y}_{t-1}; \boldsymbol{\theta})} \end{aligned} \tag{1.4}$$

where η_{jt} can be directly calculated by (1.3). The second term in the numerator of (1.4) is equal to

$$\begin{aligned}
& P\left(s_t = j|\boldsymbol{Y}_{t-1};\boldsymbol{\theta}\right) \\
= & \sum_{i=1}^{N} P\left(s_t = j, s_{t-1} = i|\boldsymbol{Y}_{t-1};\boldsymbol{\theta}\right) \\
= & \sum_{i=1}^{N} P\left(s_t = j|s_{t-1} = i, \boldsymbol{Y}_{t-1};\boldsymbol{\theta}\right) P\left(s_{t-1} = i|\boldsymbol{Y}_{t-1};\boldsymbol{\theta}\right) \\
= & \sum_{i=1}^{N} p_{ij}\xi_{i,t-1} \qquad (1.5)
\end{aligned}$$

The denominator of (1.4) can be calculated as

$$\begin{aligned}
& f\left(y_t|\boldsymbol{Y}_{t-1};\boldsymbol{\theta}\right) \\
= & \sum_{k=1}^{N} f\left(y_t, s_t = k|\boldsymbol{Y}_{t-1};\boldsymbol{\theta}\right) \\
= & \sum_{k=1}^{N} f\left(y_t|s_t = k, \boldsymbol{Y}_{t-1};\boldsymbol{\theta}\right) P\left(s_t = k|\boldsymbol{Y}_{t-1};\boldsymbol{\theta}\right) \\
= & \sum_{k=1}^{N} \eta_{kt} \sum_{i=1}^{N} p_{ik}\xi_{i,t-1} \qquad (1.6)
\end{aligned}$$

Substituting (1.5) and (1.6) back to (1.4) gives

$$\xi_{jt} = \frac{\eta_{jt}\sum_{i=1}^{N} p_{ij}\xi_{i,t-1}}{\sum_{k=1}^{N}\eta_{kt}\sum_{i=1}^{N} p_{ik}\xi_{i,t-1}} \qquad (1.7)$$

If $\boldsymbol{\theta}$ is given, the calculation of ξ_{jt} only depends on $\xi_{i,t-1}$ and η_{it} for $i = 1, \ldots, N$. This implies an algorithm to iteratively calculate ξ_{jt}, which was first proposed in *Hamilton* (1989). One of the advantages of the Hamilton Filter is that the log likelihood can be derived as a by-product in the calculation of ξ_{jt}. The log likelihood function of the regime switching model can be written as

$$
\begin{aligned}
& \log f\left(y_1, \ldots, y_n | \boldsymbol{\theta}\right) \\
= & \sum_{t=1}^{n} \log f\left(y_t | \boldsymbol{Y}_{t-1}; \boldsymbol{\theta}\right) \\
= & \sum_{t=2}^{n} \log f\left(y_t | \boldsymbol{Y}_{t-1}; \boldsymbol{\theta}\right) + \log f\left(y_1 | y_0; \boldsymbol{\theta}\right) \\
= & \sum_{t=2}^{n} \log f\left(y_t | \boldsymbol{Y}_{t-1}; \boldsymbol{\theta}\right) + \log \sum_{j=1}^{N} f\left(y_1 | s_1 = j, y_0; \boldsymbol{\theta}\right) P(s_1 = j | y_0; \boldsymbol{\theta})
\end{aligned}
$$

It is in general not feasible to calculate $P(s_1|y_0;\boldsymbol{\theta})$ or $f(y_0;\boldsymbol{\theta})$. In practice, one can replace $P(s_1|y_0;\boldsymbol{\theta})$ by any nonnegative vector $\boldsymbol{\pi}$ whose entries sum up to 1 and work with the modified log likelihood

$$
\sum_{t=2}^{n} \log f\left(y_t | \boldsymbol{Y}_{t-1}; \boldsymbol{\theta}\right) + \log \sum_{j=1}^{N} f\left(y_1 | s_1 = j, y_0; \boldsymbol{\theta}\right) \pi_j \tag{1.8}
$$

where $\boldsymbol{\pi}$ is often chosen to be the stationary distribution of the transition matrix $\boldsymbol{P}$. As indicated by *Krishnamurthy and Ryden* (1998), the choice of $\boldsymbol{\pi}$ will not change the asymptotic properties of the maximum-likelihood estimator (MLE). Therefore its impact on parameter estimation will be minor when the number of observations is large. In this thesis, initial state probabilities are equally weighted i.e $\pi_i = \pi_j = 1/N$ for all i, j. A detailed description of the Hamilton Filter with likelihood calculation is outlined in Algorithm 1.

1.3 Parameter Estimation and Consistency of MLE

The parameter $\boldsymbol{\theta}$ can be estimated by maximizing the modified loglikelihood function (1.8). The first theoretical result regarding the consistency of the MLE of general hidden Markov models was given by *Leroux* (1992). Later *Bickel et al.* (1998) proved the asymptotically normality of the MLE for general hidden Markov models. It was also shown that the observed information matrix is a consistent estimator of the Fisher information matrix. An important assumption in those proofs is that the conditional distribution of y_t depends on

Algorithm 1 Hamilton Filter

Require: y_t where $t = 0 \ldots n-1$ and $p_{ij}, c_i, \phi_i, \sigma_i$ with $i, j = 1 \ldots N$

Set $\xi_{i,0} = 1/N, i = 1 \ldots N$
$loglike = 0$
for t from 1 to n **do**
    for i from 1 to N **do**
        Set $\eta_{it} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y_t - c_i - \phi_i y_{t-1})^2}{2\sigma_i^2}\right]$
    end for
    Set $sum = 0$
    for j from 1 to N **do**
        $f_j = \eta_{jt} \sum_{i=1}^{N} p_{ij} \cdot \xi_{i,t-1}$
        $sum = sum + f_j$
    end for
    for j from 1 to N **do**
        $\xi_{jt} = f_j / sum$
    end for
    $loglike = \log(sum)$
end for

and only on the hidden state x_t i.e the following condition is assumed to be satisfied:

$$P\left(y_t | \mathbf{Y}_{t-1}, \boldsymbol{X}_t\right) = P\left(y_t | x_t\right)$$

This assumption is clearly not true in the regime-switching model specified by (1.2), where y_t also depends on y_{t-1}. For that reason, regime-switching model (1.2) is also called Markov-switching autoregression. *Krishnamurthy and Ryden* (1998) first proved the consistency of the MLE of finite state regime switching models. *Douc et al.* (2004) proved later the consistency and asymptotic normality of the MLE in the more general case when the state space X_n is compact. For completeness of this thesis, the main results of *Krishnamurthy and Ryden* (1998) are summarized in Theorem 1.3.1.

Theorem 1.3.1 (*Krishnamurthy and Ryden* (1998))**.** Let $\{X_k\}_{k=1}^{\infty}$ be a finite state Markov chain with values $\{1, \ldots, r\}$. Here r is assumed to be known and fixed. Denote the transition probabilities as $a_{ij} = P\left(X_n = j | X_{n-1} = i\right)$ for $i, j = 1, \ldots, r$, where $a_{ij} = a_{ij}(\boldsymbol{\phi})$ is a function of a parameter vector $\boldsymbol{\phi}$ in a compact Euclidean space $\boldsymbol{\Phi}$. For each $\boldsymbol{\phi}$, let $Q_{\boldsymbol{\phi}}$ be the transition kernel of this Markov chain. Denote $\{e_k\}_{k=1}^{\infty}$ a sequence of i.i.d random variables with known marginal distribution. Let d be a positive integer and

$\{Y_k\}_{k=-d+1}^{\infty}$ be a real-valued (nonlinear) autoregressive series of lag d such that

$$Y_n = g\left(Y_{n-1}, \ldots, Y_{n-d}, e_n; \boldsymbol{\theta}_{X_n}(\boldsymbol{\phi})\right)$$

where g is a real-valued function and $\boldsymbol{\theta}_i$ are functions from $\boldsymbol{\Phi}$ to a Euclidean space $\boldsymbol{\Theta}$. Finally, let $\boldsymbol{\phi}^0$ denote the true parameter. Assume the following conditions hold

- $A(\boldsymbol{\phi}^0)$ is irreducible and $Q_{\boldsymbol{\phi}^0}$ has a unique stationary distribution.
- The Markov chain $\{Z_k\}_{k=1}^{\infty} = \{(X_k, Y_k, \ldots, Y_{k-d+1})\}_{k=1}^{\infty}$ is ergodic under $\boldsymbol{\phi}^0$.
- $a_{ij}(\boldsymbol{\phi})$ and $\boldsymbol{\theta}_i(\boldsymbol{\phi})$ are continuous and $f\left(y_1|y_0, \ldots, y_{-d+1}; \boldsymbol{\theta}_i(\boldsymbol{\phi})\right)$ is continuous.
- For each i and $\boldsymbol{\phi} \in \boldsymbol{\Phi}$,

  $$E_{\boldsymbol{\phi}^0}\left\{\sup_{|\phi'-\phi|\leq\delta}|\log f(Y_1|Y_0, \ldots, Y_{-d+1}; |\boldsymbol{\theta}_i(\boldsymbol{\phi}')\right\} < \infty$$

  for some $\delta > 0$

Then for each $\boldsymbol{\phi} \in \boldsymbol{\Phi}$, there exists a constant $H(\boldsymbol{\phi})$ such that

$$\frac{1}{n}\log\max_i f\left(Y_1, \ldots, Y_n|Y_0, \ldots, Y_{-d+1}, X_1 = i; \boldsymbol{\phi}\right) \to H(\boldsymbol{\phi})$$

as $n \to \infty$. Furthermore, define $K(\boldsymbol{\phi}) = H(\boldsymbol{\phi}^0) - H(\boldsymbol{\phi})$ and $\boldsymbol{\Phi}^0 = \{\boldsymbol{\phi} \in \boldsymbol{\Phi} : K(\boldsymbol{\phi}) = 0\}$. Denote $\hat{\phi}_n$ the MLEs (a set of estimates that maximize the conditional likelihood) with n observations. We have $\sup_{\phi\in\hat{\phi}_n} \inf_{\phi'\in\boldsymbol{\Phi}^0} |\boldsymbol{\phi} - \boldsymbol{\phi}'| \to 0$ almost surely.

From a practice perspective, the MLE can be obtained by maximizing the modified conditional log-likelihood (1.8). Note that this is a constrained optimization problem with $0 < p_{ij} < 1$, $\sum_{i=1}^{N} p_{ij} = 1$, $\sigma_i > 0$ and $0 < \phi_i < 1$ for all $i, j = 1 \ldots N$. In a more general case when the ϕ_i are bounded within interval $[a, b]$, appropriate transformations can be applied so that $0 < \phi_i < 1$. Given these constraints, we can transform parameters to convert the maximization of (1.8) into an unconstrained problem. For σ_i and ϕ_i, let us define $l_i = \log(\sigma_i), r_i = \text{logit}(\phi_i)$ and optimize with respect to l_i and r_i instead. In order to optimize $\boldsymbol{P}$, we define an ordered list for each regime i

$$\{-\infty < \gamma_{i,1} < \gamma_{i,2} < \ldots < \gamma_{i,N-1} < \infty\}$$

such that

$$\begin{aligned} P(z < \gamma_{i,1}) &= p_{i,1} \\ P(\gamma_{i,1} < z < \gamma_{i,2}) &= p_{i,2} \\ &\cdots \\ P(\gamma_{i,N-1} < z) &= p_{i,N} \end{aligned}$$

where $z \sim N(0,1)$. In this way, we can re-parameterize the model using

$$\gamma_{i,1}, \log(\gamma_{i,2} - \gamma_{i,1}), \ldots, \log(\gamma_{i,N-1} - \gamma_{i,N-2})$$

and convert the optimization problem into an unconstrained one.

Two types of methods are commonly used to find the global maximum of (1.8). The first approach is to use Expectation-Maximization (EM) algorithms. An EM method is an iterative algorithm for finding MLE of parameters in statistical models with latent variables. A single iteration of an EM algorithm consists of two steps: an expectation step to calculate the expectation of the loglikelihood of the complete data and a maximization step to maximize this quantity with respect to the unknown parameters. The EM algorithm is usually credited to *Dempster et al.* (1977). It is well known that the EM algorithm converges to the nearest local maximum (*Wu* (1983)). For a detailed discussion of theory and applications of EM algorithms, one can refer to *McLachlan and Krishnan* (2008).

Another approach to obtain MLE is via direct optimization of (1.8) using a numerical optimization solver. Loosely speaking, numerical optimization techniques can be classified into two groups, either as deterministic methods or stochastic methods. Deterministic optimization algorithms have no ingredient of randomness in their designs. Their underlying mechanisms often assume certain properties of the objective function or rely on geometrical structure of the search space. Popular deterministic optimization frameworks and methods include line-search algorithms, trust-region algorithms and Nelder-Mead methods etc. Stochastic optimization algorithms, on the other hand, allow (intelligent) random search of the parameter space. Simulated annealing (*Kirkpatrick et al.* (1983)), particle swarm optimization (*Kennedy and Eberhart* (1995)) and evolutionary algorithms (*Bäck* (1996)) are examples of popular stochastic optimization techniques.

In this thesis, the Nelder-Mead algorithm is used to optimize (1.8). The Nelder-Mead algorithm is a deterministic optimization method which was proposed by *Nelder and Mead* (1965). It is probably one of the most popular optimization algorithms for the past 40 years. Due to its simplicity and robustness, Nelder-Mead is still being extensively used in many applications such as optical engineering (*Tse et al.* (2010)), biostatistics (*Minard et al.* (2011)) and kinematics (*Wang et al.* (2011)). Nelder-Mead utilizes no derivative information. These type of methods are often called derivative-free or direct search methods.

We now examine a simple illustrative example which helps to reveal the basic properties of a regime switching model and the structure of the associated loglikelihood function. Consider a two regime model with the following generating parameters

$$\boldsymbol{\phi} = (0.97, 0.92)\,, \boldsymbol{c} = (12, 28)\,, \boldsymbol{\sigma} = (5, 8)\,, p = \begin{pmatrix} 0.75 & 0.25 \\ 0.1 & 0.9 \end{pmatrix} \tag{1.9}$$

From this model, we sample a sequence of 500 data points as observations. The simulated sequence and true regimes are shown in Figure 1.2. The portions of the sequence that are generated from regime 2 are highlighted in red. Suppose all parameters are known except for $\phi_1 = 0.97$ and $c_1 = 12$, which should be estimated from the data. In Figure 1.3, we plot the loglikelihood function and its contour. Note the loglikelihood function has larger values over a long and narrow region. Along a line passing through the center of this region, the loglikelihood gradually decreases as it moves away from the global maximum. There are also many local maxima along the same direction. On the other hand, the loglikelihood value drops sharply in other directions. The structure of this loglikelihood function indicates a possible correlation between ϕ and c.

In this special example, the only unknowns are ϕ_1 and c_1, which are the multiplicative coefficient and the constant in the AR(1) process. It is likely that ϕ_1 and c_1 interplay in a similar way as in a AR(1) model. For a stationary AR(1) model, the mean of the series is $\frac{c}{1-\phi}$. It is then reasonable to assume that the gradient (i.e the direction of the steepest decrease of loglikelihood) is perpendicular to

$$\frac{c}{1-\phi} = \frac{12}{1-0.97} = 400.$$

Or equivalently,

$$\phi = 1 - \frac{c}{400} \tag{1.10}$$

The contour plot in Figure 1.3b seems to confirm our hypothesis. The black line is as defined in (1.10) and the circle represents the true parameters. Along (1.10), the loglikelihood function remains large even at points far away from the true values. Moving in other directions, particularly along the line perpendicular to (1.10), the value of the loglikelihood function quickly drops. This implies that the shape of the loglikelihood function is not only determined by ϕ_i and c_i but also by means of the underlying time series. The optimization of loglikelihood function (1.8) is generally very hard as we can observe from Figure (1.3a), in which many local minima exist around the neighborhood of the global maximum.

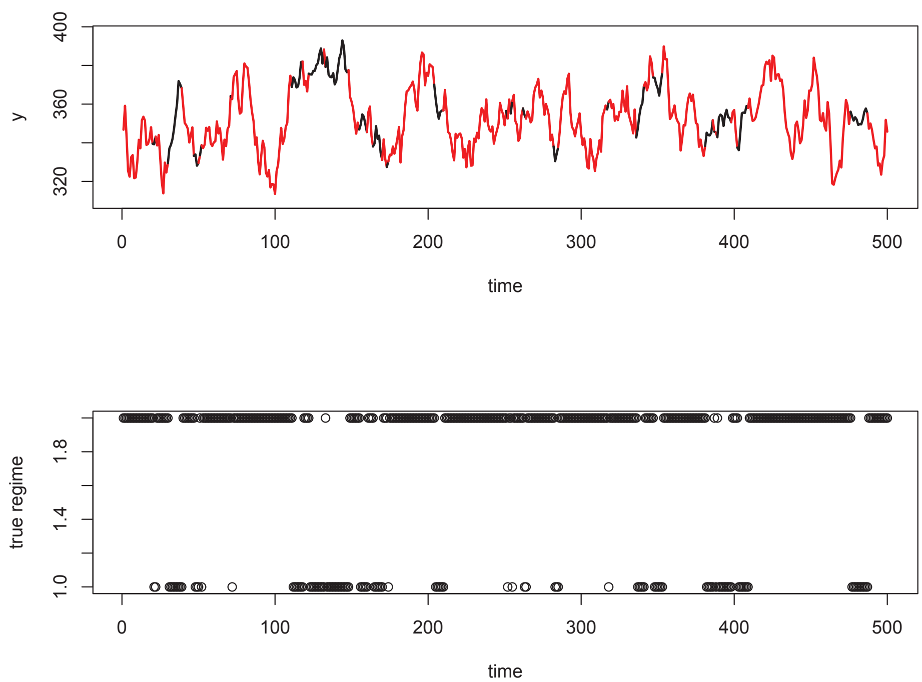
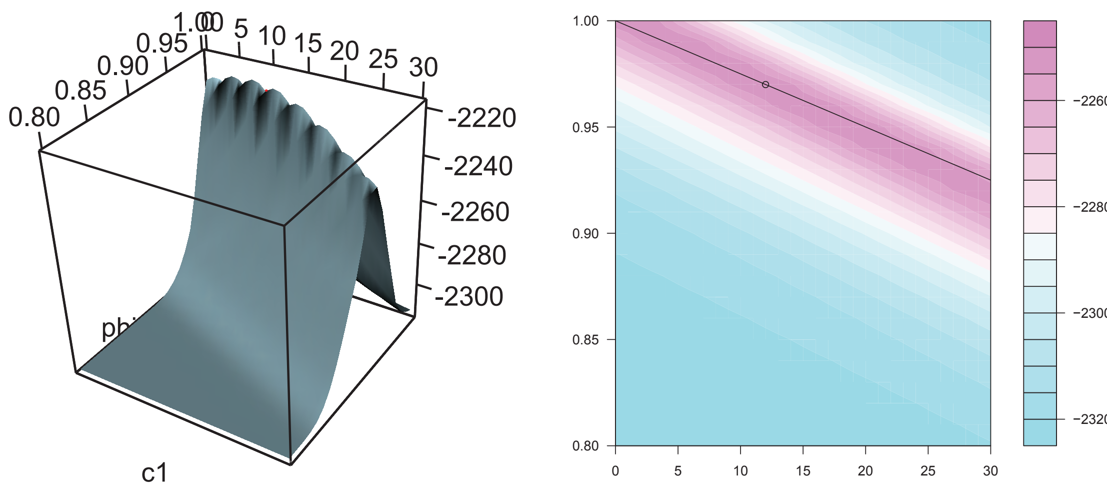


Figure 1.2: A simulated sequence with true regimes. Data from regime 2 is colored in red in the sequence plot.



(a) loglikelihood

(b) contour plot

Figure 1.3: In (a), the loglikelihood function of the regime switching model specified in (1.9) is plotted as a function of ϕ_1 and c_1. All other parameters were assumed to be known. A total of 500 points was simulated. The contour of the loglikelihood function is plotted in (b). The circle indicates the true parameters. The points on the line define the same sequence mean as the true parameters.

CHAPTER 2

PARAMETER ESTIMATION OF REGIME-SWITCHING SUBMODELS

At the end of Chapter 1, we examined an example of a regime-switching model in order to illustrate the difficulties in parameter estimation for this type of models. We showed that there might be correlation between ϕ and c, which introduces many local maxima to the loglikelihood function. In order to further investigate this hypothesis, we study two regime-switching submodels which are more restrictive than (1.2). In these two models, we assume that either ϕ_i or c_i is known and try to estimate all other parameters under those circumstances. Those two models are easier to estimate than (1.2). Our goal is to provide some empirical evidence to understand how parameter estimation is influenced by ϕ_i and c_i in a separate manner. As our experimental results show, these two restrictive models are already very hard to estimate by themselves. It should be noted that more restrictive regime-switching models are not unusual in applied work.

