A New Regime Switching Model for Econometric Time Series

A Dissertation Presented

by

Ning Sun

to

The Graduate School in Partial Fulfillment of the Requirements for the

Degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

August 2011

Stony Brook University

The Graduate School

Ning Sun

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Haipeng Xing - Dissertation Advisor
Assistant Professor, Applied Mathematics and Statistics

Stephen Finch - Chairperson of Defense
Professor, Applied Mathematics and Statistics

Wei Zhu
Deputy Chair, Professor, Applied Mathematics and Statistics

Yixin Fang
Assistant Professor, Division of Biostatistics, School of Medicine,
New York University

This dissertation is accepted by the Graduate School.

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

A New Regime Switching Model for Econometric Time Series

by

Ning Sun

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

2011

Ever since the publication of Hamilton's (1989) seminal work on regime switching model, a large amount of its applications have been found in econometric and statistical problems. Despite its enormous popularity, one shortcoming of the model is that the model describes the regime qualitatively instead of quantitatively. In this dissertation research, we first review the classic regime switching model and its broad applications in different problems. Then we introduce a stochastic regime switching model in which the parameter in each regime is a random variable following certain distribution. A forward filtering procedure shows the posterior distribution of the parameter as a mixture distribution with explicit weights which can be calculated recursively. Furthermore, based on the reversibility of the hidden Markov chain, a backward filtering procedure can be conducted in a similar way.

Based on Bayes' theorem, both the smoothing estimate of parameter and probabilities can be calculated explicitly. We also develop an expectation-maximization algorithm to estimate the hyperparameters in the model. Furthermore, we propose a bounded complexity mixture (BCMIX) approximation, which has much lower computational complexity yet comparable to the Bayes estimates in statistical efficiency. We perform intensive simulation studies to evaluate the Bayes and BCMIX estimates of time-varying parameters, in terms of the sum of squared errors, L_2 errors of the estimates, the Kullback-Leibler divergence, and the identification ratios of true regimes. We also apply this model to analyze some economic time series.

To my parents, Luning and Nicholas with all my love

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my advisor, Prof. Haipeng Xing, for the exciting topics suggested, for interesting discussions and for his guidance and continuous support.

I would also like to thank Prof. Stephen Finch, Prof. Wei Zhu and Dr. Jianxin Shi for sharing their valuable experience and insights. It is my great honor to have them on my dissertation committee.

I would like to thank all my groupmates and friends, especially Tongfei Guo, Yifan Mo, Tong Shen and Yang Yu, for all their support and help.

Chapter 1

Introduction

There has been great interest in modeling stochastic systems with time-varying parameters. More specifically, the regime switching model (Hamilton 1989, 1994) has proven to be very useful for modeling economic and financial time series. The model assumes the parameters of an autoregressive regression as the hidden states of a finite-state Markov chain. It has been generalized to analyze many economic time series associated with events such as financial crises (Jeanne and Masson, 2000; Cerra and Saxena, 2005; Hamilton, 2005; Guo et al., 2011) or changes in government policy (Hamilton, 1988; Davig, 2004; Sims and Zha, 2006; Hauzenberger, 2010).

In this chapter, we review some theoretical works on the regime switching model and its applications in different problems in the first section. Then a summary of the classic regime switching model based on Hamilton (1994) is given in the second section. New observations based on the analysis of real econometric time series are shown in the third section as the motivation of our study. The last section gives the outline of this dissertation.

1.1 Literature on Regime Switching Problems

Consider a linear regression model with dependent variable y_t and the independent variable $\boldsymbol{x}_t$ which is a $(d \times 1)$ vector for $t = 1, ..., T$. For any t, the observations are generated by one of two regimes:

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta}_1 + \mu_{1t},$$

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta}_2 + \mu_{2t},$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are $(d \times 1)$ vectors of coefficients, μ_{1t} and μ_{2t} are assumed to be distributed as $N(0, \sigma_1^2)$ and $N(0, \sigma_2^2)$ respectively. If it is further assumed that $(\boldsymbol{\beta}_1, \sigma_1^2) \neq (\boldsymbol{\beta}_2, \sigma_2^2)$, the regression is called a switching regression. There are different possible assumptions on the structure for the switching regression. The simplest assumption is that there is at most one switch in the data series, that is, the first m observations in a time series are generated by regime 1 and the remaining $T - m$ observations by regime 2. Problems of this type have been analyzed in different ways by Quandt (1958, 1960), Brown and Durbin (1968), and Farley and Hinich (1970).

This simple model permitting only one switch is not realistic and useful in some economic contexts. A more complex situation arises if it is assumed that the system may switch back and forth between the two regimes. For example, the first $m(1)$ observations are from regime 1, the following $m(2)$ are from regime 2, and the next $m(3)$ are from regime 1 again. However, the total number of switches M and $m(j)$ for $1 \leq j \leq M$ are not known.

Quandt (1972) introduces a λ-method assuming that nature chooses between regimes 1 and 2 with probabilities λ and $1 - \lambda$. The conditional density of y_t is

$$\begin{aligned} h(y_t|\boldsymbol{x}_t) &= \lambda f_1(y_t|\boldsymbol{x}_t) + (1-\lambda) f_2(y_t|\boldsymbol{x}_t) \\ &= \frac{\lambda}{\sqrt{2\pi}\sigma_1} \exp\left\{ -\frac{1}{2} \frac{(y_t - \boldsymbol{x}_t'\beta_1)^2}{\sigma_1^2} \right\} + \frac{1-\lambda}{\sqrt{2\pi}\sigma_2} \exp\left\{ -\frac{1}{2} \frac{(y_t - \boldsymbol{x}_t'\beta_2)^2}{\sigma_2^2} \right\} \end{aligned}$$

and therefore the likelihood function is

$$l = \sum_{t=1}^{n} \log h(y_t|\boldsymbol{x}_t),$$

which can be maximized with respect to $\boldsymbol{\beta}$'s, σ's, and λ.

The essence of this method is that the choice of regime 1 or 2 at stage t is a Bernoulli trial by nature and independent of what state the system was in on the previous trial. This assumption can be relaxed by introducing a transition matrix P, where P_{ij} being the probability that the system will make a transition from state i to state j. This interpretation makes the regime switching process governed by a Markov chain.

Goldfeld and Quandt (1973) propose a different τ-method adopting this Markov switching structure. Denote the probability that the system is initially in one or the other of the two regimes by $\lambda_0' = (\lambda_{1,0}, \lambda_{2,0})$ where $\lambda_{1,0} = 1 - \lambda_{2,0}$, and the probability at stage t is $\lambda_t' = (\lambda_{1,t}, \lambda_{2,t})$. Then $\lambda_t' = \lambda_{t-1}' A$ and $\lambda_t' = \lambda_0' P^t$. The conditional density is

$$h(y_t|\boldsymbol{x}_t) = \lambda_{1,t} f_1(y_t|\boldsymbol{x}_t) + (1 - \lambda_{1,t}) f_2(y_t|\boldsymbol{x}_t) = \lambda_t' f_t,$$

where $f_t = (f_1(y_t|\boldsymbol{x}_t), f_2(y_t|\boldsymbol{x}_t))'$. Hence the log likelihood function is

$$l = \sum_{t=1}^{n} \log(\lambda_t' f_t).$$

Let the transition matrix be

$$\begin{pmatrix} \tau_1 & 1-\tau_1 \\ 1-\tau_2 & \tau_2 \end{pmatrix}.$$

Define p_{ij}^t as the ijth element of P^t, then $p_{i1}^1 = \tau_1$, $p_{21}^1 = 1 - \tau_2$ and we have the difference equation system

$$p_{11}^t = \tau_1 p_{11}^{t-1} + (1-\tau_1) p_{21}^{t-1},$$

$$p_{21}^t = (1-\tau_2) p_{11}^{t-1} + \tau_2 p_{21}^{t-1}.$$

The general solution of the system is

$$\begin{pmatrix} p_{11}^t \\ p_{21}^t \end{pmatrix} = \frac{(\tau_1+\tau_2-1)^t}{\tau_1+\tau_2-2} \begin{pmatrix} \tau_1 - 1 \\ 1-\tau_2 \end{pmatrix} - \frac{1}{\tau_1+\tau_2-2} \begin{pmatrix} 1-\tau_2 \\ 1-\tau_2 \end{pmatrix}.$$

The substitution of $\lambda_t' = (\lambda_{1,0} p_{11}^t + \lambda_{2,0} p_{12}^t, \lambda_{1,0} p_{12}^t + \lambda_{2,0} p_{22}^t)$ in the likelihood function makes it a function of the $\boldsymbol{\beta}$'s, σ's, $\lambda_{1,0}$, τ_1 and τ_2.

Goldfeld and Quandt (1973) applied this method to a model for a housing market in disequilibrium proposed by Fair and Jaffee (1972) in which the demand function is

$$D_t = \boldsymbol{x}_t' \boldsymbol{\alpha}_t + \mu_t^D$$

and the supply function is

$$S_t = \boldsymbol{x}_t' \boldsymbol{\beta}_t + \mu_t^S,$$

where $y_t = \min\{D_t, S_t\}$ is the actually observed number of housing starts in month t and bx_t is a vector of independent variables. If there is an excess demand, the observed point lies on the supply function and if there is an excess demand, it lies on the demand function. Therefore this is a two-regime problem.

Cosslett and Lee (1985) suggest a recursive algorithm that is computationally tractable for the evaluation of the likelihood function. Hamilton (1989) extends the above Markov switching model to the analysis of gross national product (GNP) and business cycle. In his work, the parameters of an autoregression are viewed as the outcome of a discrete-state Markov process. Building upon ideas developed by Cosslett and Lee (1985), a nonlinear filter and smoother are presented for uncovering optimal statistical estimates of the state of the economy based on observations of output. An empirical application of this technique to postwar U.S. real GNP suggests that the periodic shift from a positive growth rate to a negative growth rate is a recurrent feature of the U.S. business cycle, and indeed could be used as an objective criterion for defining and measuring economic recessions. The estimated parameter values suggest that a typical economic recession is associated with a 3% permanent drop in the level of GNP.

The work of Hamilton (1989) is an early application of regime switching model in financial and economic analysis. Moreover, it helps to popularize the regime switching model. Hamilton (1990) further introduces a vector autoregression model subject to occasional discrete shifts, where a discrete-valued Markov process governs the shifts. It is observed that the usual numerical maximum of the likelihood functions is subject to computational difficulties associated with the often ill-behaved likelihood surface (multiple local maxima, essential singularities, and local increases as boundary conditions are approached). He suggests a numerically robust Expectation-Maximization (EM) algorithm to overcome the numerical difficulties. This algorithm permits potential application of the approach to large vector systems.

A lot of literature uses variations of the standard Markov regime switching model to describe the time series behavior of U.S. short-term interest rates. Cecchetti et al.(1993) specify aggregate consumption by a regime switching process to explain the first and second moments of the risk-free rate and the return to equity. Cai (1994), Hamilton and Susmel (1994) and So et al.(1998) generalize the usual ARCH/GARCH models by encompassing regime switching properties to model the stochastic volatilities. Garcia and Perron (1996) identify shifts in the time series of the U.S. real interest rate from 1961 to 1986, using a regime switching framework by allowing three possible regimes affecting both the mean and variance. Gray (1996) develops a generalized regime switching model of the short-term interest rate, allowing the short rate to exhibit both mean reversion and conditional heteroskedasticity and nesting the GARCH and square root process specifications. Dai et al. (2007) develop a dynamic term structure model with "priced" factor and regime-shift risks in which the shifts are governed by a discrete-time Markov process with state-dependent transition probabilities.

Furthermore, regime shifts in foreign exchange rates are also analyzed. Engel and Hamilton (1990) develop a new statistical model of exchange rate dynamics as a sequence of stochastic segmented time trends, and show that a regime switching model outperforms a random-walk model as a forecaster. Bekaert and Hodrick (1993) estimate a two-regime model and find that the variances of forward premiums are nine to ten times larger in the more volatile regime. Engel (1994) fits a Markov switching model for eighteen exchange rates at quarterly and monthly frequencies and shows some evidence that the forecast of the Markov model are superior at predicting the direction of change of the exchange rate. Bollen et al. (2000) estimate a Markov-switching model with two regimes for log exchange rate changes and their variances, but with mean and variance regimes allowed to switch independently, and show that the standard deviations of exchange rate returns are two to three times higher in the more volatile regime. Frömmel et al. (2005) extend the real interest differential (RID) model by introducing Markov regime switches for three exchange rates and show that the key fundamental which determines regimes is the interest rate. Fiess and Shankar (2009) apply regime switching methods to two simple indices of central bank exchange rate policy to generate likelihoods of high and low intervention and show strong evidence that the economy's balance sheet and economic performance determine the likelihood of switching in the exchange rate regime.

The regime switching model is also applied to asset pricing. Schwert (1989) models stock volatility using a two-regime model and analyzes the relation of stock volatility with real and nominal macroeconomic volatility, economic activity, financial leverage, and stock trading activity. Turner et al. (1989) examine a variety of models in which the variance of a portfolio's excess return depends on a state variable generated by a first-order Markov process, with the state both known and unknown. Cecchetti et al. (1990) consider a Lucas asset pricing model in which the economy's endowment switches between high economic growth and low economic growth, and such switching accounts for a number of features of stock market returns. Abel (1994) derives simple closed-form solutions for expected rates of return on stocks and riskless one-period bills under the assumption that shocks to the growth rates of consumption and dividends are generated by a Markov regime switching process. Abel (1999) incorporates this assumption to a general equilibrium model to analyze term premia and the risk premia. Veronesi (1999) presents a dynamic, rational expectations equilibrium model of asset prices where the drift of fundamentals (dividends) shifts between two unobservable states at random times and shows that in equilibrium, investors' willingness to hedge against changes in their own "uncertainty" on the true state makes stock prices overreact to bad news in good times and under react to good news in bad times. Whitelaw (2000) investigates the empirical contradiction that expected stock returns are weakly related to volatility at the market level in a general equilibrium exchange economy characterized by a regime switching consumption process with time-varying transition probabilities between regimes. Lettau et al. (2008) estimate a two-state regime switching model for the volatility and mean of consumption growth and show the relationship between the falling macroeconomic risk and the aggregate stock prices.

Asset allocation and portfolio decisions are also modeled in a regime switching framework. Ang and Bekaert (2002) solve the dynamic portfolio choice problem of a U.S. investor faced with a time-varying investment opportunity set modeled using a regime switching process which may be characterized by correlations and volatilities that increase in bad times. Guidolin and Timmermann (2006) study asset allocation decisions in the presence of regime switching in asset returns, using four separate regimes to capture the joint distribution of stock and bond returns, and identify the optimal asset allocations. Guidolin and Timmermann (2008) investigate the international asset allocation effects of time-variations in higher-order moments of stock returns such as skewness and kurtosis. Tu (2010) provides a Bayesian framework for making portfolio decisions that takes the regime switching between upturn and downturn into account, and reveal that the economic value of accounting for regimes is substantially independent of whether or not model and parameter uncertainties are incorporated.

1.2 A Classic Regime Switching Model

In this section we will restate the classic regime switching model based on Hamilton (1994).

Let $\boldsymbol{y}_t$ be an $(n \times 1)$ vector of observed endogenous variables and $\boldsymbol{x}_t$ a $(d \times 1)$ vector of observed exogenous variables. Let $\mathcal{F}_t = (\boldsymbol{y}'_t, \boldsymbol{y}'_{t-1}, ..., \boldsymbol{y}'_{-m}, \boldsymbol{x}'_t, \boldsymbol{x}'_{t-1}, ..., \boldsymbol{x}'_{-m})'$ be a vector containing all observations obtained through date t. If there are K different regimes, and the process is governed by regime $s_t = j$ at date t, then the conditional density of y_t is assumed to be given by

$$f(\boldsymbol{y}_t | s_t = j, \boldsymbol{x}_t, \mathcal{F}_{t-1}; \boldsymbol{\alpha}), \tag{1.2.1}$$

where $\boldsymbol{\alpha}$ is a vector of parameters characterizing the conditional density, and $j = 1, 2, ..., K$. These densities will be collected in an $(K \times 1)$ vector denoted $\boldsymbol{\eta}_t$. For example, consider a first-order autoregression in which both the constant term and the autoregressive coefficient might be different for different subsamples:

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

where $\epsilon_t \sim$ i.i.d.$N(0, \sigma^2)$. The proposal will be to model the regime s_t as the outcome of an unobserved K-state Markov chain with s_t independent of ϵ_τ for all t and τ. In this example, y_t is a scalar ($n = 1$), the exogenous variables consist only of a constant term ($\boldsymbol{x}_t = 1$), and the unknown parameters in $\boldsymbol{\alpha}$ consist of $c_1, ..., c_K, \phi_1, ..., \phi_K$ and σ^2. With $K = 2$ regimes the two densities represented by (1.2.1) are

$$\boldsymbol{\eta}_t = \begin{pmatrix} f(y_t|s_t = 1, y_{t-1}; \boldsymbol{\alpha}) \\ f(y_t|s_t = 2, y_{t-1}; \boldsymbol{\alpha}) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{-(y_t - c_1 - \phi_1 y_{t-1})^2}{2\sigma^2}\right\} \\ \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{-(y_t - c_2 - \phi_2 y_{t-1})^2}{2\sigma^2}\right\} \end{pmatrix} \tag{1.2.2}$$

It is assumed that s_t evolves according to a Markov chain that is independent of past observations on y_t or current or past x_t:

$$P(s_t = j|s_{t-1} = i, s_{t-2} = k, ..., \boldsymbol{x}_t, \mathcal{F}_{t-1}) = P(s_t = j|s_{t-1} = i) = p_{ij}. \tag{1.2.3}$$

The population parameters that describe a time series governed by (1.2.1) and (1.2.3) consist of $\boldsymbol{\alpha}$ and the various transition probabilities p_{ij}, which are collected in a vector $\boldsymbol{\theta}$.

Suppose $\boldsymbol{\theta}$ is known, let $P(s_t = j|\mathcal{F}_t; \boldsymbol{\theta})$ denote the inference on the value of s_t based on data obtained through date t and based on knowledge of the population parameter $\boldsymbol{\theta}$. Collect the conditional probabilities $P(s_t = j|\mathcal{F}_t; \boldsymbol{\theta})$ for $j = 1, ..., K$ in an $(K \times 1)$ vector denoted $\boldsymbol{\xi}_{t|t}$, and the forecasts $P(s_{t+1} = j|\mathcal{F}_t; \boldsymbol{\theta})$ in an $(K \times 1)$ vector denoted $\boldsymbol{\xi}_{t+1|t}$, then the optimal inference and forecast for each date t in the sample can be found by iterating on

$$\hat{\boldsymbol{\xi}}_{t|t} = \frac{\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t}{\mathbf{1}'\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t} \tag{1.2.4}$$

$$\hat{\boldsymbol{\xi}}_{t+1|t} = P \cdot \hat{\boldsymbol{\xi}}_{t|t}, \tag{1.2.5}$$

where P represents the $(K \times K)$ transition matrix defined in (1.2.3), $\mathbf{1}$ represents an $(K \times 1)$ vector of 1s, and the symbol $\odot$ denotes element-by-element multiplication. Given a starting value $\hat{\boldsymbol{\xi}}_{1|0}$ and an assumed value for the population parameter vector $\boldsymbol{\theta}$, one can iterate on (1.2.4) and (1.2.5) for $t = 1, ..., T$ to calculate the values of $\hat{\boldsymbol{\xi}}_{t|t}$ and $\hat{\boldsymbol{\xi}}_{t|t-1}$ for each date t in the sample. The log likelihood function $l(\boldsymbol{\theta})$ for the observed data $\mathcal{F}_T$ evaluated at the value of $\boldsymbol{\theta}$ that was used to perform the iterations can also be calculated as a by-product of this algorithm from

$$l(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log f(\boldsymbol{y}_t | \boldsymbol{x}_t, \mathcal{F}_{t-1}; \boldsymbol{\theta}) = \sum_{t=1}^{T} \log(\mathbf{1}'\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t). \tag{1.2.6}$$

The proof of (1.2.4) to (1.2.6) can be found on Page 693 in Hamilton (1994).

Let $\hat{\boldsymbol{\xi}}_{t|\tau}$ represent the $(K \times 1)$ vector whose jth element is $P(s_t = j | \mathcal{F}_\tau; \boldsymbol{\theta})$. For $t > \tau$, this represents a forecast about the regime for some future period; for $t < \tau$, this represents the smoothed inference about the regime the process was in at date t based on data obtained through some later date τ. The optimal m-period-ahead forecast of $\boldsymbol{\xi}_{t+m}$ is

$$\hat{\boldsymbol{\xi}}_{t+m|t} = P^m \cdot \hat{\boldsymbol{\xi}}_{t|t},$$

and the smoothed inferences can be calculated using an algorithm developed by Kim (1994) as

$$\hat{\boldsymbol{\xi}}_{t|T} = \hat{\boldsymbol{\xi}}_{t|t} \odot \{P'[\hat{\boldsymbol{\xi}}_{t+1|T} (\div) \hat{\boldsymbol{\xi}}_{t+1|t}]\}, \tag{1.2.7}$$

where the sign $(\div)$ denotes element-by-element division. The smoothed probabilities $\hat{\boldsymbol{\xi}}_{t|T}$ are found by iteration on (1.2.7) backward for $t = T-1, ..., 1$. This iteration is started with $\hat{\boldsymbol{\xi}}_{T|T}$, which is obtained from (1.2.4) for $t = T$. This algorithm is valid only when s_t follows a first-order Markov chain as in (1.2.3), when the conditional density (1.2.1) depends on $s_t, s_{t-1}, ...$ only through the current state s_t, and when $\boldsymbol{x}_t$, the vector of explanatory variables other than the lagged values of $\boldsymbol{y}$, is strictly exogenous, meaning that $\boldsymbol{x}_t$ is independent of s_τ for all t and τ.

In the iteration on (1.2.4) and (1.2.5), the parameter vector $\boldsymbol{\theta}$ was taken to be a fixed, known vector. The parameters can be estimated by an EM algorithm. Once the iteration has been completed for $t = 1, ..., T$ for a given fixed $\boldsymbol{\theta}$, the value of the log likelihood implied by that value of $\boldsymbol{\theta}$ is then known from (1.2.6). The value of $\boldsymbol{\theta}$ that maximized the log likelihood can be found numerically. It is shown in Hamilton (1990) that the maximum likelihood estimates for the transition probabilities satisfy

$$\hat{p}_{ij} = \frac{\sum_{t=1}^{T} P(s_t = j, s_{t-1} = i|\mathcal{F}_T; \hat{\boldsymbol{\theta}})}{\sum_{t=1}^{T} P(s_{t-1} = i|\mathcal{F}_T; \hat{\boldsymbol{\theta}})}, \tag{1.2.8}$$

where $\hat{\boldsymbol{\theta}}$ denotes the full vector of maximum likelihood estimates. The maximum likelihood estimate of the vector $\boldsymbol{\alpha}$ that governs the conditional density (1.2.1) is characterized by

$$\sum_{t=1}^{T} \left(\frac{\partial \log \boldsymbol{\eta}_t}{\partial \boldsymbol{\alpha}'} \right)' \hat{\boldsymbol{\xi}}_{t|T} = 0. \tag{1.2.9}$$

The details can be found in Hamilton (1990).

The above approach readily generalizes to processes in which the probabilities of $s_t = j$ depend not only on the value of s_{t-1} but also on lagged values of $\boldsymbol{y}_t$ or strictly exogenous explanatory variables, as in Diebold et al.(1994), Filardo (1994), and Peria (2002). However, often there are relatively few transitions among regimes, making it difficult to estimate such parameters accurately, and most applications have assumed a time-invariant Markov chain. For the same reason, most applications assume only $K = 2$ or 3 different regimes.

Several tests have been developed for identifying the number of regimes. Most of them are based on the likelihood ratio (LR) technique. Hansen (1992) revisits the model of Hamilton (1989) and uses empirical process theory to bound the asymptotic distribution of the LR test statistic. His method sets a grid over different values of the transition probabilities and the parameters describing the second state. The constrained likelihood is optimized with respect to the nuisance parameters at each point of the grid. This method gives a bound for the likelihood ratio test statistic and not a critical value. Hamilton (1996) applies the score function technique for different tests of model misspecification. As he has mentioned, conducting proper inference is regime switching models is especially challenging. When formulated in the natural way, testing the null hypothesis that there is a single regime versus the alternative of two regimes can involve a nuisance parameter not identified under the null hypothesis. Thus standard LR tests, the related Lagrange multiplier and Wald tests cannot be used in the usual manner. Garcia (1998) reviews Hansen's problem. He treats the transition probabilities as nuisance parameters and derives the asymptotic distribution of the Sup LR test in terms of the remaining parameters of the model and uses Monte Carlo experiments to show that these asymptotic distributions are very good approximations to their empirical counterparts. Cho and White (2007) include the boundary of the parameter space when developing their limit theory. But their method needs to specify a parameter space for the coefficients that vary over regimes. So far, there is no satisfying test for the number of regimes.

In a Bayesian approach, both the parameters $\boldsymbol{\theta}$ and the values of the states $\boldsymbol{s} = (s_1, ..., s_T)'$ are viewed as random variables. A particular Markov Chain Monte Carlo method, the Gibbs sampler, is widely applicable to obtaining the marginal posterior distributions of interest. The method , in the ith iteration, involves generating a realization $\boldsymbol{\theta}(i)$ from the posterior distribution of $\boldsymbol{\theta}|\{\mathcal{F}_T, \boldsymbol{s}(i-1)\}$, and generating $\boldsymbol{s}(i)$ from the posterior distribution of $\boldsymbol{s}|\{\boldsymbol{\theta}(i), \mathcal{F}_T\}$. An algorithm for generating a draw from the second distribution, $\boldsymbol{s}|\{\boldsymbol{\theta}(i), \mathcal{F}_T\}$, was developed by Albert and Chib (1993).