2.1 A Mixture Model with Switching Feature

2.1.1 Model and Experimental Settings

The first model to be considered is a simplification of (1.2) where the multiplicative coefficients ϕ_i are zero. More specifically, we study

$$y_t = c_{s_t} + \epsilon_t \tag{2.1}$$

N=2			
	$\boldsymbol{c}$	$\boldsymbol{\sigma}$	$\boldsymbol{P}$
M1	$\boldsymbol{c} = (0, 10)$	$\boldsymbol{\sigma} = (1, 2)$	$\boldsymbol{P} = \begin{pmatrix} 0.75 & 0.25 \\ 0.1 & 0.9 \end{pmatrix}$
M2	$\boldsymbol{c} = (0, 5)$	$\boldsymbol{\sigma} = (1, 2)$	$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix}$
M3	$\boldsymbol{c} = (0, 2)$	$\boldsymbol{\sigma} = (1, 2)$	$\boldsymbol{P} = \begin{pmatrix} 0.85 & 0.15 \\ 0.1 & 0.9 \end{pmatrix}$
N=3			
M4	$\boldsymbol{c} = (0, 10, 20)$	$\boldsymbol{\sigma} = (1, 2, 3)$	$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.75 \end{pmatrix}$
M5	$\boldsymbol{c} = (0, 5, 10)$	$\boldsymbol{\sigma} = (1, 2, 3)$	$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.75 \end{pmatrix}$
M6	$\boldsymbol{c} = (0, 2, 4)$	$\boldsymbol{\sigma} = (1, 2, 3)$	$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.75 \end{pmatrix}$

Table 2.1: Parameters in the simulation of model (2.1)

where s_t is the index of N possible regimes $1, 2, \ldots, N$ and

$$\epsilon_t \sim N\left(0, \sigma_{s_t}^2\right).$$

The sequence $\{s_t\}$ follows a Markov chain with transition matrix $\boldsymbol{P}$. Associated with each regime i is a pair of parameters (c_i, σ_i), which are to be inferred from the data together with $\boldsymbol{P}$. In the rest of this thesis, we refer to this model as the *switching mixture model*.

The switching mixture model (2.1) can be considered as an extension of finite mixture models, which are well studied. A finite mixture model is defined in exactly the same way as in (2.1) except s_t is assumed to follow a multinomial distribution. The general theory and applications of finite mixture models are well presented in *McLachlan and Peel* (2005).

In our simulation, data is generated from six different models whose parameters are listed in Table 2.1. Three of the six models (M1-M3) have two regimes and the other three (M4-M6) have three regimes. The models are numbered in an order of increasing difficulty. The absolute differences between the c_i increase in model M1, M2 and M3, while the standard deviation of the error is kept the same. This makes M1 the easiest model and M3 the hardest model to fit. Similarly, parameters of M6 are the hardest to estimate, whereas M5 is relatively easier to fit and M4 is the easiest model. In Figure 2.1, three sequences simulated from models M1, M2 and M3 are plotted. It is apparent from the graph that M1 is the easiest model among the three. The sequence from M1 has a clear pattern and data points from two different regimes are well separated. On the other hand, it is very difficult to tell the number of regimes in M2 and M3 by looking directly at the data. Three sequences simulated from model M4 - M6 are plotted in Figure 2.2. Similar patterns can be observed from those sequences, although separation of clusters is not clear even in the easiest model M4.

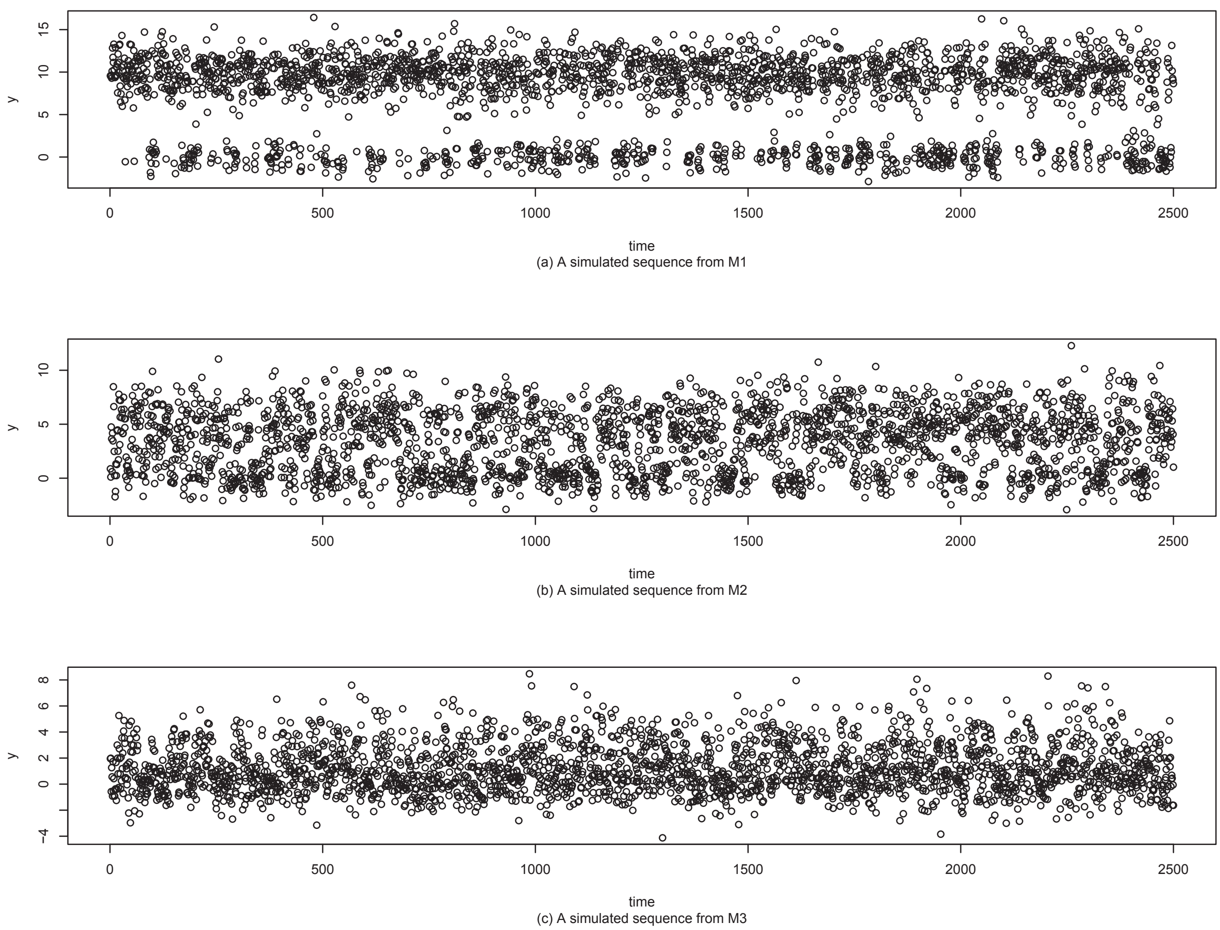


Figure 2.1: Simulated sequences from switching mixture model M1 - M3 are plotted in (a) - (c) respectively. Data points from two different regimes of M1 are clearly separated. It is difficult to tell the number of regimes by directly looking at the sequences from M2 and M3.

From each model, 100 sequences with 2500 data points were sampled. For each sequence, parameters were estimated on the first 100, 400, 900, 1600 and 2500 points. The number of regimes was assumed to be known in parameter estimation.

In order to obtain the MLE, it is critical to choose appropriate starting points. In

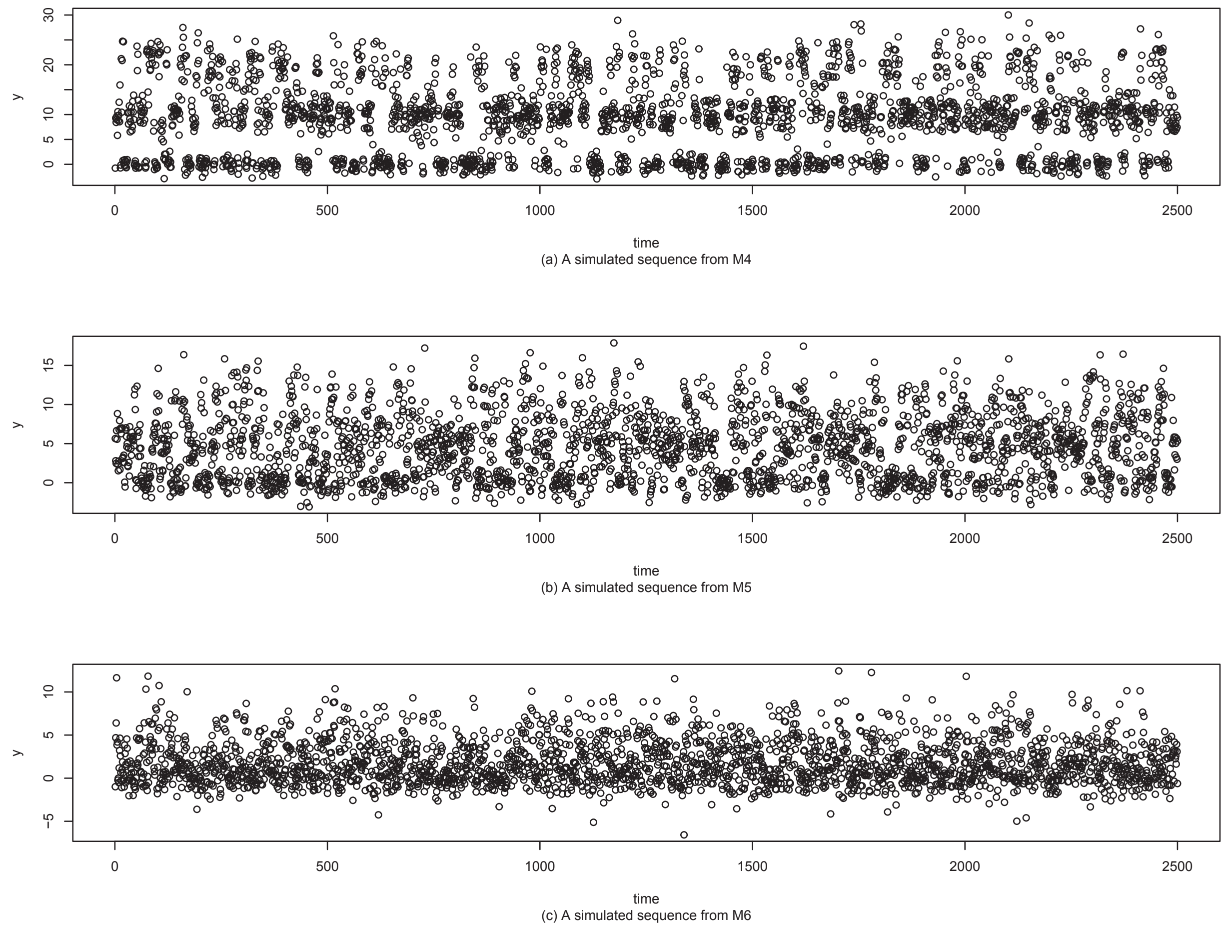


Figure 2.2: Simulated sequences from switching mixture model M4 - M6 are plotted in (a) - (c) respectively.

numerical optimization, the choice of initial search values has an great impact on the performance of the optimizer. The closer the initial values are to the true parameters, the more likely an optimizer is able to converge to the global optimum and thus provide a better estimate. For derivative-free black box optimization problems, random sampling is the only choice to generate initial values without knowing anything about the contours of the objective function. The focus is often on restart strategies, which aim to intelligently restart the optimizer at different regions of the search space to avoid unnecessary calculation. In most analyses of financial data using regime switching models, optimizers are simply restarted from a large pool of initial values. The choices of those initial values are usually subjective and reflect the author's belief about where the global maximum resides.

Most optimization problems in statistical modeling are aimed at estimating model parameters. Their objective functions are usually loglikelihood functions. Although derivative information is not available in general, these optimization problems are not completely "black box". Rather, it is possible to extract some information from the data to generate initial values that may be in the neighborhood of the true parameters. As shown above, the only difference between mixture models and switching mixture models is in their switching mechanisms. The Markov switching feature introduces a few more parameters and makes things slightly more complicated. However, we believe the differences between these two models to be minor. As a demonstration, we plot the loglikelihood function of model M3 with respect to c_1 (i.e all parameters except for c_1 are fixed) for a simulated sequence in Figure 2.3. Plotted in the same graph are the loglikelihood functions of the corresponding mixture model for the same sequence but with a different number of points. Mixing coefficients of the mixture model are calculated as the limiting probability of the transition matrix of M3. The loglikelihood functions are shifted so that they can be incorporated in the same figure. All loglikelihood functions peak around 0, which is equal to the true c_1. Although the sequence is simulated from M3, the MLE of the mixture model is very close to the true parameter. This suggests that we might be able to use mixture models as approximates to switching mixture models in parameter estimation.

In order to generate initial values in our experimentation, a mixture model is fitted to the simulated data using the R library 'mclust' (*Fraley et al.* (2012)). The fitting outputs include estimates of c_i and σ_i for $i = 1, \ldots, N$, which are used as initial values in the switching mixture model. Each observation y_i will also be assigned to one of the N regimes by 'mclust', from which n_{ij}, the number of times $\{y_t\}$ switch between regime i to regime j can be obtained. The initial transition probability p_{ij} can be calculated as $n_{ij}/(n-1)$, where n is the length of the sequence.

2.1.2 Experimental Results

The results of parameter estimation are listed in Table 2.2 - Table 2.7. In these tables, we report the percentage of the 100 runs in which the optimizer converges (denoted as cp) and the percentage for which the estimated parameters have a larger loglikelihood value than for that of the true parameters (denoted as bp). An estimated parameter will be considered as a good estimate if it has a higher loglikelihood than the true parameter. In the tables, we also report the mean and standard deviation of the estimated parameters. It appears that the switching mixture model (2.1) can be well estimated in almost all cases, implying that

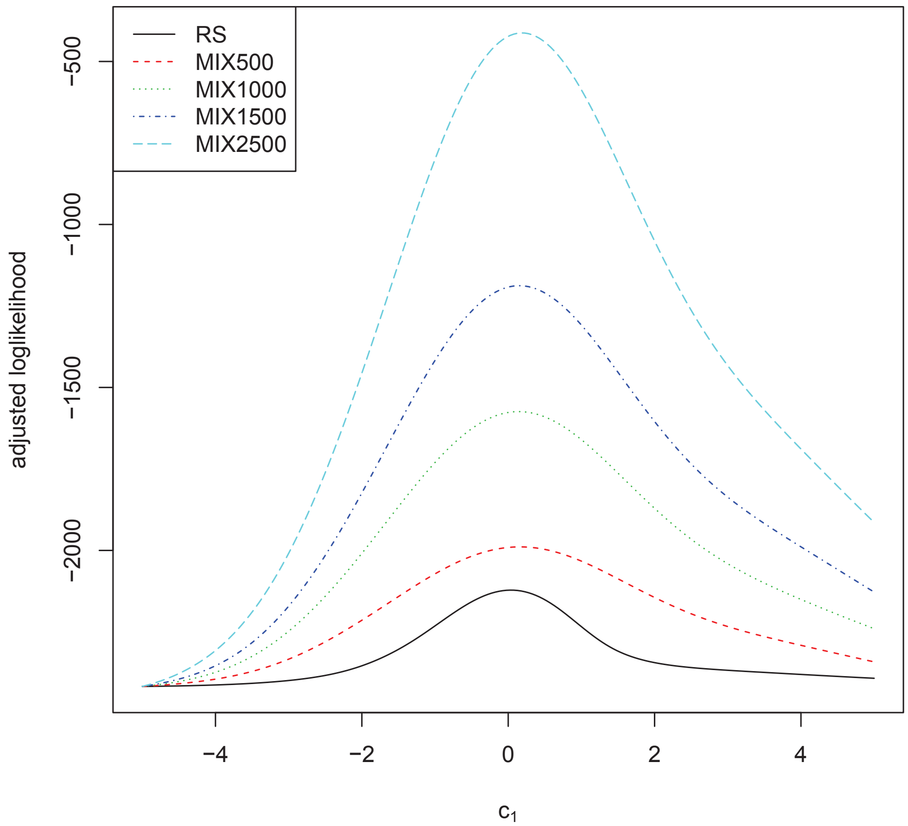


Figure 2.3: Loglikelihood functions of model M3 and the associated mixture model with respect to c_1. Data is simulated from M3. Mixing coefficients of the mixture model are calculated as the limiting probability of the transition matrix of M3. The loglikelihood functions are shifted to start from the same value.

the mixture model is able to provide good initial values. This confirms our assumption that switching mixture models are very well approximated by mixture models.

In all three two-regime models M1 - M3, both the percentage of convergence and the percentage of better likelihood are close to 100% in most cases. The only exception is when 100 data points were used in the estimation of M3. Even in this case, both bp and cp are larger than 90%. Boxplots of estimate errors in M1, M2 and M3, which were calculated as the difference between estimated parameters and true parameters, are given in Figure 2.4. In all cases, the estimated value converges to the true parameter and the standard deviation of errors decreases as the number of points used for estimation increases. Out of the three models, M3 appears to be the hardest model to fit. However when n is large, the

N=2						
	cp	bp		$\boldsymbol{c} = (0, 10)$	$\boldsymbol{\sigma} = (1, 2)$	$\boldsymbol{P} = \begin{pmatrix} 0.75 & 0.25 \\ 0.1 & 0.9 \end{pmatrix}$
$n = 100$	100%	100%	Mean	(0.05,10)	(0.94,1.97)	$\begin{pmatrix} 0.72 & 0.28 \\ 0.097 & 0.903 \end{pmatrix}$
			Std	(0.2,0.26)	(0.14,0.17)	$\begin{pmatrix} 0.097 & 0.097 \\ 0.033 & 0.033 \end{pmatrix}$
$n = 400$	100%	100%	Mean	(0.003,10.02)	(0.995,1.99)	$\begin{pmatrix} 0.745 & 0.255 \\ 0.099 & 0.901 \end{pmatrix}$
			Std	(0.08,0.12)	(0.069,0.079)	$\begin{pmatrix} 0.041 & 0.041 \\ 0.019 & 0.019 \end{pmatrix}$
$n = 900$	100%	100%	Mean	(0.0023,10.013)	(1.003,1.998)	$\begin{pmatrix} 0.746 & 0.254 \\ 0.101 & 0.899 \end{pmatrix}$
			Std	(0.058,0.076)	(0.041,0.055)	$\begin{pmatrix} 0.029 & 0.029 \\ 0.012 & 0.012 \end{pmatrix}$
$n = 1600$	100%	100%	Mean	(-0.00027,10.011)	(1.004,1.998)	$\begin{pmatrix} 0.747 & 0.253 \\ 0.101 & 0.899 \end{pmatrix}$
			Std	(0.042,0.062)	(0.029,0.038)	$\begin{pmatrix} 0.022 & 0.022 \\ 0.0085 & 0.0085 \end{pmatrix}$
$n = 2500$	100%	100%	Mean	(0.00093,10.013)	(1.006,1.997)	$\begin{pmatrix} 0.747 & 0.253 \\ 0.101 & 0.899 \end{pmatrix}$
			Std	(0.0033,0.052)	(0.024,0.035)	$\begin{pmatrix} 0.017 & 0.017 \\ 0.0063 & 0.0063 \end{pmatrix}$

Table 2.2: Estimation Result of model M1

differences in the standard deviation of estimation errors for the three models becomes small. Moreover, $n = 400$ seems to be a threshold in the performance of the estimator. When $n = 100$, none of the three models can be estimated well. The absolute values and variances of the estimation errors are quite large in this case. When $n \geq 400$, the differences in the error means and variances among models become small.

Parameters in three-regime models M4 - M6 can be estimated reasonably well. In all cases, the cp value is close to 100%. The value of bp remains high for different n when fitting M4 and M5. On the other hand, bp drops as n increases for model M6. It seems that as more information becomes available, the more likely it is that the optimizer gets trapped in a local maximum. Note M6 is the hardest model to fit among M4 - M6. The corresponding mixture model is also difficult to fit by itself. This would result in initial values that are far away from the true parameters even when n is large. Boxplots of errors in the parameter estimation of M4 - M6 are given in Figure 2.5. Errors in the estimates of M4 and M5 are comparable, whereas estimation errors in M6 have significant larger variances. For all 3 models, at least 900 data points are required in order to get reasonable estimates.

N=2						
	cp	bp		$\boldsymbol{c} = (0, 5)$	$\boldsymbol{\sigma} = (1, 2)$	$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix}$
$n = 100$	100%	100%	Mean	(0.03,5.01)	(1,1.97)	$\begin{pmatrix} 0.78 & 0.22 \\ 0.11 & 0.89 \end{pmatrix}$
			Std	(0.22,0.33)	(0.17,0.25)	$\begin{pmatrix} 0.091 & 0.091 \\ 0.049 & 0.049 \end{pmatrix}$
$n = 400$	100%	100%	Mean	(0.005,5.02)	(0.998,1.99)	$\begin{pmatrix} 0.802 & 0.198 \\ 0.103 & 0.897 \end{pmatrix}$
			Std	(0.099,0.16)	(0.068,0.12)	$\begin{pmatrix} 0.042 & 0.042 \\ 0.023 & 0.023 \end{pmatrix}$
$n = 900$	100%	100%	Mean	(-0.0056,5)	(0.996,1.99)	$\begin{pmatrix} 0.801 & 0.199 \\ 0.102 & 0.898 \end{pmatrix}$
			Std	(0.065,0.108)	(0.048,0.077)	$\begin{pmatrix} 0.026 & 0.026 \\ 0.016 & 0.016 \end{pmatrix}$
$n = 1600$	100%	100%	Mean	(-0.002,5.01)	(0.997,1.99)	$\begin{pmatrix} 0.8 & 0.2 \\ 0.102 & 0.898 \end{pmatrix}$
			Std	(0.049,0.081)	(0.035,0.056)	$\begin{pmatrix} 0.018 & 0.018 \\ 0.011 & 0.011 \end{pmatrix}$
$n = 2500$	100%	100%	Mean	(-0.0026,5)	(1.001,1.992)	$\begin{pmatrix} 0.802 & 0.198 \\ 0.101 & 0.899 \end{pmatrix}$
			Std	(0.039,0.064)	(0.029,0.042)	$\begin{pmatrix} 0.014 & 0.014 \\ 0.0086 & 0.0086 \end{pmatrix}$

Table 2.3: Estimation Result of model M2

N=2						
	cp	bp		$\boldsymbol{c} = (0, 2)$	$\boldsymbol{\sigma} = (1, 2)$	$\boldsymbol{P} = \begin{pmatrix} 0.85 & 0.15 \\ 0.1 & 0.9 \end{pmatrix}$
$n = 100$	93%	91%	Mean	(0.13,2.96)	(1.12,1.61)	$\begin{pmatrix} 0.81 & 0.19 \\ 0.34 & 0.66 \end{pmatrix}$
			Std	(0.44,1.18)	(0.31,0.47)	$\begin{pmatrix} 0.14 & 0.14 \\ 0.3 & 0.3 \end{pmatrix}$
$n = 400$	100%	95%	Mean	(-0.017,2.08)	(0.99,1.96)	$\begin{pmatrix} 0.84 & 0.16 \\ 0.12 & 0.88 \end{pmatrix}$
			Std	(0.13,0.3)	(0.098,0.14)	$\begin{pmatrix} 0.059 & 0.059 \\ 0.076 & 0.076 \end{pmatrix}$
$n = 900$	100%	99%	Mean	(-0.0036,2.05)	(1,1.99)	$\begin{pmatrix} 0.846 & 0.154 \\ 0.107 & 0.893 \end{pmatrix}$
			Std	(0.076,0.13)	(0.062,0.07)	$\begin{pmatrix} 0.033 & 0.033 \\ 0.026 & 0.026 \end{pmatrix}$
$n = 1600$	100%	100%	Mean	(-0.0038,2.03)	(1,1.99)	$\begin{pmatrix} 0.85 & 0.15 \\ 0.103 & 0.897 \end{pmatrix}$
			Std	(0.054,0.104)	(0.045,0.052)	$\begin{pmatrix} 0.022 & 0.022 \\ 0.018 & 0.018 \end{pmatrix}$
$n = 2500$	100%	99%	Mean	(0,2.02)	(1,2)	$\begin{pmatrix} 0.85 & 0.15 \\ 0.103 & 0.897 \end{pmatrix}$
			Std	(0.05,0.079)	(0.034,0.044)	$\begin{pmatrix} 0.016 & 0.016 \\ 0.0015 & 0.0015 \end{pmatrix}$

Table 2.4: Estimation Result of model M3

N=3						
	cp	bp		$\boldsymbol{c} = (0, 10, 20)$	$\boldsymbol{\sigma} = (1, 2, 3)$	$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.75 \end{pmatrix}$
$n = 100$	100%	99%	Mean	$(0.018, 9.99, 20)$	$(0.96, 1.94, 2.78)$	$\begin{pmatrix} 0.78 & 0.11 & 0.11 \\ 0.05 & 0.89 & 0.07 \\ 0.23 & 0.06 & 0.7 \end{pmatrix}$
			Std	$(0.19, 0.44, 0.81)$	(0.12,0.32,0.57)	$\begin{pmatrix} 0.08 & 0.07 & 0.07 \\ 0.04 & 0.06 & 0.04 \\ 0.12 & 0.07 & 0.14 \end{pmatrix}$
$n = 400$	100%	100%	Mean	$(0.014, 9.99, 19.99)$	$(0.99, 1.98, 2.95)$	$\begin{pmatrix} 0.79 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.21 & 0.05 & 0.74 \end{pmatrix}$
			Std	$(0.08, 0.16, 0.34)$	(0.06,0.12,0.22)	$\begin{pmatrix} 0.035 & 0.024 & 0.026 \\ 0.018 & 0.025 & 0.021 \\ 0.048 & 0.027 & 0.052 \end{pmatrix}$
$n = 900$	100%	100%	Mean	$(0.01, 10, 19.98)$	$(1, 1.98, 3)$	$\begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.75 \end{pmatrix}$
			Std	$(0.057, 0.1, 0.23)$	(0.046,0.08,0.16)	$\begin{pmatrix} 0.024 & 0.019 & 0.017 \\ 0.01 & 0.015 & 0.014 \\ 0.025 & 0.017 & 0.027 \end{pmatrix}$
$n = 1600$	100%	100%	Mean	$(0.01, 9.99, 19.99)$	$(1, 1.98, 2.99)$	$\begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.75 \end{pmatrix}$
			Std	$(0.045, 0.073, 0.18)$	(0.031,0.058,0.12)	$\begin{pmatrix} 0.019 & 0.015 & 0.016 \\ 0.008 & 0.012 & 0.009 \\ 0.021 & 0.012 & 0.022 \end{pmatrix}$
$n = 2500$	100%	98%	Mean	$(0, 9.99, 19.99)$	$(0.99, 1.99, 2.99)$	$\begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.75 \end{pmatrix}$
			Std	$(0.036, 0.058, 0.13)$	$(0.023, 0.046, 0.085)$	$\begin{pmatrix} 0.014 & 0.011 & 0.01 \\ 0.006 & 0.009 & 0.007 \\ 0.019 & 0.01 & 0.019 \end{pmatrix}$

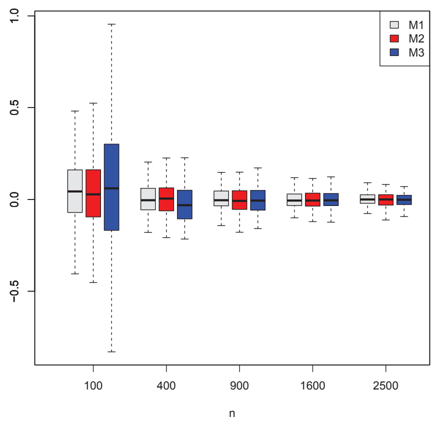
Table 2.5: Estimation Result of model M4

N=3						
	cp	bp		$\boldsymbol{c} = (0, 5, 10)$	$\boldsymbol{\sigma} = (1, 2, 3)$	$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.75 \end{pmatrix}$
$n = 100$	99%	98%	Mean	$(0.006, 4.87, 10.33)$	$(0.95, 1.93, 2.67)$	$\begin{pmatrix} 0.74 & 0.17 & 0.09 \\ 0.06 & 0.86 & 0.08 \\ 0.24 & 0.1 & 0.66 \end{pmatrix}$
			Std	$(0.28, 0.93, 1.73)$	(0.21,0.39,0.8)	$\begin{pmatrix} 0.15 & 0.16 & 0.07 \\ 0.08 & 0.1 & 0.064 \\ 0.16 & 0.17 & 0.22 \end{pmatrix}$
$n = 400$	100%	99%	Mean	$(0.018, 5.03, 9.98)$	$(0.98, 1.99, 2.94)$	$\begin{pmatrix} 0.79 & 0.11 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.21 & 0.05 & 0.74 \end{pmatrix}$
			Std	$(0.1, 0.18, 0.45)$	(0.068,0.15,0.27)	$\begin{pmatrix} 0.048 & 0.039 & 0.03 \\ 0.021 & 0.03 & 0.028 \\ 0.051 & 0.035 & 0.056 \end{pmatrix}$
$n = 900$	100%	99%	Mean	$(0.003, 5.01, 10.01)$	$(0.99, 2, 2.97)$	$\begin{pmatrix} 0.79 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.21 & 0.05 & 0.74 \end{pmatrix}$
			Std	$(0.065, 0.13, 0.3)$	(0.049,0.11,0.2)	$\begin{pmatrix} 0.026 & 0.023 & 0.021 \\ 0.015 & 0.018 & 0.015 \\ 0.027 & 0.022 & 0.035 \end{pmatrix}$
$n = 1600$	100%	100%	Mean	$(0, 4.99, 9.99)$	$(1, 2, 2.99)$	$\begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.74 \end{pmatrix}$
			Std	$(0.043, 0.094, 0.22)$	(0.038,0.072,0.14)	$\begin{pmatrix} 0.019 & 0.018 & 0.016 \\ 0.011 & 0.014 & 0.01 \\ 0.02 & 0.017 & 0.027 \end{pmatrix}$
$n = 2500$	100%	99%	Mean	$(0, 4.99, 9.99)$	$(0.99, 1.99, 2.99)$	$\begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.74 \end{pmatrix}$
			Std	$(0.034, 0.077, 0.16)$	$(0.028, 0.054, 0.12)$	$\begin{pmatrix} 0.016 & 0.013 & 0.013 \\ 0.009 & 0.011 & 0.009 \\ 0.017 & 0.016 & 0.021 \end{pmatrix}$

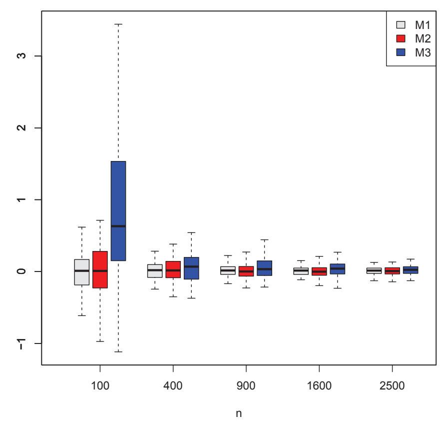
Table 2.6: Estimation Result of model M5

N=3						
	cp	bp		$\boldsymbol{c} = (0, 2, 4)$	$\boldsymbol{\sigma} = (1, 2, 3)$	$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.75 \end{pmatrix}$
$n = 100$	99%	98%	Mean	$(-0.31, 1.35, 4.89)$	$(1.42, 1.39, 1.88)$	$\begin{pmatrix} 0.5 & 0.31 & 0.18 \\ 0.22 & 0.52 & 0.25 \\ 0.25 & 0.32 & 0.43 \end{pmatrix}$
			Std	$(1.28, 1.05, 2.43)$	$(3.26, 0.77, 1.3)$	$\begin{pmatrix} 0.34 & 0.3 & 0.27 \\ 0.28 & 0.32 & 0.26 \\ 0.29 & 0.28 & 0.3 \end{pmatrix}$
$n = 400$	100%	92%	Mean	$(-0.05, 1.9, 4.48)$	$(1.06, 1.65, 2.68)$	$\begin{pmatrix} 0.74 & 0.17 & 0.09 \\ 0.13 & 0.68 & 0.19 \\ 0.18 & 0.22 & 0.61 \end{pmatrix}$
			Std	$(0.22, 0.67, 1.56)$	$(0.86, 0.48, 0.54)$	$\begin{pmatrix} 0.13 & 0.12 & 0.13 \\ 0.17 & 0.28 & 0.23 \\ 0.14 & 0.23 & 0.25 \end{pmatrix}$
$n = 900$	100%	88%	Mean	$(0.01, 1.94, 4.02)$	$(1.01, 1.88, 2.92)$	$\begin{pmatrix} 0.77 & 0.13 & 0.1 \\ 0.09 & 0.78 & 0.13 \\ 0.19 & 0.1 & 0.71 \end{pmatrix}$
			Std	$(0.11, 0.48, 0.6)$	(0.09,0.71,0.26)	$\begin{pmatrix} 0.11 & 0.12 & 0.07 \\ 0.16 & 0.23 & 0.16 \\ 0.07 & 0.08 & 0.11 \end{pmatrix}$
$n = 1600$	100%	88%	Mean	$(0, 1.97, 4.06)$	$(1, 1.95, 2.94)$	$\begin{pmatrix} 0.79 & 0.12 & 0.09 \\ 0.05 & 0.85 & 0.1 \\ 0.2 & 0.09 & 0.72 \end{pmatrix}$
			Std	$(0.07, 0.31, 0.5)$	(0.05,0.19,0.21)	$\begin{pmatrix} 0.06 & 0.07 & 0.04 \\ 0.06 & 0.14 & 0.11 \\ 0.06 & 0.07 & 0.08 \end{pmatrix}$
$n = 2500$	100%	79%	Mean	$(0.01, 1.98, 4.01)$	$(1, 1.97, 2.97)$	$\begin{pmatrix} 0.79 & 0.11 & 0.1 \\ 0.05 & 0.86 & 0.1 \\ 0.2 & 0.08 & 0.72 \end{pmatrix}$
			Std	$(0.05, 0.18, 0.5)$	$(0.04, 0.11, 0.16)$	$\begin{pmatrix} 0.03 & 0.05 & 0.04 \\ 0.04 & 0.14 & 0.13 \\ 0.05 & 0.08 & 0.08 \end{pmatrix}$

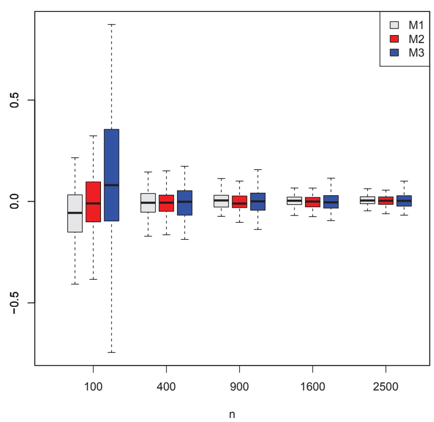
Table 2.7: Estimation Result of model M6



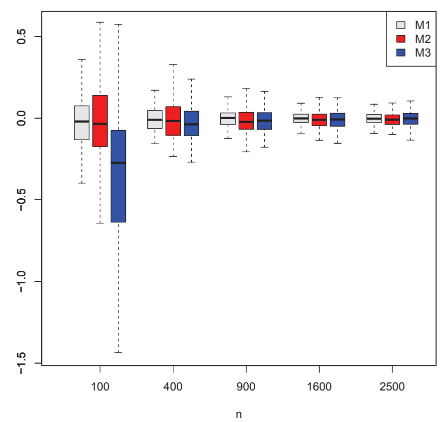
(a) c_1



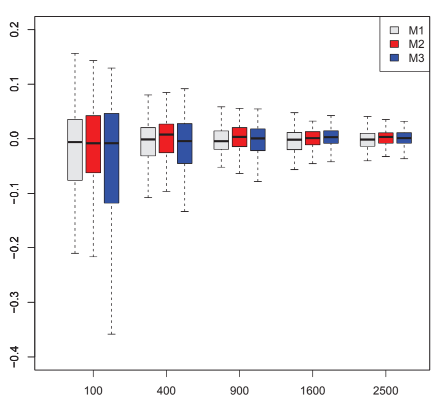
(b) c_2



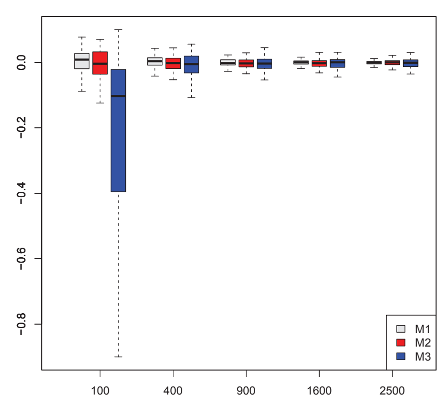
(c) σ_1



(d) σ_2

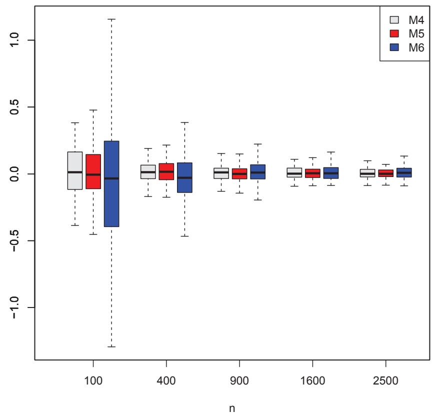


(e) p_{11}

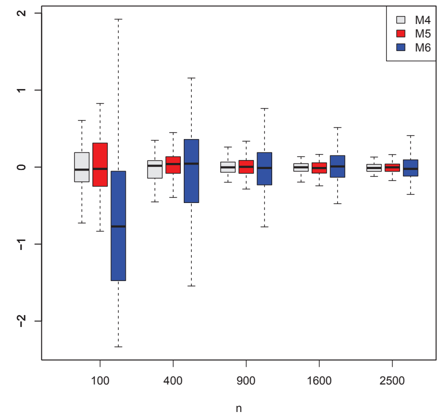


(f) p_{22}

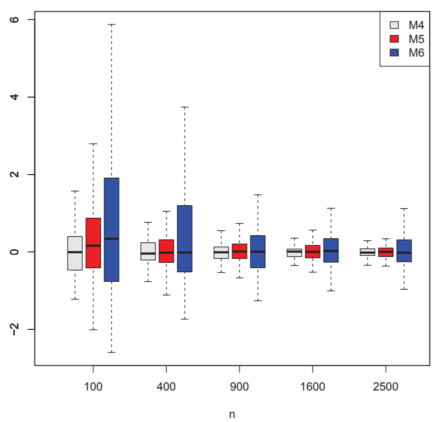
Figure 2.4: Boxplots of estimation errors in M1, M2 and M3



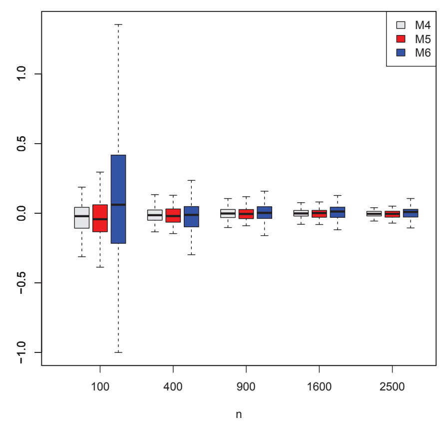
(a) c_1



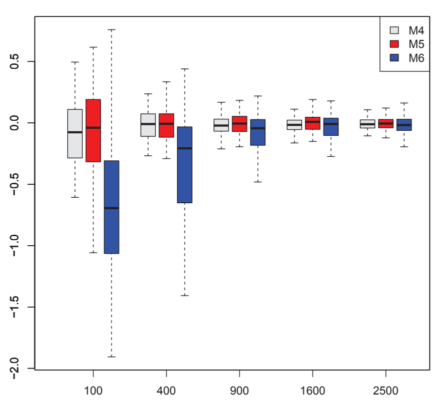
(b) c_2



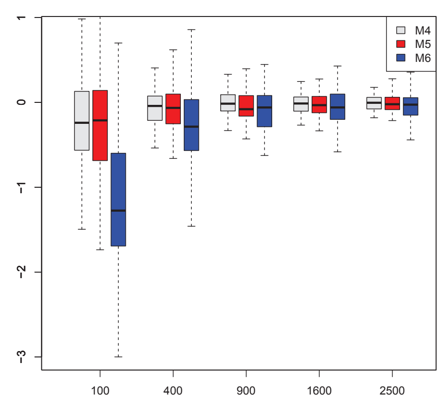
(c) c_3



(d) σ_1

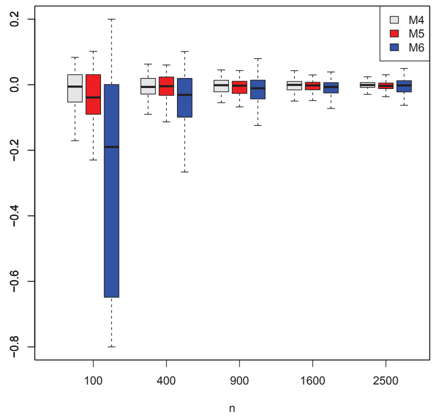


(e) σ_2

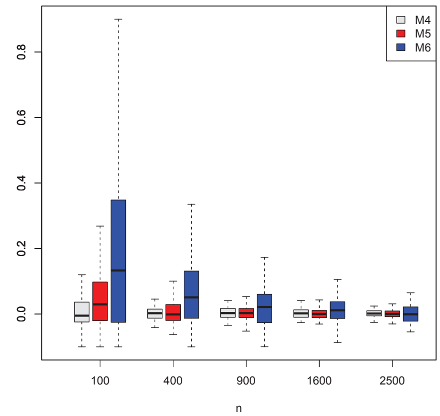


(f) σ_3

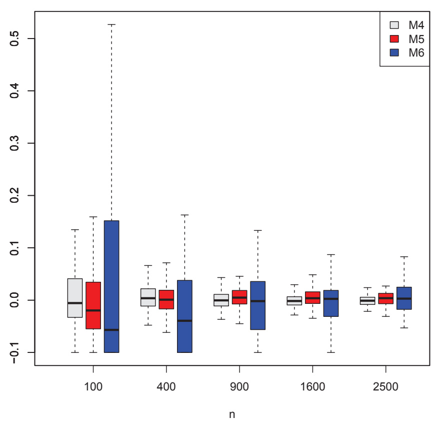
Figure 2.5: Boxplots of estimation errors in M4 - M6