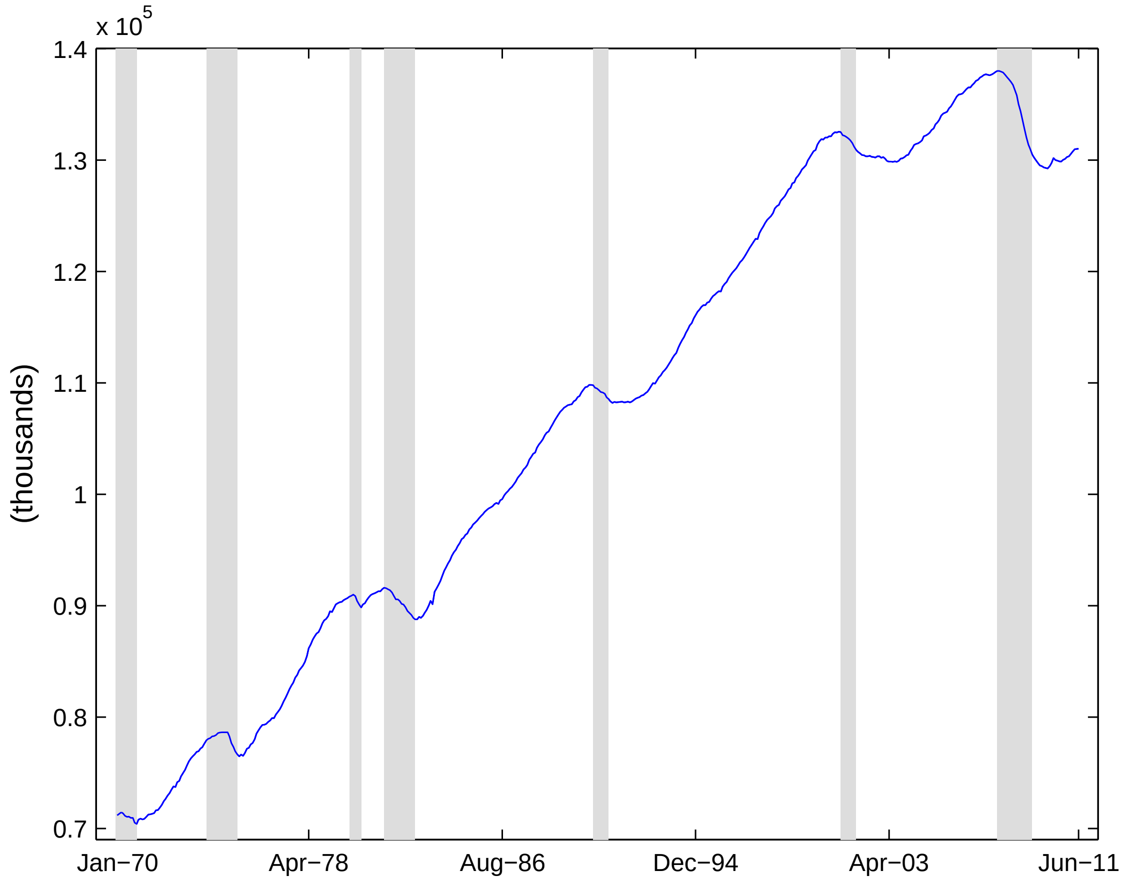
1.3 A Motivating Question

Although the assumption of a finite (and small) set of regimes in the classic regime switching model seems adequate for many applications, it is too restrictive for many econometric time series that undergo structural changes occasionally so that the model parameters at different periods have different values, even though they belong to the same regime.

For instance, to describe business cycles, one tends to classify qualitatively the economic states as two regimes ("expansion" and "recession"); that is, the model parameters are represented by a two state hidden Markov chain. Similarly, when modeling the stock market, one tends to decompose market fluctuations into "bull" and "bear" markets. The classic regime switching models assume that the model parameters take two values, corresponding to two states. However, even in the same state, the model parameters at different periods might not necessarily be the same.

To illustrate the motivation of this dissertation study, we show a real data example. We use "All Employees: Total nonfarm (PAYEMS)" available from the website of Federal Reserve Bank of St. Louis at `http://research.stlouisfed.org/fred2/series/PAYEMS/`. This data comes from U.S. Department of Labor: Bureau of Labor Statistics. The nonfarm payrolls figure is an extremely important coincident indicator. It is the benchmark labor statistic used to determine the health of the job market because of its large sample size and historical significance in relation to predicting business cycles accurately. We use the monthly data (in thousands) from January 1970 to June 2011. Figure 1.1 shows the

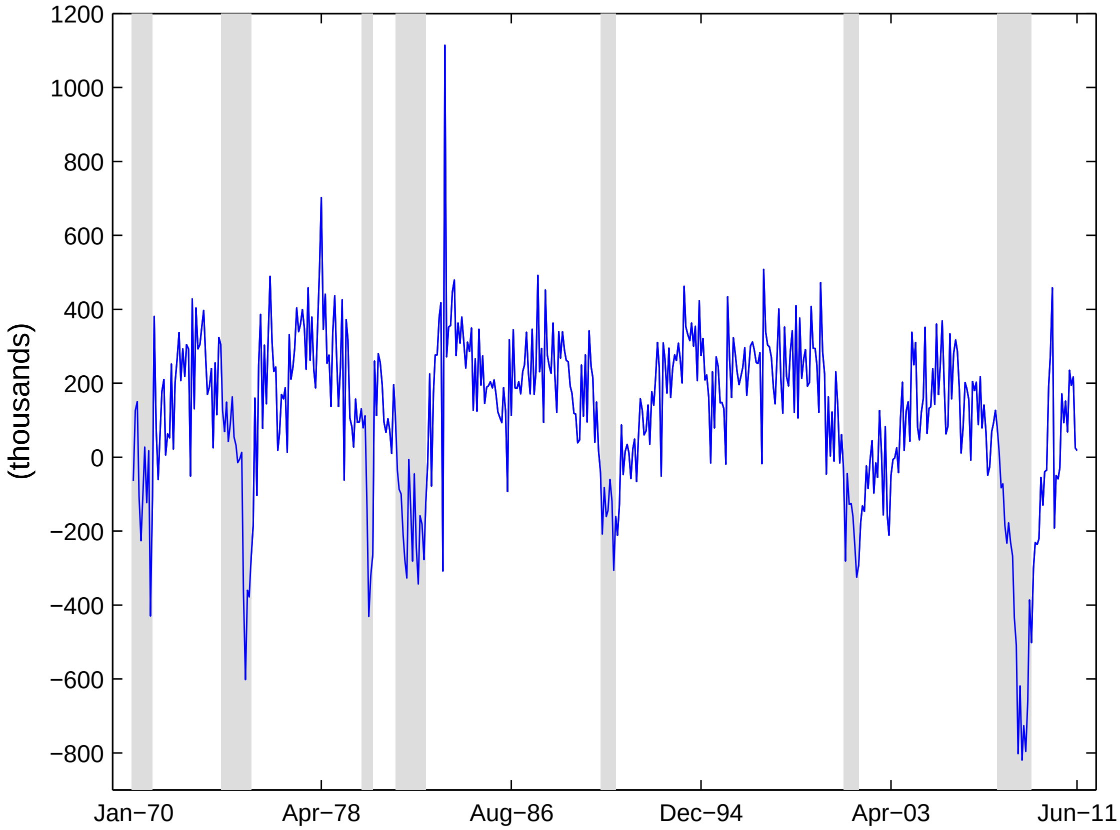
Figure 1.1: PAYEMS series: In thousands. NBER recessions are shown as shaded areas.



original series. The series shows an increasing trend over time. The shaded areas denote the recessions identified by NBER. Not surprisingly, each recession is overlapped by a decrease in the PAYEMS series, indicating a rising unemployment rate during a recession. After each recession, the employment situation begins to recover to the pre-slump level.

To make the series stationary, we use the change in PAYEMS (in thousands) which is calculated as the differenced series $y_t = y'_t - y'_{t-1}$, where y'_t is the original PAYEMS series. The series is shown in Figure 1.2. It can be seen that during each NBER recession the differenced series reaches a local minimum and then rebounds. This series is stationary. The p-value of an augmented Dickey-Fuller test is less than 0.01 for the null hypothesis that the series is non-stationary.

Figure 1.2: PAYEMS series: Change in thousands. NBER recessions are shown as shaded areas.



We assume a classic regime switching autogressive model

$$y_t = \alpha_t + \beta_t y_{t-1} + \epsilon_t, \tag{1.3.1}$$

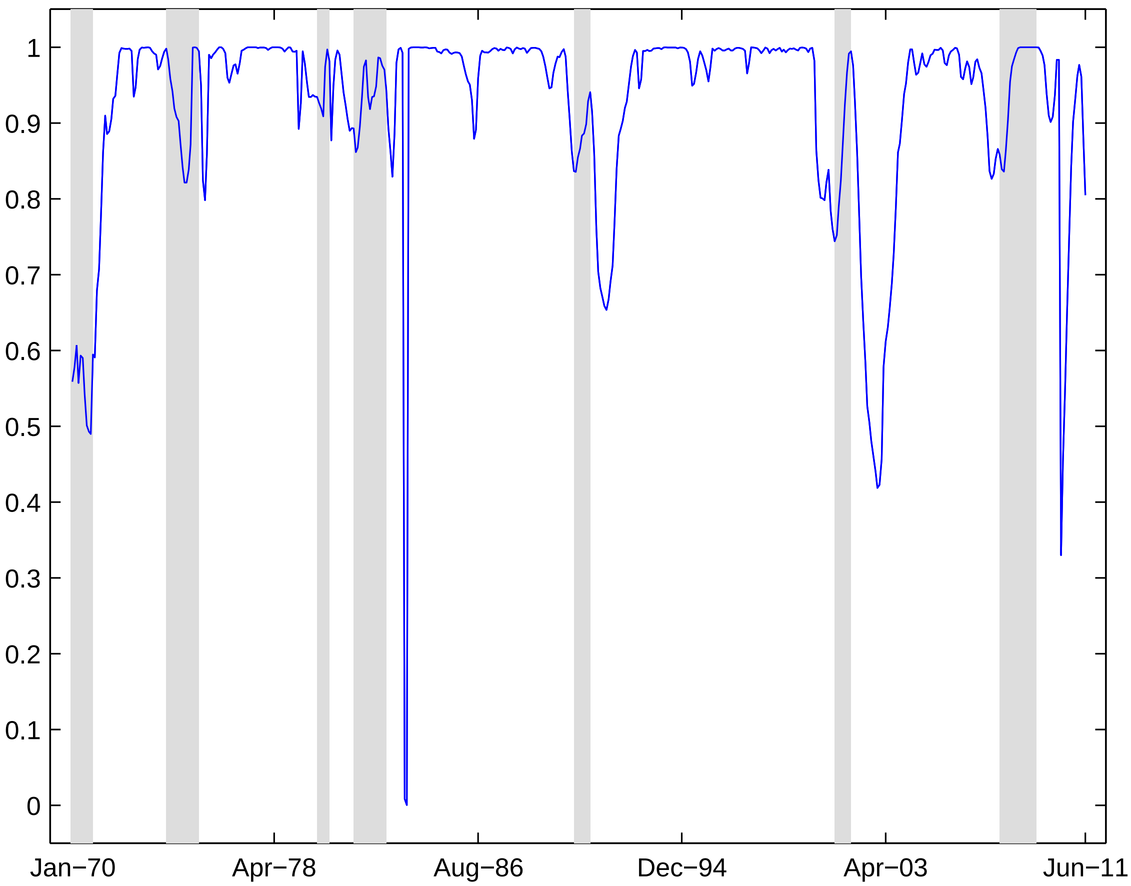
where $\alpha_t = \alpha(s_t)$ and $\beta_t = \beta(s_t)$ are changing from regime to regime. But within each regime, both coefficients are constant. The random variable ϵ_t is normally distributed with mean zero and variance σ^2. To implement the classic regime switching model, we use a Matlab package written by Marcelo Perlin (`http://www.mathworks.com/matlabcentral/fileexchange/15789-msregress-a-package-for-markov-regime-switching-models-in-matlab`). As mentioned in Perlin (2010), the package estimates the classic regime switching model with a two-step approximation, and it does not adopt the EM algorithm to estimate the hyperparameters. We assume that there are two regimes, $K = 2$. The estimated probability of $P(s_t = 1)$ is shown in Figure 1.3. The recessions identified by NBER are shown as shaded areas in the figure. For most of time, the identified regime is 1. There are some transitions from regime 1 to regime 2. Most of them occur after a NBER recession. To test whether the coefficients are constant within each regime, we pick two subperiods, August 1986 to May 1990 and December 1994 to December 1999, during both of which the estimated probabilities $P(s_t = 1)$ are greater than 0.9. We estimate a regular AR(1) model $y_t = \alpha + \beta y_{t-1} + \epsilon_t$ over each subperiod. The estimated α and β are 223.02 and 0.01 respectively for the first subperiod. The estimated α and β are 275.38 and -0.14 respectively for the second subperiod. The significant difference indicates the possibility that within each regime, the coefficients are not constant. This is the issue we want to address in this dissertation. In Section 2.1, a new stochastic regime switching model will be proposed in which within each regime the coefficient is defined as a random variable instead of a constant, following some regime-specific distribution.

Since we do not use the exact method proposed by Hamilton (1989) to estimate the classic regime switching model, it is possible that the approximating algorithm utilized in the Matlab package is not accurate in identifying the regimes. In Section 3.2.2, we will apply the model we propose in Chapter 2 to the same PAYEMS series to compare the results with Figure 1.3.

1.4 Outline

This dissertation research is motivated by the concerns mentioned above. It studies the estimation of parameters in a stochastic regime switching model, exploring its applications to some econometric time series. In Chapter 2, we propose the stochastic regime switching model with the associated smoothing estimates of parameters and inference on regimes.

Figure 1.3: PAYEMS: Estimates of $P(s_t = 1)$ in a classic regime switching autoregressive model. NBER recessions are shown as shaded areas.



The proposed model uses a Bayesian framework, and hence contains certain hyperparameters. Their estimation is considered. Furthermore, to improve the computational efficiency of the estimation, a bounded complexity mixture approximation (BCMIX) is considered in Chapter 3 where other implementation details are also discussed. In Chapter 4 we conduct extensive numerical simulation studies to test the accuracy and efficiency of our proposed Bayes estimate and BCMIX estimate. After that, our model is applied to analyze "Total Financial Assets - Assets - Balance Sheet of Nonfarm Nonfinancial Corporate Business (TFAABSNNCB)" series and "All Employees: Total nonfarm (PAYEMS)" series to illustrate its usefulness. Some concluding remarks are given in Chapter 5.

Chapter 2

Estimation in a Stochastic Regime Switching Model

2.1 A Stochastic Regime Switching Model

The classic regime switching model described in Section 1.2 can be used to model business cycles. For instance, economic expansion and recession can be described via a two-state Markov chain. However, in practice, the values of the unknown parameters in $\boldsymbol{\alpha}$ during two different expansion (or recession) periods are not necessary the same. In this section, we will consider a stochastic regime switching model to incorporate this feature.

Assume the observations $\{y_t\}$, $t = 1, ..., T$ follow the stochastic regression model

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta}_t + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma^2), \tag{2.1.1}$$

where $\boldsymbol{x}_t$ is the $(d \times 1)$ stochastic regressor consisting of the historical observations and exogenous variables. So in our model, the lagged variable y_{t-1} is included in $\boldsymbol{x}_t$. The values of the piecewise constant parameter $\boldsymbol{\beta}_t$ depend on a hidden state s_t which satisfy the following assumptions:

(A1) The Markov chain $\{s_t \in \{1, \ldots, K\} | t \geq 0\}$ is irreducible, and follows the transition probability matrix $P = (p_{ij})_{1 \leq i,j \leq K}$; i.e., $p_{ij} = P\{s_t = j | s_{t-1} = i\}$.

(A2) The Markov chain $\{s_t, t \geq 0\}$ has a stationary distribution $\pi = (\pi_1, \ldots, \pi_K)'$.

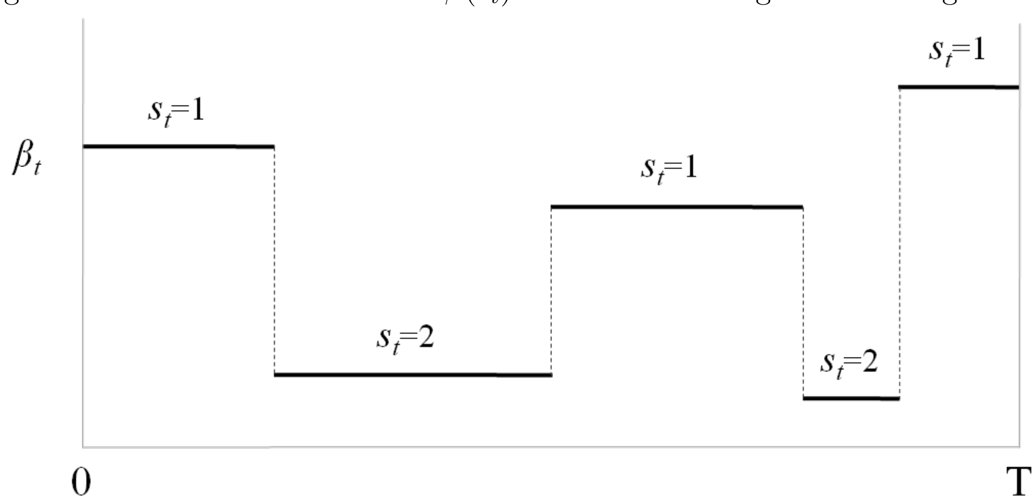
(A3) Define $s_0 \neq s_1$. $\boldsymbol{\beta}_t = \mathbf{1}_{\{s_t = s_{t-1}\}} \boldsymbol{\beta}_{t-1} + \mathbf{1}_{\{s_t = k \neq s_{t-1}\}} \boldsymbol{z}_t$, in which $\boldsymbol{z}_t$ are independent and identically distributed random variables and follow

$$\boldsymbol{z}_t \sim N(\boldsymbol{z}(k), \boldsymbol{V}(k)), \tag{2.1.2}$$

Note that with the assumption (A2), $\{\boldsymbol{\beta}_t\}$ has a stationary distribution and hence a reversible Markov chain can be defined. The classic regime switching model described in Section 1.2 does not satisfy (A3). Instead, the classic regime switching model assumes that given s_t, $\boldsymbol{\beta}_t = \boldsymbol{\beta}(s_t)$, a constant depending on the regime. (A3) adds the new feature to our model by assuming the parameter in each regime is a random variable following a certain distribution.

Assume there are two regimes; that is, $K = 2$. Figure 2.1 visually demonstrates the assumptions, showing an example of possible values of a one-dimensional β_t. Four transitions occur during the period $0 \leq t \leq T$. Within each regime, β_t take different values. The values are realizations from the regime-specific distribution of $\beta(s_t)$. The transitions are governed by some hidden Markov chain.

Figure 2.1: Illustration: Values of $\beta(s_t)$ in a stochastic regime switching model.

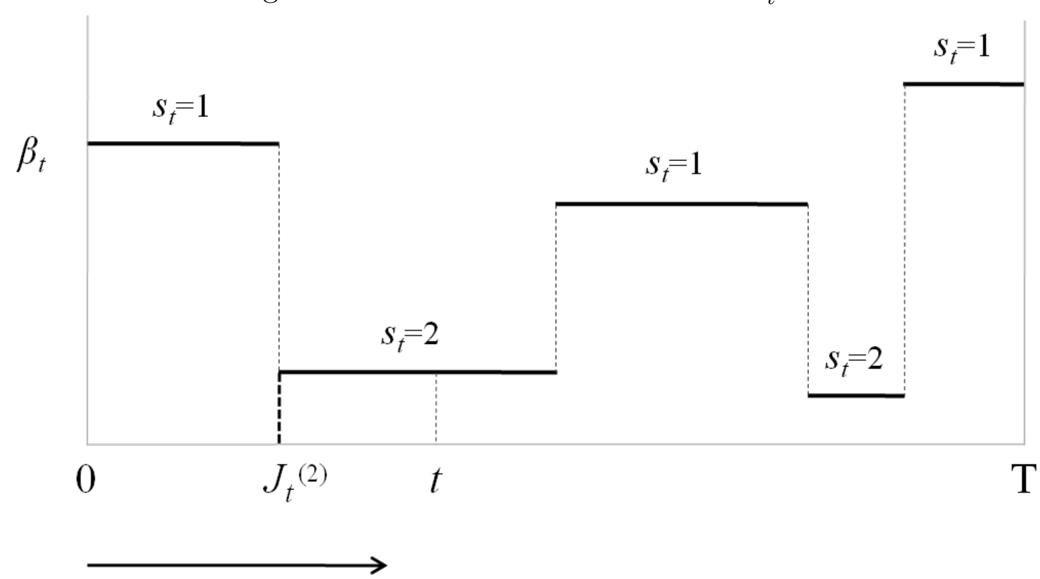


The model is motivated by Lai, Liu and Xing (2005), Lai, Xing and Zhang (2008), Lai and Xing (2011) in which regression models with piecewise constant parameters are studied. In this dissertation, we combine the Bayesian approach developed in their works with regime switching framework. In the following sections, we will derive the filtering and smoothing estimates of parameters following Lai and Xing (2011).

2.2 The Forward Filtering Estimate of Parameters

First let us see the forward filtering estimate of $\boldsymbol{\beta}_t$; that is, the estimate of $\boldsymbol{\beta}_t$ for any time t given all the historical information from the beginning to t. Let $\boldsymbol{y}_{ij} = (\boldsymbol{y}_i, ..., \boldsymbol{y}_j)$, $\boldsymbol{x}_{ij} = (\boldsymbol{x}_i, ..., \boldsymbol{x}_j)$, $\mathcal{F}_t = (\boldsymbol{x}_{1t}, \boldsymbol{y}_{1t})$ and $\mathcal{F}_{ij} = (\boldsymbol{x}_{ij}, \boldsymbol{y}_{ij})$. Let $J_t^{(k)} = \max\{i \leq t : s_{i-1} \neq s_i = \cdots = s_t = k\}$ be the most recent switching time less than or equal to t when s_t switches from another regime to regime k. Figure 2.2 illustrates the definition of $J_t^{(k)}$. At time t, $s_t = 2$, and the most recent switching occurs before t is at time $J_t^{(2)}$ as shown in the figure.

Figure 2.2: Illustration: Definition of $J_t^{(k)}$.



Define

$$\xi_t^{(k)} = P(s_t = k|\mathcal{F}_t), \qquad \xi_{i,t}^{(k)} = P(J_t^{(k)} = i|\mathcal{F}_t)$$

for $1 \le i \le t$ and $1 \le k \le K$. The quantity $\xi_t^{(k)}$ is the conditional probability that the current regime is k, $\xi_{i,t}^{(k)}$ is the conditional probability that the current regime is k and the recent transition occurs at time i. Thus $\xi_t^{(k)} = \sum_{i=1}^t P(J_t^{(k)}\mathcal{F}_t) = \sum_{i=1}^t \xi_{i,t}^{(k)}$. If we know all the historical information up to time t, $\mathcal{F}_t$, and that the recent transition occurs at time i from some regime to regime k, we just need to use the information after this transition to estimate the current value of $\boldsymbol{\beta}_t$.