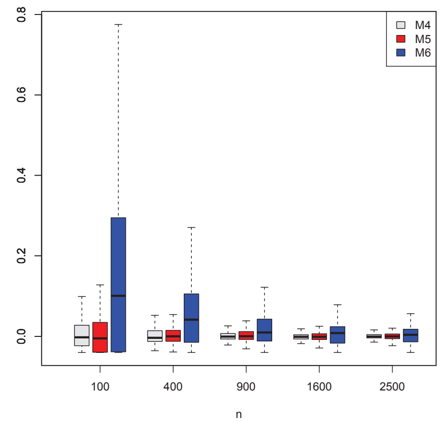
(g) p_{11}



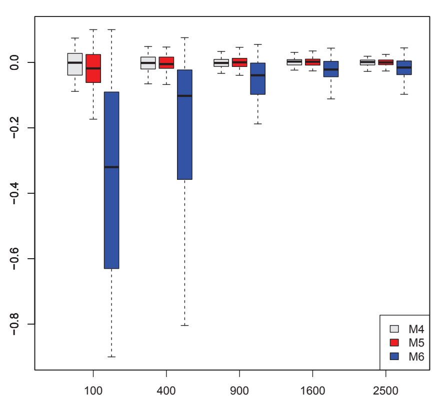
(h) p_{12}



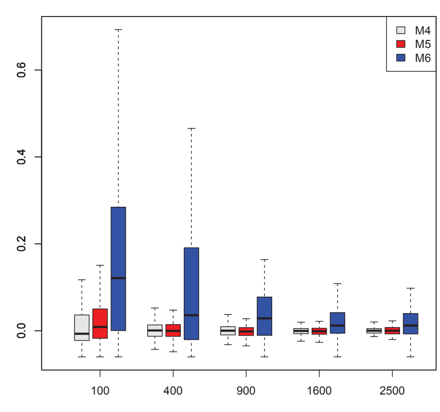
(i) p_{13}



(j) p_{21}

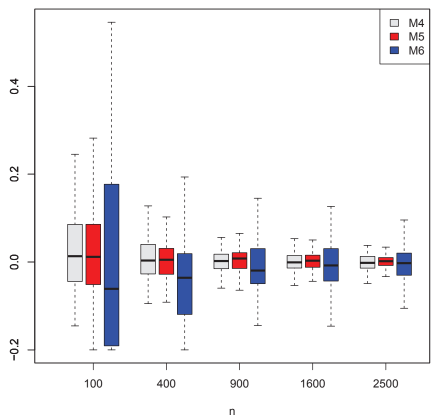


(k) p_{22}

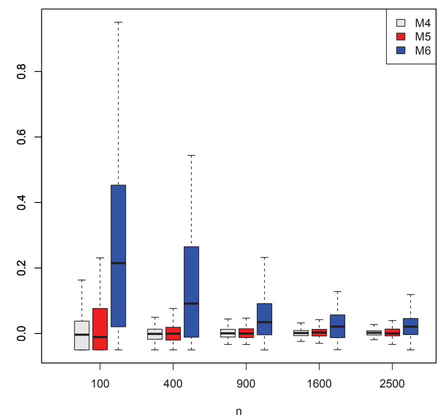


(l) p_{23}

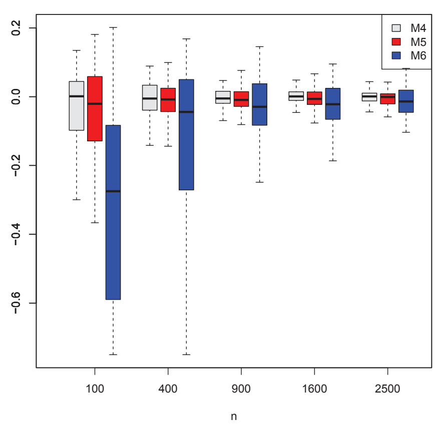
Figure 2.5: Boxplots of estimation errors in M4 - M6 (Continued)



(m) p_{31}



(n) p_{32}



(o) p_{33}

Figure 2.5: Boxplots of estimation errors in M4 - M6 (Continued)

2.2 A Regime-Switching Model with Known Constant

2.2.1 Model Specification and Experimental Settings

Next, we consider another simplified version of (1.2) where the constants are all known i.e,

$$y_t = \phi_{s_t} y_{t-1} + c_{s_t} + \epsilon_t \tag{2.2}$$

where $\epsilon_t \sim N\left(0, \sigma_{s_t}^2\right)$ and $c_1, \ldots c_N$ are given. We refer to this model as the *fixed intercepts switching model* (FISM).

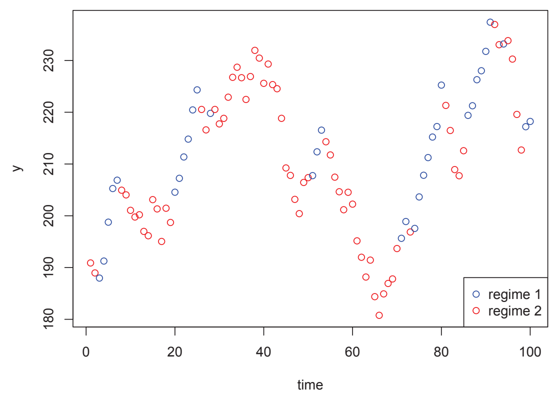
If we rewrite y_{t-1} as x_t in (2.2) and drop the Markov switching property, the resulting model can be considered as a mixture of N regression lines. Hereafter, we refer to this simplified model as the *mixing regression model* (MRM). The regime s_t at time t in the MRM follows a multinomial distribution with

$$\boldsymbol{P} = (p_1, \ldots, p_N)$$

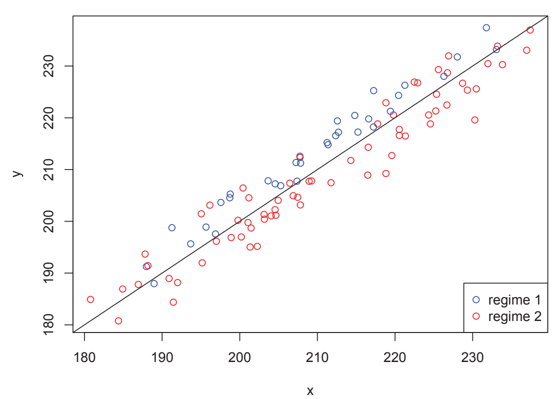
As a demonstration, a sequence of data sampled from (2.2) is plotted in Figure 2.6a. The two-dimensional data points (y_t, y_{t-1}) are shown in Figure 2.6b. Data points from regime 1 are highlighted in blue and values from regime 2 are colored in red. A point (y_t, y_{t-1}) is considered from regime i if y_t is in regime i. In the last section, we have shown that mixture models can be used to approximate switching mixture models and to generate initial search values. In order to estimate parameters in (2.2), we can take a similar approach by fitting a MRM and use its output as initial search values. One problem with this approach is that the MRM is hard to estimate by itself. Instead of writing down the likelihood function for the MRM, we can further reduce it to a clustering problem. The objective is to classify points (y_t, y_{t-1}) into N clusters (lines) by minimizing their sum of least square terms. More specifically, we want to assign points (y_t, y_{t-1}) to N lines so that

$$\sum_{t=1}^{n} \left(y_t - \phi_{s_t} y_{t-1} - c_{s_t}\right)^2 \tag{2.3}$$

is minimized. Note that this is similar to the traditional clustering problem. The difference is that here we use the least square penalty rather than Euclidian distances in the optimization. To solve this optimization problem, we propose a deterministic iterative algorithm



(a) A sequence simulated from a regime switching model with two regimes



(b) Mixture of two lines

Figure 2.6: In (a), a simulated sequence of a regime switching model with two regimes is plotted. The generating parameters are: $\phi = (0.97, 0.92)$, $c = (10, 15)$, $\sigma = (2, 4)$, $p_{11} = 0.75$ and $p_{22} = 0.9$. In (b), the sequence is converted to a mixture of two regression lines by letting $x_t = y_{t-1}$

that is similar to the k-mean algorithm. We call this method the *k-line* algorithm.

Suppose at step $k+1$, a set of $\phi_1^k, \ldots, \phi_N^k$ are estimated from the last iteration k. For each point (y_t, y_{t-1}), we assign it to one of the N lines such that the least square term is minimized. After all points are classified, a new coefficient ϕ_i^{k+1} is updated by minimizing (2.3) only over those points that have been assigned to line i. This algorithm continues to iterate until the difference between the sums of residual squares from two consecutive steps is small. A graphical example demonstrating one iteration of the k-line algorithm is shown in Figure 2.7.

Similar to the k-mean problem, the k-line clustering is computationally difficult and the k-line algorithm will not guarantee to converge to a global minimum. It is well known that k-mean clustering is NP hard and so is k-line clustering. However our goal is not to find a single global minimum for this problem. Instead, we are interested in a pool of estimates of ϕ_is with which (2.3) are small but not necessary minimal. Our hope is that these estimates will correspond to different regions of the original search space of (2.2), so that the probability that the optimizer stops at a local minimum is small. In order to generate initial values for the k-clustering algorithm, we propose the following two methods:

1. **RSTART1:** As we have shown above, the regime switching model can be considered as a mixture of regression lines. If the regime s_t at time t is given for all t, ϕ_i and c_i can be estimated as easily as fitting a regression line to the data points belonging to regime i only. Although s_t is generally not known, we can still get a rough estimate of ϕ_i for each regime i by fitting a regression line (with given c_i) to all data. The estimates $\hat{\phi}_1, \ldots, \hat{\phi}_N$ will not converge to the true values as n increases. However, they should provide a reasonable approximate to the true values, especially when the differences among ϕ_is are minor. More specifically, in RSTART1, $\hat{\phi}_i$ is estimated by minimizing the sum of least squares

$$\sum_{t=1}^{n} \left(y_t - \hat{\phi}_i y_{t-1} - c_i \right)^2,$$

where c_i is known. $\hat{\phi}_i$ has a closed form solution, which is equal to

$$\frac{\sum_{t=1}^{n} (y_t - c_i) y_{t-1}}{\sum_{t=1}^{n} y_{t-1}^2}.$$

Estimates $\hat{\phi}_1, \ldots, \hat{\phi}_N$ will be used as initial values in the k-clustering algorithm.

2. **RSTART2:** The basic idea of RSTART2 is to introduce some randomness in generating initial values so that different regions of the search space can be explored. In RSTART2, first solve a linear regression for all data points. Suppose the regression line is $y = \hat{\phi}x + \hat{c}$ and the standard deviation of the error term is $\hat{\sigma}$. Next, calculate the mean of x or y_{t-1} i.e $\bar{x} = \sum_{i=1}^{n-1} y_i/(n-1)$. Let $\bar{y} = \hat{\phi}\bar{x} + \hat{c}$. We then sample N values $y_i, i = 1, \ldots, N$ from $N(\bar{y}, \hat{\sigma})$. The slope of the line by connecting the two points $(0, c_i)$ and $(\bar{x}, y_i)$ is used as initial estimate of ϕ_i i.e $\hat{\phi}_i$ is calculated as $(y_i - c_i)/\bar{x}$.

The first method above is a deterministic algorithm. The second method is a probabilistic algorithm that will give different initial values each time. In the estimation of the mixture regression model (2.2), we restarted the optimizer 5 times for each sequence. Initial values were generated using the first method once and the second method was used in the other four runs. We refer to this initial-value-sampling scheme as *R.IVS1*.

Data was simulated from five different models and the generating parameters are listed in Table 2.8. Model R1-R3 have two regimes and the other two models (M4 and M5) have three regimes. As in Section 2.1, 100 sequences with 2500 data points are sampled from those models. Parameters were estimated on the first 100, 400, 900, 1600 and 2500 points and the number of regimes was assumed to be known.

2.2.2 Experimental Results

The simulation results are given in Table 2.9 - Table 2.13. cp remains high in all models with differing numbers of data points. bp versus n is plotted in Figure 2.10, in which different patterns can be observed for R1 and R2 - R5. The bp value in the estimation of R1 drops first, reaches its minimum at $n = 900$ and bounces back after that. The difference between the minimal and maximal bp values is small for R1. On the other hand, the bp values for R2 - R5 show steady decline with increasing n. Among these 4 models, the optimizer performed best in R5 and worst in R3. However, the bp value is above 80$\%$ even in the worst scenario, indicating the proposed strategy to select initial values is successful.

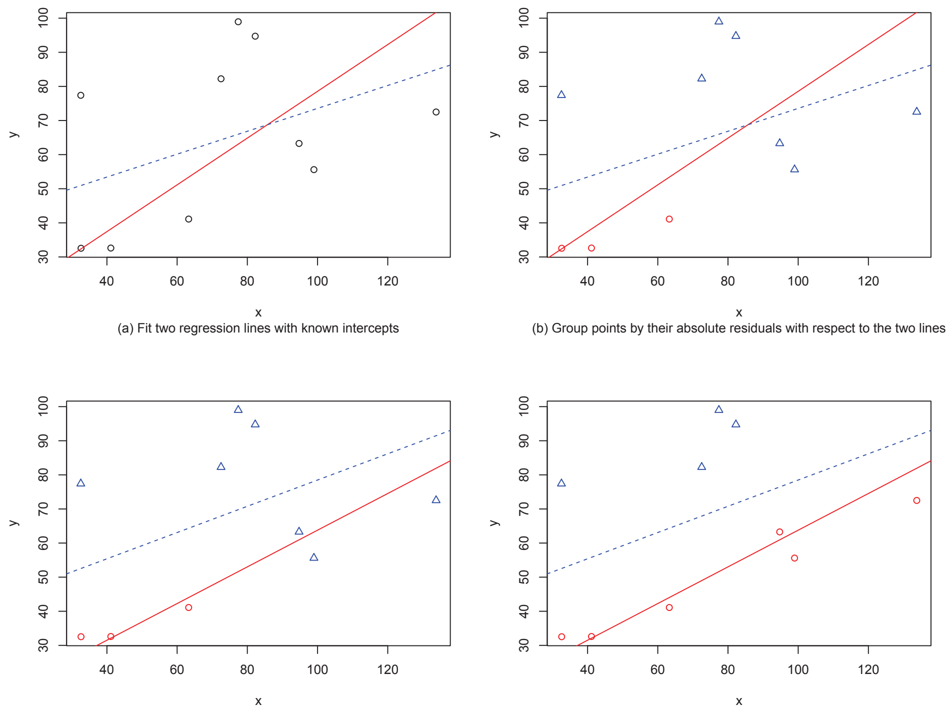
The boxplots of estimation errors are plotted in Figure 2.11 and 2.12 for two dimensional and three dimensional models respectively. The multiplicative coefficients ϕ_is can be estimated very well in all cases. The absolute error in the estimation of ϕ_i for all models

N=2				
	$\boldsymbol{c}$	$\boldsymbol{\phi}$	$\boldsymbol{\sigma}$	$\boldsymbol{P}$
R1	$\boldsymbol{c} = (5, 30)$	$\boldsymbol{\phi} = (0.98, 0.95)$	$\boldsymbol{\sigma} = (1, 2)$	$\boldsymbol{P} = \begin{pmatrix} 0.75 & 0.25 \\ 0.1 & 0.9 \end{pmatrix}$
R2	$\boldsymbol{c} = (10, 20)$	$\boldsymbol{\phi} = (0.98, 0.96)$	$\boldsymbol{\sigma} = (2, 3)$	$\boldsymbol{P} = \begin{pmatrix} 0.75 & 0.25 \\ 0.1 & 0.9 \end{pmatrix}$
R3	$\boldsymbol{c} = (10, 15)$	$\boldsymbol{\phi} = (0.98, 0.97)$	$\boldsymbol{\sigma} = (3, 6)$	$\boldsymbol{P} = \begin{pmatrix} 0.75 & 0.25 \\ 0.1 & 0.9 \end{pmatrix}$
N=3				
R4	$\boldsymbol{c} = (5, 15, 30)$	$\boldsymbol{\phi} = (0.98, 0.95, 0.92)$	$\boldsymbol{\sigma} = (2, 2, 3)$	$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.75 \end{pmatrix}$
R5	$\boldsymbol{c} = (10, 15, 20)$	$\boldsymbol{\phi} = (0.98, 0.96, 0.94)$	$\boldsymbol{\sigma} = (3, 3, 5)$	$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.75 \end{pmatrix}$

Table 2.8: Parameters in the simulation of model (2.2)

quickly vanishes as n increases. Estimates of entries of the transition matrix $\boldsymbol{P}$ are very slow to converge except in R1. Even with 2500 points, the error mean and variance in the estimation of individual transition probability p_{ij} can be large. For example, the mean errors are larger than 0.1 for most estimates $\hat{p}_{ii}$ of the diagonal entries of transition matrix $\boldsymbol{P}$ for model R2 - R5.

The standard deviations of the error terms $\boldsymbol{\sigma}$ were estimated well in R1 and R4, when the true $\boldsymbol{\sigma}$ is small. Estimates of $\boldsymbol{\sigma}$ were reasonable in R2 and R5. In contrast, $\hat{\sigma}$ is far away from the true value in R3, where the variances are the largest amongst all models. Recall that R3 also has the worst *bp* values. This implies that the magnitude of error terms has a great impact on the performance of the optimizer. A large variance of error could make a low dimensional model (i.e with small number of regimes) harder to fit than a higher dimensional model.



(a) Fit two regression lines with known intercepts

(b) Group points by their absolute residuals with respect to the two lines

(c) Update slopes of regression lines based on points assigned to them

(d) Repeat step (b) and (c) until convergence

Figure 2.7: An illustration of how to choose initial search values.

N=2, $\boldsymbol{c} = (5, 30)$						
	cp	bp		$\boldsymbol{\phi} = (0.98, 0.95)$	$\boldsymbol{\sigma} = (1, 2)$	$\boldsymbol{P} = \begin{pmatrix} 0.75 & 0.25 \\ 0.1 & 0.9 \end{pmatrix}$
$n = 100$	100%	96%	Mean	$(0.981, 0.949)$	$(1.01, 1.98)$	$\begin{pmatrix} 0.751 & 0.249 \\ 0.116 & 0.884 \end{pmatrix}$
			Std	$(0.003, 0.0025)$	$(0.31, 0.37)$	$\begin{pmatrix} 0.095 & 0.095 \\ 0.091 & 0.091 \end{pmatrix}$
$n = 400$	100%	95%	Mean	$(0.981, 0.949)$	$(1.09, 2.04)$	$\begin{pmatrix} 0.751 & 0.249 \\ 0.11 & 0.89 \end{pmatrix}$
			Std	$(0.003, 0.003)$	$(0.43, 0.43)$	$\begin{pmatrix} 0.052 & 0.052 \\ 0.029 & 0.029 \end{pmatrix}$
$n = 900$	100%	92%	Mean	$(0.981, 0.949)$	$(1.12, 1.99)$	$\begin{pmatrix} 0.76 & 0.24 \\ 0.11 & 0.89 \end{pmatrix}$
			Std	$(0.004, 0.004)$	$(0.43, 0.08)$	$\begin{pmatrix} 0.049 & 0.049 \\ 0.037 & 0.037 \end{pmatrix}$
$n = 1600$	100%	99%	Mean	$(0.98, 0.95)$	$(1.02, 1.99)$	$\begin{pmatrix} 0.75 & 0.25 \\ 0.1 & 0.9 \end{pmatrix}$
			Std	$(0.001, 0.002)$	$(0.19, 0.06)$	$\begin{pmatrix} 0.028 & 0.028 \\ 0.019 & 0.019 \end{pmatrix}$
$n = 2500$	100%	97%	Mean	$(0.98, 0.95)$	$(1.05, 2)$	$\begin{pmatrix} 0.76 & 0.24 \\ 0.1 & 0.9 \end{pmatrix}$
			Std	$(0.003, 0.003)$	$(0.28, 0.05)$	$\begin{pmatrix} 0.031 & 0.031 \\ 0.026 & 0.026 \end{pmatrix}$

Table 2.9: Estimation Result of model R1

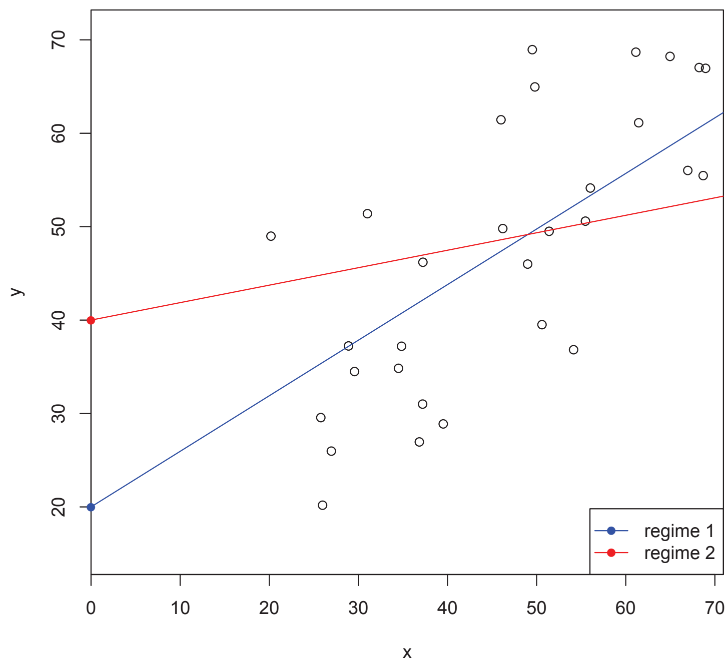
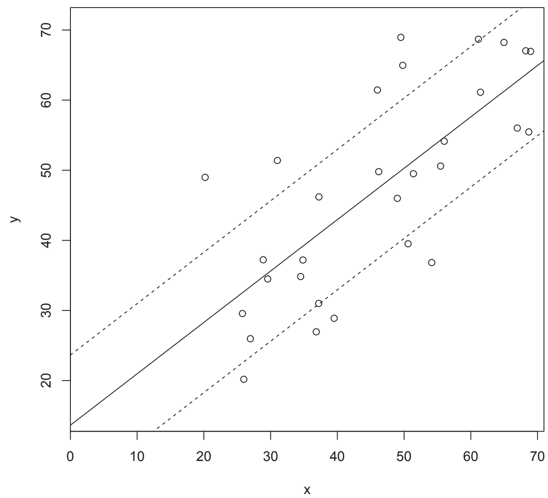


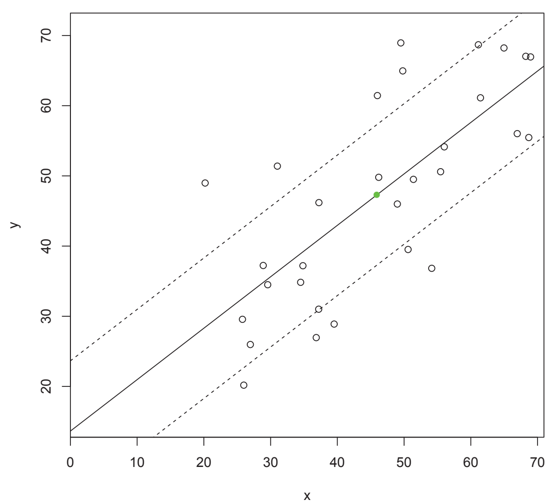
Figure 2.8: In RSTART1, two regression lines (with given constant terms) are fitted to the whole data set. The resulting slope terms ϕ_i are used as initial values in the k-clustering algorithm.

N=2, $\boldsymbol{c} = (10, 20)$						
	cp	bp		$\boldsymbol{\phi} = (0.98, 0.96)$	$\boldsymbol{\sigma} = (2, 3)$	$\boldsymbol{P} = \begin{pmatrix} 0.75 & 0.25 \\ 0.1 & 0.9 \end{pmatrix}$
$n = 100$	99%	100%	Mean	$(0.98, 0.96)$	$(7.3, 2.19)$	$\begin{pmatrix} 0.36 & 0.64 \\ 0.39 & 0.61 \end{pmatrix}$
			Std	$(0.007, 0.004)$	$(54.23, 0.86)$	$\begin{pmatrix} 0.36 & 0.36 \\ 0.35 & 0.35 \end{pmatrix}$
$n = 400$	100%	99%	Mean	$(0.98, 0.96)$	$(2.29, 2.47)$	$\begin{pmatrix} 0.43 & 0.57 \\ 0.35 & 0.65 \end{pmatrix}$
			Std	$(0.005, 0.002)$	$(0.93, 0.62)$	$\begin{pmatrix} 0.34 & 0.34 \\ 0.35 & 0.35 \end{pmatrix}$
$n = 900$	100%	95%	Mean	$(0.98, 0.96)$	$(2.46, 2.54)$	$\begin{pmatrix} 0.54 & 0.46 \\ 0.29 & 0.71 \end{pmatrix}$
			Std	$(0.003, 0.001)$	$(0.85, 0.56)$	$\begin{pmatrix} 0.33 & 0.33 \\ 0.31 & 0.31 \end{pmatrix}$
$n = 1600$	100%	93%	Mean	$(0.98, 0.96)$	$(2.49, 2.67)$	$\begin{pmatrix} 0.53 & 0.47 \\ 0.25 & 0.75 \end{pmatrix}$
			Std	$(0.003, 0.0001)$	$(0.86, 0.48)$	$\begin{pmatrix} 0.33 & 0.33 \\ 0.28 & 0.28 \end{pmatrix}$
$n = 2500$	100%	85%	Mean	$(0.98, 0.96)$	$(2.52, 2.7)$	$\begin{pmatrix} 0.6 & 0.4 \\ 0.23 & 0.77 \end{pmatrix}$
			Std	$(0.002, 0.0003)$	$(0.75, 0.42)$	$\begin{pmatrix} 0.27 & 0.27 \\ 0.22 & 0.22 \end{pmatrix}$

Table 2.10: Estimation Result of model R2

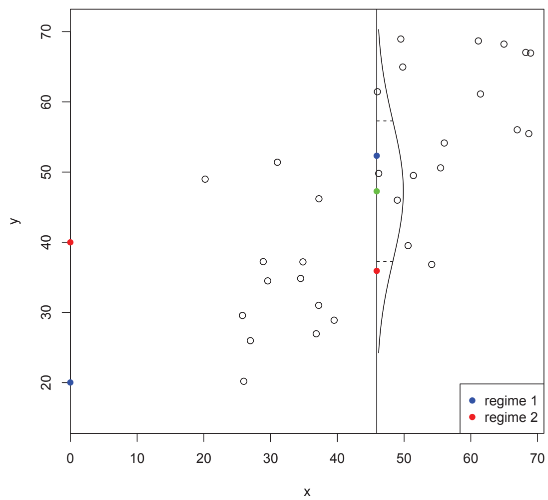


(a) First, a regression line is fitted to all data. Suppose the estimated standard deviation of residual is $\hat{\sigma}$.

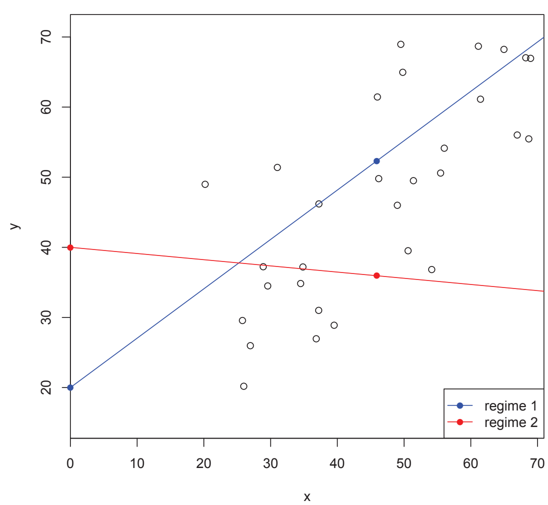


(b) The center of data points (the green dot) is calculated with coordinates $(\bar{x}, \bar{y})$.

Figure 2.9: An illustration of RSTART2 on a two-regime model.



(c) Generate two points from the vertical line passing through $(\bar{x}, \bar{y})$. The y-coordinates are sampled from $N(\bar{y}, \hat{\sigma})$.



(d) Connect the two sampled points with the intercepts $(0, c_i)$ for $i = 1, 2$. The slopes of the two lines are used as initial values in the k-line algorithm.

Figure 2.9: An illustration of RSTART2 on a two-regime model.

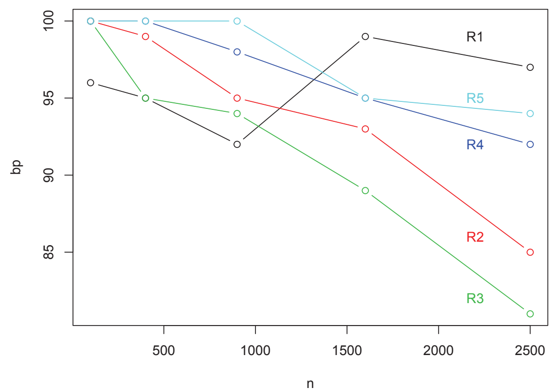


Figure 2.10: Percentage of better loglikelihood in estimations of R1 - R5

N=2, $\boldsymbol{c} = (10, 15)$						
	cp	bp		$\boldsymbol{\phi} = (0.98, 0.97)$	$\boldsymbol{\sigma} = (3, 6)$	$\boldsymbol{P} = \begin{pmatrix} 0.75 & 0.25 \\ 0.1 & 0.9 \end{pmatrix}$
$n = 100$	95%	100%	Mean	$(0.98, 0.97)$	$(5.03, 3.65)$	$\begin{pmatrix} 0.56 & 0.44 \\ 0.42 & 0.58 \end{pmatrix}$
			Std	$(0.008, 0.006)$	$(2.22, 1.78)$	$\begin{pmatrix} 0.35 & 0.35 \\ 0.36 & 0.36 \end{pmatrix}$
$n = 400$	100%	95%	Mean	$(0.98, 0.97)$	$(5.43, 4)$	$\begin{pmatrix} 0.72 & 0.28 \\ 0.29 & 0.71 \end{pmatrix}$
			Std	$(0.003, 0.003)$	$(1.78, 1.49)$	$\begin{pmatrix} 0.26 & 0.26 \\ 0.28 & 0.28 \end{pmatrix}$
$n = 900$	100%	94%	Mean	$(0.98, 0.97)$	$(5.54, 3.6)$	$\begin{pmatrix} 0.8 & 0.2 \\ 0.28 & 0.72 \end{pmatrix}$
			Std	$(0.002, 0.003)$	$(1.3, 1.38)$	$\begin{pmatrix} 0.19 & 0.19 \\ 0.22 & 0.22 \end{pmatrix}$
$n = 1600$	99%	89%	Mean	$(0.98, 0.97)$	$(5.38, 3.74)$	$\begin{pmatrix} 0.81 & 0.19 \\ 0.22 & 0.78 \end{pmatrix}$
			Std	$(0.003, 0.0005)$	$(1.46, 1.27)$	$\begin{pmatrix} 0.18 & 0.18 \\ 0.11 & 0.11 \end{pmatrix}$
$n = 2500$	100%	81%	Mean	$(0.98, 0.97)$	$(5.24, 3.91)$	$\begin{pmatrix} 0.82 & 0.18 \\ 0.22 & 0.78 \end{pmatrix}$
			Std	$(0.002, 0.0004)$	$(1.45, 1.35)$	$\begin{pmatrix} 0.12 & 0.12 \\ 0.11 & 0.11 \end{pmatrix}$

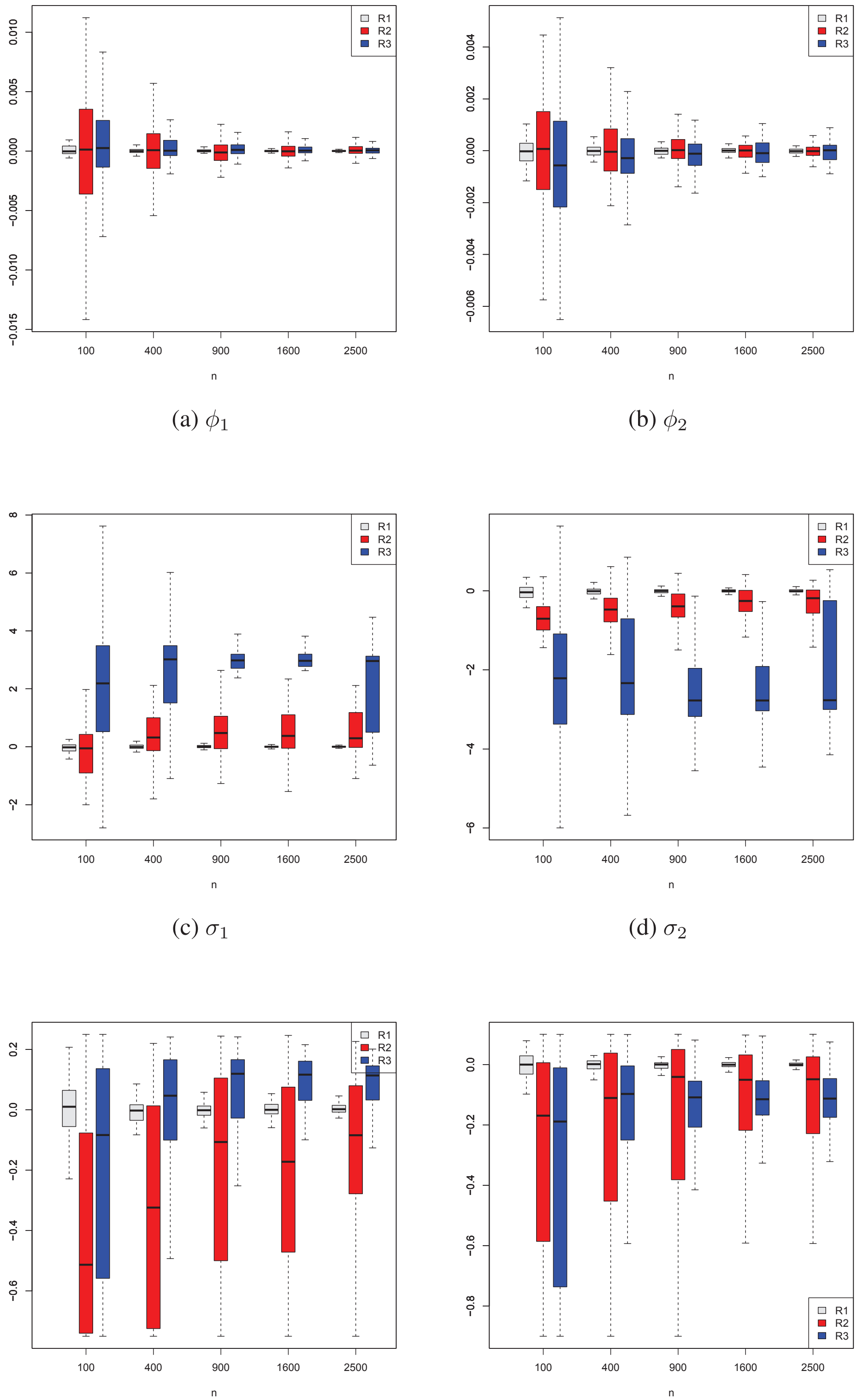
Table 2.11: Estimation Result of model R3

N=3,$\boldsymbol{c} = (5, 15, 30)$						
	cp	bp		$\boldsymbol{\phi} = (0.98, 0.95, 0.92)$	$\boldsymbol{\sigma} = (2, 2, 3)$	$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.75 \end{pmatrix}$
$n = 100$	99%	100%	Mean	$(0.984, 0.957, 0.913)$	$(1.84, 1.78, 1.79)$	$\begin{pmatrix} 0.58 & 0.21 & 0.21 \\ 0.3 & 0.55 & 0.15 \\ 0.25 & 0.26 & 0.49 \end{pmatrix}$
			Std	$(0.008, 0.011, 0.013)$	$(0.66, 0.86, 0.95)$	$\begin{pmatrix} 0.35 & 0.27 & 0.27 \\ 0.3 & 0.33 & 0.24 \\ 0.29 & 0.29 & 0.32 \end{pmatrix}$
$n = 400$	100%	100%	Mean	$(0.982, 0.953, 0.917)$	$(1.92, 1.96, 2.56)$	$\begin{pmatrix} 0.51 & 0.36 & 0.13 \\ 0.3 & 0.62 & 0.08 \\ 0.22 & 0.14 & 0.64 \end{pmatrix}$
			Std	$(0.006, 0.008, 0.008)$	$(0.58, 0.59, 0.74)$	$\begin{pmatrix} 0.3 & 0.29 & 0.18 \\ 0.27 & 0.28 & 0.13 \\ 0.23 & 0.14 & 0.23 \end{pmatrix}$
$n = 900$	100%	98%	Mean	$(0.981, 0.953, 0.918)$	$(2.04, 2.07, 2.73)$	$\begin{pmatrix} 0.58 & 0.3 & 0.12 \\ 0.28 & 0.64 & 0.08 \\ 0.19 & 0.13 & 0.68 \end{pmatrix}$
			Std	$(0.005, 0.007, 0.006)$	$(0.57, 0.48, 0.51)$	$\begin{pmatrix} 0.27 & 0.26 & 0.17 \\ 0.24 & 0.26 & 0.14 \\ 0.18 & 0.11 & 0.19 \end{pmatrix}$
$n = 1600$	100%	95%	Mean	$(0.981, 0.95, 0.92)$	$(2.01, 1.95, 2.9)$	$\begin{pmatrix} 0.55 & 0.36 & 0.09 \\ 0.28 & 0.64 & 0.08 \\ 0.14 & 0.14 & 0.73 \end{pmatrix}$
			Std	$(0.003, 0.004, 0.003)$	$(0.35, 0.28, 0.31)$	$\begin{pmatrix} 0.26 & 0.26 & 0.14 \\ 0.22 & 0.23 & 0.08 \\ 0.11 & 0.11 & 0.11 \end{pmatrix}$
$n = 2500$	100%	92%	Mean	$(0.98, 0.95, 0.92)$	$(1.98, 1.99, 2.98)$	$\begin{pmatrix} 0.58 & 0.34 & 0.08 \\ 0.25 & 0.67 & 0.08 \\ 0.13 & 0.13 & 0.74 \end{pmatrix}$
			Std	$(0.002, 0.002, 0.002)$	$(0.21, 0.22, 0.24)$	$\begin{pmatrix} 0.24 & 0.24 & 0.06 \\ 0.23 & 0.23 & 0.1 \\ 0.08 & 0.09 & 0.08 \end{pmatrix}$

Table 2.12: Estimation Result of model R4

N=3, $\boldsymbol{c} = (10, 15, 20)$						
	cp	bp		$\boldsymbol{\phi} = (0.98, 0.96, 0.94)$	$\boldsymbol{\sigma} = (3, 3, 5)$	$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.75 \end{pmatrix}$
$n = 100$	100%	100%	Mean	$(0.974, 0.96, 0.946)$	$(2.55, 2.39, 2.73)$	$\begin{pmatrix} 0.45 & 0.25 & 0.3 \\ 0.27 & 0.48 & 0.25 \\ 0.24 & 0.23 & 0.53 \end{pmatrix}$
			Std	$(0.011, 0.011, 0.011)$	$(1.64, 1.32, 1.23)$	$\begin{pmatrix} 0.34 & 0.27 & 0.34 \\ 0.31 & 0.32 & 0.29 \\ 0.32 & 0.28 & 0.36 \end{pmatrix}$
$n = 400$	100%	100%	Mean	$(0.97, 0.96, 0.945)$	$(3.25, 2.83, 3.56)$	$\begin{pmatrix} 0.64 & 0.17 & 0.19 \\ 0.16 & 0.66 & 0.18 \\ 0.13 & 0.16 & 0.71 \end{pmatrix}$
			Std	$(0.011, 0.01, 0.009)$	$(1.14, 0.87, 1.24)$	$\begin{pmatrix} 0.28 & 0.22 & 0.23 \\ 0.2 & 0.29 & 0.25 \\ 0.16 & 0.2 & 0.23 \end{pmatrix}$
$n = 900$	100%	100%	Mean	$(0.973, 0.96, 0.946)$	$(3.51, 3.4, 3.71)$	$\begin{pmatrix} 0.77 & 0.14 & 0.09 \\ 0.13 & 0.76 & 0.11 \\ 0.11 & 0.11 & 0.78 \end{pmatrix}$
			Std	$(0.007, 0.008, 0.007)$	$(0.98, 0.92, 1.16)$	$\begin{pmatrix} 0.15 & 0.15 & 0.08 \\ 0.11 & 0.16 & 0.13 \\ 0.09 & 0.13 & 0.13 \end{pmatrix}$
$n = 1600$	100%	95%	Mean	$(0.975, 0.96, 0.944)$	$(3.47, 3.43, 3.96)$	$\begin{pmatrix} 0.77 & 0.13 & 0.1 \\ 0.12 & 0.81 & 0.07 \\ 0.11 & 0.09 & 0.8 \end{pmatrix}$
			Std	$(0.007, 0.006, 0.005)$	$(0.95, 0.9, 1.07)$	$\begin{pmatrix} 0.1 & 0.09 & 0.07 \\ 0.1 & 0.1 & 0.05 \\ 0.09 & 0.08 & 0.1 \end{pmatrix}$
$n = 2500$	100%	94%	Mean	$(0.976, 0.96, 0.944)$	$(3.44, 3.43, 4.08)$	$\begin{pmatrix} 0.8 & 0.11 & 0.09 \\ 0.1 & 0.83 & 0.07 \\ 0.11 & 0.09 & 0.8 \end{pmatrix}$
			Std	$(0.006, 0.005, 0.005)$	$(0.85, 0.86, 1.05)$	$\begin{pmatrix} 0.08 & 0.06 & 0.05 \\ 0.08 & 0.09 & 0.06 \\ 0.08 & 0.06 & 0.09 \end{pmatrix}$