Since $y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_t + \epsilon_t$, where $\epsilon_t \sim N(0, \sigma^2)$, and $\boldsymbol{\beta}_t \sim N(\boldsymbol{z}(k), \boldsymbol{V}(k))$. The posterior distribution of $\boldsymbol{\beta}_t$ given $\mathcal{F}_{it}$ is:

$$\begin{aligned}
f(\boldsymbol{\beta}_t|\mathcal{F}_{it}) &\propto \prod_{j=i}^{t} f(y_t|\boldsymbol{\beta}_t)\cdot f(\boldsymbol{\beta}_t)\\
&\propto \prod_{j=i}^{t} \exp(-\frac{(y_j - \boldsymbol{x}_j'\boldsymbol{\beta}_t)^2}{2\sigma^2})\cdot \exp(\frac{1}{2}(\boldsymbol{\beta}_t - \boldsymbol{z}(k))'\boldsymbol{V}(k)^{-1}(\boldsymbol{\beta}_t - \boldsymbol{z}(k)))\\
&\propto \big(\boldsymbol{\beta}_t - \boldsymbol{z}_{i,t}^{(k)}\big)'\big(\boldsymbol{V}_{i,t}^{(k)}\big)^{-1}\big(\boldsymbol{\beta}_t - \boldsymbol{z}_{i,t}^{(k)}\big),
\end{aligned}$$

where

$$\boldsymbol{V}_{i,t}^{(k)} = \big(\boldsymbol{V}(k)^{-1} + \frac{\sum_{j=i}^t \boldsymbol{x}_j\boldsymbol{x}_j'}{\sigma^2}\big)^{-1}, \qquad \boldsymbol{z}_{i,t}^{(k)} = \boldsymbol{V}_{i,j}\big(\boldsymbol{V}(k)^{-1}\boldsymbol{z}(k) + \frac{\sum_{j=i}^t \boldsymbol{x}_j y_j}{\sigma^2}\big).$$

Thus the conditional distribution of $\boldsymbol{\beta}_t$ is $g_{i,t}^{(k)}(\boldsymbol{\beta}_t)$ which is defined as

$$\boldsymbol{\beta}_t|\{\mathcal{F}_t, J_t^{(k)} = i\} \sim N(\boldsymbol{z}_{i,t}^{(k)}, \boldsymbol{V}_{i,t}^{(k)}), \tag{2.2.1}$$

If no historical information is given and only the event $s_t = k$ is known, the conditional distribution of $\boldsymbol{\beta}_t$ is $g_{0,0}^{(k)}(\boldsymbol{\beta}_t)$ which is defined as

$$\boldsymbol{\beta}_t|\{s_t = k\} \sim N(\boldsymbol{z}(k), \boldsymbol{V}(k)). \tag{2.2.2}$$

It follows that the posterior distribution of $\boldsymbol{\beta}_t$ given $\mathcal{F}_t$ is a mixture of normal distributions:

$$\begin{aligned}\boldsymbol{\beta}_t|\mathcal{F}_t &\sim \sum_{k=1}^{K}\sum_{i=1}^{t} P(J_t^{(k)} = i|\mathcal{F}_t) f(\boldsymbol{\beta}_t|\mathcal{F}_t, J_t^{(k)} = i) \\ &= \sum_{k=1}^{K}\sum_{i=1}^{t} \xi_{i,t}^{(k)} g_{i,t}^{(k)}(\boldsymbol{\beta}_t).\end{aligned} \tag{2.2.3}$$

Let us see how to derive the mixture weight $\xi_{i,t}^{(k)}$. First note that

$$f(\boldsymbol{\beta}_t, y_t, s_{t-1} = k|\mathcal{F}_{t-1}) = \sum_{l=1}^{K} f(\boldsymbol{\beta}_t, y_t, s_{t-1} = k, s_t = l|\mathcal{F}_{t-1}).$$

When $l \neq k$,

$$\begin{aligned}&f(\boldsymbol{\beta}_t, y_t, s_{t-1} = k, s_t = l|\mathcal{F}_{t-1}) \\ =&f(\boldsymbol{\beta}_t, y_t|\mathcal{F}_{t-1}, s_{t-1} = k, s_t = l)P(s_{t-1} = k, s_t = l|\mathcal{F}_{t-1}) \\ =&f(y_t|\mathcal{F}_{t-1}, J_t^{(l)} = t) f(\boldsymbol{\beta}_t|\mathcal{F}_t, J_t^{(l)} = t) P(s_t = l|s_{t-1} = k) P(s_{t-1} = k|\mathcal{F}_{t-1}) \\ =&f(y_t|\mathcal{F}_{t-1}, J_t^{(l)} = t) g_{t,t}^{(l)}(\boldsymbol{\beta}_t) p_{k,l} \xi_{t-1}^{(k)}.\end{aligned}$$

When $l = k$,

$$f(\boldsymbol{\beta}_t, y_t, s_{t-1} = k, s_t = k|\mathcal{F}_{t-1}) = \sum_{i=1}^{t-1} f(J_t^{(k)} = i, \boldsymbol{\beta}_t, y_t|\mathcal{F}_{t-1})$$

$$= \sum_{i=1}^{t-1} f(\boldsymbol{\beta}_t, y_t|\mathcal{F}_{t-1}, J_t^{(k)} = i) P(s_{t-1} = k, s_t = k|\mathcal{F}_{t-1})$$

$$= \sum_{i=1}^{t-1} f(y_t|\mathcal{F}_{t-1}, J_t^{(k)} = i) f(\boldsymbol{\beta}_t|\mathcal{F}_t, J_t^{(k)} = i) P(s_t = k|s_{t-1} = k) P(s_{t-1} = k|\mathcal{F}_{t-1})$$

$$= \sum_{i=1}^{t-1} f(y_t|\mathcal{F}_{t-1}, J_t^{(l)} = t) g_{i,t}^{(k)}(\boldsymbol{\beta}_t) p_{k,k} \xi_{i,t-1}^{(k)}.$$

Define

$$\xi_{i,t}^{(k)*} = \begin{cases} \left(\sum_{l \neq k} \xi_{t-1}^{(l)} p_{lk}\right) f(y_t|J_t^{(k)} = t) & i = t, \\ p_{kk} \xi_{i,t-1}^{(k)} f(y_t|\mathcal{F}_{t-1}, J_t^{(k)} = i) & i < t, \end{cases}$$

then

$$f(\boldsymbol{\beta}_t, y_t, s_{t-1} = k|\mathcal{F}_{t-1})$$

$$= \sum_{k \neq l} f(y_t|\mathcal{F}_{t-1}, J_t^{(l)} = t) g_{t,t}^{(l)}(\boldsymbol{\beta}_t) p_{k,l} \xi_{t-1}^{(k)} + \sum_{i=1}^{t-1} f(y_t|\mathcal{F}_{t-1}, J_t^{(l)} = t) g_{i,t}^{(k)}(\boldsymbol{\beta}_t) p_{k,k} \xi_{i,t-1}^{(k)}$$

$$= \xi_{t,t}^{(k)*} g_{t,t}^{(k)}(\boldsymbol{\beta}_t) + \sum_{i=1}^{t-1} \xi_{i,t}^{(k)*} g_{i,t}^{(k)}(\boldsymbol{\beta}_t).$$

Thus

$$f(\boldsymbol{\beta}_t|\mathcal{F}_t) \propto \sum_{k=1}^{K} f(\boldsymbol{\beta}_t, y_t, s_{t-1} = k|\mathcal{F}_{t-1})$$

$$= \sum_{k=1}^{K} \xi_{t,t}^{(k)*} g_{t,t}^{(k)}(\boldsymbol{\beta}_t) + \sum_{k=1}^{K} \sum_{i=1}^{t-1} \xi_{i,t}^{(k)*} g_{i,t}^{(k)}(\boldsymbol{\beta}_t).$$

So the mixture weight $\xi_{i,t}^{(k)}$ is the conditional probability which can be determined by the recursions

$$\xi_{i,t}^{(k)} \propto \xi_{i,t}^{(k)*} := \begin{cases} \big(\sum_{l\neq k} \xi_{t-1}^{(l)} p_{lk}\big) f(y_t | J_t^{(k)} = t) & i = t, \\ p_{kk}\xi_{i,t-1}^{(k)} f(y_t|\mathcal{F}_{t-1}, J_t^{(k)} = i) & i < t. \end{cases} \tag{2.2.4}$$

Define

$$\begin{aligned} f_{0,0}^{(k)} &= |\boldsymbol{V}(k)|^{1/2} \exp\left\{\frac{1}{2}\boldsymbol{z}(k)'\boldsymbol{V}(k)^{-1}\boldsymbol{z}(k)\right\}, \\ f_{i,j}^{(k)} &= |\boldsymbol{V}_{ij}^{(k)}|^{1/2} \exp\left\{\frac{1}{2}\boldsymbol{z}_{ij}^{(k)\prime}(\boldsymbol{V}_{ij}^{(k)})^{-1}\boldsymbol{z}_{ij(k)}\right\}. \end{aligned} \tag{2.2.5}$$

Let us use $f_{0,0}^{(k)}$ and $f_{i,j}^{(k)}$ to present the posterior densities $f(y_t|J_t^{(k)} = t)$ and $f(y_t|\mathcal{F}_{t-1}, J_t^{(k)} = i)$. For the first posterior density $f(y_t|J_t^{(k)} = t)$, note that

$$f(y_t|J_t^{(k)} = t) = \int f(y_t|\boldsymbol{\beta}_t, J_t^{(k)} = t) f(\boldsymbol{\beta}_t|J_t^{(k)} = t) d\boldsymbol{\beta}_t.$$

Using $\phi_{\boldsymbol{z},\boldsymbol{V}}(\boldsymbol{\beta})$ to denote the density function of a normal distribution with mean $\boldsymbol{z}$ and variance $\boldsymbol{V}$; that is, $\phi_{\boldsymbol{z},\boldsymbol{V}}(\boldsymbol{\beta}) = ((2\pi)^d|\boldsymbol{V}|)^{-1/2}\exp\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{z})'\boldsymbol{V}^{-1}(\boldsymbol{\beta}-\boldsymbol{z})\}$, where d is the dimension of $\boldsymbol{\beta}$. Note that

$$\begin{aligned} & f(y_t|\boldsymbol{\beta}_t, J_t^{(k)} = t) f(\boldsymbol{\beta}_t|J_t^{(k)} = t) \\ =& \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big\{ - \frac{(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_t)^2}{2\sigma^2}\Big\} ((2\pi)^d |\boldsymbol{V}(k)|)^{-\frac{1}{2}} \exp\Big\{ - \frac{1}{2}(\boldsymbol{\beta}_t - \boldsymbol{z}(k))'\boldsymbol{V}(k)^{-1}(\boldsymbol{\beta}_t - \boldsymbol{z}(k))\Big\} \\ =& \frac{1}{\sqrt{2\pi\sigma^2}} ((2\pi)^d |\boldsymbol{V}(k)|)^{-\frac{1}{2}} \exp\Big\{ - \frac{(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_t)^2}{2\sigma^2} + -\frac{1}{2}(\boldsymbol{\beta}_t - \boldsymbol{z}(k))'\boldsymbol{V}(k)^{-1}(\boldsymbol{\beta}_t - \boldsymbol{z}(k))\Big\} \\ =& \frac{1}{\sqrt{2\pi\sigma^2}} ((2\pi)^d |\boldsymbol{V}(k)|)^{-\frac{1}{2}} \exp\Big\{ - \frac{1}{2}(\boldsymbol{\beta}_t - \widetilde{\boldsymbol{z}})'\widetilde{\boldsymbol{V}}^{-1}(\boldsymbol{\beta}_t - \widetilde{\boldsymbol{z}}) \\ & \qquad - \frac{1}{2}\boldsymbol{z}(k)'\boldsymbol{V}(k)^{-1}\boldsymbol{z}(k) - \frac{y_t^2}{2\sigma^2} + \frac{1}{2}\widetilde{\boldsymbol{z}}'\widetilde{\boldsymbol{V}}^{-1}\widetilde{\boldsymbol{z}}\Big\}, \end{aligned}$$

where

$$\begin{aligned} \widetilde{\boldsymbol{V}} =& \big(\boldsymbol{V}(k)^{-1} + \frac{\boldsymbol{x}_t\boldsymbol{x}_t'}{\sigma^2}\big)^{-1} = \boldsymbol{V}_{t,t}^{(k)}, \\ \widetilde{\boldsymbol{z}} =& \widetilde{\boldsymbol{V}}\big(\boldsymbol{V}(k)^{-1}\boldsymbol{z}(k) + \frac{\boldsymbol{x}_t y_t}{\sigma^2}\big) = \boldsymbol{z}_{t,t}^{(k)}. \end{aligned}$$

Thus

$$f(y_t|\boldsymbol{\beta}_t, J_t^{(k)}=t)f(\boldsymbol{\beta}_t|J_t^{(k)}=t) = \frac{\phi_{\boldsymbol{z}_{t,t}^{(k)},\boldsymbol{V}_{t,t}^{(k)}}(\boldsymbol{\beta}_t)\phi_{\boldsymbol{z}(k),\boldsymbol{V}(k)}(\mathbf{0})\phi_{0,\sigma^2}(y_t)}{\phi_{\boldsymbol{z}_{t,t}^{(k)},\boldsymbol{V}_{t,t}^{(k)}}(\mathbf{0})}.$$

Therefore

$$\begin{aligned}f(y_t|J_t^{(k)}=t) &= \int \frac{\phi_{\boldsymbol{z}_{t,t}^{(k)},\boldsymbol{V}_{t,t}^{(k)}}(\boldsymbol{\beta}_t)\phi_{\boldsymbol{z}(k),\boldsymbol{V}(k)}(\mathbf{0})\phi_{0,\sigma^2}(y_t)}{\phi_{\boldsymbol{z}_{t,t}^{(k)},\boldsymbol{V}_{t,t}^{(k)}}(\mathbf{0})}d\boldsymbol{\beta}_t \\ =&\frac{\phi_{\boldsymbol{z}(k),\boldsymbol{V}(k)}(\mathbf{0})\phi_{0,\sigma^2}(y_t)}{\phi_{\boldsymbol{z}_{t,t}^{(k)},\boldsymbol{V}_{t,t}^{(k)}}(\mathbf{0})} = \frac{f_{t,t}^{(k)}}{f_{0,0}^{(k)}}\phi_{0,\sigma^2}(y_t).\end{aligned}$$

The second conditional density can be transferred to a similar integral

$$f(y_t|\mathcal{F}_{t-1}, J_t^{(k)}=i) = \int f(y_t|\boldsymbol{\beta}_t, \mathcal{F}_{t-1}, J_t^{(k)}=i)f(\boldsymbol{\beta}_t|\mathcal{F}_{t-1}, J_t^{(k)}=i)d\boldsymbol{\beta}_t,$$

where $f(\boldsymbol{\beta}_t|\mathcal{F}_{t-1}, J_t^{(k)}=i) = g_{i,t-1}^{(k)} = \phi_{\boldsymbol{z}_{i,t-1}^{(k)},\boldsymbol{V}_{i,t-1}^{(k)}}(\boldsymbol{\beta}_t)$. The integrand can be rewritten as

$$\begin{aligned}&f(y_t|\boldsymbol{\beta}_t, \mathcal{F}_{t-1}, J_t^{(k)}=i)f(\boldsymbol{\beta}_t|\mathcal{F}_{t-1}, J_t^{(k)}=i)\\ =&\frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big\{-\frac{(y_t-\boldsymbol{x}_t'\boldsymbol{\beta}_t)^2}{2\sigma^2}\Big\}((2\pi)^d|\boldsymbol{V}(k)|)^{-\frac{1}{2}}\exp\Big\{-\frac{1}{2}(\boldsymbol{\beta}_t-\boldsymbol{z}_{i,t-1}^{(k)})'(\boldsymbol{V}_{i,t-1}^{(k)})^{-1}(\boldsymbol{\beta}_t-\boldsymbol{z}_{i,t-1}^{(k)})\Big\}\\ =&\frac{1}{\sqrt{2\pi\sigma^2}}((2\pi)^d|\boldsymbol{V}(k)|)^{-\frac{1}{2}}\exp\Big\{-\frac{1}{2}(\boldsymbol{\beta}_t-\widetilde{\boldsymbol{z}})'\widetilde{\boldsymbol{V}}^{-1}(\boldsymbol{\beta}_t-\widetilde{\boldsymbol{z}})\\ &\qquad -\frac{1}{2}\boldsymbol{z}_{i,t-1}^{(k)\prime}(\boldsymbol{V}_{i,t-1}^{(k)})^{-1}\boldsymbol{z}_{i,t-1}^{(k)}-\frac{y_t^2}{2\sigma^2}+\frac{1}{2}\widetilde{\boldsymbol{z}}'\widetilde{\boldsymbol{V}}^{-1}\widetilde{\boldsymbol{z}}\Big\},\end{aligned}$$

where

$$\begin{aligned}\widetilde{\boldsymbol{V}} =&\big((\boldsymbol{V}_{i,t-1}^{(k)})^{-1}+\frac{\boldsymbol{x}_t\boldsymbol{x}_t'}{\sigma^2}\big)^{-1} = \boldsymbol{V}_{i,t}^{(k)},\\ \widetilde{\boldsymbol{z}} =&\widetilde{\boldsymbol{V}}\big((\boldsymbol{V}_{i,t-1}^{(k)})^{-1}\boldsymbol{z}_{i,t-1}^{(k)}+\frac{\boldsymbol{x}_t y_t}{\sigma^2}\big) = \boldsymbol{z}_{i,t}^{(k)}.\end{aligned}$$

Thus

$$f(y_t|\mathcal{F}_{t-1}, J_t^{(k)}=i) = \frac{\phi_{\boldsymbol{z}_{i,t-1}^{(k)},\boldsymbol{V}_{i,t-1}^{(k)}}(\mathbf{0})\phi_{0,\sigma^2}(y_t)}{\phi_{\boldsymbol{z}_{i,t}^{(k)},\boldsymbol{V}_{i,t}^{(k)}}(\mathbf{0})} = \frac{f_{i,t}^{(k)}}{f_{i,t-1}^{(k)}}\phi_{0,\sigma^2}(y_t).$$

Then

$$\frac{f(y_t|J_t^{(k)}=t)}{f(y_t|\mathcal{F}_{t-1},J_t^{(k)}=i)}=\frac{f_{t,t}^{(k)}/f_{0,0}^{(k)}}{f_{i,t}^{(k)}/f_{i,t-1}^{(k)}}. \tag{2.2.6}$$

which is not a function of y_t and $\boldsymbol{x}_t$. Plugging (2.2.6) into (2.2.4) yielding $\xi_{i,t}^{(k)}=\frac{\xi_{i,t}^{(k)*}}{\sum_{k=1}^K\sum_{i=1}^t\xi_{i,t}^{(k)*}}$, where

$$\xi_{i,t}^{(k)}\propto\xi_{i,t}^{(k)*}:=\begin{cases}\left(\sum_{l\neq k}\xi_{t-1}^{(l)}p_{lk}\right)f_{t,t}^{(k)}/f_{0,0}^{(k)} & i=t,\\ p_{kk}\xi_{i,t-1}^{(k)}f_{i,t}^{(k)}/f_{i,t-1}^{(k)} & i<t.\end{cases} \tag{2.2.7}$$

From (2.2.1) and (2.2.3), we know that the posterior distribution of $\boldsymbol{\beta}_t$ given $\mathcal{F}_t$ is a mixture of normal distributions. So the filtering estimate can be calculated by the posterior expectation as

$$\hat{\boldsymbol{\beta}}_{t|t}:=E(\boldsymbol{\beta}_t|\mathcal{F}_t)=\sum_{k=1}^K\sum_{i=1}^t\xi_{i,t}^{(k)}\boldsymbol{z}_{i,t}^{(k)}. \tag{2.2.8}$$

2.3 The Backward Filtering Estimate of Parameters

As indicated at the end of Section 2.1, $\{\boldsymbol{\beta}_t\}$ is a reversible Markov chain. Therefore we can obtain a backward filter that is analogous to (2.2.3). That is, we reverse the time, starting with time T and estimating $\boldsymbol{\beta}_t$ for any time t given the "historical" information from time T to t.

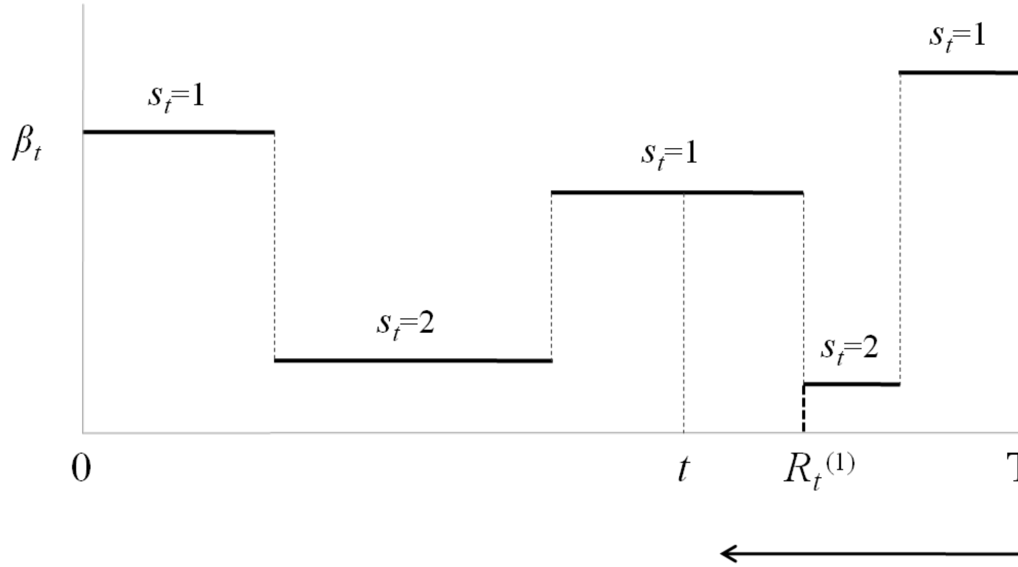
Define $R_t^{(k)}=\min\{j\geq t:k=s_t\cdots=s_{j-1}\neq s_j\}$ be the most recent switching time larger than or equal to t when s_t switches from the regime k to another regime. Figure 2.3 illustrates the definition of $R_t^{(k)}$. At time t, the regime is $s_t=1$, and the most recent transition occurs after t is at $R_t^{(1)}$ as shown in Figure 2.3.

Define

$$\eta_t^{(k)}=P(s_t=k|\mathcal{F}_{t,T}),\qquad\eta_{j,t}^{(k)}=P(R_t^{(k)}=j|\mathcal{F}_{t,T})$$

for $t\leq j\leq T$ and $1\leq k\leq K$. The quantity $\eta_t^{(k)}$ is the conditional probability that the

Figure 2.3: Illustration: Definition of $R_t^{(k)}$.



current state is k given information $\mathcal{F}_{t,T}$ $\eta_{i,t}^{(k)}$ is the conditional probability that the current regime is k and the next transition occurs at time j given $\mathcal{F}_{t,T}$. Thus $\eta_t^{(k)} = \sum_{j=t}^{T} \eta_{t,j}^{(k)}$. If we know all the information from time t to T and that the next transition occurs at time j, we just need to use the information before the switch to estimate the current value of $\boldsymbol{\beta}_t$. Similar to (2.2.1), the conditional distribution of $\boldsymbol{\beta}_t$, given $\mathcal{F}_{t,T}$ and the event $R_t^{(k)} = j$, is $g_{t,j}^{(k)}(\boldsymbol{\beta}_t)$. Thus the backward filter is defined as

$$\begin{aligned} \boldsymbol{\beta}_{t+1}|\mathcal{F}_{t+1,T} &\sim \sum_{k=1}^{K} \sum_{j=t+1}^{T} P(R_{t+1}^{(k)} = j|\mathcal{F}_{t+1,T}) f(\boldsymbol{\beta}_{t+1}|\mathcal{F}_{t+1,T}, R_{t+1}^{(k)} = j) \\ &= \sum_{k=1}^{K} \sum_{j=t+1}^{T} \eta_{t+1,j}^{(k)} g_{t+1,j}^{(k)}(\boldsymbol{\beta}_{t+1}), \end{aligned} \tag{2.3.1}$$

in which the weights $\eta_{t+1,j}^{(k)}$ can be obtained by backward induction using the time-reversed counterpart of (2.2.7):

$$\eta_{t+1,j}^{(k)} \propto \eta_{t+1,j}^{(k)*} := \begin{cases} \left(\sum_{l \neq k} \eta_{t+2}^{(l)} \widetilde{p}_{lk}\right) f_{t+1,t+1}^{(k)} / f_{0,0}^{(k)} & j = t+1, \\ \widetilde{p}_{kk} \eta_{t+2,j}^{(k)} f_{t+1,j}^{(k)} / f_{t+2,j}^{(k)} & j > t+1, \end{cases} \tag{2.3.2}$$

where $\widetilde{P} = (\widetilde{p}_{lk})$ is the transition matrix of the reversed chain of $\{s_t\}$; that is, $\widetilde{p}_{lk} = P(s_t = k | s_{t+1} = l)$.