Table 2.13: Estimation Result of model R5

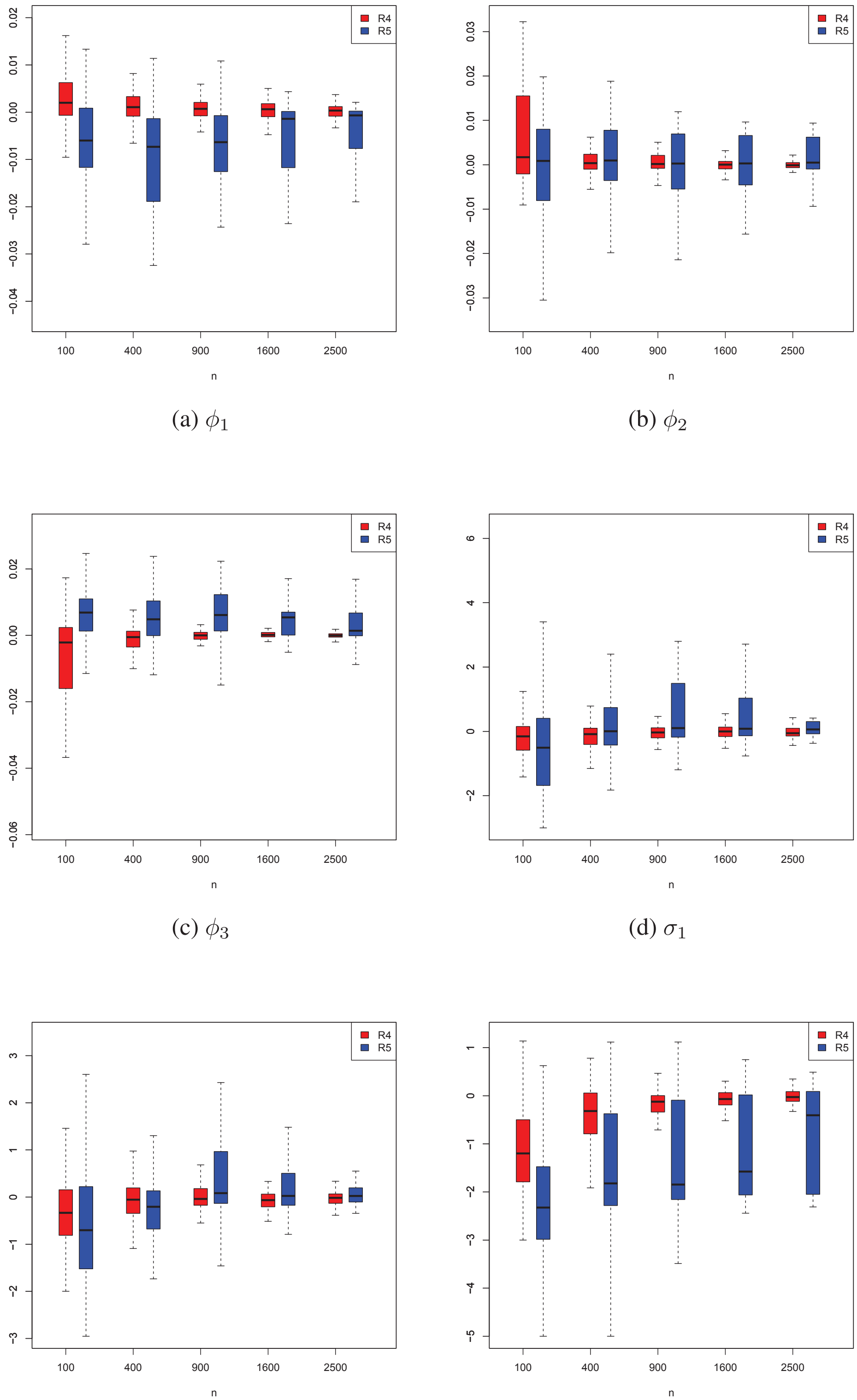


(a) ϕ_1 (b) ϕ_2

(c) σ_1 (d) σ_2

(e) p_{11} (f) p_{22}

Figure 2.11: Boxplots of estimation errors in R1 - R3

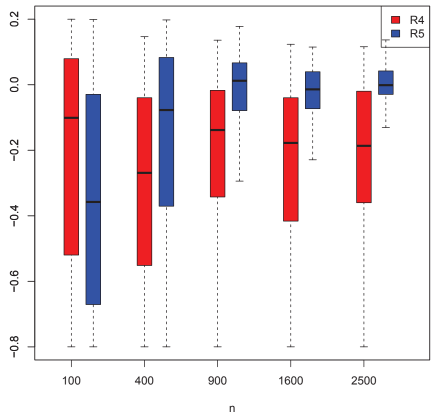


(a) ϕ_1 (b) ϕ_2

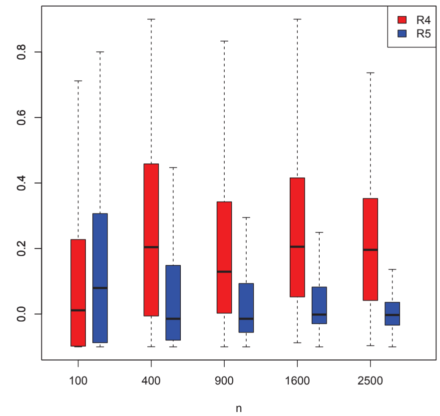
(c) ϕ_3 (d) σ_1

(e) σ_2 (f) σ_3

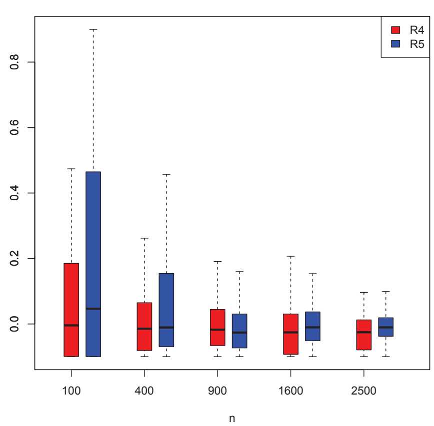
Figure 2.12: Boxplots of estimation errors in R4 and R5



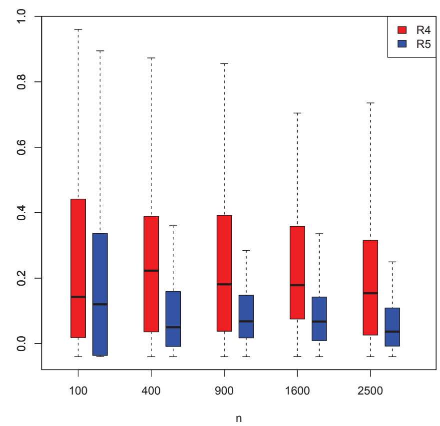
(g) p_{11}



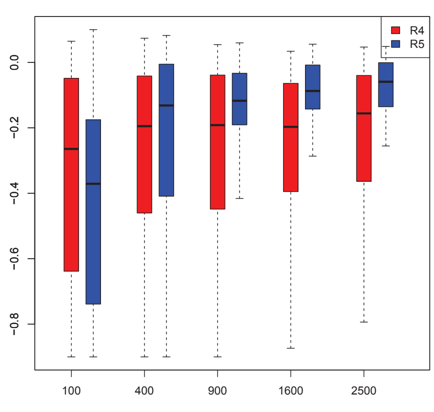
(h) p_{12}



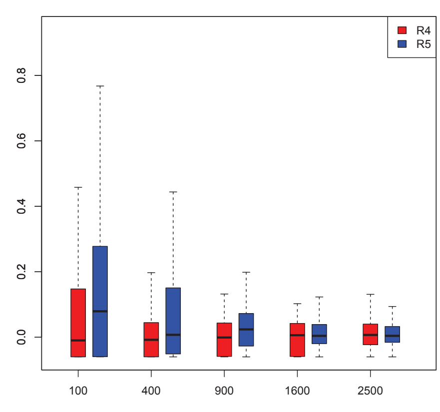
(i) p_{13}



(j) p_{21}

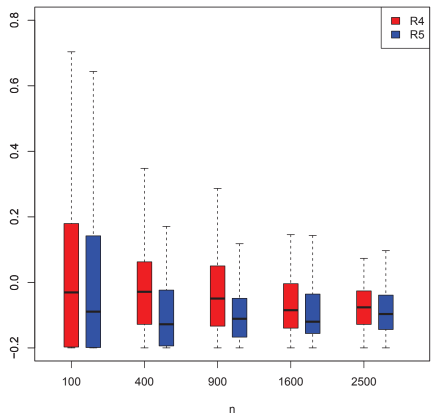


(k) p_{22}

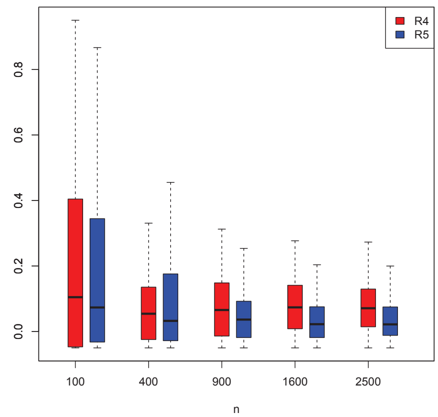


(l) p_{23}

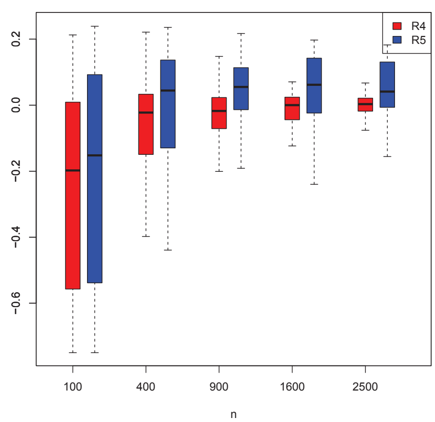
Figure 2.12: Boxplots of estimation errors in R4 and R5 (Continued)



(m) p_{31}



(n) p_{32}



(o) p_{33}

Figure 2.12: Boxplots of estimation errors in R4 and R5 (Continued)

CHAPTER 3

PARAMETER ESTIMATION OF REGIME-SWITCHING MODELS

3.1 Experimentation

In the last chapter, two models which are more restrictive than regime-switching models were carefully studied. Now let us revisit the regime-switching AR(1) model that was defined in Chapter 1:

$$y_t = c_{s_t} + \phi_{s_t} y_{t-1} + \epsilon_t \tag{3.1}$$

where s_t can switch between multiple regimes $1, \ldots, N$ according to a Markov chain and

$$\epsilon_t \sim N\left(0, \sigma_{s_t}^2\right)$$

The goal of this chapter is to examine parameter estimation for this more general model. We propose a mechanism to generate initial search values based on results from the last chapter. We show that this scheme can perform reasonably well and reduce the computational cost dramatically.

It should be noted that the formulation of the regime-switching model (3.1) is the same as the regression switching model (2.2). The only difference is whether the constants c_i are known or not. Therefore, we use the same generating parameters and sample datasets as in Section 2.2. The true parameters are listed in Table 2.8, where models are given different names to distinguish them from fixed intercepts switching models. Here the constants are unknown and need to be inferred from the data.

In the simulation study of the fixed intercepts switching models, we proposed two ways

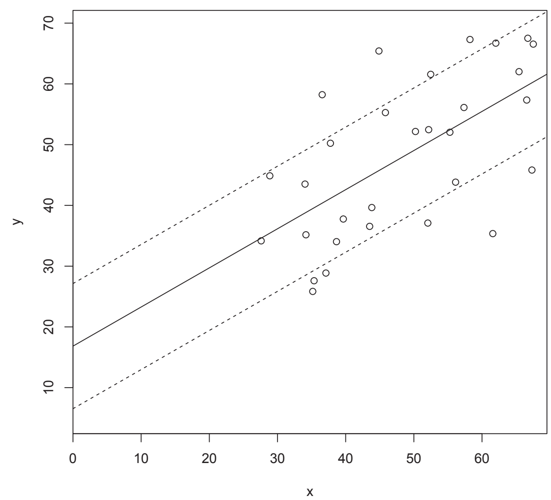
N=2				
	$\boldsymbol{c}$	$\boldsymbol{\phi}$	$\boldsymbol{\sigma}$	$\boldsymbol{P}$
F1	$\boldsymbol{c} = (5, 30)$	$\boldsymbol{\phi} = (0.98, 0.95)$	$\boldsymbol{\sigma} = (1, 2)$	$\boldsymbol{P} = \begin{pmatrix} 0.75 & 0.25 \\ 0.1 & 0.9 \end{pmatrix}$
F2	$\boldsymbol{c} = (10, 20)$	$\boldsymbol{\phi} = (0.98, 0.96)$	$\boldsymbol{\sigma} = (2, 3)$	$\boldsymbol{P} = \begin{pmatrix} 0.75 & 0.25 \\ 0.1 & 0.9 \end{pmatrix}$
F3	$\boldsymbol{c} = (10, 15)$	$\boldsymbol{\phi} = (0.98, 0.97)$	$\boldsymbol{\sigma} = (3, 6)$	$\boldsymbol{P} = \begin{pmatrix} 0.75 & 0.25 \\ 0.1 & 0.9 \end{pmatrix}$
N=3				
F4	$\boldsymbol{c} = (5, 15, 30)$	$\boldsymbol{\phi} = (0.98, 0.95, 0.92)$	$\boldsymbol{\sigma} = (2, 2, 3)$	$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.75 \end{pmatrix}$
F5	$\boldsymbol{c} = (10, 15, 20)$	$\boldsymbol{\phi} = (0.98, 0.96, 0.94)$	$\boldsymbol{\sigma} = (3, 3, 5)$	$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.75 \end{pmatrix}$

Table 3.1: Generating parameters of model (3.1) in the simulation. The values are the same as in Table 2.8

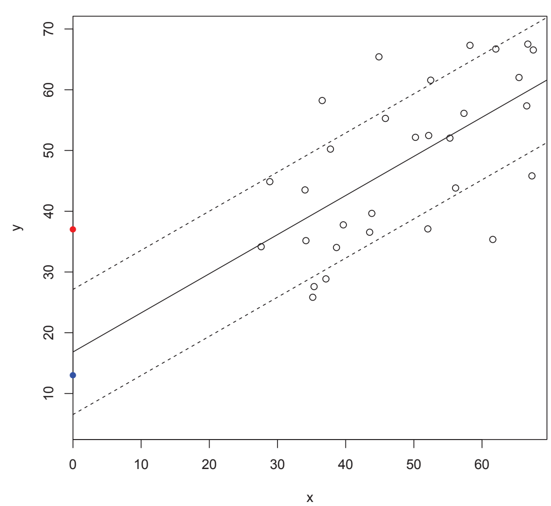
to generate initial values and used them in combination in the optimization of the likelihood function. This initial-value-sampling scheme is called R.IVS1.Our experimental results showed R.IVS1 was able to provide good initial search values. Therefore we want to adapt this scheme to fit the regime switching model (3.1). Because the only difference between regime switching models and mixing regression models rests with whether the c_i are known, we can first sample $\hat{c}_i$ for all i and treat them as given. RSTART1 or RSTART2 can then be used to sample initial values for other parameters. To be more specific, a regression line is first fitted to all data points (y_t, y_{t-1}) for $t = 2, \ldots, n_t$. Suppose that the estimated constant and standard deviation are $\hat{c}$ and $\hat{\sigma}$. Initial values of c_i^0 are sampled from the normal distribution $N(\hat{c}, \hat{\sigma})$. Initial values for other parameters are generated using RSTART1 and RSTART2 where the c_i^0 are treated as given. Please see Figure 3.1 for a graphical demonstration of how to sample initial constants.

3.2 Experimental Results

The estimation results are listed in Table 3.2 - Table 3.6. The boxplots of estimation error are given in Figure 3.3 and 3.4, where outliers are not shown. First, notice that there are a few unusually large numbers in Table 3.2 - Table 3.6, which have been displayed in bold font. The estimates of σ_1 in F1 with $n = 100$ have mean 578 and standard deviation 5754. A careful examination of the data finds two outliers with value 57 and 57543. After removing these two outliers, the mean and standard deviation are reduced to 1.75 and 1.18 respectively. There are four other instances of extreme values. All of them come from



(a) First, a regression line is fitted to all data. Suppose the estimated constant is $\hat{c}$ and estimated error standard deviation is $\hat{\sigma}$.



(b) Sample constants c_1 and c_2 from $N(\hat{c}, \hat{\sigma}^2)$ and treat them as given.

Figure 3.1: An illustration of how to sample initial intercepts in the parameter estimation of a regime-switching model.

estimations of model F4 and can be found in Table 3.5. Estimates of σ_3 with $n = 1600$ and $n = 2500$ all have large mean and standard deviation. There is one outlier in each case with value 164725 ($n = 1600$) and 5446 ($n = 2500$). Removing outliers reduces their mean and standard deviation to 2.72 and 0.66 for $n = 1600$ and to 2.74 and 0.64 for $n = 2500$. Another pair of extreme values appears in the estimation of σ_1 with $n = 1600$. This is due to an outlier which equals 5336. The adjusted mean and variance are 2.33 and 2.61. The last pair of extreme values from F4 are from estimates of c_3 with $n = 2500$, where the mean and standard deviation are 39848 and 398231. It turns out there is one outlier which equals to 3982343. The adjusted mean and standard deviation are 25.2 and 8.2. The existence of those outliers suggests the optimizer failed to reach the neighborhood of true parameters in those cases. This may be caused by bad initial search values. It is also possible that the log-likelihood function has certain structure such as a sharp valley, forcing the optimizer to search along one direction. The true reason causing this phenomenon would need further investigation. Nevertheless, extreme values are signs of poor performance of the optimizer and illustrate the difficulties in the parameter estimation of regime-switching models.

The percentage of getting better likelihood in 100 runs is plotted with respect to n in Figure 3.2 for all models. The bp values for F2, F3 and F5 are all roughly 90% and above, indicating good estimation results in those cases. The bp values for F1 and F4 exhibit different patterns. It should be noted that F1 and F4 are the two models with extreme values in their estimations. Their low bp values confirm our claim above about the poor performance of the optimizer in those cases.

When the number of data points used in optimization is small ($n = 100, 400$), the bp value for F1 is very low. The performance of the optimizer improves as n increases and bp reaches its maximum with 2500 data points. From Table 3.2, we can see that the parameters of F1 can be well estimated. There is a clear trend of convergency in all parameters of F1, which suggests F1 is not a very hard model to estimate. More importantly, this indicates the proposed method to generate initial values failed to provide good start search points in the estimation of F1, especially when n is small.

The bp value for F4 drops sharply as n increases. From Table 3.5, we can see that the parameter estimation of F4 is rather poor compared with other models. As we have discussed above, there are quite a few outliers in the estimates of c and σ, especially with large n. The estimation of transition matrix $\boldsymbol{P}$ was also not satisfactory. For example, the

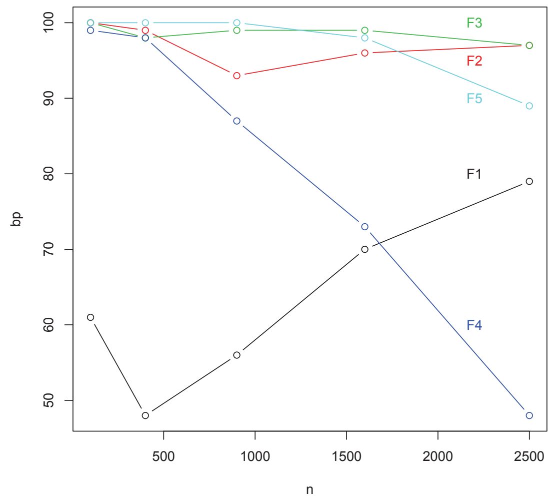


Figure 3.2: Percentage of better loglikelihood in the estimations of F1 - F5

mean of $\hat{p}_{11}$ for $n = 2500$ is equal to 0.48 and the true value is 0.8.

For all models, the multiplicative coefficients $\boldsymbol{\phi}$ can be estimated very well even with small n. $\boldsymbol{\phi}$ can be estimated with two digits of precision for all models and all n. This is the same as the result from estimating R1 - R5, where the constant term c is assumed to be given. In other words, $\boldsymbol{\phi}$ can always be estimated well regardless of whether knowing $\boldsymbol{c}$. $\boldsymbol{c}$, $\boldsymbol{\sigma}$ and $\boldsymbol{P}$ can be reasonably well estimated in F2, F3 and F5. Although very slow, there is a tendency to converge. However, those estimates are still relatively far away from the true values even with all 2500 data points.

Our experimental results demonstrate the difficulties of estimating regime-switching models. Even in some easy cases, parameter estimates show very slow convergent speed. Often in those models, estimated parameters have a better loglikelihood value than the true parameter. This indicates the regime-switching models are very hard to fit and will need more data in order to obtain accurate estimates. The proposed scheme to sample initial values is successful. The total number of parameters in a N regime model is $N^2 + 2N$. Except in F4, the parameters can be reasonably estimated with only 5 restarts, which is very unusual in high dimensional optimization.

N=2							
	cp	bp		$\boldsymbol{\phi} = (0.98, 0.95)$	$\boldsymbol{c} = (5, 30)$	$\boldsymbol{\sigma} = (1, 2)$	$\boldsymbol{P} = \begin{pmatrix} 0.75 & 0.25 \\ 0.1 & 0.9 \end{pmatrix}$
$n = 100$	98%	61%	Mean	$(0.966, 0.955)$	$(10.62, 23.51)$	$(\mathbf{577.71}, 2.12)$	$\begin{pmatrix} 0.78 & 0.22 \\ 0.17 & 0.83 \end{pmatrix}$
			Std	$(0.099, 0.017)$	$(10.95, 10.68)$	$(\mathbf{5754.05}, 1.37)$	$\begin{pmatrix} 0.17 & 0.17 \\ 0.12 & 0.12 \end{pmatrix}$
$n = 400$	100%	48%	Mean	$(0.977, 0.954)$	$(10.34, 24.25)$	$(1.68, 1.88)$	$\begin{pmatrix} 0.82 & 0.18 \\ 0.17 & 0.83 \end{pmatrix}$
			Std	$(0.011, 0.012)$	$(8.06, 8.98)$	$(0.85, 0.71)$	$\begin{pmatrix} 0.08 & 0.08 \\ 0.08 & 0.08 \end{pmatrix}$
$n = 900$	98%	56%	Mean	$(0.98, 0.955)$	$(7.76, 24.68)$	$(1.47, 1.74)$	$\begin{pmatrix} 0.8 & 0.2 \\ 0.15 & 0.85 \end{pmatrix}$
			Std	$(0.007, 0.009)$	$(5.64, 8.2)$	$(0.69, 0.39)$	$\begin{pmatrix} 0.074 & 0.074 \\ 0.076 & 0.076 \end{pmatrix}$
$n = 1600$	100%	70%	Mean	$(0.98, 0.953)$	$(6.61, 27.19)$	$(1.22, 1.88)$	$\begin{pmatrix} 0.78 & 0.22 \\ 0.12 & 0.88 \end{pmatrix}$
			Std	$(0.005, 0.008)$	$(3.89, 6.65)$	$(0.5, 0.31)$	$\begin{pmatrix} 0.06 & 0.06 \\ 0.06 & 0.06 \end{pmatrix}$
$n = 2500$	100%	79%	Mean	$(0.98, 0.951)$	$(5.99, 28.57)$	$(1.12, 1.92)$	$\begin{pmatrix} 0.77 & 0.23 \\ 0.11 & 0.89 \end{pmatrix}$
			Std	$(0.003, 0.005)$	$(2.76, 5.12)$	$(0.38, 0.25)$	$\begin{pmatrix} 0.05 & 0.05 \\ 0.05 & 0.05 \end{pmatrix}$

Table 3.2: Estimation Result of model F1

N=2							
	cp	bp		$\boldsymbol{\phi} = (0.98, 0.96)$	$\boldsymbol{c} = (10, 20)$	$\boldsymbol{\sigma} = (2, 3)$	$\boldsymbol{P} = \begin{pmatrix} 0.75 & 0.25 \\ 0.1 & 0.9 \end{pmatrix}$
$n = 100$	100%	100%	Mean	$(0.92, 0.91)$	$(39.64, 47.42)$	$(2.03, 2.13)$	$\begin{pmatrix} 0.5 & 0.5 \\ 0.46 & 0.54 \end{pmatrix}$
			Std	$(0.058, 0.058)$	$(28.72, 28.88)$	$(0.81, 0.84)$	$\begin{pmatrix} 0.35 & 0.35 \\ 0.35 & 0.35 \end{pmatrix}$
$n = 400$	100%	99%	Mean	$(0.958, 0.946)$	$(20.59, 26.67)$	$(2.36, 2.54)$	$\begin{pmatrix} 0.52 & 0.48 \\ 0.41 & 0.59 \end{pmatrix}$
			Std	$(0.018, 0.022)$	$(8.96, 10.94)$	$(0.65, 0.64)$	$\begin{pmatrix} 0.37 & 0.37 \\ 0.33 & 0.33 \end{pmatrix}$
$n = 900$	100%	93%	Mean	$(0.963, 0.953)$	$(18.64, 23.25)$	$(2.48, 2.55)$	$\begin{pmatrix} 0.62 & 0.38 \\ 0.37 & 0.63 \end{pmatrix}$
			Std	$(0.011, 0.014)$	$(5.35, 6.88)$	$(0.67, 0.64)$	$\begin{pmatrix} 0.33 & 0.33 \\ 0.32 & 0.32 \end{pmatrix}$
$n = 1600$	100%	96%	Mean	$(0.965, 0.959)$	$(17.03, 20.49)$	$(2.42, 2.51)$	$\begin{pmatrix} 0.63 & 0.37 \\ 0.27 & 0.73 \end{pmatrix}$
			Std	$(0.008, 0.009)$	$(3.93, 4.26)$	$(0.78, 0.71)$	$\begin{pmatrix} 0.31 & 0.31 \\ 0.28 & 0.28 \end{pmatrix}$
$n = 2500$	100%	97%	Mean	$(0.967, 0.96)$	$(16.55, 20.11)$	$(2.53, 2.5)$	$\begin{pmatrix} 0.73 & 0.27 \\ 0.27 & 0.73 \end{pmatrix}$
			Std	$(0.007, 0.009)$	$(3.4, 4.29)$	$(0.65, 0.63)$	$\begin{pmatrix} 0.24 & 0.24 \\ 0.25 & 0.25 \end{pmatrix}$

Table 3.3: Estimation Result of model F2

N=2							
	cp	bp		$\boldsymbol{\phi} = (0.98, 0.97)$	$\boldsymbol{c} = (10, 15)$	$\boldsymbol{\sigma} = (3, 6)$	$\boldsymbol{P} = \begin{pmatrix} 0.75 & 0.25 \\ 0.1 & 0.9 \end{pmatrix}$
$n = 100$	98%	100%	Mean	$(0.928, 0.89)$	$(35.72, 54.09)$	$(4.12, 4.18)$	$\begin{pmatrix} 0.55 & 0.45 \\ 0.41 & 0.59 \end{pmatrix}$
			Std	$(0.055, 0.096)$	$(26.45, 46.77)$	$(2.34, 1.67)$	$\begin{pmatrix} 0.37 & 0.37 \\ 0.37 & 0.37 \end{pmatrix}$
$n = 400$	100%	98%	Mean	$(0.968, 0.948)$	$(15.99, 25.54)$	$(5.12, 4.08)$	$\begin{pmatrix} 0.73 & 0.27 \\ 0.35 & 0.65 \end{pmatrix}$
			Std	$(0.019, 0.03)$	$(9.24, 14.47)$	$(1.61, 1.7)$	$\begin{pmatrix} 0.26 & 0.26 \\ 0.32 & 0.32 \end{pmatrix}$
$n = 900$	100%	99%	Mean	$(0.973, 0.962)$	$(13.75, 18.9)$	$(4.97, 4.11)$	$\begin{pmatrix} 0.81 & 0.19 \\ 0.25 & 0.75 \end{pmatrix}$
			Std	$(0.01, 0.014)$	$(5.01, 6.96)$	$(1.49, 1.65)$	$\begin{pmatrix} 0.16 & 0.16 \\ 0.2 & 0.2 \end{pmatrix}$
$n = 1600$	100%	99%	Mean	$(0.975, 0.965)$	$(12.6, 17.47)$	$(4.65, 4.43)$	$\begin{pmatrix} 0.81 & 0.19 \\ 0.2 & 0.8 \end{pmatrix}$
			Std	$(0.009, 0.009)$	$(4.27, 4.37)$	$(1.55, 1.59)$	$\begin{pmatrix} 0.14 & 0.14 \\ 0.11 & 0.11 \end{pmatrix}$
$n = 2500$	100%	97%	Mean	$(0.976, 0.968)$	$(12.2, 16)$	$(4.59, 4.47)$	$\begin{pmatrix} 0.81 & 0.19 \\ 0.19 & 0.81 \end{pmatrix}$
			Std	$(0.007, 0.008)$	$(3.29, 4.02)$	$(1.54, 1.56)$	$\begin{pmatrix} 0.1 & 0.1 \\ 0.09 & 0.09 \end{pmatrix}$

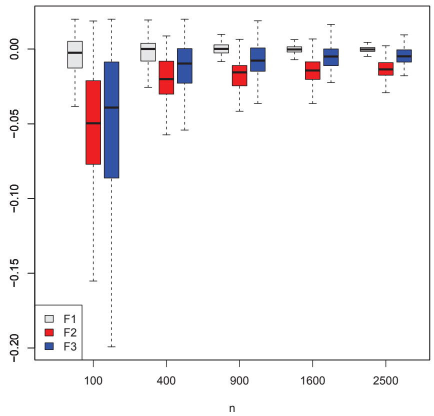
Table 3.4: Estimation Result of model F3

N=3							
n	cp	bp		$\boldsymbol{\phi} = (0.98, 0.95, 0.92)$	$\boldsymbol{c} = (5, 15, 30)$	$\boldsymbol{\sigma} = (2, 2, 3)$	$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.75 \end{pmatrix}$
100	100%	99%	Mean	$(0.97, 0.954, 0.928)$	$(9.74, 15.61, 28.27)$	$(1.63, 1.6, 1.85)$	$\begin{pmatrix} 0.5 & 0.31 & 0.19 \\ 0.32 & 0.52 & 0.16 \\ 0.22 & 0.24 & 0.54 \end{pmatrix}$
			Std	$(0.028, 0.025, 0.065)$	$(8.67, 10.44, 21.08)$	$(0.76, 0.79, 0.89)$	$\begin{pmatrix} 0.31 & 0.29 & 0.26 \\ 0.33 & 0.34 & 0.21 \\ 0.27 & 0.29 & 0.32 \end{pmatrix}$
400	100%	98%	Mean	$(0.977, 0.959, 0.934)$	$(7.18, 13.25, 24.07)$	$(2.07, 1.92, 2.23)$	$\begin{pmatrix} 0.56 & 0.3 & 0.14 \\ 0.27 & 0.59 & 0.14 \\ 0.18 & 0.19 & 0.63 \end{pmatrix}$
			Std	$(0.02, 0.02, 0.03)$	$(4.47, 6.71, 10.64)$	$(0.74, 0.66, 0.88)$	$\begin{pmatrix} 0.29 & 0.29 & 0.17 \\ 0.3 & 0.32 & 0.22 \\ 0.21 & 0.22 & 0.26 \end{pmatrix}$
900	100%	87%	Mean	$(0.977, 0.956, 0.934)$	$(7.05, 14.43, 24.2)$	$(2.17, 2.11, 2.53)$	$\begin{pmatrix} 0.56 & 0.29 & 0.15 \\ 0.27 & 0.62 & 0.11 \\ 0.2 & 0.15 & 0.65 \end{pmatrix}$
			Std	$(0.01, 0.02, 0.03)$	$(3.9, 7.61, 9.82)$	$(2.09, 0.79, 0.74)$	$\begin{pmatrix} 0.32 & 0.3 & 0.21 \\ 0.32 & 0.33 & 0.17 \\ 0.23 & 0.19 & 0.25 \end{pmatrix}$
1600	100%	73%	Mean	$(0.98, 0.956, 0.926)$	$(6.34, 14.11, 26.86)$	$(\mathbf{55.67}, 1.98, \mathbf{1649.95})$	$\begin{pmatrix} 0.55 & 0.26 & 0.19 \\ 0.26 & 0.66 & 0.08 \\ 0.18 & 0.14 & 0.68 \end{pmatrix}$
			Std	$(0.01, 0.02, 0.04)$	$(3.89, 6.48, 8.57)$	$(\mathbf{533}, 0.64, \mathbf{16472})$	$\begin{pmatrix} 0.37 & 0.33 & 0.27 \\ 0.32 & 0.32 & 0.11 \\ 0.23 & 0.17 & 0.22 \end{pmatrix}$
2500	100%	48%	Mean	$(0.978, 0.954, 0.922)$	$(5.84, 14.07, \mathbf{39848})$	$(3.27, 2.14, \mathbf{57.18})$	$\begin{pmatrix} 0.48 & 0.31 & 0.21 \\ 0.2 & 0.71 & 0.09 \\ 0.17 & 0.14 & 0.69 \end{pmatrix}$
			Std	$(0.02, 0.04, 0.1)$	$(3.83, 6.43, \mathbf{398231})$	$(11.64, 0.62, \mathbf{544})$	$\begin{pmatrix} 0.36 & 0.35 & 0.29 \\ 0.28 & 0.29 & 0.12 \\ 0.21 & 0.14 & 0.21 \end{pmatrix}$

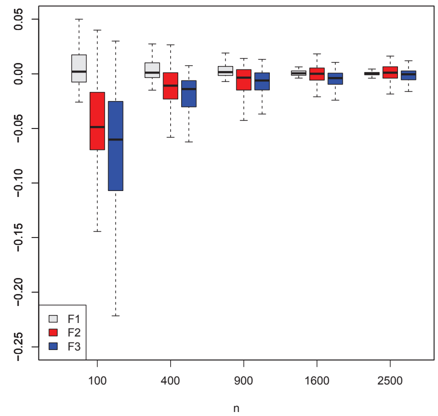
Table 3.5: Estimation Result of model F4

N=3							
n	cp	bp		$\boldsymbol{\phi} = (0.98, 0.96, 0.94)$	$\boldsymbol{c} = (10, 15, 20)$	$\boldsymbol{\sigma} = (3, 3, 5)$	$\boldsymbol{P} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.04 & 0.9 & 0.06 \\ 0.2 & 0.05 & 0.75 \end{pmatrix}$
100	100%	100%	Mean	$(0.961, 0.947, 0.924)$	$(13.32, 19.39, 28.96)$	$(2.73, 3.08, 2.36)$	$\begin{pmatrix} 0.51 & 0.22 & 0.27 \\ 0.3 & 0.47 & 0.23 \\ 0.28 & 0.22 & 0.5 \end{pmatrix}$
			Std	$(0.03, 0.04, 0.07)$	$(12.79, 14.15, 23.9)$	$(1.46, 7.97, 1.41)$	$\begin{pmatrix} 0.33 & 0.26 & 0.32 \\ 0.33 & 0.33 & 0.28 \\ 0.34 & 0.3 & 0.37 \end{pmatrix}$
400	100%	100%	Mean	$(0.971, 0.957, 0.94)$	$(10.45, 15.94, 21.88)$	$(3.34, 3.08, 3.34)$	$\begin{pmatrix} 0.64 & 0.17 & 0.19 \\ 0.18 & 0.78 & 0.14 \\ 0.18 & 0.17 & 0.65 \end{pmatrix}$
			Std	$(0.02, 0.02, 0.03)$	$(6.21, 7.72, 9.96)$	$(1.27, 1.11, 1.26)$	$\begin{pmatrix} 0.26 & 0.18 & 0.23 \\ 0.23 & 0.29 & 0.24 \\ 0.23 & 0.22 & 0.29 \end{pmatrix}$
900	100%	100%	Mean	$(0.974, 0.963, 0.944)$	$(9.51, 13.9, 20.6)$	$(3.55, 3.42, 3.56)$	$\begin{pmatrix} 0.76 & 0.12 & 0.12 \\ 0.13 & 0.76 & 0.11 \\ 0.15 & 0.09 & 0.76 \end{pmatrix}$
			Std	$(0.01, 0.02, 0.03)$	$(4.12, 5.47, 9.17)$	$(1.04, 1.03, 1.14)$	$\begin{pmatrix} 0.19 & 0.12 & 0.15 \\ 0.16 & 0.16 & 0.12 \\ 0.19 & 0.08 & 0.18 \end{pmatrix}$
1600	100%	98%	Mean	$(0.973, 0.967, 0.952)$	$(9.52, 12.7, 17.96)$	$(3.86, 3.51, 3.48)$	$\begin{pmatrix} 0.77 & 0.14 & 0.09 \\ 0.12 & 0.79 & 0.09 \\ 0.1 & 0.08 & 0.82 \end{pmatrix}$
			Std	$(0.01, 0.01, 0.02)$	$(3.06, 3.3, 5.6)$	$(1.07, 0.89, 0.97)$	$\begin{pmatrix} 0.11 & 0.12 & 0.07 \\ 0.1 & 0.1 & 0.08 \\ 0.08 & 0.07 & 0.09 \end{pmatrix}$
2500	100%	89%	Mean	$(0.974, 0.966, 0.957)$	$(9.62, 12.66, 16.15)$	$(3.71, 3.61, 3.63)$	$\begin{pmatrix} 0.79 & 0.11 & 0.1 \\ 0.11 & 0.8 & 0.09 \\ 0.1 & 0.08 & 0.82 \end{pmatrix}$
			Std	$(0.01, 0.01, 0.01)$	$(2.78, 2.58, 4.1)$	$(1.03, 0.95, 0.99)$	$\begin{pmatrix} 0.08 & 0.07 & 0.07 \\ 0.08 & 0.09 & 0.06 \\ 0.07 & 0.07 & 0.08 \end{pmatrix}$

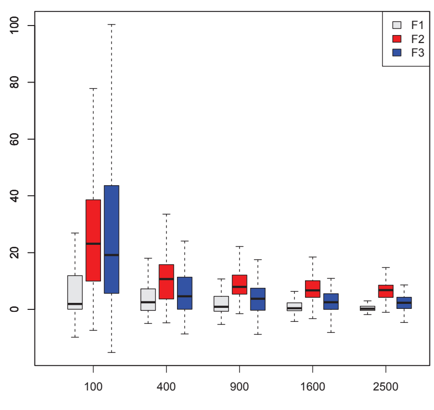
Table 3.6: Estimation Result of model F5



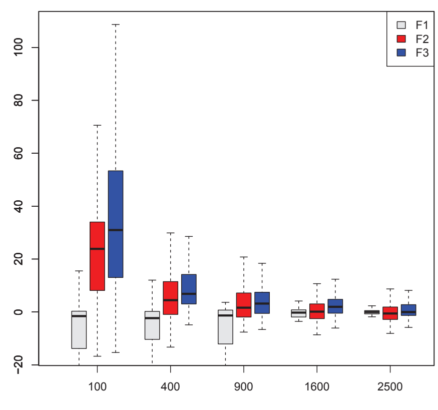
(a) ϕ_1



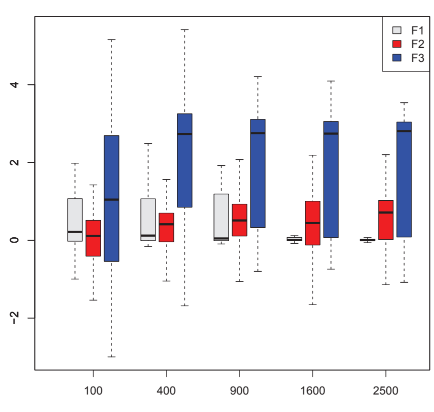
(b) ϕ_2



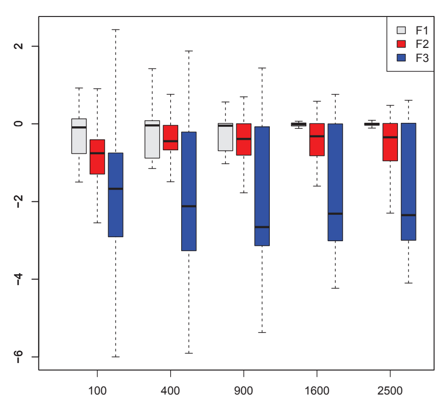
(c) c_1



(d) c_2

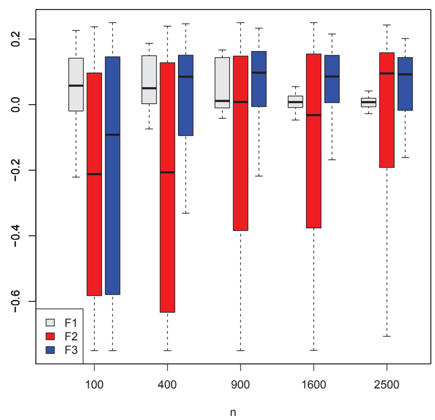


(e) σ_1

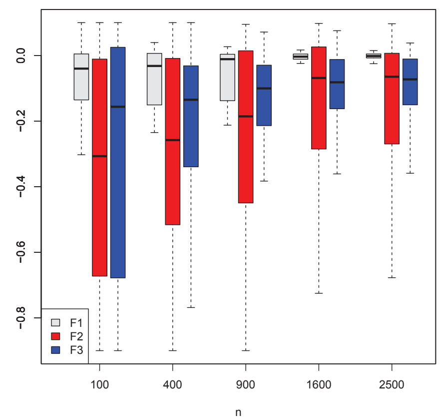


(f) σ_2

Figure 3.3: Boxplots of estimation errors in F1 - F3

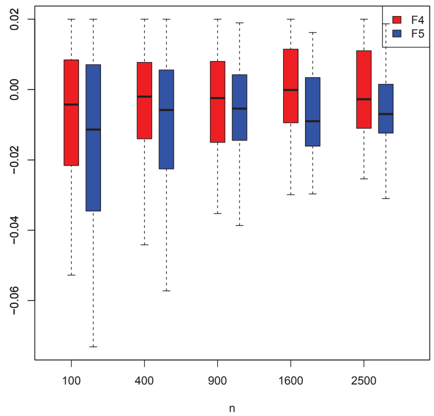


(g) p_{11}

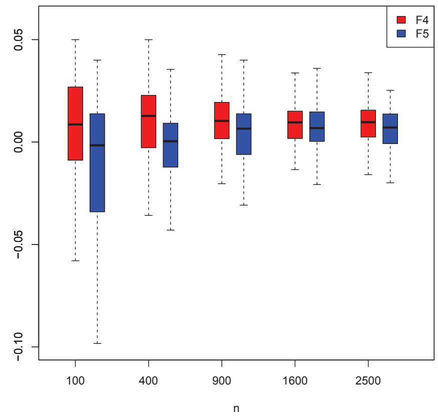


(h) p_{22}

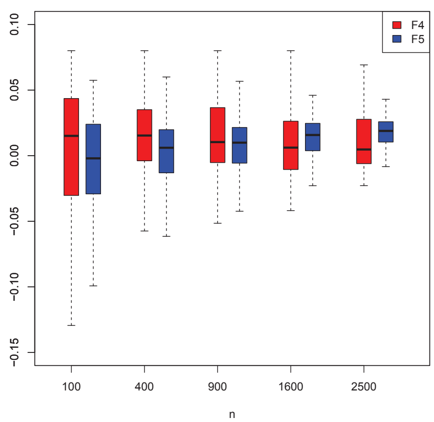
Figure 3.3: Boxplots of estimation errors in F1 - F3 (Continued)