We can go one step further to calculate $f(\boldsymbol{\beta}_t | \mathcal{F}_{t+1,T})$. Following (2.3.1) and the reversibility of $\{\boldsymbol{\beta}_t\}$,

$$\begin{aligned} f(\boldsymbol{\beta}_t | \mathcal{F}_{t+1,T}) &= \sum_{k=1}^{K} f(\boldsymbol{\beta}_t, s_{t+1} = k | \mathcal{F}_{t+1,T}) = \sum_{k=1}^{K} P(s_{t+1} = k | \mathcal{F}_{t+1,T}) f(\boldsymbol{\beta}_t | s_{t+1} = k, \mathcal{F}_{t+1,T}) \\ &= \sum_{k=1}^{K} \sum_{l=1}^{K} P(s_{t+1} = k | \mathcal{F}_{t+1,T}) f(\boldsymbol{\beta}_t, s_t = l | s_{t+1} = k, \mathcal{F}_{t+1,T}) \\ &= \sum_{k=1}^{K} \sum_{l=1}^{K} P(s_{t+1} = k | \mathcal{F}_{t+1,T}) P(s_t = l | s_{t+1} = k) f(\boldsymbol{\beta}_t | s_{t+1} = k, s_t = l, \mathcal{F}_{t+1,T}) \\ &= \sum_{k=1}^{K} \sum_{l=1}^{K} P(s_{t+1} = k | \mathcal{F}_{t+1,T}) \widetilde{p}_{kl} f(\boldsymbol{\beta}_t | s_{t+1} = k, s_t = l, \mathcal{F}_{t+1,T}). \end{aligned}$$

When $k = l$,

$$\begin{aligned} &\widetilde{p}_{kk} P(s_{t+1} = l | \mathcal{F}_{t+1,T}) f(\boldsymbol{\beta}_t | s_{t+1} = k, s_t = k, \mathcal{F}_{t+1,T}) \\ =& \widetilde{p}_{kk} f(\boldsymbol{\beta}, s_{t+1} = k | \mathcal{F}_{t+1,T}) \Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}_t} \\ =& \widetilde{p}_{kk} \sum_{j=t+1}^{T} f(\boldsymbol{\beta}, R_t^{(k)} = j | \mathcal{F}_{t+1,T}) \Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}_t} \\ =& \widetilde{p}_{kk} \sum_{j=t+1}^{T} P(R_t^{(k)} = j | \mathcal{F}_{t+1,T}) f(\boldsymbol{\beta} | R_t^{(k)} = j, \mathcal{F}_{t+1,T}) \Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}_t} \\ =& \widetilde{p}_{kk} \sum_{j=t+1}^{T} \eta_{t+1,j}^{(k)} g_{t+1,j}^{(k)}(\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}_t}; \end{aligned}$$

when $k \neq l$,

$$\begin{aligned}
&\widetilde{p}_{kl} P(s_{t+1} = k | \mathcal{F}_{t+1,T}) f(\boldsymbol{\beta}_t | s_{t+1} = k, s_t = l, \mathcal{F}_{t+1,T}) \\
=&\widetilde{p}_{kl} \eta_{t+1}^{(k)} f(\boldsymbol{\beta} | s_t = l)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_t} = \widetilde{p}_{kl} \eta_{t+1}^{(k)} g_{0,0}^{(l)}(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_t}.
\end{aligned}$$

Thus

$$\begin{aligned}
f(\boldsymbol{\beta}_t | \mathcal{F}_{t+1,T}) &= \sum_{k=1}^{K} f(\boldsymbol{\beta}_t, s_{t+1} = k | \mathcal{F}_{t+1,T}) = \sum_{k=1}^{K} P(s_{t+1} = k | \mathcal{F}_{t+1,T}) \\
&= \sum_{k=1}^{K} \sum_{l=1}^{K} P(s_{t+1} = k | \mathcal{F}_{t+1,T}) \widetilde{p}_{kl} f(\boldsymbol{\beta}_t | s_{t+1} = k, s_t = l, \mathcal{F}_{t+1,T}) \\
&= \sum_{k=1}^{K} \widetilde{p}_{kk} \sum_{j=t+1}^{T} \eta_{t+1,j}^{(k)} g_{t+1,j}^{(k)}(\boldsymbol{\beta}) + \sum_{k=1}^{K} \sum_{l \neq k} \widetilde{p}_{kl} \eta_{t+1}^{(k)} g_{0,0}^{(l)}(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_t}.
\end{aligned}$$

So we have

$$f(\boldsymbol{\beta}_t | \mathcal{F}_{t+1,T}) = \sum_{k=1}^{K} \left\{ \widetilde{p}_{kk} \sum_{j=t+1}^{T} \eta_{t+1,j}^{(k)} g_{t+1,j}^{(k)}(\boldsymbol{\beta}) + \Big(\sum_{l \neq k} \widetilde{p}_{kl} \eta_{t+1}^{(k)} \Big) g_{0,0}^{(l)}(\boldsymbol{\beta}) \right\}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_t}. \qquad (2.3.3)$$

2.4 The Smoothing Estimate of Parameters

In this section, we will show how to estimate $\boldsymbol{\beta}_t$ for any time t when all the information $\mathcal{F}_T$ is given. Using Bayes' theorem, we can combine the forward filter (2.2.3) with its backward variant (2.3.1)

$$f(\boldsymbol{\beta}_t | \mathcal{F}_T) = \sum_{k=1}^{K} f(\boldsymbol{\beta}_t, s_t = k | \mathcal{F}_T) \propto \sum_{k=1}^{K} f(\boldsymbol{\beta}_t, s_t = k | \mathcal{F}_t) f(\boldsymbol{\beta}_t, s_t = k | \mathcal{F}_{t+1,T}) \big/ f(\boldsymbol{\beta}, s_t = k).$$

From this we can derive the posterior distribution of $\boldsymbol{\beta}_t$ given $\mathcal{F}_T$. Let $g_t(\cdot|\mathcal{F}_T)$, $g_t(\cdot|\mathcal{F}_t)$, and $g_t(\cdot|\mathcal{F}_{t+1,T})$ denote the density functions of the absolutely continuous components of $\boldsymbol{\beta}_t$

given $\mathcal{F}_T$, $\mathcal{F}_t$, and $\mathcal{F}_{t+1,T}$ respectively. Applying Bayes' theorem,

$$g_t(\boldsymbol{\beta}|\mathcal{F}_T) = \sum_{k=1}^{K} g_t(\boldsymbol{\beta}, s_t = k|\mathcal{F}_T) \propto \sum_{k=1}^{K} g_t(\boldsymbol{\beta}, s_t = k|\mathcal{F}_t) g_t(\boldsymbol{\beta}, s_t = k|\mathcal{F}_{t+1,T}) \big/ f(\boldsymbol{\beta}, s_t = k).$$

The right hand side is a mixture of different states. Following (2.2.3) and the proof of (2.3.3), we have

$$\begin{aligned}
& g_t(\boldsymbol{\beta}, s_t = k|\mathcal{F}_t) g_t(\boldsymbol{\beta}, s_t = k|\mathcal{F}_{t+1,T}) \big/ f(\boldsymbol{\beta}, s_t = k) \\
=& \frac{\Big\{ \sum_{i=1}^{t} P(J_t^{(k)} = i|\mathcal{F}_t) f(\boldsymbol{\beta}_t|\mathcal{F}_t, J_t^{(k)} = i) \Big\} \Big\{ \sum_{l=1}^{K} P(s_{t+1} = l|\mathcal{F}_{t+1,T}) f(\boldsymbol{\beta}_t, s_t = k|s_{t+1} = l, \mathcal{F}_{t+1,T}) \Big\}}{P(s_t = k) f(\boldsymbol{\beta}_t|s_t = k)} \\
=& \frac{\Big\{ \sum_{i=1}^{t} \xi_{i,t}^{(k)} g_{i,t}^{(k)}(\boldsymbol{\beta}) \Big\} \Big\{ \widetilde{p}_{kk} \sum_{j=t+1}^{T} \eta_{t+1,j}^{(k)} g_{t+1,j}^{(k)}(\boldsymbol{\beta}) + \Big(\sum_{l \neq k} \widetilde{p}_{lk} \eta_{t+1}^{(l)} \Big) g_{0,0}^{(k)}(\boldsymbol{\beta}) \Big\}}{\pi_k g_{0,0}^{(k)}(\boldsymbol{\beta})} \\
=& \sum_{i=1}^{t} \xi_{i,t}^{(k)} \Big(\sum_{l \neq k} \frac{\widetilde{p}_{lk}}{\pi_k} \eta_{t+1}^{(l)} \Big) g_{i,t}^{(k)}(\boldsymbol{\beta}) + \frac{\widetilde{p}_{kk}}{\pi_k} \sum_{1 \le i \le t < j \le T} \xi_{i,t}^{(k)} \eta_{t+1,j}^{(k)} \frac{g_{i,t}^{(k)}(\boldsymbol{\beta}) g_{t+1,j}^{(k)}(\boldsymbol{\beta})}{g_{0,0}^{(k)}(\boldsymbol{\beta})}.
\end{aligned}$$

Based on the reversibility of P,

$$\begin{aligned}
\widetilde{p}_{kk} =& P(s_t = k|s_{t+1} = k) = \frac{P(s_t = k, s_{t+1} = k)}{P(s_{t+1} = k)} \\
=& \frac{P(s_t = k, s_{t+1} = k)}{P(s_t = k)} = P(s_{t+1} = k|s_t = k) = p_{kk}.
\end{aligned}$$

The definitions of $g_{i,j}^{(k)}$ and $f_{i,j}^{(k)}$ are

$$\begin{aligned}
g_{0,0}^{(k)}(\boldsymbol{\beta}) &= ((2\pi)^d |\boldsymbol{V}(k)|)^{-1/2} \exp\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{z}(k))' \boldsymbol{V}(k)^{-1} (\boldsymbol{\beta} - \boldsymbol{z}(k))\}, \\
g_{i,j}^{(k)}(\boldsymbol{\beta}) &= ((2\pi)^d |\boldsymbol{V}_{i,j}^{(k)}|)^{-1/2} \exp\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{z}_{i,j}^{(k)})' (\boldsymbol{V}_{i,j}^{(k)})^{-1} (\boldsymbol{\beta} - \boldsymbol{z}_{i,j}^{(k)})\}, \\
f_{0,0}^{(k)} &= |\boldsymbol{V}(k)|^{1/2} \exp\left\{ \frac{1}{2} \boldsymbol{z}(k)' \boldsymbol{V}(k)^{-1} \boldsymbol{z}(k) \right\}, \\
f_{i,j}^{(k)} &= |\boldsymbol{V}_{ij}^{(k)}|^{1/2} \exp\left\{ \frac{1}{2} \boldsymbol{z}_{ij}^{(k)\prime} (\boldsymbol{V}_{ij}^{(k)})^{-1} \boldsymbol{z}_{ij}^{(k)} \right\}.
\end{aligned}$$

And we have

$$\begin{aligned}
&\frac{g_{i,t}^{(k)}(\boldsymbol{\beta})g_{t+1,j}^{(k)}(\boldsymbol{\beta})}{g_{i,j}^{(k)}(\boldsymbol{\beta})g_{0,0}^{(k)}(\boldsymbol{\beta})}\\
=&\frac{|\boldsymbol{V}_{i,t}^{(k)}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{z}_{i,t}^{(k)})'(\boldsymbol{V}_{i,t}^{(k)})^{-1}(\boldsymbol{\beta}-\boldsymbol{z}_{i,t}^{(k)})\}|\boldsymbol{V}_{t+1,j}^{(k)}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{z}_{t+1,j}^{(k)})'(\boldsymbol{V}_{t+1,j}^{(k)})^{-1}(\boldsymbol{\beta}-\boldsymbol{z}_{t+1,j}^{(k)})\}}{|\boldsymbol{V}_{i,j}^{(k)}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{z}_{i,j}^{(k)})'(\boldsymbol{V}_{i,j}^{(k)})^{-1}(\boldsymbol{\beta}-\boldsymbol{z}_{i,j}^{(k)})\}|\boldsymbol{V}(k)|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{z}(k))'\boldsymbol{V}(k)^{-1}(\boldsymbol{\beta}-\boldsymbol{z}(k))\}}\\
=&\Big(\frac{|\boldsymbol{V}_{i,j}^{(k)}||\boldsymbol{V}(k)|}{|\boldsymbol{V}_{i,t}^{(k)}||\boldsymbol{V}_{t+1,j}^{(k)}|}\Big)^{\frac{1}{2}}\exp\Big\{\frac{1}{2}\Big((\boldsymbol{\beta}-\boldsymbol{z}_{i,j}^{(k)})'(\boldsymbol{V}_{i,j}^{(k)})^{-1}(\boldsymbol{\beta}-\boldsymbol{z}_{i,j}^{(k)})+(\boldsymbol{\beta}-\boldsymbol{z}(k))'\boldsymbol{V}(k)^{-1}(\boldsymbol{\beta}-\boldsymbol{z}(k))\\
&\qquad-(\boldsymbol{\beta}-\boldsymbol{z}_{i,t}^{(k)})'(\boldsymbol{V}_{i,t}^{(k)})^{-1}(\boldsymbol{\beta}-\boldsymbol{z}_{i,t}^{(k)})-(\boldsymbol{\beta}-\boldsymbol{z}_{t+1,j}^{(k)})'(\boldsymbol{V}_{t+1,j}^{(k)})^{-1}(\boldsymbol{\beta}-\boldsymbol{z}_{t+1,j}^{(k)})\Big)\Big\}.
\end{aligned}$$

Expanding the part inside the parentheses of the second term yields

$$\begin{aligned}
&(\boldsymbol{\beta}-\boldsymbol{z}_{i,j}^{(k)})'(\boldsymbol{V}_{i,j}^{(k)})^{-1}(\boldsymbol{\beta}-\boldsymbol{z}_{i,j}^{(k)})+(\boldsymbol{\beta}-\boldsymbol{z}(k))'\boldsymbol{V}(k)^{-1}(\boldsymbol{\beta}-\boldsymbol{z}(k))\\
&\quad-(\boldsymbol{\beta}-\boldsymbol{z}_{i,t}^{(k)})'(\boldsymbol{V}_{i,t}^{(k)})^{-1}(\boldsymbol{\beta}-\boldsymbol{z}_{i,t}^{(k)})-(\boldsymbol{\beta}-\boldsymbol{z}_{t+1,j}^{(k)})'(\boldsymbol{V}_{t+1,j}^{(k)})^{-1}(\boldsymbol{\beta}-\boldsymbol{z}_{t+1,j}^{(k)})\\
=&\boldsymbol{\beta}'\Big((\boldsymbol{V}_{i,j}^{(k)})^{-1}+\boldsymbol{V}(k)^{-1}-(\boldsymbol{V}_{i,t}^{(k)})^{-1}-(\boldsymbol{V}_{t+1,j}^{(k)})^{-1}\Big)\boldsymbol{\beta}\\
&\quad-2\boldsymbol{\beta}'\Big((\boldsymbol{V}_{i,j}^{(k)})^{-1}\boldsymbol{z}_{i,j}^{(k)}+\boldsymbol{V}(k)^{-1}\boldsymbol{z}(k)-(\boldsymbol{V}_{i,t}^{(k)})^{-1}\boldsymbol{z}_{i,t}^{(k)}-(\boldsymbol{V}_{t+1,j}^{(k)})^{-1}\boldsymbol{z}_{t+1,j}^{(k)}\Big)\\
&\quad+\boldsymbol{z}_{i,j}^{(k)\prime}(\boldsymbol{V}_{i,j}^{(k)})^{-1}\boldsymbol{z}_{i,j}^{(k)}+\boldsymbol{z}(k)'\boldsymbol{V}(k)^{-1}\boldsymbol{z}(k)-\boldsymbol{z}_{i,t}^{(k)\prime}(\boldsymbol{V}_{i,t}^{(k)})^{-1}\boldsymbol{z}_{i,t}^{(k)}-\boldsymbol{z}_{t+1,j}^{(k)}{}'(\boldsymbol{V}_{t+1,j}^{(k)})^{-1}\boldsymbol{z}_{t+1,j}^{(k)}.
\end{aligned}$$

As later shown in (2.9.2) in Section 2.8, we have

$$\begin{aligned}
&(\boldsymbol{V}_{i,j}^{(k)})^{-1}+\boldsymbol{V}(k)^{-1}-(\boldsymbol{V}_{i,t}^{(k)})^{-1}-(\boldsymbol{V}_{t+1,j}^{(k)})^{-1}=0,\\
&(\boldsymbol{V}_{i,j}^{(k)})^{-1}\boldsymbol{z}_{i,j}^{(k)}+\boldsymbol{V}(k)^{-1}\boldsymbol{z}(k)-(\boldsymbol{V}_{i,t}^{(k)})^{-1}\boldsymbol{z}_{i,t}^{(k)}-(\boldsymbol{V}_{t+1,j}^{(k)})^{-1}\boldsymbol{z}_{t+1,j}^{(k)}=0.
\end{aligned}$$

So

$$\begin{aligned}
&\frac{g_{i,t}^{(k)}(\boldsymbol{\beta})g_{t+1,j}^{(k)}(\boldsymbol{\beta})}{g_{i,j}^{(k)}(\boldsymbol{\beta})g_{0,0}^{(k)}(\boldsymbol{\beta})}\\
=&\Big(\frac{|\boldsymbol{V}_{i,j}^{(k)}||\boldsymbol{V}(k)|}{|\boldsymbol{V}_{i,t}^{(k)}||\boldsymbol{V}_{t+1,j}^{(k)}|}\Big)^{\frac{1}{2}}\frac{\exp\left\{\frac{1}{2}\boldsymbol{z}_{i,j}^{(k)\prime}(\boldsymbol{V}_{i,j}^{(k)})^{-1}\boldsymbol{z}_{i,j}^{(k)}\right\}\exp\left\{\frac{1}{2}\boldsymbol{z}(k)'\boldsymbol{V}(k)^{-1}\boldsymbol{z}(k)\right\}}{\exp\left\{\frac{1}{2}\boldsymbol{z}_{i,t}^{(k)\prime}(\boldsymbol{V}_{i,t}^{(k)})^{-1}\boldsymbol{z}_{i,t}^{(k)}\right\}\exp\left\{\frac{1}{2}\boldsymbol{z}_{t+1,j}^{(k)\prime}(\boldsymbol{V}_{t+1,j}^{(k)})^{-1}\boldsymbol{z}_{i,t}^{(k)}\right\}}\\
=&\frac{f_{i,j}^{(k)}f_{0,0}^{(k)}}{f_{i,t}^{(k)}f_{t+1,j}^{(k)}},
\end{aligned}$$

that is,

$$g_{i,t}^{(k)}(\boldsymbol{\beta})g_{t+1,j}^{(k)}(\boldsymbol{\beta})\Big/g_{0,0}^{(k)}(\boldsymbol{\beta})=\frac{f_{i,j}^{(k)}f_{0,0}^{(k)}}{f_{i,t}^{(k)}f_{t+1,j}^{(k)}}g_{i,j}^{(k)}(\boldsymbol{\beta}).$$

So the posterior distribution of $\boldsymbol{\beta}_t$ given $\mathcal{F}_T$ is a mixture of normal distributions:

$$\begin{aligned}
\boldsymbol{\beta}_t|\mathcal{F}_T\sim&\sum_{k=1}^{K}\sum_{1\le i\le t\le j\le T}P(J_t^{(k)}=i,R_{t+1}^{(k)}=j|\mathcal{F}_T)f(\boldsymbol{\beta}_t|\mathcal{F}_T,J_t^{(k)}=i,R_t^{(k)}=j)\\
=&\sum_{k=1}^{K}\sum_{1\le i\le t\le j\le T}\alpha_{ijt}^{(k)}g_{i,j}^{(k)}(\boldsymbol{\beta}_t),
\end{aligned}\tag{2.4.1}$$

where the mixture weight $\alpha_{ijt}^{(k)}$ is the conditional probability which can be calculated recursively as

$$\begin{aligned}
&\alpha_{ijt}^{(k)}=\alpha_{ijt}^{(k)*}\Big/D_t,\qquad D_t=\sum_{k=1}^{K}\sum_{1\le i\le t\le j\le T}\alpha_{ijt}^{(k)*},\\
&\alpha_{ijt}^{(k)*}=\begin{cases}\xi_{i,t}^{(k)}\big(\sum_{l\neq k}\eta_{t+1}^{(l)}p_{kl}/\pi_l\big) & i\le t=j,\\ p_{kk}\xi_{i,t}^{(k)}\eta_{t+1,j}^{(k)}f_{i,j}^{(k)}f_{0,0}^{(k)}/(\pi_k f_{i,t}^{(k)}f_{t+1,j}^{(k)}) & i\le t<j.\end{cases}
\end{aligned}\tag{2.4.2}$$

The advantage of our method is that the posterior distribution is given explicitly. Thus the smoothing estimate of the parameters $\boldsymbol{\beta}_t$ can be calculated directly based on the posterior distribution. There is no need to use complicated numerical methods involving numbers of recursive computations. So this method is more accurate and time saving compared with existing Bayesian methods. From (2.4.1), $\boldsymbol{\beta}_t$ can be estimated by the posterior mean of $\boldsymbol{\beta}_t$

given $\mathcal{F}_T$, which is

$$\hat{\boldsymbol{\beta}}_{t|T} := E(\boldsymbol{\beta}_t|\mathcal{F}_T) = \sum_{k=1}^{K} \sum_{1\le i\le t\le j\le T} \alpha_{ijt}^{(k)} \boldsymbol{z}_{i,j}^{(k)}. \tag{2.4.3}$$

2.5 Inference on Regimes

We are also interested in the unknown regime at each time t. The $\alpha_{ijt}^{(k)}$ in (2.4.2) are posterior probabilities that are useful for the inference. The derivation of (2.4.2) shows that, for $i \le t \le j$,

$$\alpha_{ijt}^{(k)} = P(C_{ij}^{(k)}|\mathcal{F}_T),$$

where

$$\begin{aligned} C_{ij}^{(k)} =& \{s_i = \cdots = s_j = k, s_i \neq s_{i-1}, s_j \neq s_{j+1}\} \\ =& \{J_t^{(k)} = i, R_t^{(k)} = j\}. \end{aligned}$$

For the problem of classifying the regime at stage t, a natural quantity is

$$\begin{aligned} P(s_t = k|\mathcal{F}_T) =& \sum_{1\le i\le t\le j\le T} P(J_t^{(k)} = i, R_t^{(k)} = j|\mathcal{F}_T) \\ =& \sum_{1\le i\le t\le j\le T} \alpha_{ijt}^{(k)}. \end{aligned} \tag{2.5.1}$$

Since the above quantity represents the probability of regime k without specifying the value of $\boldsymbol{\beta}_t$, it is more robust than the quantity $\hat{\xi}_{t|T}$ in Section 1.2. Moreover, the computation of the above probability does not need the complicated iteration of (1.2.7) which further involves repeated calculation of (1.2.4) and (1.2.5). Thus the proposed quantity in (2.5.1) is also computationally more efficient.

2.6 Forecast of Parameters

Other than estimating $\boldsymbol{\beta}_t$ given $\mathcal{F}_T$, we are also interested in forecasting $\boldsymbol{\beta}_{t+1}$ given $\mathcal{F}_t$. The forward filiter shows the posterior distribution $f(\boldsymbol{\beta}_t|\mathcal{F}_t)$. We can go one step further to calculate $f(\boldsymbol{\beta}_{t+1}|\mathcal{F}_t)$. Note that

$$\begin{aligned}
f(\boldsymbol{\beta}_{t+1}|\mathcal{F}_t) &= \sum_{k=1}^{K} f(\boldsymbol{\beta}_{t+1}, s_t = k|\mathcal{F}_t) = \sum_{k=1}^{K} P(s_t = k|\mathcal{F}_t) f(\boldsymbol{\beta}_{t+1}|s_t = k, \mathcal{F}_t) \\
&= \sum_{k=1}^{K}\sum_{l=1}^{K} P(s_t = k|\mathcal{F}_t) f(\boldsymbol{\beta}_{t+1}, s_{t+1} = l|s_t = k, \mathcal{F}_t) \\
&= \sum_{k=1}^{K}\sum_{l=1}^{K} P(s_t = k|\mathcal{F}_t) P(s_{t+1} = l|s_t = k) f(\boldsymbol{\beta}_{t+1}|s_t = k, s_{t+1} = l, \mathcal{F}_t) \\
&= \sum_{k=1}^{K}\sum_{l=1}^{K} P(s_t = k|\mathcal{F}_t) p_{kl} f(\boldsymbol{\beta}_{t+1}|s_t = k, s_{t+1} = l, \mathcal{F}_t).
\end{aligned}$$

When $k = l$,

$$\begin{aligned}
& p_{kk} P(s_t = k|\mathcal{F}_t) f(\boldsymbol{\beta}_{t+1}|s_t = k, s_{t+1} = k, \mathcal{F}_t) \\
=& p_{kk} f(\boldsymbol{\beta}, s_t = k|\mathcal{F}_t)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{t+1}} \\
=& p_{kk} \sum_{i=1}^{t} f(\boldsymbol{\beta}, J_t^{(k)} = i|\mathcal{F}_t)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{t+1}} \\
=& p_{kk} \sum_{i=1}^{t} P(J_t^{(k)} = i|\mathcal{F}_t) f(\boldsymbol{\beta}|J_t^{(k)} = i, \mathcal{F}_t)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{t+1}} \\
=& p_{kk} \sum_{i=1}^{t} \xi_{i,t}^{(k)} g_{i,t}^{(k)}(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{t+1}};
\end{aligned}$$

when $k \neq l$,

$$\begin{aligned}
& p_{kl} P(s_t = k|\mathcal{F}_t) f(\boldsymbol{\beta}_{t+1}|s_t = k, s_{t+1} = l, \mathcal{F}_t) \\
=& p_{kl} \xi_t^{(k)} f(\boldsymbol{\beta}|s_{t+1} = l)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{t+1}} = p_{kl} \xi_t^{(k)} g_{0,0}^{(l)}(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{t+1}}.
\end{aligned}$$

Thus

$$
\begin{aligned}
f(\boldsymbol{\beta}_{t+1}|\mathcal{F}_t) &= \sum_{k=1}^{K} f(\boldsymbol{\beta}_{t+1}, s_t = k|\mathcal{F}_t) = \sum_{k=1}^{K} P(s_t = k|\mathcal{F}_t) \\
&= \sum_{k=1}^{K}\sum_{l=1}^{K} P(s_t = k|\mathcal{F}_t) p_{kl} f(\boldsymbol{\beta}_{t+1}|s_t = k, s_{t+1} = l, \mathcal{F}_t) \\
&= \sum_{k=1}^{K} p_{kk} \sum_{i=1}^{t} \xi_{i,t}^{(k)} g_{i,t}^{(k)}(\boldsymbol{\beta}) + \sum_{k=1}^{K}\sum_{l\neq k} p_{kl}\xi_t^{(k)} g_{0,0}^{(l)}(\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{t+1}}.
\end{aligned}
$$

So we have

$$
f(\boldsymbol{\beta}_{t+1}|\mathcal{F}_t) = \sum_{k=1}^{K} \Big\{ p_{kk} \sum_{i=1}^{t} \xi_{i,t}^{(k)} g_{i,t}^{(k)}(\boldsymbol{\beta}) + \Big(\sum_{l\neq k} p_{kl}\xi_t^{(k)}\Big) g_{0,0}^{(l)}(\boldsymbol{\beta}) \Big\}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{t+1}}. \tag{2.6.1}
$$

We can use the posterior expectation to predict $\boldsymbol{\beta}_{t+1}$ given $\mathcal{F}_t$ as

$$
\hat{\boldsymbol{\beta}}_{t+1|t} := E(\boldsymbol{\beta}_{t+1}|\mathcal{F}_t) = \sum_{k=1}^{K} \Big\{ p_{kk} \sum_{i=1}^{t} \xi_{i,t}^{(k)} \boldsymbol{z}_{i,t}^{(k)} + \Big(\sum_{l\neq k} p_{kl}\xi_t^{(k)}\Big) \boldsymbol{z}(l) \Big\}. \tag{2.6.2}
$$

In the simulation studies and real data analysis, we will use the forecast shown in (2.6.2) to predict $\boldsymbol{\beta}_{t+1}$, and therefore $\boldsymbol{y}_{t+1}$.

2.7 Bounded Complexity Mixture Approximation

Although the forward filter (2.2.3) uses a recursive updating formula (2.2.7) for the weights $\xi_{i,t}^{(k)}$ $(1 \le i \le t, 1 \le k \le K)$, the number of weights increases dramatically with t, resulting in rapidly increasing computational complexity and memory requirements in estimating $\boldsymbol{\beta}_t$ as t keeps increasing. To address the issue of computational efficiency, we consider an approximation procedure with much lower computational complexity yet comparable to the Bayes estimates in statistical efficiency. The procedure follows Lai, Liu and Xing (2005), Lai, Xing and Zhang (2008), and Lai and Xing (2011).

The approximation is to keep only a fixed number M of weights at every stage t (which is tantamount to setting the other weights to be 0). Following Lai, Liu and Xing (2005) who consider a change point autoregressive model, we keep the most recent m weights $\xi_{i,t}^{(k)}$ (with $t-m < i \le t$) and the largest $M-m$ of the remaining weights, where $1 \le m < M$. Specifically, the updating formula (2.2.3) for the weights $\xi_{i,t}^{(k)}$ is modified as follows to obtain a bounded complexity mixture (BCMIX) approximation.

Let $\mathcal{K}_{t-1}^{(k)}$ denote the set of indices i for which $\xi_{i,t-1}^{(k)}$ in (2.2.7) is kept at stage $t-1$ for regime k; thus there are M indices in $\mathcal{K}_{t-1}^{(k)}$ and $\mathcal{K}_{t-1}^{(k)} \supset \{t-1, \cdots, t-m\}$. At stage t when a new observation arrives, define $\xi_{i,t}^{(k)*}$ by (2.2.7) for $i \in \{t\} \cup \mathcal{K}_{t-1}^{(k)}$ and let i_t be the index not belonging to the most recent m stages, $\{t, t-1, \cdots, t-m+1\}$ such that

$$\xi_{i_t,t}^{(k)*} = \min\{\xi_{i,t}^{(k)*} : i \in \mathcal{K}_{t-1}^{(k)} \quad \text{and} \quad i \le t-m\}, \tag{2.7.1}$$

choosing $i_t^{(k)}$ to be the one farthest from t if the minimizing set in (2.7.1) has more than one element. Define $\mathcal{K}_t^{(k)} = \{t\} \cup (\mathcal{K}_{t-1}^{(k)} - \{i_t^{(k)}\})$ and let

$$\xi_{i,t}^{(k)} = \left(\xi_{i,t}^{(k)*} \Big/ \sum_{j \in \mathcal{K}_t^{(k)}} \xi_{j,t}^{(k)*}\right), \qquad i \in \mathcal{K}_t^{(k)}, \tag{2.7.2}$$

which yields a BCMIX approximation to the forward filter.

Similarly, to obtain a BCMIX approximation to the backward filter defined in (2.3.2), let $\widetilde{\mathcal{K}}_{t+1}^{(k)}$ denote the set of indices j for which $\eta_{j,t+1}^{(k)}$ in (2.3.2) is kept at stage $t+1$ for regime k; thus, $\widetilde{\mathcal{K}}_{t+1}^{(k)} \supset \{t+1, \cdots, t+m\}$. At stage t, define $\eta_{j,t}^{(k)}$ by (2.3.2) for $j \in \{t\} \cup \mathcal{K}_{t+1}^{(k)}$ and let j_t be the index not belonging to the most recent m stages, $\{t, t+1, \cdots, t+m-1\}$ such that

$$\eta_{j_t,t}^{(k)*} = \min\{\eta_{j,t}^{(k)*} : j \in \widetilde{\mathcal{K}}_{t+1}^{(k)} \quad \text{and} \quad j \ge t+m\}, \tag{2.7.3}$$

choosing $j_t^{(k)}$ to be the one farthest from t if the minimizing set in (2.7.3) has more than one element. Define $\widetilde{\mathcal{K}}_t^{(k)} = \{t\} \cup (\mathcal{K}_{t+1}^{(k)} - \{i_t^{(k)}\})$ and let

$$\eta_{j,t}^{(k)} = \Big(\eta_{j,t}^{(k)*} \Big/ \sum_{j \in \widetilde{\mathcal{K}}_t^{(k)}} \eta_{j,t}^{(k)*}\Big), \qquad j \in \widetilde{\mathcal{K}}_t^{(k)}, \tag{2.7.4}$$

which yields a BCMIX approximation to the backward filter.

For the smoothing estimate $E(\boldsymbol{\beta}_t|\mathcal{F}_T)$ and its associated posterior distribution, we can construct BCMIX approximations by combining the preceding forward and backward BCMIX filters, which have index sets $\mathcal{K}_t^{(k)}$ for the forward filter and $\widetilde{\mathcal{K}}_{t+1}^{(k)}$ for the backward filter at stage t. The BCMIX approximation to (2.4.2) is defined by

$$\begin{aligned} &\widetilde{\alpha}_{ijt} = \alpha_{ijt}^{*}/\widetilde{D}_t, \qquad \widetilde{D}_t = \sum_{i \in \mathcal{K}_t^{(k)}, j \in \{t\} \cup \widetilde{\mathcal{K}}_{t+1}^{(k)}} \alpha_{ijt}^{*}, \\ &\alpha_{ijt}^{(k)*} = \begin{cases} \xi_{i,t}^{(k)} \big(\sum_{l \neq k} \eta_{t+1}^{(l)} p_{kl}/\pi_l \big) & i \in \mathcal{K}_t^{(k)}, \\ p_{kk} \xi_{i,t}^{(k)} \eta_{t+1,j}^{(k)} f_{i,j}^{(k)} f_{0,0}^{(k)} \big/ (\pi_k f_{i,t}^{(k)} f_{t+1,j}^{(k)}) & i \in \mathcal{K}_t^{(k)}, j \in \{t\} \cup \widetilde{\mathcal{K}}_{t+1}^{(k)}. \end{cases} \end{aligned} \tag{2.7.5}$$

The BCMIX approximation to the posterior distribution of $\boldsymbol{\beta}_t$ given $\mathcal{F}_T$ in (2.4.1) is

$$f(\boldsymbol{\beta}_t|\mathcal{F}_T) \approx \sum_{k=1}^{K} \sum_{i \in \mathcal{K}_t^{(k)}, j \in \{t\} \cup \widetilde{\mathcal{K}}_{t+1}^{(k)}} \widetilde{\alpha}_{ijt}^{(k)} g_{i,j}^{(k)}(\boldsymbol{\beta}_t),$$

the BCMIX approximation to $E(\boldsymbol{\beta}_t|\mathcal{F}_T)$in (2.4.3) is therefore

$$\hat{\boldsymbol{\beta}}_{t|T} \approx \sum_{k=1}^{K} \sum_{i \in \mathcal{K}_t^{(k)}, j \in \{t\} \cup \widetilde{\mathcal{K}}_{t+1}^{(k)}} \widetilde{\alpha}_{ijt}^{(k)} \boldsymbol{z}_{i,j}^{(k)}, \tag{2.7.6}$$

and the BCMIX approximation to the probability of $P(s_t = k|\mathcal{F}_T)$ in (2.5.1) is

$$\hat{r}_{t|T}^{(k)} := P(s_t = k|\mathcal{F}_T) \approx \sum_{k=1}^{K} \sum_{i \in \mathcal{K}_t^{(k)}, j \in \{t\} \cup \widetilde{\mathcal{K}}_{t+1}^{(k)}} \widetilde{\alpha}_{ijt}^{(k)}.$$

The BCMIX procedure fixes the number of filters as M at every stage, keeping the m closest weights and the other $M-m$ largest weights. Clearly the results depend on the specification of M and m: If they are too small, computing time is saved at the cost of discarding some important weights; if they are too large, estimating accuracy cannot be improved by excess computation. In Section 3.1.1, we will show the effect of M and m on the estimation results.

2.8 Estimation of Hyperparameters

It is shown in Appendix that the conditional density function of y_t give $\mathcal{F}_{1,t-1}$ is

$$f(y_t|\mathcal{F}_{t-1}) = \sum_{k=1}^{K} \sum_{i=1}^{t} \xi_{it}^{(k)*}, \tag{2.8.1}$$

where $\xi_{it}^{(k)*}$ are given by (2.2.7) and are functions of hyperparameter vector $\boldsymbol{\theta} = (P, \boldsymbol{z}(k), \boldsymbol{V}(k), \sigma^2);$ $1 \leq k \leq K)$. Given $\boldsymbol{\theta}$ and the observed data $\mathcal{F}_T$, the log likelihood function is

$$l(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log f(y_t|\mathcal{F}_{t-1}) = \sum_{t=1}^{T} \log \Big\{ \sum_{k=1}^{K} \sum_{i=1}^{t} \xi_{it}^{(k)*} \Big\}, \tag{2.8.2}$$

in which $f(\cdot|\cdot)$ denotes conditional density function. Maximizing (2.8.2) over $\boldsymbol{\theta}$ yields the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$.

Since $\boldsymbol{\theta}$ is a $[(K+2)K+1]$-dimensional vector and the functions $\xi_{it}^{(k)}$ have to be computed recursively for $1 \leq t \leq T$, direct maximization of (2.8.2) may be computationally expensive due to the curse of dimensionality. In this section, we will follow the procedure in Lai, Xing and Zhang (2008) to use the EM algorithm which exploits the much simpler structure of the log likelihood $l_c(\boldsymbol{\theta})$ of the complete data $\{(y_t, s_t, \boldsymbol{\beta}_t), 1 \le t \le T\}$:

$$\begin{aligned} l_c(\boldsymbol{\theta}) =& \sum_{t=1}^{T} \Big\{ \log f(y_t|\boldsymbol{\beta}_t) + \sum_{k=1}^{K} f(\boldsymbol{\beta}_t|s_t = k)\mathbf{1}_{\{s_t=k\}} + \sum_{k,l=1}^{K} \log(p_{kl})\mathbf{1}_{\{s_{t-1}=k,s_t=l\}} \Big\} \\ =& -\frac{1}{2}\sum_{t=1}^{T}\Big\{\frac{(y_t - \boldsymbol{\beta}_t'\boldsymbol{x}_t)^2}{\sigma^2} + \log(2\pi\sigma^2)\Big\} \\ & -\frac{1}{2}\sum_{t=1}^{T}\sum_{k=1}^{K}\Big\{(\boldsymbol{\beta}_t - \boldsymbol{z}(k))'\boldsymbol{V}(k)^{-1}(\boldsymbol{\beta}_t - \boldsymbol{z}(k)) + \log((2\pi)^d|\boldsymbol{V}(k)|)\Big\}\mathbf{1}_{\{s_t=k\}} \\ & + \sum_{t=1}^{T}\sum_{k,l=1}^{K}\log(p_{kl})\mathbf{1}_{\{s_{t-1}=k,s_t=l\}}. \end{aligned} \tag{2.8.3}$$

The E-step of the EM algorithm calculates $E(l_c(\boldsymbol{\theta})|\mathcal{F}_T)$ which is

$$\begin{aligned} E(l_c(\boldsymbol{\theta})|\mathcal{F}_T) =& -\frac{1}{2\sigma^2}\sum_{t=1}^{T} E[(y_t - \boldsymbol{\beta}_t'\boldsymbol{x}_t)^2|\mathcal{F}_T] - \frac{T}{2}\log(2\pi\sigma^2) \\ & -\frac{1}{2}\sum_{t=1}^{T}\sum_{k=1}^{K} E[(\boldsymbol{\beta}_t - \boldsymbol{z}(k))'\boldsymbol{V}(k)^{-1}(\boldsymbol{\beta}_t - \boldsymbol{z}(k))\mathbf{1}_{\{s_t=k\}}|\mathcal{F}_T] \\ & -\frac{T}{2}\sum_{k=1}^{K}\log((2\pi)^d|\boldsymbol{V}(k)|)E(\mathbf{1}_{\{s_t=k\}}|\mathcal{F}_T) + \sum_{t=1}^{T}\sum_{k,l=1}^{K}\log(p_{kl})E(\mathbf{1}_{\{s_{t-1}=k,s_t=l\}}|\mathcal{F}_T). \end{aligned} \tag{2.8.4}$$

It involves $E[(y_t - \boldsymbol{\beta}_t'\boldsymbol{x}_t)^2|\mathcal{F}_T]$, $E[(\boldsymbol{\beta}_t - \boldsymbol{z}(k))'\boldsymbol{V}(k)^{-1}(\boldsymbol{\beta}_t - \boldsymbol{z}(k))\mathbf{1}_{\{s_t=k\}}|\mathcal{F}_T]$, and the conditional probabilities $E(\mathbf{1}_{\{s_t=k\}}|\mathcal{F}_T) = P(s_t = k|\mathcal{F}_T)$ and $E(\mathbf{1}_{\{s_{t-1}=k,s_t=l\}}|\mathcal{F}_T) = P(s_{t-1} = k, s_t = l|\mathcal{F}_T)$. For the first conditional probability,

$$\begin{aligned} P(s_t = k|\mathcal{F}_T) &= \sum_{i=1}^{t} P(J_t^{(k)} = i|\mathcal{F}_T) = \sum_{i=1}^{t}\sum_{j=t}^{T} P(J_t^{(k)} = i, R_t^{(k)} = j|\mathcal{F}_T) \\ &= \sum_{1\le i\le t\le j\le T} P(C_{ij}^{(k)}|\mathcal{F}_T) = \sum_{1\le i\le t\le j\le T}\alpha_{ijt}^{(k)}. \end{aligned}$$

For the second conditional probability,

$$P(s_{t-1} = k, s_t = l|\mathcal{F}_T) = P(s_t = l|s_{t-1} = k, \mathcal{F}_T)P(s_{t-1} = k|\mathcal{F}_T). \tag{2.8.5}$$

From the above derivation, we know that

$$P(s_{t-1} = k|\mathcal{F}_T) = \sum_{1\leq i\leq t-1\leq j\leq T} \alpha^{(k)}_{i,j,t-1}.$$

Furthermore,

$$\begin{aligned}
&P(s_t = j|s_{t-1} = i, \mathcal{F}_T) = \frac{P(s_t = j, s_{t-1} = i, \mathcal{F}_T)}{P(s_{t-1} = i, \mathcal{F}_T)} \\
=&\frac{P(s_t = j, s_{t-1} = i, \mathcal{F}_t|\mathcal{F}_{t+1,T})}{P(s_{t-1} = i, \mathcal{F}_t|\mathcal{F}_{t+1,T})} \\
=&\frac{P(s_{t-1} = i, \mathcal{F}_t|s_t = j)P(s_t = j|\mathcal{F}_{t+1,T})}{P(s_{t-1} = i, \mathcal{F}_t|\mathcal{F}_{t+1,T})} \\
=&\frac{P(s_{t-1} = i, \mathcal{F}_t)P(s_t = j|s_{t-1} = i, \mathcal{F}_t)}{P(s_t = j)}\frac{P(s_t = j|\mathcal{F}_{t+1,n})}{P(s_{t-1} = i, \mathcal{F}_t|\mathcal{F}_{t+1,T})} \\
=&\frac{P(s_{t-1} = i, \mathcal{F}_t)}{P(s_{t-1} = i, \mathcal{F}_t|\mathcal{F}_{t+1,T})}\frac{P(s_t = j|s_{t-1} = i, y_t)P(s_t = j|\mathcal{F}_{t+1,T})}{P(s_t = j)} \\
\propto&\frac{P(s_t = j, y_t|s_{t-1} = i)P(s_t = j|\mathcal{F}_{t+1,T})}{P(s_t = j)} \\
=&\frac{f(y_t|s_t = j, s_{t-1} = i)P(s_t = j|s_{t-1} = i)\sum_{k=1}^K P(s_t = j, s_{t+1} = k|\mathcal{F}_{t+1,T})}{P(s_t = j)} \\
=&\frac{f(y_t|s_t = j, s_{t-1} = i)P(s_t = j|s_{t-1} = i)\sum_{k=1}^K P(s_t = j|s_{t+1} = k, \mathcal{F}_{t+1,T})P(s_{t+1} = k|\mathcal{F}_{t+1,T})}{P(s_t = j)} \\
=&\frac{f(y_t|s_t = j)P(s_t = j|s_{t-1} = i)\sum_{k=1}^K P(s_t = j|s_{t+1} = k)P(s_{t+1} = k|\mathcal{F}_{t+1,T})}{P(s_t = j)} \\
=&\frac{f^{(j)}_{t,t}/f^{(j)}_{0,0}p_{ij}\sum_{k=1}^K \widetilde{p}_{kj}\eta^k_{t+1}}{\pi_j}.
\end{aligned}$$

Thus

$$P(s_t = l | s_{t-1} = k, \mathcal{F}_T) = \frac{f_{t,t}^{(l)} / f_{0,0}^{(l)} p_{kl} \widetilde{P}_l' \eta_{t+1} / \pi_l}{\sum_{i=1}^K \left[f_{t,t}^{(i)} / f_{0,0}^{(i)} p_{ki} \widetilde{P}_i' \eta_{t+1} / \pi_i \right]}. \tag{2.8.6}$$

Plugging (2.8.6) into (2.8.5), we have

$$P(s_t = l, s_{t-1} = k | \mathcal{F}_T) = \frac{f_{t,t}^{(l)} / f_{0,0}^{(l)} p_{kl} \widetilde{P}_l' \eta_{t+1} / \pi_l}{\sum_{i=1}^K \left[f_{t,t}^{(i)} / f_{0,0}^{(i)} p_{ki} \widetilde{P}_i' \eta_{t+1} / \pi_i \right]} \sum_{1 \le i \le t-1 \le j \le T} \alpha_{i,j,t-1}^{(k)}.$$

Then the conditional probabilities are:

$$\begin{aligned} E(\mathbf{1}_{\{s_t = k\}} | \mathcal{F}_T) &= \sum_{1 \le i \le t \le j \le T} \alpha_{ijt}^{(k)}, \\ E(\mathbf{1}_{\{s_{t-1} = k, s_t = l\}} | \mathcal{F}_T) &= \frac{f_{t,t}^{(l)} / f_{0,0}^{(l)} p_{kl} \widetilde{P}_l' \eta_{t+1} / \pi_l}{\sum_{i=1}^K \left[f_{t,t}^{(i)} / f_{0,0}^{(i)} p_{ki} \widetilde{P}_i' \eta_{t+1} / \pi_i \right]} \sum_{1 \le i \le t-1 \le j \le T} \alpha_{i,j,t-1}^{(k)}. \end{aligned} \tag{2.8.7}$$

The M-step of the EM algorithm involves calculating the partial derivatives of (2.8.4) with respect to $\boldsymbol{\theta}$. The closed-form updating formulas are

$$\begin{aligned} \hat{p}_{kl,new} &= \frac{\sum_{t=2}^T P(s_{t-1} = k, s_t = l | \mathcal{F}_T, \hat{\boldsymbol{\theta}}_{old})}{\sum_{t=2}^T P(s_{t-1} = k | \mathcal{F}_T, \hat{\boldsymbol{\theta}}_{old})}, \\ \hat{\boldsymbol{z}}(k)_{new} &= \frac{\sum_{t=1}^T E(\boldsymbol{\beta}_t \mathbf{1}\{s_t = k\} | \mathcal{F}_T, \hat{\boldsymbol{\theta}}_{old})}{\sum_{t=1}^T P(s_t = k | \mathcal{F}_T, \hat{\boldsymbol{\theta}}_{old})}, \\ \hat{\boldsymbol{V}}(k)_{new} &= \frac{\sum_{t=1}^T E[(\boldsymbol{\beta}_t - \hat{\boldsymbol{z}}(k)_{old})(\boldsymbol{\beta}_t - \hat{\boldsymbol{z}}(k)_{old})' \mathbf{1}\{s_t = k\} | \mathcal{F}_T, \hat{\boldsymbol{\theta}}_{old}]}{\sum_{t=1}^T P(s_t = k | \mathcal{F}_T, \hat{\boldsymbol{\theta}}_{old})}, \\ \hat{\sigma^2}_{new} &= \frac{\sum_{t=1}^T E[(y_t - \boldsymbol{\beta}_t' \boldsymbol{x}_t)^2 | \mathcal{F}_T, \hat{\boldsymbol{\theta}}_{old}]}{T}. \end{aligned} \tag{2.8.8}$$

The conditional probabilities (2.8.7) can be used to calculate $\hat{p}_{kl,new}$ in (2.8.8). The other three posterior expectations needed in the updating formulas are $E(\boldsymbol{\beta}_t \mathbf{1}\{s_t = k\} | \mathcal{F}_T)$, $E[(\boldsymbol{\beta}_t -$

$\boldsymbol{z}(k))(\boldsymbol{\beta}_t - \boldsymbol{z}(k))'\mathbf{1}\{s_t = k\}|\mathcal{F}_T]$ and $E[(y_t - \boldsymbol{\beta}_t'\boldsymbol{x}_t)^2|\mathcal{F}_T]$. Note that

$$E(\boldsymbol{\beta}_t\mathbf{1}\{s_t = k\}|\mathcal{F}_T) = \sum_{1\le i\le t\le j\le T} E(\boldsymbol{\beta}_t|P(J_t^{(k)} = i, R_t^{(k)} = j, \mathcal{F}_T)P(J_t^{(k)} = i, R_t^{(k)} = j|\mathcal{F}_T),$$

in which $P(J_t^{(k)} = i, R_t^{(k)} = j|\mathcal{F}_T) = P(C_{ij}^{(k)}|\mathcal{F}_T) = \alpha_{ijt}^{(k)}$. Given $C_{ij}^{(k)}$ and $\mathcal{F}_T$, the conditional density of $\boldsymbol{\beta}_t$ is $g_{i,j}^{(k)}(\boldsymbol{\beta}_t)$, which is a normal distribution as given (2.2.1) with mean of $\boldsymbol{z}_{i,j}^{(k)}$ and variance of $\boldsymbol{V}_{i,j}^{(k)}$. Thus

$$E(\boldsymbol{\beta}_t\mathbf{1}\{s_t = k\}|\mathcal{F}_T) = \sum_{1\le i\le t\le j\le T} \alpha_{ijt}^{(k)}\boldsymbol{z}_{i,j}^{(k)}.$$

Similarly, for the second posterior expectation,

$$\begin{aligned}
&E[(\boldsymbol{\beta}_t - \boldsymbol{z}(k))(\boldsymbol{\beta}_t - \boldsymbol{z}(k))'\mathbf{1}\{s_t = k\}|\mathcal{F}_T] \\
=& \sum_{1\le i\le t\le j\le T} E[(\boldsymbol{\beta}_t - \boldsymbol{z}(k))(\boldsymbol{\beta}_t - \boldsymbol{z}(k))'|C_{ij}^{(k)}, \mathcal{F}_T]P(C_{ij}^{(k)}|\mathcal{F}_T) \\
=& \sum_{1\le i\le t\le j\le T} \alpha_{ijt}^{(k)} E[\boldsymbol{\beta}_t\boldsymbol{\beta}_t' - \boldsymbol{z}(k)\boldsymbol{\beta}_t' - \boldsymbol{\beta}_t\boldsymbol{z}(k)' + \boldsymbol{z}(k)\boldsymbol{z}(k)'|C_{ij}^{(k)}, \mathcal{F}_T] \\
=& \sum_{1\le i\le t\le j\le T} \alpha_{ijt}^{(k)}\Big\{E(\boldsymbol{\beta}_t\boldsymbol{\beta}_t'|C_{ij}^{(k)}, \mathcal{F}_T) - \boldsymbol{z}(k)E(\boldsymbol{\beta}_t'|C_{ij}^{(k)}, \mathcal{F}_T) - E(\boldsymbol{\beta}_t|C_{ij}^{(k)}, \mathcal{F}_T)\boldsymbol{z}(k)' + \boldsymbol{z}(k)\boldsymbol{z}(k)'\Big\}
\end{aligned}$$

Here $E(\boldsymbol{\beta}_t|C_{ij}^{(k)}, \mathcal{F}_T) = \boldsymbol{z}_{i,j}^{(k)}$ and

$$\begin{aligned}
&E(\boldsymbol{\beta}_t\boldsymbol{\beta}_t'|C_{ij}^{(k)}, \mathcal{F}_T) = var(\boldsymbol{\beta}_t|C_{ij}^{(k)}, \mathcal{F}_T) + E(\boldsymbol{\beta}_t|C_{ij}^{(k)}, \mathcal{F}_T)E(\boldsymbol{\beta}_t'|C_{ij}^{(k)}, \mathcal{F}_T) \\
=&\Big(\boldsymbol{V}_{i,j}^{(k)} + \boldsymbol{z}_{i,j}^{(k)}\boldsymbol{z}_{i,j}^{(k)\prime}\Big).
\end{aligned}$$