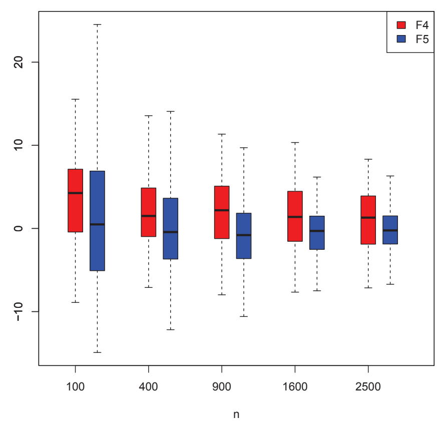
(a) ϕ_1



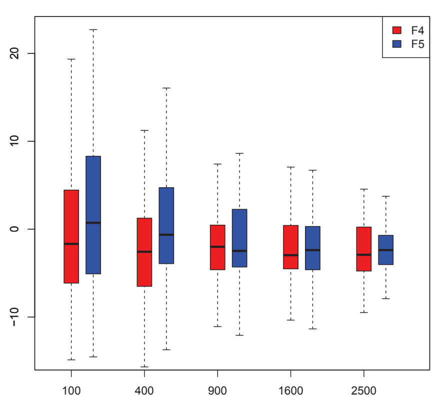
(b) ϕ_2



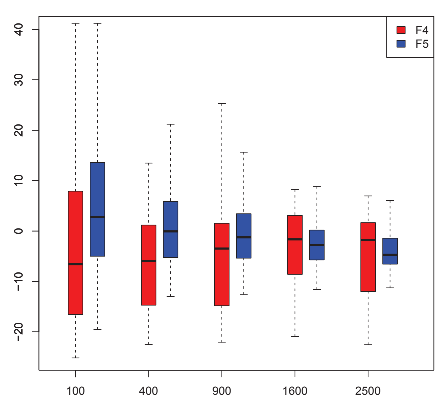
(c) ϕ_3



(d) c_1

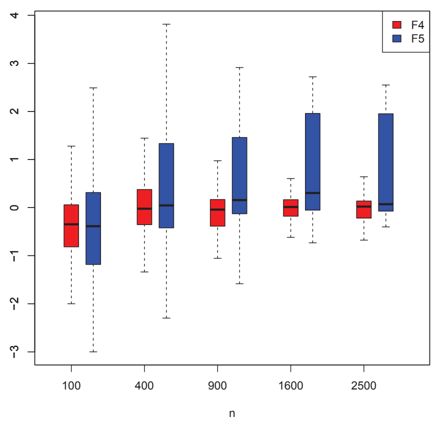


(e) c_2

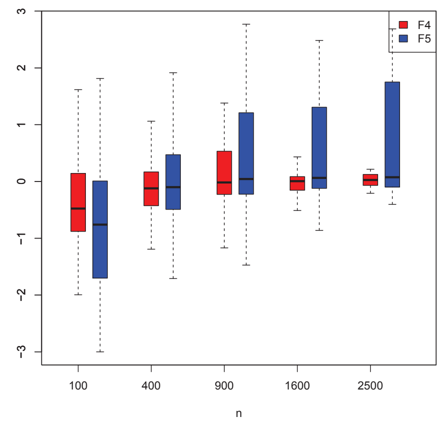


(f) c_3

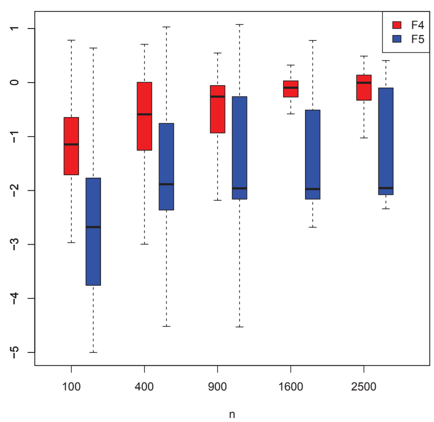
Figure 3.4: Boxplots of estimation errors in F4 and F5



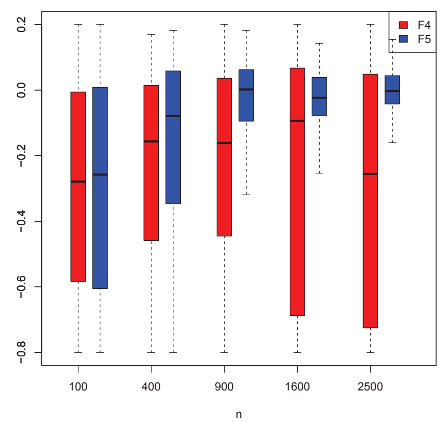
(g) σ_1



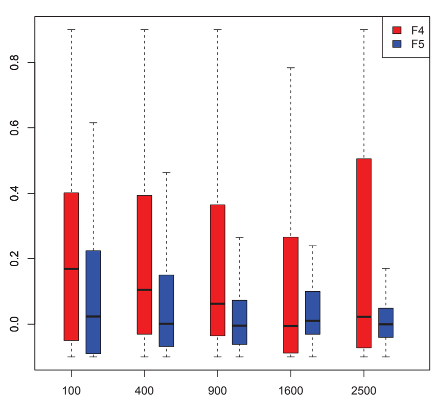
(h) σ_2



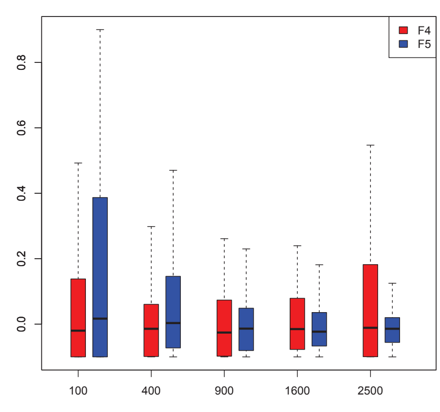
(i) σ_3



(j) p_{11}

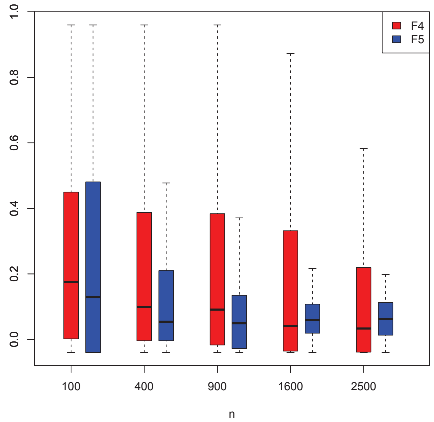


(k) p_{12}

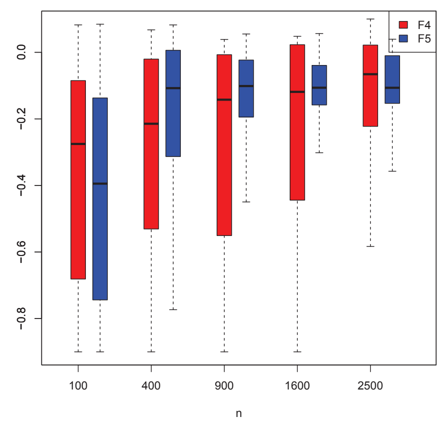


(l) p_{13}

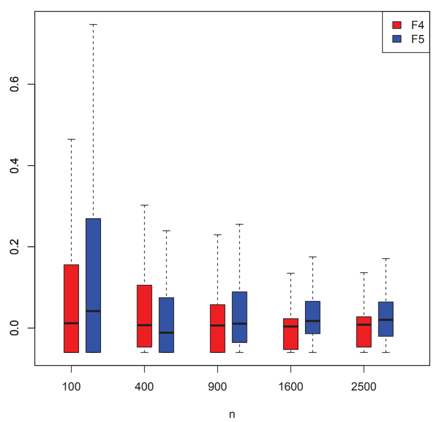
Figure 3.4: Boxplots of estimation errors in F4 and F5 (Continued)



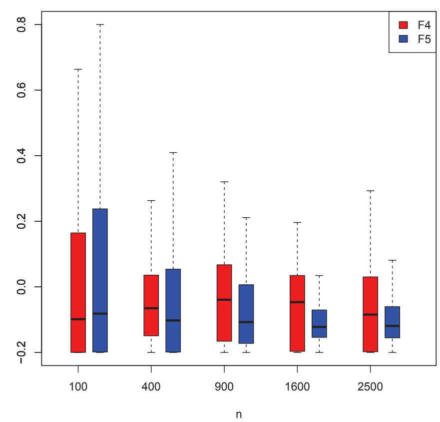
(m) p_{21}



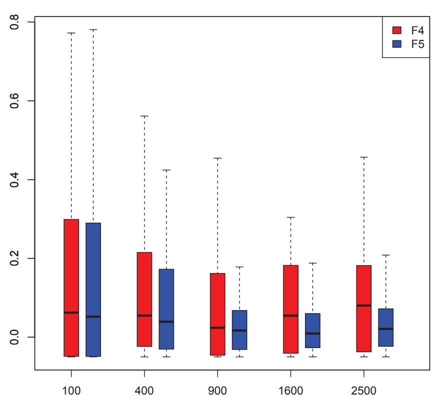
(n) p_{22}



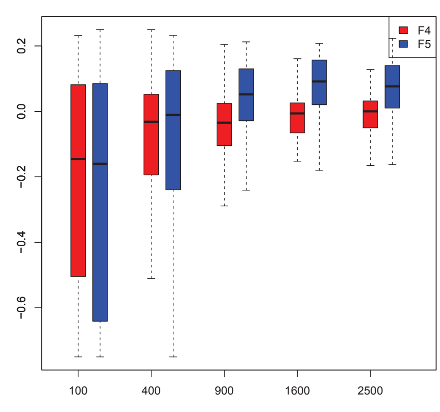
(o) p_{23}



(p) p_{31}



(q) p_{32}



(r) p_{33}

Figure 3.4: Boxplots of estimation errors in F4 and F5 (Continued)

CHAPTER 4

CONCLUSIONS AND FUTURE WORK

In this thesis, we conducted a simulation study of a regime switching model, in order to better understand the (empirical) properties of the loglikelihood function and MLE. It was shown that the interaction of the slope coefficient ϕ and the constant c introduces many local maxima to the loglikelihood function, which makes parameter estimation a very hard optimization problem. To further investigate the interaction between ϕ and c, two sub-models that are easier to fit than the regime switching model (1.2) were studied. The first model is called the switching mixture model (2.1), in which ϕ is assumed to be zero. We showed that this model can be considered as an extension of a mixture model. Each switching mixture model has an associated mixture model by dropping the Markov property. Our experimentation indicated switching mixture models can be well approximated by mixture models. A good strategy for fitting a switching mixture model is to first fit the corresponding mixture model. The estimated parameters of the mixture model can then be used as initial search values in the parameter estimation of the switching mixture model.

The second sub-model studied in this thesis is called the fixed intercepts switching model, in which the constant vector $\mathbf{c}$ is assumed to be given. In order to fit this model, we suggest initially approximating intercepts switching models by so-called k-line clustering problems. A deterministic algorithm was proposed to find approximate solutions of k-line clustering problems, which can be used as initial search values in optimizing the loglikelihood function of the corresponding fixed intercepts switching models. An initial-value-sampling scheme R.IVS1 was proposed to generate initial values in order to solve the k-line clustering problem. Experimental results demonstrated the effectiveness of this method. Parameters in most fixed intercepts switching models were estimated very well.

Finally, the technique for deriving initial values in fitting fixed intercepts switching models was extended in order to solve regime switching models. The proposed method was shown to provide good initial search values and effectively reduce the number of restarts of the optimizer. Even in 3-regime models (with 15 unknown parameters), parameters can be estimated reasonably well with only 5 restarts.

The work in this thesis can be extended in many aspects. Although the initial-value-sampling method was proven successful, there were situations in which its performance was not as good as in other cases. The potential reasons are not clear and need further investigation. Also, the optimizer in our experimentation was only restarted 5 times due to limited time and computational capabilities. It will be interesting to increase the number of restarts of the optimizer and observe whether the probability of getting estimates with better loglikelihood will increase as well. An increasing bp value indicates the proposed initial-value-sampling method is able to generate points that spread over the entire region around the MLE.

For simplicity, the regime-switching model considered in this thesis is a one dimensional AR(1) model. In order to further study parameter estimation and the convergence properties of regime switching models, one could consider multidimensional AR(k) models i.e

$$\mathbf{y}_t = \mathbf{A}_{s_t}\mathbf{y}_{t-1}^{(k)} + \mathbf{c}_{s_t} + \epsilon_t$$

where $\mathbf{y}_t$, $\mathbf{c}_{s_t}$ and ϵ_t are m dimensional vectors. $\mathbf{y}_{t-1}^{(k)}$ is a mk dimensional vector of past observations up to time $t-k$ and $\mathbf{A}$ is a m by mk matrix. Note that the proposed method to generate initial search values can be extended to m dimensions using multiple regression techniques. Parameter estimation in multidimensional AR(k) models is expected to be much harder. Furthermore, we only considered models with two or three regimes in this thesis. The properties of regime switching models with more regimes should also be studied. It would be interesting to test whether the proposed initial-value-sampling method could succeed in higher dimensional cases.

In this thesis, the Nelder-Mead algorithm was used to optimize loglikelihood functions. Although a popular optimizer, Nelder-Mead does have some drawbacks as shown by *McKinnon* (1998). Other alternative optimizers could also be used. Of particular interest to the author is a special type of evolutionary strategy called the *Covariance Matrix Adaptation Evolution Strategy* (CMA-ES). Evolutionary strategies are a class of optimization techniques based on the idea of evolution and adaptation. CMA-ES was proposed by *Hansen and Ostermeier* (2001). It is often applied when traditional optimizers (such as quasi-Newton or conjugate gradient) fail due to local maxima, discontinuities or noises in objective functions. However, CMA-ES is rarely known or used among statisticians. It would be interesting to benchmark and compare the performance of CMA-ES and Nelder-Mead in fitting regime-switching models.

Bibliography

Altug, S., and M. Bildirici, Business cycles around the globe: A regime switching approach, *TUSIAD-Koc University Economic Research Forum Working Papers 1009*, TUSIAD-Koc University Economic Research Forum, 2010.

Ang, A., and G. Bekaert, Regime switches in interest rates, *Journal of Business & Economic Statistics*, *20*, 163–182, 2002.

Bäck, T., *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*, Oxford University Press, Oxford, UK, 1996.

Bekaert, G., and R. J. Hodrick, Characterizing predictable components in excess returns on equity and foreign exchange markets, *The Journal of Finance*, *47*, 467–509, 1992.

Bickel, P. J., Y. Ritov, and T. Ryden, Asymptotic normality of the maximum-likelihood estimator for general hidden markov models, *The Annals of Statistics*, *26*, 1614–1635, 1998.

Bollen, N. P., S. F. Gray, and R. E. Whaley, Regime switching in foreign exchange rates:: Evidence from currency option prices, *Journal of Econometrics*, *94*, 239 – 276, 2000.

Dempster, A., L. N., and R. D., Maximum likelihood estimation from incomplete data via the em algorithm, *Journal of the Royal Statistical Society*, *39*, 1–38, 1977.

Douc, R., E. Moulines, and T. Ryden, Asymptotic properties of the maximum likelihood estimator in autoregressive models with markov regime, *The Annals of Statistics*, *32*, 2254–2304, 2004.

Fraley, C., A. E. Raftery, T. B. Murphy, and L. Scrucca, Mclust version 4 for r: Normal mixture modeling for model-based clustering, classification, and density estimation, *Tech. Rep. 597*, Department of Statistics, University of Washington, 2012.

Garcia, R., and P. Perron, An analysis of the real interest rate under regime shifts, *The Review of Economics and Statistics*, *78*, 111–25, 1996.

Goldfield, S., and R. Quandt, A markov model for switching regressions, *Journal of Econometrics*, *1*, 3–15, 1973.

Hamilton, J., A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, *57*, 357–384, 1989.

Hamilton, J., Regime-switching models, *Palgrave Dictionary of Economics*, 2005.

Hamilton, J., and R. Susmel, Autoregressive conditional heteroskedasticity and changes in regime, *Journal of Econometrics*, *64*, 307–333, 1994.

Hansen, N., and A. Ostermeier, Completely derandomized self-adaptation in evolution strategies, *Evol. Comput.*, *9*, 159–195, 2001.

Kennedy, J., and R. Eberhart, Particle swarm optimization, in *Proceedings of International Conference on Neural Networks*, vol. 4, pp. 1942–1948, IEEE, Perth, WA, Australia, 1995.

Kim, C.-J., and C. R. Nelson, Has the u.s. economy become more stable? a bayesian approach based on a markov-switching model of the business cycle, *The Review of Economics and Statistics*, *81*, 608–616, 1999.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi, Optimization by simulated annealing, *Science*, *220*, 671–680, 1983.

Krishnamurthy, V., and T. Ryden, Consistent estimation of linear and non-linear autoregressive models with markov regime, *Journal of Time Series Analysis*, *19*, 291–307, 1998.

Leroux, B. G., Maximum-likelihood estimation for hidden markov models, *Stochastic Processes and their Applications*, *40*, 127 – 143, 1992.

Lindgren, G., Markov regime models for mixed distributions and switching regressions, *Scandinavian Journal of Statistics*, *5*, 89–91, 1978.

McKinnon, K. I. M., Convergence of the nelder–mead simplex method to a nonstationary point, *SIAM J. on Optimization*, *9*, 148–158, 1998.

McLachlan, G., and D. Peel, *Finite Mixture Models*, John Wiley & Sons, Inc., 2005.

McLachlan, G. J., and T. Krishnan, *The EM Algorithm and Extensions*, vol. 382, Wiley-Interscience, 2008.

Minard, C. G., W. Chan, D. W. Wetter, and C. J. Etzel, Trends in smoking cessation: a markov model approach, *Journal of Applied Statistics*, *39*, 113–127, 2011.

Nelder, J. A., and R. Mead, A simplex method for function minimization, *Computer Journal*, *7*, 308–313, 1965.

Olivier, C., E. Moulines, and T. Ryden, *Inference In Hidden Markov Models*, Springer, 2005.

Owyang, M. T., and G. Ramey, Regime switching and monetary policy measurement, *Journal of Monetary Economics*, *51*, 1577–1597, 2004.

Piger, J., J. Morley, and C.-J. Kim, Nonlinearity and the permanent effects of recessions, *Journal of Applied Econometrics*, *20*, 291–309, 2005.

Schwert, G., Why does stock market volatility change over time?, *Journal of Finance*, *44*, 1115–1153, 1989.

Sims, C. A., and T. Zha, Were there regime switches in u.s. monetary policy?, *The American Economic Review*, *96*, 54–81, 2006.

Tse, C. H., M. Tang, P. P. Shum, and R. F. Wu, Nelder-Mead simplex method for modeling of cascaded continuous-wave multiple-Stokes raman fiber lasers, *Optical Engineering*, *49*, 091009–091009–6, 2010.

Tu, J., Is regime switching in stock returns important in portfolio decisions?, *Management Science*, *56*, 1198–1215, 2010.

Valente, G., Monetary policy rules and regime shifts, *Applied Financial Economics*, *13*, 525–535, 2003.

Wang, Z., J. He, and H. Gu, Forward kinematics analysis of a Six-Degree-of-Freedom stewart platform based on independent component analysis and Nelder-Mead algorithm, *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, *41*, 589 –597, 2011.

Ward, M., Target velocity identification using 3-D matched filter with Nelder-Mead optimization, in *IEEE Aerospace Conference*, pp. 1–7, 2011.

Wu, C. F. J., On the convergence properties of the em algorithm, *The Annals of Statistics*, *11*, 95–103, 1983.

Zucchini, W., and I. L. MacDonald, *Hidden Markov Models for Time Series: An Introduction Using R*, 1st ed., Chapman and Hall/CRC, 2009.