So

$$\begin{aligned}
&E[(\boldsymbol{\beta}_t - \boldsymbol{z}(k))(\boldsymbol{\beta}_t - \boldsymbol{z}(k))'\mathbf{1}\{s_t = k\}|\mathcal{F}_T] \\
=& \sum_{1\le i\le t\le j\le T} \alpha_{ijt}^{(k)}\Big(\boldsymbol{z}_{i,j}^{(k)}\boldsymbol{z}_{i,j}^{(k)\prime} + \boldsymbol{V}_{i,j}^{(k)} - \boldsymbol{z}(k)\boldsymbol{z}_{i,j}^{(k)\prime} - \boldsymbol{z}_{i,j}^{(k)}\boldsymbol{z}(k)' + \boldsymbol{z}(k)\boldsymbol{z}(k)'\Big)
\end{aligned}$$

For the last posterior expectation, according to (2.4.1) and the above proof,

$$
\begin{aligned}
&E[(y_t - \boldsymbol{\beta}_t'\boldsymbol{x}_t)^2|\mathcal{F}_T] \\
=&\sum_{k=1}^{K}\sum_{1\le i\le t\le j\le T}\alpha_{ijt}^{(k)}E[(y_t - \boldsymbol{\beta}_t'\boldsymbol{x}_t)^2|C_{ij}^{(k)},\mathcal{F}_T] \\
=&\sum_{k=1}^{K}\sum_{1\le i\le t\le j\le T}\alpha_{ijt}^{(k)}\Big\{y_t^2 - 2E(\boldsymbol{\beta}_t'|C_{ij}^{(k)},\mathcal{F}_T)\boldsymbol{x}_t + \boldsymbol{x}_t'E(\boldsymbol{\beta}_t\boldsymbol{\beta}_t'|C_{ij}^{(k)},\mathcal{F}_T)\boldsymbol{x}_t\Big\} \\
=&\sum_{k=1}^{K}\sum_{1\le i\le t\le j\le T}\alpha_{ijt}^{(k)}\Big\{y_t^2 - 2\boldsymbol{z}_{i,j}^{(k)\prime}\boldsymbol{x}_t + \boldsymbol{x}_t'(\boldsymbol{z}_{i,j}^{(k)}\boldsymbol{z}_{i,j}^{(k)\prime} + \boldsymbol{V}_{i,j}^{(k)})\boldsymbol{x}_t\Big\}
\end{aligned}
$$

In summary, the posterior expectations necessary for the updating formulas can be calculated as

$$
\begin{aligned}
&E(\boldsymbol{\beta}_t\mathbf{1}\{s_t=k\}|\mathcal{F}_T) = \sum_{1\le i\le t\le j\le T}\alpha_{ijt}^{(k)}\boldsymbol{z}_{i,j}^{(k)}, \\
&E[(\boldsymbol{\beta}_t - \boldsymbol{z}(k))(\boldsymbol{\beta}_t - \boldsymbol{z}(k))'\mathbf{1}\{s_t=k\}|\mathcal{F}_T] \\
&\quad= \sum_{1\le i\le t\le j\le T}\alpha_{ijt}^{(k)}\Big(\boldsymbol{z}_{i,j}^{(k)}\boldsymbol{z}_{i,j}^{(k)\prime} + \boldsymbol{V}_{i,j}^{(k)} - \boldsymbol{z}(k)\boldsymbol{z}_{i,j}^{(k)\prime} - \boldsymbol{z}_{i,j}^{(k)}\boldsymbol{z}(k)' + \boldsymbol{z}(k)\boldsymbol{z}(k)'\Big), \\
&E[(y_t - \boldsymbol{\beta}_t'\boldsymbol{x}_t)^2|\mathcal{F}_T] = \sum_{k=1}^{K}\sum_{1\le i\le t\le j\le T}\alpha_{ijt}^{(k)}\Big\{y_t^2 - 2\boldsymbol{z}_{i,j}^{(k)\prime}\boldsymbol{x}_t + \boldsymbol{x}_t'\boldsymbol{z}_{i,j}^{(k)}\boldsymbol{z}_{i,j}^{(k)\prime}\boldsymbol{x}_t + \boldsymbol{x}_t'\boldsymbol{V}_{i,j}^{(k)}\boldsymbol{x}_t\Big\},
\end{aligned}
\tag{2.8.9}
$$

which can be used to calculate $\hat{z}(k)_{new}$, $\hat{V}(k)_{new}$ and $\hat{\sigma^2}_{new}$ in (2.8.8). The iterative scheme (2.8.8) is carried out until convergence or until some prescribed upper bound on the number of iterations is reached.

To speed up the computations involved in the EM algorithm, one can use the BCMIX approximations in Section 2.6 instead of the full recursions to determine $\xi_{i,t}^{(k)}$, $\eta_{j,t}^{(k)}$, $\alpha_{ijt}^{(k)}$, etc. Applications to the simulation studies have shown that the EM estimates converge quite fast.

2.9 Implementation

We have shown the posterior distribution of parameter $\boldsymbol{\beta}_t$ is mixture of distributions. In this section, we describe in detail how to implement the algorithms, presenting explicit formulas. Let us start with a description of Bayes algorithm.

Step 1 Calculating $\boldsymbol{V}_{i,j}^{(k)}$ and $\boldsymbol{z}_{i,j}^{(k)}$. Similar to (2.2.1), given $\mathcal{F}_T$ and $C_{ij}^{(k)}, i < j$ we use

$$\boldsymbol{V}_{i,j}^{(k)} = \big(\boldsymbol{V}(k)^{-1} + \frac{\sum_{t=i}^{j} \boldsymbol{x}_t \boldsymbol{x}_t'}{\sigma^2}\big)^{-1},$$
$$\boldsymbol{z}_{i,j}^{(k)} = \boldsymbol{V}_{i,j}\big(\boldsymbol{V}(k)^{-1}\boldsymbol{z}(k) + \frac{\sum_{t=i}^{j} \boldsymbol{x}_j y_j}{\sigma^2}\big).$$

The results can be saved in two three-dimensional matrices for future calculation. More specifically, $g_{i,j}^{(k)}(\boldsymbol{\beta}_t)$ is calculated by (2.2.1). If there is no information other than $s_t = k$ is given, the conditional distribution is $g_{0,0}^{(k)}(\boldsymbol{\beta}_t)$ as in (2.2.2). Using $\boldsymbol{V}_{i,j}^{(k)}$, $\boldsymbol{z}_{i,j}^{(k)}$ and (2.2.5), we can also calculate the conditional densities $f_{0,0}^{(k)}$ and $f_{i,j}^{(k)}$. They are also used to calculate the smoothing estimate of $\boldsymbol{\beta}_t$ by (2.4.3).

Step 2 Calculating the forward filter (2.2.7) in a recursive manner.

(A) Start with $t = 1$. According to (2.2.7), we have

$$\xi_{1,1}^{(k)} \propto \xi_{1,1}^{(k)*} = \big(\sum_{l \neq k} \xi_0^{(l)} p_{lk}\big) f_{1,1}^{(k)} / f_{0,0}^{(k)}.$$

Substitute $\xi_0^{(l)}$ for $l \neq k$ by the stationary distribution π_l, use $f_{1,1}^{(k)}$ and $f_{0,0}^{(k)}$ to calculate $\big(\sum_{l \neq k} \pi_l p_{lk}\big) f_{1,1}^{(k)} / f_{0,0}^{(k)}$, which gives the value of $\xi_{1,1}^{(k)*}$, and therefore $\xi_{1,1}^{(k)} = \frac{\xi_{1,1}^{(k)*}}{\sum_{k=1}^{K} \xi_{1,1}^{(k)*}}$.

(B) At $t > 1$, calculate $\xi_{t,t}^{(k)*} = \big(\sum_{l \neq k} \xi_{t-1}^{(l)} p_{lk}\big) f_{t,t}^{(k)} / f_{0,0}^{(k)}$ directly. Use $\xi_{i,t-1}^{(k)}$ to calculate $\xi_{i,t}^{(k)*} = p_{kk} \xi_{i,t-1}^{(k)} f_{i,t}^{(k)} / f_{i,t-1}^{(k)}$ for $i < t$. Normalize $\xi_{i,t}^{(k)*}$ by dividing $\sum_{1 \leq i \leq t} \xi_{i,t}^{(k)*}$ to get $\xi_{i,t}^{(k)}$. Keep doing (B) until $t = T$.

Step 3 Calculating the backward filter (2.3.2) in a recursive manner. The backward filter $\eta_{j,t+1}^{(k)}$ can be calculated similarly by starting with $t = T$.

Step 4 Calculating the smoothing mixture weight (2.4.2) and the smoothing estimate (2.4.3).

One main challenge for this procedure is the computational complexity due to the space needed to save the matrices and number of weights which is increasing with t. There are two ways to increase the computational efficiency of this procedure.

The first modification is to implement the BCMIX approximation so that number of weights will be a fixed number M. The cost associated with the method is to keep the index set $\mathcal{K}_t^{(k)}$ for forward filter $\xi_{i,t}^{(k)}$ and $\widetilde{\mathcal{K}}_{t+1}^{(k)}$ for backward filter $\eta_{j,t+1}^{(k)}$. The basic procedure is similar to the preceding one with calculation of up to $M + 1$ weights for each stage t. The detailed procedure is as follows.

Step 1 Calculating $\boldsymbol{V}_{i,j}^{(k)}$ and $\boldsymbol{z}_{i,j}^{(k)}$.

Step 2 Calculating the BCMIX forward filter (2.7.2) in a recursive manner.

(A) For $1 \le t \le M$, use the Bayes procedure to calculate $\xi_{i,t}^{(k)*}$, $\xi_{i,t}^{(k)}$. The index set $\mathcal{K}_t^{(k)}$ at stage t is $\{1, \cdots, t\}$.

(B) At $t > M$, use new information at stage t to calculate $f_{t,t}^{(k)}$ and therefore $\xi_{t,t}^{(k)*} = \big(\sum_{l \neq k} \xi_{t-1}^{(l)} p_{lk}\big) f_{t,t}^{(k)} / f_{0,0}^{(k)}$. Use $\xi_{i,t-1}^{(k)}$ to calculate $\xi_{i,t}^{(k)*} = p_{kk} \xi_{i,t-1}^{(k)} f_{i,t}^{(k)} / f_{i,t-1}^{(k)}$ for $i \in \mathcal{K}_{t-1}^{(k)}$. Compare the weights in $\mathcal{K}_{t-1}^{(k)} - \{i_t^{(k)}\}$ and drop the smallest one. The remaining M weights form the new index set $\mathcal{K}_t^{(k)}$, and $\xi_{i,t}^{(k)} = \frac{\xi_{i,t}^{(k)*}}{\sum_{j \in \mathcal{K}_t^{(k)}} \xi_{j,t}^{(k)*}}$. Keep doing (B) until $t = T$, saving both the index sets and the BCMIX forward filters for future calculation.

Step 3 Calculating the BCMIX backward filter (2.7.4) in a recursive manner starting with $t = T$.

Step 4 Calculating the BCMIX smoothing mixture weight (2.4.2) and the smoothing estimate (2.4.3).

When one takes a second look at the BCMIX procedure, it is easy to find that only a small portion of the huge precalculated matrices $\boldsymbol{z}_{i,j}^{(k)}$ and $\boldsymbol{V}_{i,j}^{(k)}$ have been used. So it wastes a lot of space and time to calculate all the elements. However, we do not know which elements to use before calculating the index sets. A better idea is to calculate $\boldsymbol{z}_{i,j}^{(k)}$ and $\boldsymbol{V}_{i,j}^{(k)}$ when we need them. One more challenge is that the formulas to calculate $\boldsymbol{z}_{i,j}^{(k)}$ and $\boldsymbol{V}_{i,j}^{(k)}$ involve matrix inversion, which will take a long time to implement. Instead of calculating $\boldsymbol{z}_{i,j}^{(k)}$ and $\boldsymbol{V}_{i,j}^{(k)}$ directly, we can calculate $VI_{i,j}^{(k)} := (\boldsymbol{V}_{i,j}^{(k)})^{-1}$ and $VIZ_{i,j}^{(k)} := (\boldsymbol{V}_{i,j}^{(k)})^{-1}\boldsymbol{z}_{i,j}^{(k)}$ by the following simple recursive formulas if we know $VI_{i,j-1}^{(k)}$ and $VIZ_{i,j-1}^{(k)}$

$$
\begin{aligned}
VI_{i,j}^{(k)} &= \boldsymbol{V}(k)^{-1} + \frac{\sum_{t=i}^{j}\boldsymbol{x}_t\boldsymbol{x}_t'}{\sigma^2} = VI_{i,j-1}^{(k)} + \frac{\boldsymbol{x}_j\boldsymbol{x}_j'}{\sigma^2}, \\
VIZ_{i,j}^{(k)} &= \boldsymbol{V}(k)^{-1}\boldsymbol{z}(k) + \frac{\sum_{t=i}^{j}\boldsymbol{x}_t y_t}{\sigma^2} = VIZ_{i,j-1}^{(k)} + \frac{\boldsymbol{x}_j y_j}{\sigma^2}.
\end{aligned}
\tag{2.9.1}
$$

So the BCMIX algorithm can be further simplified by adding this recursive updating feature. The detailed procedure is as follows.

Step 1 Calculating the BCMIX forward filter (2.7.2) in a recursive manner from $t = 1$. Follow Step 2 in the above BCMIX algorithm. Assume at stage $t-1$ we have finished calculating $\xi_{i,t-1}^{(k)}$ and $\mathcal{K}_{t-1}^{(k)}$, and saved all the $VI_{i,t-1}^{(k)}$ and $VIZ_{i,t-1}^{(k)}$ for $i \in \mathcal{K}_{t-1}^{(k)}$. At stage t, $VI_{i,t}^{(k)}$ and $VIZ_{i,t}^{(k)}$ for $i \in \mathcal{K}_{t-1}^{(k)}$ can be calculated by (2.9.1). $VI_{t,t}^{(k)} = \boldsymbol{V}(k)^{-1} + \frac{\boldsymbol{x}_t\boldsymbol{x}_t'}{\sigma^2}$, and $VIZ_{t,t}^{(k)} = \boldsymbol{V}(k)^{-1}\boldsymbol{z}(k) + \frac{\boldsymbol{x}_t y_t}{\sigma^2}$. They are used to calculate $f_{i,t}^{(k)}$, $f_{t,t}^{(k)}$ by

$$
\begin{aligned}
f_{i,t}^{(k)} &= |\boldsymbol{V}_{it}^{(k)}|^{1/2}\exp\{\frac{1}{2}\boldsymbol{z}_{it}(k)'(\boldsymbol{V}_{it}^{(k)})^{-1}\boldsymbol{z}_{it}^{(k)}\} \\
&= |VI_{it}^{(k)}|^{-1/2}\exp\{\frac{1}{2}VIZ_{i,t}^{(k)\prime}(VI_{it}^{(k)})^{-1}VIZ_{i,t}^{(k)}\},
\end{aligned}
$$

and therefore $\xi_{i,t}^{(k)*}$ are calculated for all $i \in \{t\} \cup \mathcal{K}_{t-1}^{(k)}$. A small weight is dropped by the BCMIX rule and the remaining index set $\mathcal{K}_t^{(k)}$, $\xi_{i,t}^{(k)}$, $VI_{i,t}^{(k)}$ and $VIZ_{i,t}^{(k)}$ are saved.

Step 2 Calculating the BCMIX backward filter (2.7.4) in a recursive manner starting with $t = T$. If we know $VI^{(k)}_{i-1,j}$ and $VIZ^{(k)}_{i-1,j}$, and want to calculate $VI^{(k)}_{i,j}$ and $VIZ^{(k)}_{i,j}$ by the recursive formulas

$$VI^{(k)}_{i,j} = VI^{(k)}_{i-1,j} + \frac{\boldsymbol{x}_i \boldsymbol{x}_i'}{\sigma^2}, \quad VIZ^{(k)}_{i,j} = VIZ^{(k)}_{i-1,j} + \frac{\boldsymbol{x}_i y_i}{\sigma^2}.$$

Using these updating formulas, we can recursively calculate $VI^{(k)}_{t+1,j}$ and $VIZ^{(k)}_{t+1,j}$ for $j \in \widetilde{\mathcal{K}}^{(k)}_{t+1}$ and conduct Step 3 in the above BCMIX algorithm.

Step 3 Calculating the BCMIX smoothing mixture weight $\widetilde{\alpha}^{(k)}_{ijt}$ and the smoothing estimate $\hat{\boldsymbol{\beta}}_{t|T}$. We can evaluate $VI^{(k)}_{i,j}$ and $VIZ^{(k)}_{i,j}$ for $i \in \mathcal{K}^{(k)}_t, j \in \widetilde{\mathcal{K}}^{(k)}_{t+1}$ by

$$\begin{aligned}
VI^{(k)}_{i,j} &= \boldsymbol{V}(k)^{-1} + \frac{\sum_{l=i}^{j} \boldsymbol{x}_l \boldsymbol{x}_l'}{\sigma^2} \\
&= (\boldsymbol{V}(k)^{-1} + \frac{\sum_{l=i}^{t} \boldsymbol{x}_l \boldsymbol{x}_l'}{\sigma^2}) + (\boldsymbol{V}(k)^{-1} + \frac{\sum_{l=t+1}^{j} \boldsymbol{x}_l \boldsymbol{x}_l'}{\sigma^2}) - \boldsymbol{V}(k)^{-1} \\
&= VI^{(k)}_{i,t} + VI^{(k)}_{t+1,j} - \boldsymbol{V}(k)^{-1}, \\
VIZ^{(k)}_{i,j} &= \boldsymbol{V}(k)^{-1}\boldsymbol{z}(k) + \frac{\sum_{l=i}^{j} \boldsymbol{x}_l y_l}{\sigma^2} \\
&= (\boldsymbol{V}(k)^{-1}\boldsymbol{z}(k) + \frac{\sum_{l=i}^{t} \boldsymbol{x}_l y_l}{\sigma^2}) + (\boldsymbol{V}(k)^{-1}\boldsymbol{z}(k) + \frac{\sum_{l=t+1}^{j} \boldsymbol{x}_l y_l}{\sigma^2}) - \boldsymbol{V}(k)^{-1}\boldsymbol{z}(k) \\
&= VIZ^{(k)}_{i,t} + VIZ^{(k)}_{t+1,j} - \boldsymbol{V}(k)^{-1}\boldsymbol{z}(k).
\end{aligned} \tag{2.9.2}$$

The smoothing estimate of $\boldsymbol{\beta}_t$ can be calculated as $\hat{\boldsymbol{\beta}}_{t|T}$ defined in (2.7.6). Furthermore, the inference on regimes can be conducted using (2.5.1) by substituting α_{ijt} by $\widetilde{\alpha}_{ijt}$ calculated in Step 3.

Chapter 3

Numerical Studies

In this chapter, we first compare the performances of the Bayes and BCMIX estimates, through Monte Carlo simulations. The BCMIX approximation is shown to be statistically and computationally efficient. We then examine the relationship between the BCMIX performance and simulation settings. The choice of hyperparameters is also discussed. The last section applies a stochastic regime switching autoregressive model to analyze "Total Financial Assets - Assets - Balance Sheet of Nonfarm Nonfinancial Corporate Business (TFAAB-SNNCB)" data and "All Employees: Total nonfarm (PAYEMS)" data which has been used to demonstrate the motivation of this study in Section 1.3.

3.1 Simulation Results

There are three criteria by which we assess the performance of the estimation of parameter $\boldsymbol{\beta}_t$: the sum of squared errors, the Kullback-Leibler divergence and the L_2 errors between the true and estimated parameters. In our model, if a time series of T observations is generated, with a series of $\hat{\boldsymbol{\beta}}_t$ estimated, the SSE is defined by

$$SSE = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2 = \sum_{t=1}^{T}(y_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}_t)^2,$$

which measures the discrepancy between the observed and explained dependent variables. The Kullback-Leibler divergence is calculated by

$$KL(\boldsymbol{\beta}_t, \hat{\boldsymbol{\beta}}_t) = \frac{(\boldsymbol{x}_t'(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t))^2}{\sigma^2},$$

which measures the discrepancy between models with $\boldsymbol{\beta}_t$ and $\hat{\boldsymbol{\beta}}_t$. We use κ, the average of KL over the whole sample period, defined by

$$\kappa := \frac{1}{T}\sum_{t=1}^{T} KL(\boldsymbol{\beta}_t, \hat{\boldsymbol{\beta}}_t)$$

as our measure. When σ^2 is 1, the main difference between SSE and κ for our model is sum of the residuals. Simply rewriting right hand side of the SSE yields that

$$\begin{aligned} SSE &= \frac{1}{T}\sum_{t=1}^{T}(y_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}_t)^2 = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{x}_t'\boldsymbol{\beta}_t + \epsilon_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}_t)^2 \\ &= \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{x}_t'(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t))^2 + \frac{1}{T}\sum_{t=1}^{T}(\epsilon_t)^2 + \frac{2}{T}\sum_{t=1}^{T}\boldsymbol{x}_t'(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t)\epsilon_t \qquad (3.1.1) \\ &= \kappa + \frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 + \frac{2}{T}\sum_{t=1}^{T}\boldsymbol{x}_t'(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t)\epsilon_t. \end{aligned}$$

The third term should be small with an expectation of zero. The main difference comes from the second term, $\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2$. From this comparison, κ should be a more appropriate criterion since it measures the divergence between the true and estimated parameters. That is, SSE also includes $\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2$, which does not depend on the estimating accuracy. The L_2 error is defined by

$$L_2 = \frac{1}{T}\sum_{t=1}^{T}||\boldsymbol{\beta}_t - \hat{\boldsymbol{\beta}}_t||_2,$$

which measures the errors between the true and estimated parameters.

We also need to evaluate the performance of the smoothed probability $\hat{r}_{t|T}^{(k)}$ as discussed in

Section 2.5. We use this probability to provide assessment of the hidden state s_t belonging to regime k. However, this is not a logical variable only taking a value of 1 or 0, but a probability theoretically close to 1 or 0. When there is a transition from some regime to another one, the probability might show some fuzziness. An intuitive and simple way to make the inference on s_t is to compare the smoothed probability $\hat{r}_{t|T}^{(k)}$ with 0.5. If for any $1 \leq k \leq K$, $\hat{r}_{t|T}^{(k)} > 0.5$, we identity $s_t = k$. More specifically, to evaluate the performance of this procedure, we define an identification ratio as

$$IR := \frac{\sum_{t=1}^{T} \sum_{k=1}^{K} \mathbf{1}_{(\hat{r}_{t|T}^{(k)} > 0.5) \cap (s_t = k)}}{T},$$

where $\mathbf{1}$ denotes an indicator function, and T is the length of the sequence. If the true regime is k, and a probability reasonably close to 1, $\hat{r}_{t|T}^{(k)} > 0.5$, is obtained from the procedure, then $(\hat{r}_{t|T}^{(k)} > 0.5) \cap (s_t = k)$ is true, and the indicator function returns 1 for stage t.

3.1.1 Comparison between Bayes and BCMIX Estimates

As mentioned in Section 2.7, the Bayes method is accurate but computationally inefficient since the number of weights increases with t, resulting in rapidly increasing computational complexity and memory requirement in estimating β_t as t keeps increasing. The BCMIX approximation is much faster and does not need to save so many variables. This section is to compare the performances of the Bayes method described in Section 2.4 and the BCMIX approximation described in Section 2.7.

The section considers a simple stochastic regime switching autoregressive model, whose observation process is written as

$$y_t = \beta_t y_{t-1} + \epsilon_t,$$

in which $\epsilon_t \sim N(0, \sigma^2)$. There are two regimes, $K = 2$, and the values of the parameter $\boldsymbol{\beta}_t$ depend on the hidden state s_t. The hyperparameters consist of $(z(1), V(1))$, $(z(2), V(2))$, P and σ^2. In all the examples shown in this section, data are generated according to hyperparameter values: $(z(1), V(1)) = (0.5, 0.16)$, $(z(2), V(2)) = (-0.5, 0.16)$, $P = \begin{pmatrix} 0.999 & 0.001 \\ 0.001 & 0.999 \end{pmatrix}$, and $\sigma^2 = 1$. Furthermore, given s_t, $\boldsymbol{\beta}_t$ is a realization from a truncated Normal$(z(s_t), V(s_t))$ distribution such that $|\boldsymbol{\beta}_t| < 1$ to make the series stationary. We generate $N = 500$ series, each of length $T = 1000$, and consider s_t changing over time in four scenarios:

Scenario 1. There is only one transition from regime 1 to regime 2. $s_t = 1$ for $1 \leq t \leq 300$; $s_t = 2$ for $301 \leq t \leq 1000$.

Scenario 2. There is only one transition from regime 1 to regime 2. $s_t = 1$ for $1 \leq t \leq 500$; $s_t = 2$ for $501 \leq t \leq 1000$.

Scenario 3. There are two transitions between regime 1 and regime 2. $s_t = 1$ for $1 \leq t \leq 350$; $s_t = 2$ for $351 \leq t \leq 700$; $s_t = 1$ for $701 \leq t \leq 1000$.

Scenario 4. There are three transitions between regime 1 and regime 2. $s_t = 1$ for $1 \leq t \leq 200$; $s_t = 2$ for $201 \leq t \leq 500$; $s_t = 1$ for $501 \leq t \leq 600$; $s_t = 2$ for $601 \leq t \leq 1000$.

In each scenario, we assume the true hyperparameters are given, and compute both the BCMIX and Bayes estimates. As mentioned in Section 2.7, the performance of the BCMIX procedure depends on the specification of M and m. This dependence is examined here, choosing $M = 2m$ and M =10, 20, 30 and 40. Furthermore, to access the performance of both methods, we consider a simple benchmark in which the hidden state is known so that the Bayes estimates of β_t between two transitions are given by the standard Bayesian formulas for normal populations (Section 2.7 of Box and Tiao (1973)). This is called a "fictitious Bayes" estimate. Tables 3.1, 3.2 and 3.3 compare fictitious Bayes estimate (fBayes), Bayes estimate (Bayes), and the BCMIX estimate (BCMIX) in terms of the SSE, κ and L_2 respectively.

The first three columns in Table 3.1 show that both the Bayes and BCMIX(10,5) es-

Table 3.1: Performance of Sum of squared errors (SSE) for fBayes, Bayes and BCMIX estimates. Standard errors are given in parentheses below the estimates.

			BCMIX			
Scenarios	*fBayes*	*Bayes*	(10,5)	(20,10)	(30,15)	(40,20)
Scenario 1	0.998	0.995	0.995	0.996	0.996	0.996
	(2.05E-03)	(2.06E-03)	(2.06E-03)	(2.06E-03)	(2.06E-03)	(2.06E-03)
Scenario 2	0.998	0.995	0.996	0.996	0.996	0.996
	(2.05E-03)	(2.06E-03)	(2.06E-03)	(2.06E-03)	(2.06E-03)	(2.06E-03)
Scenario 3	0.997	0.993	0.993	0.993	0.993	0.993
	(2.05E-03)	(2.04E-03)	(2.04E-03)	(2.04E-03)	(2.04E-03)	(2.04E-03)
Scenario 4	0.996	0.990	0.990	0.990	0.990	0.990
	(2.06E-03)	(2.05E-03)	(2.05E-03)	(2.05E-03)	(2.05E-03)	(2.05E-03)

timates show smaller SSE than the fictitious Bayes estimate. As discussed in Section 3.1, SSE is not an appropriate criterion for evaluating the performance of different estimation procedures. But this comparison illustrates the effectiveness of both Bayes and BCMIX estimates. Furthermore, the relative difference between BCMIX(10,5) estimate and Bayes estimate in SSE is less than 0.05%. But the Bayes estimate takes far more time to compute. The last four columns in Table 3.1 show that the average SSE over 500 sequences changes with respect to the different values of M and m. As mentioned in Section 2.7, the approximation should improve as M and m become larger since more filters are kept at each stage. However, based on Table 3.1, we cannot see the trend clearly although in each scenario all the BCMIX estimates have similar SSE. This observation shows two things. First, SSE is not an accurate measure. Second, the BCMIX procedure is very robust for this model. The estimation results do not change dramatically when M and m are getting larger.

Table 3.2 shows the comparison in terms of Kullback-Leibler divergence, which is a more accurate measure of the difference between the true and estimated parameters. As the benchmark, the fictitious Bayes estimate gives significantly better result than the other

Table 3.2: Performance of Kullback-Leibler divergence ($10^3\kappa$) for fBayes, Bayes and BCMIX estimates. Standard errors are given in parentheses below the estimates.

Scenarios	*fBayes*	*Bayes*	*BCMIX* (10,5)	(20,10)	(30,15)	(40,20)
Scenario 1	2.019 (9.13E-02)	3.963 (1.40E-01)	3.945 (1.39E-01)	3.934 (1.41E-01)	3.935 (1.41E-01)	3.932 (1.41E-01)
Scenario 2	2.031 (9.29E-02)	4.053 (1.60E-01)	4.045 (1.60E-01)	4.025 (1.61E-01)	4.021 (1.61E-01)	4.023 (1.61E-01)
Scenario 3	3.025 (1.07E-01)	6.580 (1.89E-01)	6.860 (2.00E-01)	6.610 (1.90E-01)	6.571 (1.89E-01)	6.568 (1.90E-01)
Scenario 4	3.925 (1.25E-01)	9.985 (2.80E-01)	10.780 (3.30E-01)	10.292 (3.10E-01)	10.151 (3.02E-01)	10.096 (2.98E-01)

two estimates. Comparing the first three columns, we can see that when there are 2 or 3 transitions, the BCMIX(10,5) estimate is less accurate than Bayes estimate with a relative difference in κ of less than 7%. Comparing Bayes estimate and BCMIX(20,10) estimate, the relative difference in κ is less than 3% in all scenarios. The last four columns in Table 3.2 show that κ becomes smaller when the values of M and m become larger. The most significant improvement of BCMIX occurs when M changes from 10 to 20.

Table 3.3 shows the comparison in terms of the L_2 errors between the estimated and true parameters. As the benchmark, the fictitious Bayes estimate gives significantly better result than the other two estimates. Comparing the first three columns, we can see that in three scenarios the BCMIX(10,5) estimate is slightly less accurate than Bayes estimate with a relative difference in L_2 of less than 4%. Comparing Bayes estimate and BCMIX(20,10) estimate, the relative difference in L_2 is less than 1% in all scenarios. The results of Table 3.3 and Table 3.2 verify that BCMIX estimate is comparable to the Bayes estimate in statistical efficiency but has much lower computational complexity. Moreover, the last four columns in Table 3.3 show that L_2 becomes smaller when the values of M and m become larger. The

Table 3.3: Performance of L_2 errors ($10^3 L_2$) for fBayes, Bayes and BCMIX estimates. Standard errors are given in parentheses below the estimates.

Scenarios	*fBayes*	*Bayes*	*BCMIX* (10,5)	(20,10)	(30,15)	(40,20)
Scenario 1	1.023	1.778	1.777	1.770	1.770	1.770
	(2.62E-02)	(3.19E-02)	(3.20E-02)	(3.21E-02)	(3.21E-02)	(3.20E-02)
Scenario 2	1.018	1.766	1.770	1.758	1.757	1.758
	(2.63E-02)	(3.36E-02)	(3.37E-02)	(3.37E-02)	(3.36E-02)	(3.36E-02)
Scenario 3	1.307	2.389	2.438	2.395	2.388	2.387
	(2.69E-02)	(3.27E-02)	(3.43E-02)	(3.29E-02)	(3.27E-02)	(3.27E-02)
Scenario 4	1.538	3.022	3.121	3.054	3.037	3.031
	(2.74E-02)	(3.97E-02)	(4.36E-02)	(4.19E-02)	(4.12E-02)	(4.09E-02)

most significant improvement of BCMIX occurs when M changes from 10 to 20. Based on the observations of Table 3.2 and Table 3.3,the combination of $M = 20$ and $m = 10$ is the best choice for the BCMIX procedure in our simulation studies.

Table 3.4 compares the Bayes and the BCMIX estimates in terms of identification ratio (IR). The first two columns show that both the Bayes and BCMIX methods give an average IR of larger than 97%, with slight absolute differences of less than 0.7%. This further justifies the effectiveness of the BCMIX procedure. The last four columns in Table 3.4 show that ratios increase when the values of M and m become larger, but become stable at some point. For example, in Scenario 1, BCMIX(10,5) gives an IR of 97%, BCMIX(20,10) improves the IR to 97.3%. After that, increasing M and m does not change IR anymore and the IR stabilizes at 97.2%. Again, in most cases the most significant improvement of BCMIX occurs when M changes from 10 to 20. In the next section, we will conduct more simulation studies using a more complex model with more scenarios, more sequences (N is larger) and longer series (T is larger). The Bayes method will be computationally prohibitive. We will only use the BCMIX procedure with $M = 20$ and $m = 10$ to estimate the smoothing parameter and make

Table 3.4: Performance of Identification Ratio (IR) for Bayes and BCMIX estimates. Standard errors are given in parentheses below the estimates.

		BCMIX			
Scenarios	*Bayes*	(10,5)	(20,10)	(30,15)	(40,20)
Scenario 1	0.973	0.970	0.973	0.972	0.972
	(6.51E-03)	(6.78E-03)	(6.54E-03)	(6.64E-03)	(6.64E-03)
Scenario 2	0.971	0.973	0.972	0.972	0.973
	(6.59E-03)	(6.52E-03)	(6.60E-03)	(6.62E-03)	(6.57E-03)
Scenario 3	0.974	0.975	0.977	0.975	0.975
	(5.36E-03)	(5.27E-03)	(5.25E-03)	(5.35E-03)	(5.34E-03)
Scenario 4	0.972	0.966	0.970	0.971	0.974
	(3.76E-03)	(4.73E-03)	(4.53E-03)	(4.14E-03)	(3.94E-03)

inference on regimes.

Let us take a second look at Tables 3.1, 3.2, 3.3 and 3.4 to compare the results of different scenarios. Scenarios 1 and 2 both experience one transition, but at different times. The transition from regime 1 to regime 2 occurs at $t = 300$, about one third of the series, for Scenario 1, but at $t = 500$, right in the middle of the series, for Scenario 2. Both Bayes and BCMIX estimates work slightly better in Scenario 1 in terms of SSE, L_2, κ and IR. In Scenario 1, although the numbers of observations from different regimes are not symmetric, the methods provide efficient estimates. The estimating errors become larger in Scenarios 3 and 4 when there are more transitions and the distances between two successive transitions become smaller. As shown in Table 3.3, both values of L_2 and the associated standard errors become larger when there are more transitions. In Scenario 3, $10^3 L_2$ of fictitious Bayes estimate is only 1.307, while those of Bayes and BCMIX(20,10) estimates are 2.389 and 2.395. More significant differences are shown in Table 3.2. In Scenario 4, $10^3\kappa$ of fictitious Bayes estimate is only 3.925, while those of Bayes and BCMIX(20,10) estimates are 9.985 and 10.292. However, as shown in Table 3.4, the methods can identify the correct hidden state and regime more efficiently when there are more transitions. The associated standard errors become smaller in the last two scenarios. So the Bayes and BCMIX procedures are robust and efficient to make inference on regimes.

To visualize the simulation results, here we show some figures. Figure 3.1 shows a randomly selected simulation path y_t in each scenario. From the figure we find some changing patterns in each series. For example, in the second plot (Scenario 2), the first half of the series shows more fluctuations in magnitude, while the second half is closer to zero, indicating a change in the pattern (regression coefficient) around $t = 500$. The first plot also shows a change in pattern, but before $t = 500$. In Scenario 4, since there are more transitions, we can see the changes in the series, but cannot tell the number and locations of the transitions by observing the series. Figure 3.2 shows the true β_t and estimated $\hat{\beta}_{t|T}$ of the corresponding series. Before we analyze the estimates, let us observe the true parameters in different regimes to have a better understanding of the model. In the last plot (Scenario 4), there are two regimes and three transitions from regime 1 to 2, then back to regime 1, and then to regime 2 again. However, values of β_t within each regime are not the same. For regime 1, $\beta_t = 0.92$ before the first transition, and $\beta_t = 0.60$ between the second and third transitions. For regime 2, $\beta_t = -0.39$ between the first and second transitions, and $\beta_t = -0.17$ after the third transition. This is the new feature of our model as specified in assumption (A3). Different from the classic regime switching model in which β_t is a constant within each regime, in our model β_t is a random variable following some distribution within each regime. Now let us look at the estimation results. In all plots, we cannot tell the difference between Bayes estimate (dotted line) and BCMIX estimate (dashed line). In the first two scenarios (top two plots), the estimated parameters are very close to the true β_t. In the last two scenarios (bottom two plots) there are significant deviations between β_t and $\hat{\beta}_{t|T}$.

Figure 3.3 shows the true and estimated $P(s_t = 1)$ of each series. Specifically, if the

Figure 3.1: A selected series y_t in Scenarios 1 (top-left), 2 (top-right), 3 (bottom-left) and 4 (bottom-right).

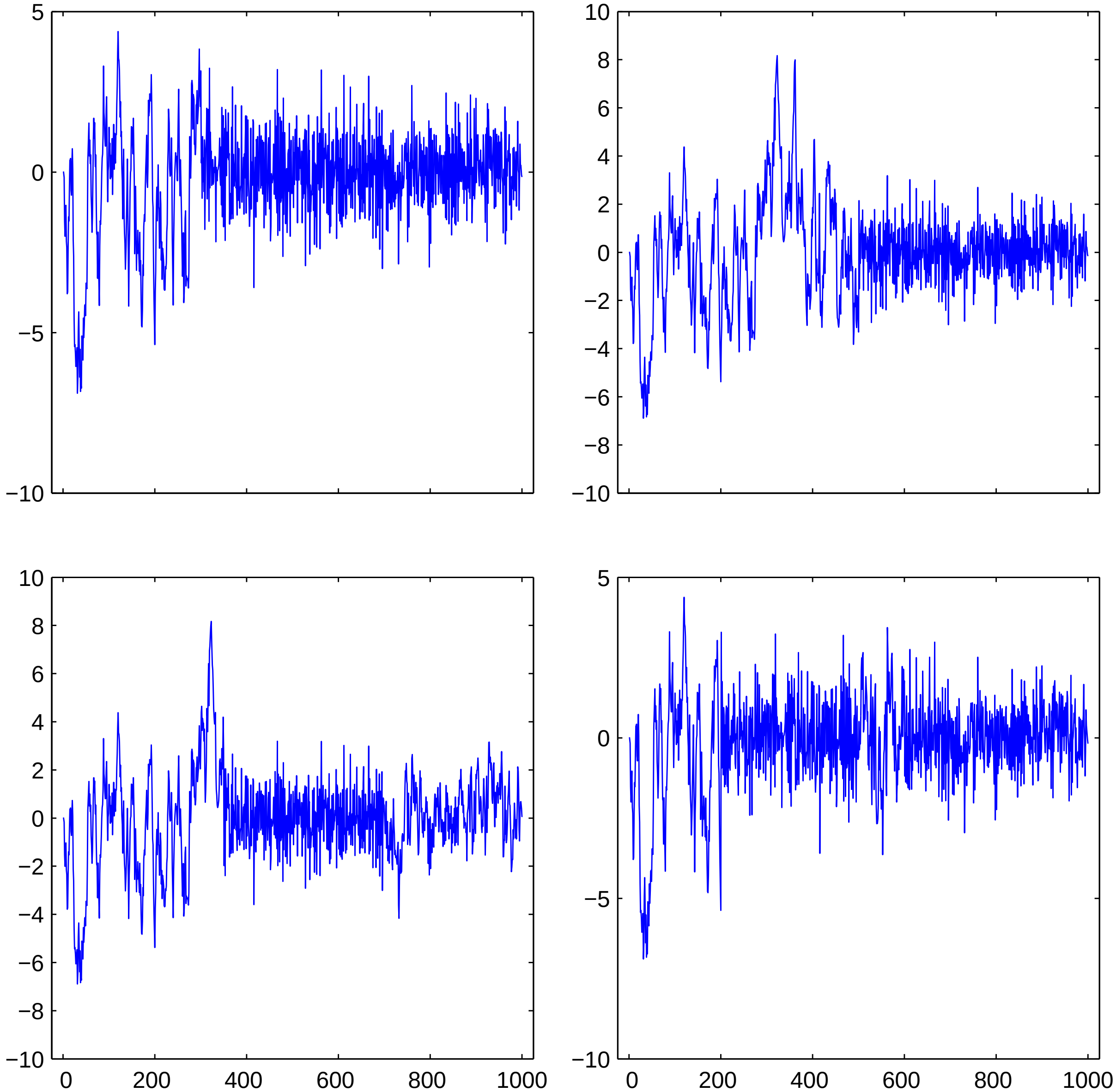


Figure 3.2: Bayes estimates (dotted line), BCMIX estimates (dashed line) of $\hat{\beta}_{t|T}$ and true β_t (solid line) of the selected series in Scenarios 1 (top-left), 2 (top-right), 3 (bottom-left) and 4 (bottom-right)

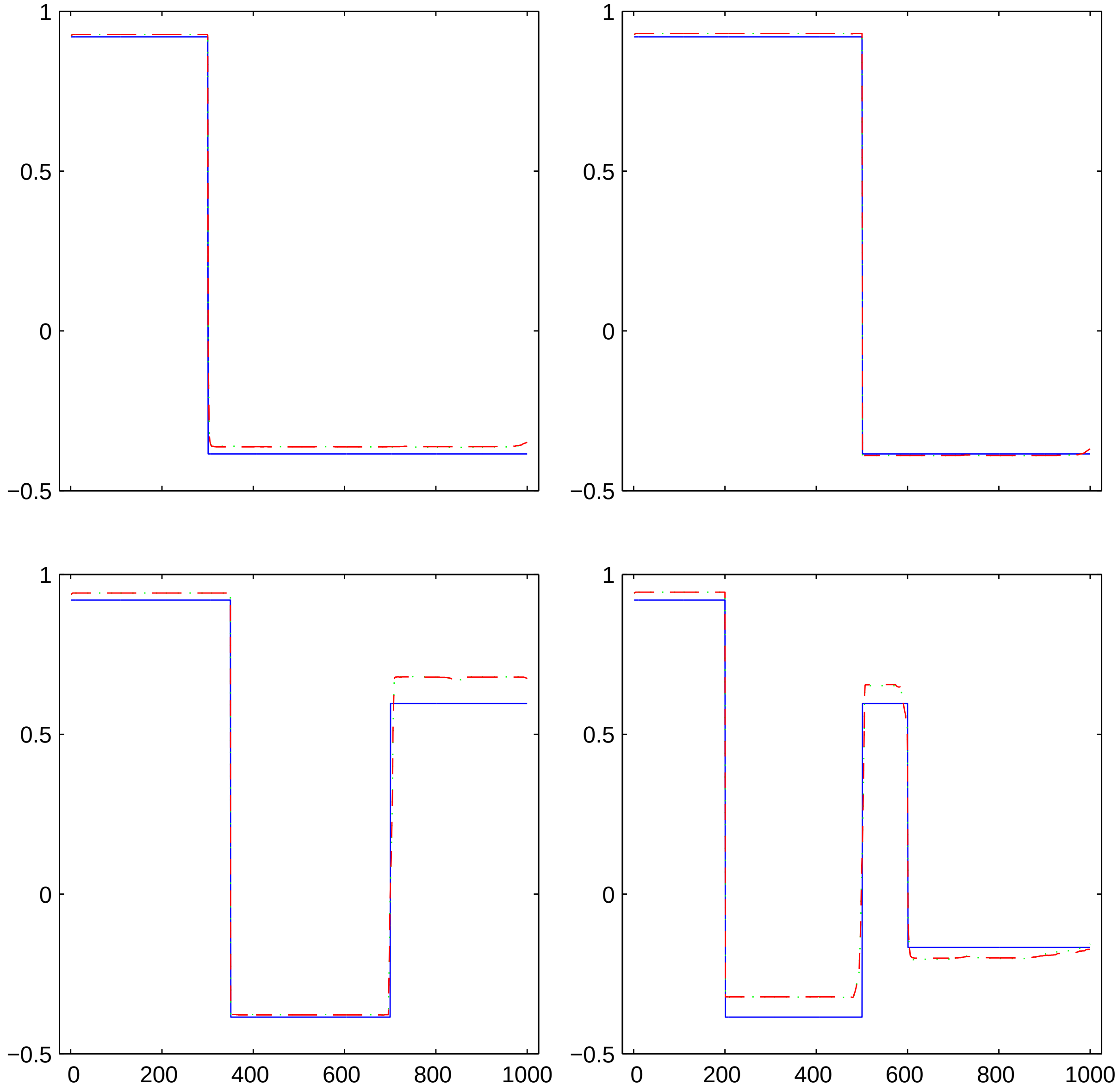
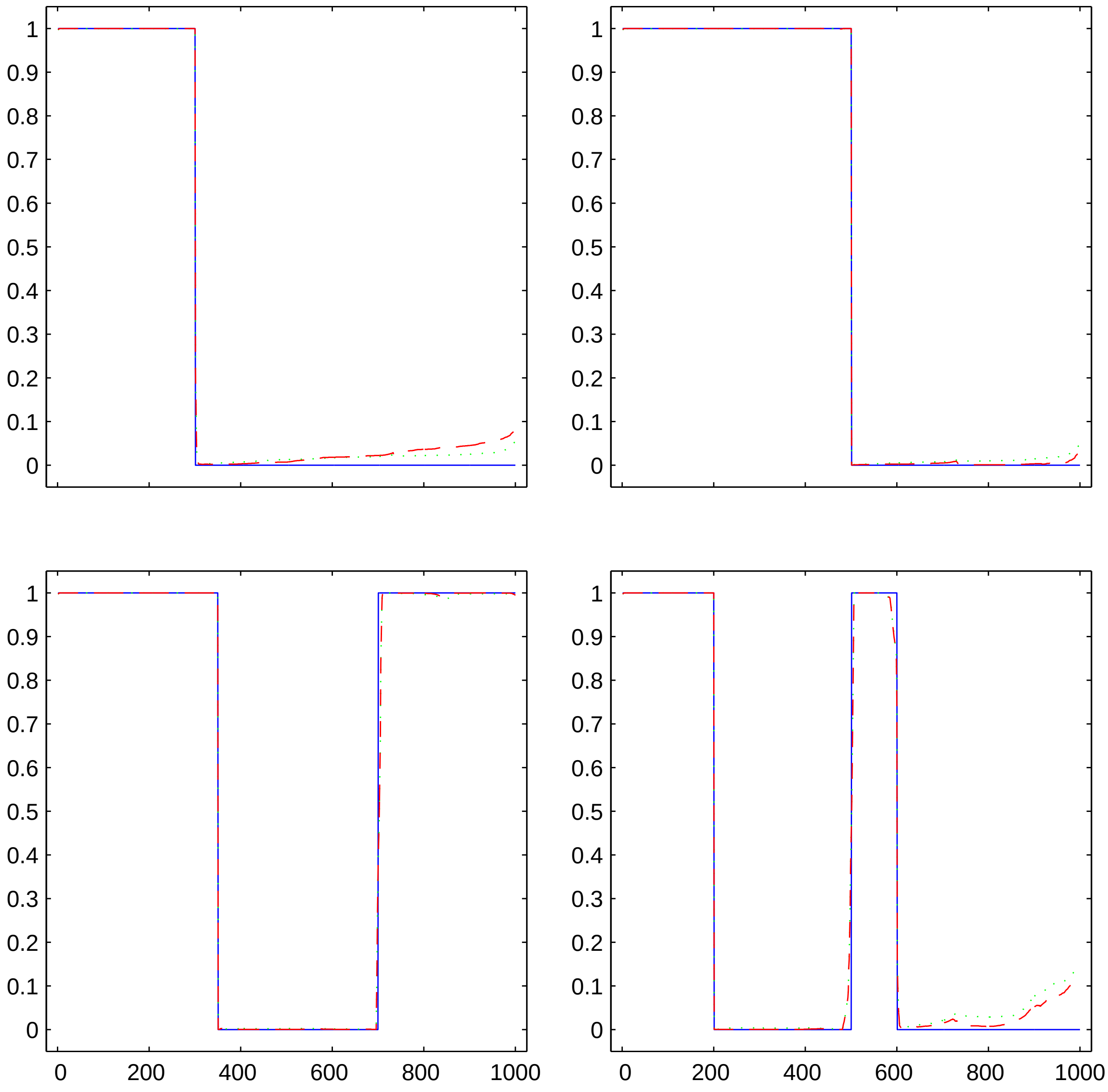


Figure 3.3: Bayes estimates (dotted line), BCMIX estimates (dashed line) of $\hat{r}_{t|T}^{(1)}$ and true $P(s_t = 1)$ (solid line) of the selected series in Scenarios 1 (top-left), 2 (top-right), 3 (bottom-left) and 4 (bottom-right)



true regime is 1, the true probability of $P(s_t = 1) = 1$; if the true regime is 2, the true probability of $P(s_t = 1) = 0$. There are two regimes in our simulation setup, hence $P(s_t = 2) = 1 - P(s_t = 1)$ for $1 \leq t \leq T$. So we only show the probability of regime 1. Slight differences between the estimated probabilities in Bayesian procedure (dotted line) and BCMIX procedure (dashed line) can be observed. When there are enough observations between two consecutive transitions, as in the first three plots (Scenarios 1, 2 and 3), both procedures capture the transitions very quickly. But when there are more frequent transitions, as in the last plot, the estimated probabilities show some fuzziness around transitions. That is why we use $P(s_t = 1) > 0.5$ to make inference on the unknown regime.

3.1.2 Simulation Settings

In this section, we will examine the effects of different simulation settings on the estimates. As mentioned in the last section, we will only use the BCMIX procedure in this section for large scale simulation studies. In this section we consider a stochastic regime switching autoregressive model

$$y_t = \alpha_t + \beta_t y_{t-1} + \epsilon_t, \tag{3.1.2}$$

in which $\epsilon_t \sim N(0, \sigma^2)$. There are two states, $K = 2$, and the parameter $(\alpha_t, \beta_t)'$ is two-dimensional. $z(1) = (0.2, -0.3)'$, $z(2) = (0.5, -0.5)'$, $V(1) = V(2) = \begin{pmatrix} 0.16 & 0 \\ 0 & 0.16 \end{pmatrix}$. Given s_t, β_t is a realization from a truncated Normal distribution such that $|\beta_t| < 1$ to make the series stationary. The transition matrix is $P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$, which has the following settings:

Scenario 1. $(p, q) = (0.001, 0.001)$.

Scenario 2. $(p, q) = (0.002, 0.001)$.

Scenario 3. $(p, q) = (0.002, 0.002)$.

Scenario 4. $(p, q) = (0.004, 0.001)$.

Scenario 5. $(p, q) = (0.004, 0.002)$.

Scenario 6. $(p, q) = (0.008, 0.004)$.

Scenario 7. $(p, q) = (0.008, 0.008)$.

Scenario 8. $(p, q) = (0.016, 0.008)$.

Scenario 9. $(p, q) = (0.016, 0.016)$.

$\sigma^2 = 1$, $N = 500$, and T takes the values of 3000, 4000, 5000, 6000, 7000 and 8000 for each scenario. In each scenario, we assume that the true hyperparameters are unknown. The hyperparameters are estimated by the EM algorithm described in Section 2.8 until convergence. Then the estimates are computed. The BCMIX procedure with $M = 20$ and $m = 10$ is used to estimate the smoothing parameters and give inference on the regime. Tables 3.5, 3.6 and 3.7 compare the estimates in different scenarios in terms of the SSE, κ and L_2 respectively.

Let us look at Tables 3.5, 3.6 and 3.7 column by column. In each column, the sample size T is fixed, but p and q are changing. Therefore the transition matrix $P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$ is different for each row. From top to bottom p and q become larger, so more transitions should be expected. Presumably the errors are getting larger when the coefficients are more volatile and experience more transitions. Table 3.6 shows a similar trend: the larger are p and q, the larger are κ. For example, when $T = 8000$, $p = 0.008$, and $q = 0.008$, $10^3\kappa$ is 13.367. The quantity $10^3\kappa$ decreases to 11.037 when p remains at 0.008 and q changes to 0.004, and decreases to 7.264 when q and q change to 0.004 and 0.002 respectively. The quantity increases to 17.267 when both p and q become 0.016. Table 3.7 illustrates the same trend:

Table 3.5: Performance of Sum of squared errors (SSE) for BCMIX estimates. Standard errors are given in parentheses.

Scenarios	$T = 3000$	$T = 4000$	$T = 5000$	$T = 6000$	$T = 7000$	$T = 8000$
$p = 0.001$ $q = 0.001$	0.995 (1.17E-03)	0.995 (1.03E-03)	0.995 (9.56E-04)	0.995 (8.70E-04)	0.995 (7.78E-04)	0.995 (7.31E-04)
$p = 0.002$ $q = 0.001$	0.994 (1.17E-03)	0.994 (1.02E-03)	0.994 (9.55E-04)	0.994 (8.65E-04)	0.994 (7.72E-04)	0.994 (7.27E-04)
$p = 0.002$ $q = 0.002$	0.993 (1.17E-03)	0.993 (1.03E-03)	0.992 (9.48E-04)	0.993 (8.66E-04)	0.993 (7.68E-04)	0.993 (7.24E-04)
$p = 0.004$ $q = 0.001$	0.994 (1.17E-03)	0.994 (1.02E-03)	0.993 (9.52E-04)	0.993 (8.65E-04)	0.994 (7.77E-04)	0.994 (7.29E-04)
$p = 0.004$ $q = 0.002$	0.992 (1.17E-03)	0.991 (1.03E-03)	0.991 (9.48E-04)	0.991 (8.63E-04)	0.991 (7.70E-04)	0.991 (7.27E-04)
$p = 0.008$ $q = 0.004$	0.988 (1.17E-03)	0.988 (1.02E-03)	0.988 (9.56E-04)	0.987 (8.74E-04)	0.988 (7.75E-04)	0.988 (7.36E-04)
$p = 0.008$ $q = 0.008$	0.986 (1.17E-03)	0.986 (1.01E-03)	0.985 (9.48E-04)	0.985 (8.67E-04)	0.985 (7.73E-04)	0.985 (7.30E-04)
$p = 0.016$ $q = 0.008$	0.984 (1.16E-03)	0.984 (1.02E-03)	0.984 (9.56E-04)	0.983 (8.66E-04)	0.984 (7.72E-04)	0.984 (7.26E-04)
$p = 0.016$ $q = 0.016$	0.982 (1.16E-03)	0.982 (1.02E-03)	0.981 (9.45E-04)	0.981 (8.68E-04)	0.982 (7.63E-04)	0.982 (7.25E-04)

the larger p and q become, the larger L_2 and standard errors are obtained. For example, when $T = 3000$, $p = 0.004$, and $q = 0.002$, $10^3 L_2$ is 1.671. The quantity $10^3 L_2$ decreases to 1.401 when p remains at 0.004 and q changes to 0.001, and decreases to 1.493 when q remains at 0.002 and p changes to 0.002. The quantity increases to 2.088 when p and q become 0.008 and 0.004 respectively. However, Table 3.5 shows an opposite trend: the larger are p and q, the smaller are SSE. In (3.1.1) it is shown that SSE equals κ plus two other terms. The second term is the sum of squared residuals, which should not change with p and q. The third term is small, but might affect SSE. One possible explanation is that in our simulation studies, the coefficients are assumed to follow two distinct Normal distributions with means of $(0.2, -0.3)'$ and $(0.5, -0.5)'$. So when there are more transitions, the coefficients change signs more frequently, and the autoregressive model should show mean-reverting tendency. Thus the magnitudes of $\boldsymbol{x}_t'\boldsymbol{\beta}_t\epsilon_t$ and $\boldsymbol{x}_t'\hat{\boldsymbol{\beta}}_t\epsilon_t$ in our simulation study should be smaller. Thus the third term is smaller when there are more transitions. This explains the opposite trend.

From Tables 3.5, 3.6 and 3.7, we can observe the effects of sample size on the performance. Table 3.5 shows that SSE is almost constant when T increases, with decreasing standard errors. As shown in Table 3.6, κ has a tendency of becoming smaller with T. When T changes from 3000 to 4000, there is a significant decrease in κ. After that, the decreasing trend is not clear. Table 3.7 shows that L_2 decreases when T increases, with decreasing standard errors. The quantity κ is the average Kullback-Leibler divergence which measures the difference between the model with true parameter $\boldsymbol{x}_t'\boldsymbol{\beta}_t$ and the model with estimated parameter $\boldsymbol{x}_t'\hat{\boldsymbol{\beta}}_{t|T}$. The quantity L_2 is the average difference between the true parameter $\boldsymbol{\beta}_t$ and the estimate $\hat{\boldsymbol{\beta}}_{t|T}$. When p and q are fixed, the errors tend to become smaller with the sample size. When the sample size is large enough, the measured divergence κ becomes stable.

Table 3.8 summarizes the identification ratio (IR) in each scenario. Most ratios are

Table 3.6: Performance of average Kullback-Leibler divergence ($10^3\kappa$) for BCMIX estimates. Standard errors are given in parentheses.

Scenarios	$T = 3000$	$T = 4000$	$T = 5000$	$T = 6000$	$T = 7000$	$T = 8000$
$p = 0.001$	3.873	3.709	3.509	3.539	3.438	3.536
$q = 0.001$	(1.33E-01)	(1.04E-01)	(8.06E-02)	(8.63E-02)	(7.48E-02)	(7.36E-02)
$p = 0.002$	4.837	4.563	4.473	4.474	4.414	4.408
$q = 0.001$	(1.48E-01)	(1.09E-01)	(9.56E-02)	(9.60E-02)	(8.29E-02)	(8.28E-02)
$p = 0.002$	6.197	6.107	6.011	5.901	5.816	5.899
$q = 0.002$	(1.58E-01)	(1.22E-01)	(1.13E-01)	(1.01E-01)	(9.20E-02)	(8.79E-02)
$p = 0.004$	5.605	5.228	5.307	5.086	5.105	5.074
$q = 0.001$	(1.51E-01)	(1.30E-01)	(1.14E-01)	(1.00E-01)	(9.17E-02)	(8.68E-02)
$p = 0.004$	7.660	7.290	7.421	7.245	7.106	7.264
$q = 0.002$	(1.67E-01)	(1.37E-01)	(1.26E-01)	(1.12E-01)	(1.03E-01)	(9.65E-02)
$p = 0.008$	11.462	10.888	11.130	11.120	10.974	11.037
$q = 0.004$	(1.86E-01)	(1.62E-01)	(1.35E-01)	(1.34E-01)	(1.16E-01)	(1.14E-01)
$p = 0.008$	13.918	13.409	13.353	13.462	13.356	13.367
$q = 0.008$	(1.94E-01)	(1.68E-01)	(1.38E-01)	(1.28E-01)	(1.22E-01)	(1.18E-01)
$p = 0.016$	15.371	15.244	15.083	15.239	14.991	15.088
$q = 0.008$	(2.04E-01)	(1.76E-01)	(1.49E-01)	(1.47E-01)	(1.26E-01)	(1.18E-01)
$p = 0.016$	17.750	17.361	17.414	17.331	17.313	17.267
$q = 0.016$	(2.12E-01)	(1.73E-01)	(1.60E-01)	(1.44E-01)	(1.38E-01)	(1.20E-01)

Table 3.7: Performance of L_2 errors ($10^3 L_2$) for BCMIX estimates. Standard errors are given in parentheses.

Scenarios	$T = 3000$	$T = 4000$	$T = 5000$	$T = 6000$	$T = 7000$	$T = 8000$
$p = 0.001$ $q = 0.001$	1.135 (2.12E-02)	0.983 (1.61E-02)	0.874 (1.11E-02)	0.801 (1.07E-02)	0.739 (8.94E-03)	0.700 (7.89E-03)
$p = 0.002$ $q = 0.001$	1.284 (2.18E-02)	1.111 (1.51E-02)	0.991 (1.22E-02)	0.918 (1.09E-02)	0.841 (8.86E-03)	0.786 (7.65E-03)
$p = 0.002$ $q = 0.002$	1.493 (2.08E-02)	1.302 (1.42E-02)	1.167 (1.12E-02)	1.067 (1.01E-02)	0.975 (8.10E-03)	0.918 (7.15E-03)
$p = 0.004$ $q = 0.001$	1.401 (2.09E-02)	1.196 (1.67E-02)	1.073 (1.28E-02)	0.980 (1.04E-02)	0.905 (9.09E-03)	0.842 (7.88E-03)
$p = 0.004$ $q = 0.002$	1.671 (1.95E-02)	1.430 (1.42E-02)	1.298 (1.12E-02)	1.183 (9.98E-03)	1.081 (8.17E-03)	1.022 (7.06E-03)
$p = 0.008$ $q = 0.004$	2.088 (1.86E-02)	1.780 (1.39E-02)	1.597 (1.02E-02)	1.477 (9.26E-03)	1.348 (7.76E-03)	1.268 (6.51E-03)
$p = 0.008$ $q = 0.008$	2.341 (1.79E-02)	1.983 (1.25E-02)	1.774 (1.03E-02)	1.627 (8.30E-03)	1.498 (7.52E-03)	1.403 (6.45E-03)
$p = 0.016$ $q = 0.008$	2.475 (1.79E-02)	2.118 (1.30E-02)	1.877 (9.87E-03)	1.730 (8.91E-03)	1.585 (6.84E-03)	1.484 (6.14E-03)
$p = 0.016$ $q = 0.016$	2.635 (1.62E-02)	2.259 (1.20E-02)	2.023 (9.61E-03)	1.846 (8.13E-03)	1.699 (7.23E-03)	1.593 (6.06E-03)

greater than 98%, showing that the procedure is very effective in identifying the transitions: when there are transitions, there is more information about different regimes, and the probability of correct identification is improved. There is a roughly positive correlation between IR and (p, q): when T is 3000 and p and q increase, there are more transitions and a slightly higher probability of correct identification. For example, $p = 0.004$, and $q = 0.002$, IR is 98.4%. The IR is smaller, which becomes 97%, when p remains at 0.004 with q changing to 0.001. The IR is larger, which is 98.8%, when both p and q become larger, both changing to 0.008. But when q remains at 0.002 with p changing to 0.002, IR remains at 98.4%. When T is larger, the ratios become stable at some values close to 99%. For example, when $T = 6000$, IR is 98.2% when $p = 0.004$, and $q = 0.001$. When p remains at 0.004 with q changing to 0.002, the ratio is 99%. From then on, no matter how much p and q increase, the ratio is between 98.8% and 99%. Observing the ratios in each row, we can see that the ratios tend to increase with T until close to 99%. The associated standard errors become smaller when T increases.

As in the last section, we will show some figures of a randomly selected simulation path in each scenario to visualize the simulation results. Figure 3.4 shows the series y_t in each scenario with $T = 3000$. Different from the series shown in Figure 3.1 in Section 3.1.1, the series in Figure 3.4 are longer with more frequent transitions between two regimes. In each scenario, the series is not evolving evenly over time, indicating changing in the autoregressive coefficient. For example, in the second plot (Scenario 2), we can find a decrease in the magnitude of y_t around $t = 1000$ and a short period right after $t = 2500$. Furthermore, we find see more fluctuations in magnitude in each series when p and q become larger. Figures 3.5 and 3.6 compare the true α_t and β_t with $\hat{\alpha}_{t|T}$ and $\hat{\beta}_{t|T}$ of the same series in each scenario. From Figure 3.5 it is clear that when p and q become larger, the series experiences more frequent transitions. For example, in the first plot with $p = 0.001$ and $q = 0.001$,

Table 3.8: Performance of identification ratio (IR) for BCMIX estimates. Standard errors are given in parentheses.

Scenarios	$T = 3000$	$T = 4000$	$T = 5000$	$T = 6000$	$T = 7000$	$T = 8000$
$p = 0.001$	0.958	0.975	0.983	0.983	0.981	0.988
$q = 0.001$	(8.21E-03)	(6.05E-03)	(4.11E-03)	(4.31E-03)	(4.43E-03)	(3.09E-03)
$p = 0.002$	0.969	0.982	0.988	0.985	0.989	0.990
$q = 0.001$	(6.90E-03)	(4.71E-03)	(3.25E-03)	(3.88E-03)	(3.05E-03)	(2.26E-03)
$p = 0.002$	0.984	0.989	0.991	0.991	0.992	0.992
$q = 0.002$	(4.33E-03)	(2.53E-03)	(1.89E-03)	(1.84E-03)	(1.49E-03)	(1.62E-03)
$p = 0.004$	0.970	0.984	0.988	0.982	0.988	0.992
$q = 0.001$	(6.77E-03)	(4.64E-03)	(3.39E-03)	(4.52E-03)	(3.06E-03)	(1.82E-03)
$p = 0.004$	0.984	0.990	0.992	0.990	0.992	0.993
$q = 0.002$	(4.07E-03)	(1.80E-03)	(1.51E-03)	(2.24E-03)	(1.49E-03)	(9.13E-04)
$p = 0.008$	0.984	0.991	0.992	0.990	0.991	0.992
$q = 0.004$	(3.26E-03)	(1.13E-03)	(8.88E-04)	(1.09E-03)	(7.30E-04)	(5.22E-04)
$p = 0.008$	0.988	0.990	0.991	0.990	0.990	0.991
$q = 0.008$	(1.71E-03)	(7.25E-04)	(7.96E-04)	(6.26E-04)	(5.59E-04)	(3.19E-04)
$p = 0.016$	0.987	0.989	0.989	0.989	0.989	0.991
$q = 0.008$	(1.40E-03)	(8.99E-04)	(8.92E-04)	(7.46E-04)	(6.83E-04)	(2.52E-04)
$p = 0.016$	0.987	0.988	0.988	0.988	0.988	0.989
$q = 0.016$	(8.53E-04)	(6.77E-04)	(5.87E-04)	(4.53E-04)	(4.70E-04)	(2.62E-04)

Figure 3.4: A selected series y_t in Scenarios 1-9 (from left to right and top to bottom).

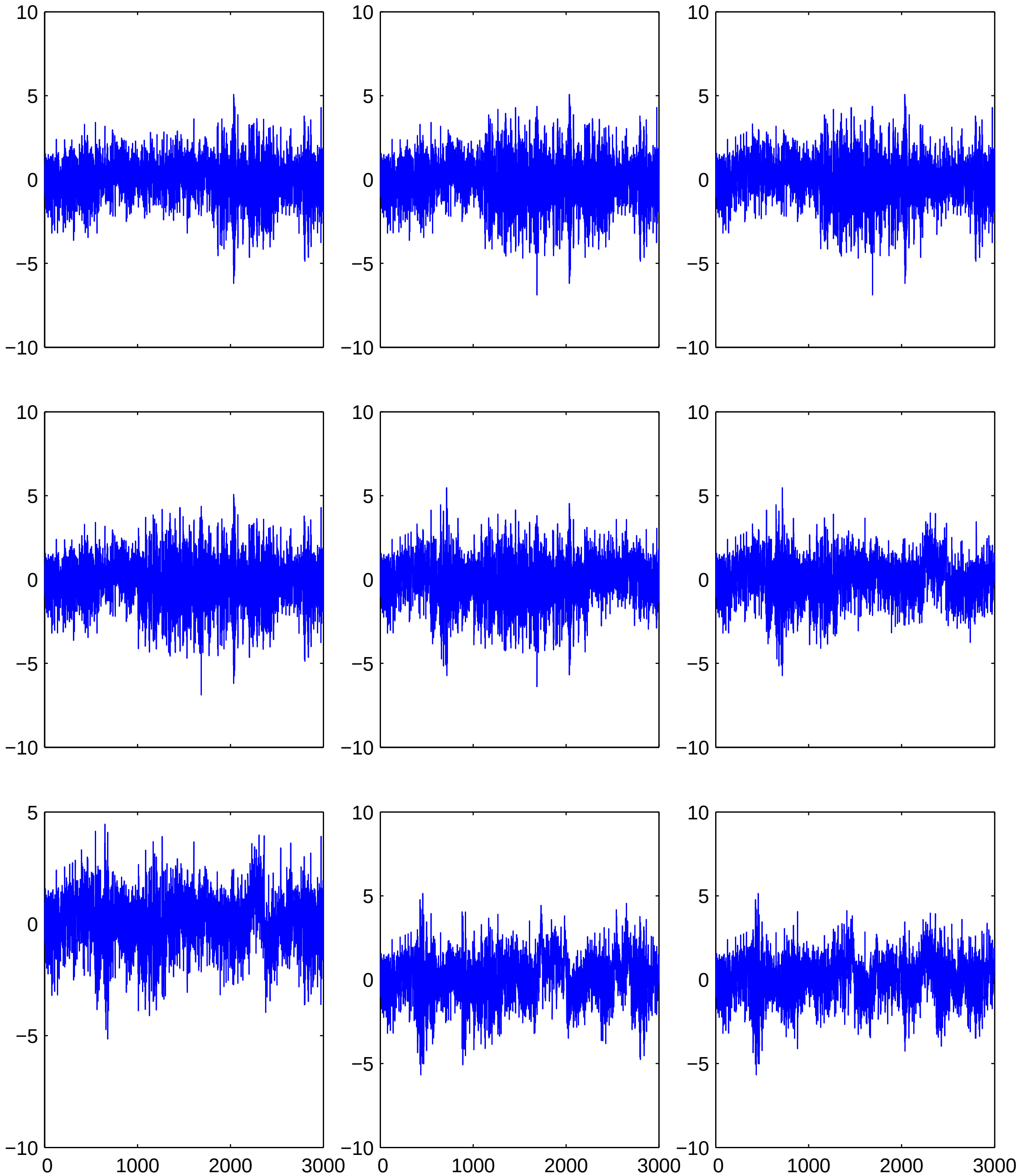
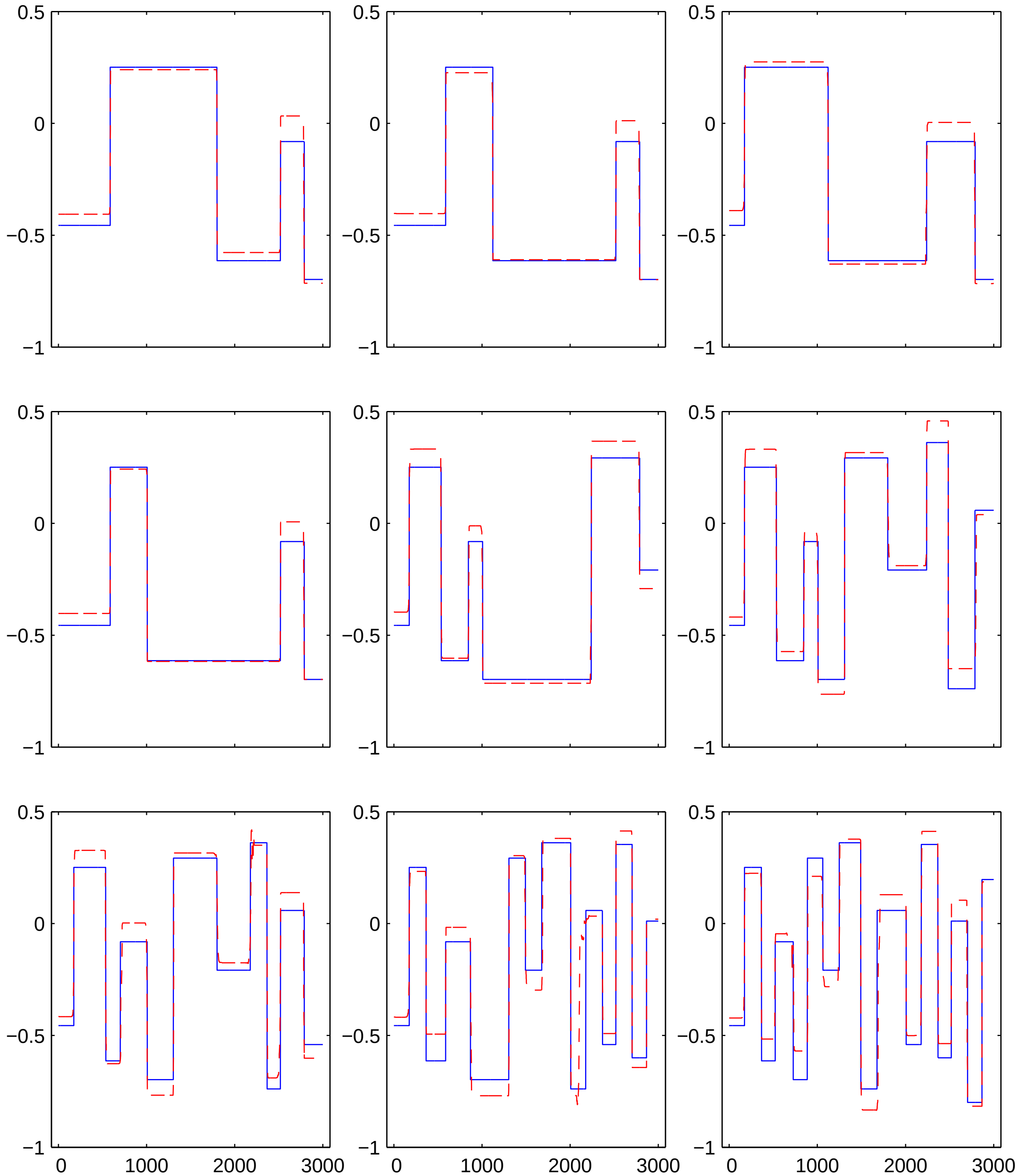


Figure 3.5: BCMIX estimates $\hat{\alpha}_{t|T}$ (dashed line) and true α_t (solid line) of the selected series in Scenarios 1-9 (from left to right and top to bottom).



there are 4 transitions in total, while in the last plot with $p = 0.016$ and $q = 0.016$, there are 15 transitions. In each plot there are two regimes, but the values of α_t within each regime are not constant. For example, in the last plot, the true α_t within regime 1 follow a Normal distribution with mean 0.2 and variance 0.16, and take realized values of 0.25, -0.08, 0.29, 0.36, 0.06, 0.35, 0.01 and 0.20 over time, while the true α_t within regime 2 follow a Normal distribution with mean -0.3 and variance 0.16, and take values of -0.46, -0.61, -0.70, -0.21, -0.74, -0.54, -0.60, -0.80. In each plot, the estimated parameter is close to the true β_t with some errors becoming more significant when there are more transitions. Looking at the ninth transition in the middle plot on the bottom, you will find that the estimating procedure identifies the transition too early and gives an estimate deviate from the true value for a while. However, the next transition is identified correctly and the deviation becomes less visible. Figure 3.6 shows $\hat{\beta}_{t|T}$ and β_t of the same series in each scenario. The results are similar to the those in Figure 3.5.

Figure 3.7 shows the true and estimated $P(s_t = 1)$ of the same series in each scenario. To clearly show the fuzziness around each transition, I use points to denote the estimated probabilities. It is clear that when there is a transition, it takes a while to recognize it. So the probability of $P(s_t = 1)$ does not jump directly from 1 to 0 or 0 to 1. Instead it adjusts step by step and takes some values in between. These "middle" points may affect the identification ratio. Moreover, there are more middle points when there are more frequent transitions, although the IR is higher.

We also try the filtering estimate $\hat{\boldsymbol{\beta}}_{t|t}$ given in (2.2.8) and the forecast $\hat{\boldsymbol{\beta}}_{t+1|t}$ given in (2.6.2). We pick a series in Scenario 1 ($p = 0.001$, $q = 0.001$) with $T = 3000$. The series y_t is shown in Figure 3.8. The middle plot shows the fitted values $\hat{y}_t$ calculated using the filtering estimates $\hat{\alpha}_{t|t}$ and $\hat{\beta}_{t|t}$. The bottom plot shows the forecasts of y_{t+1} at each stage t, which are calculated using the forecasts $\hat{\alpha}_{t+1|t}$ and $\hat{\beta}_{t+1|t}$. We can easily identify one transition right

Figure 3.6: BCMIX estimates $\hat{\beta}_{t|T}$ (dashed line) and true β_t (solid line) of the selected series in Scenarios 1-9 (from left to right and top to bottom).

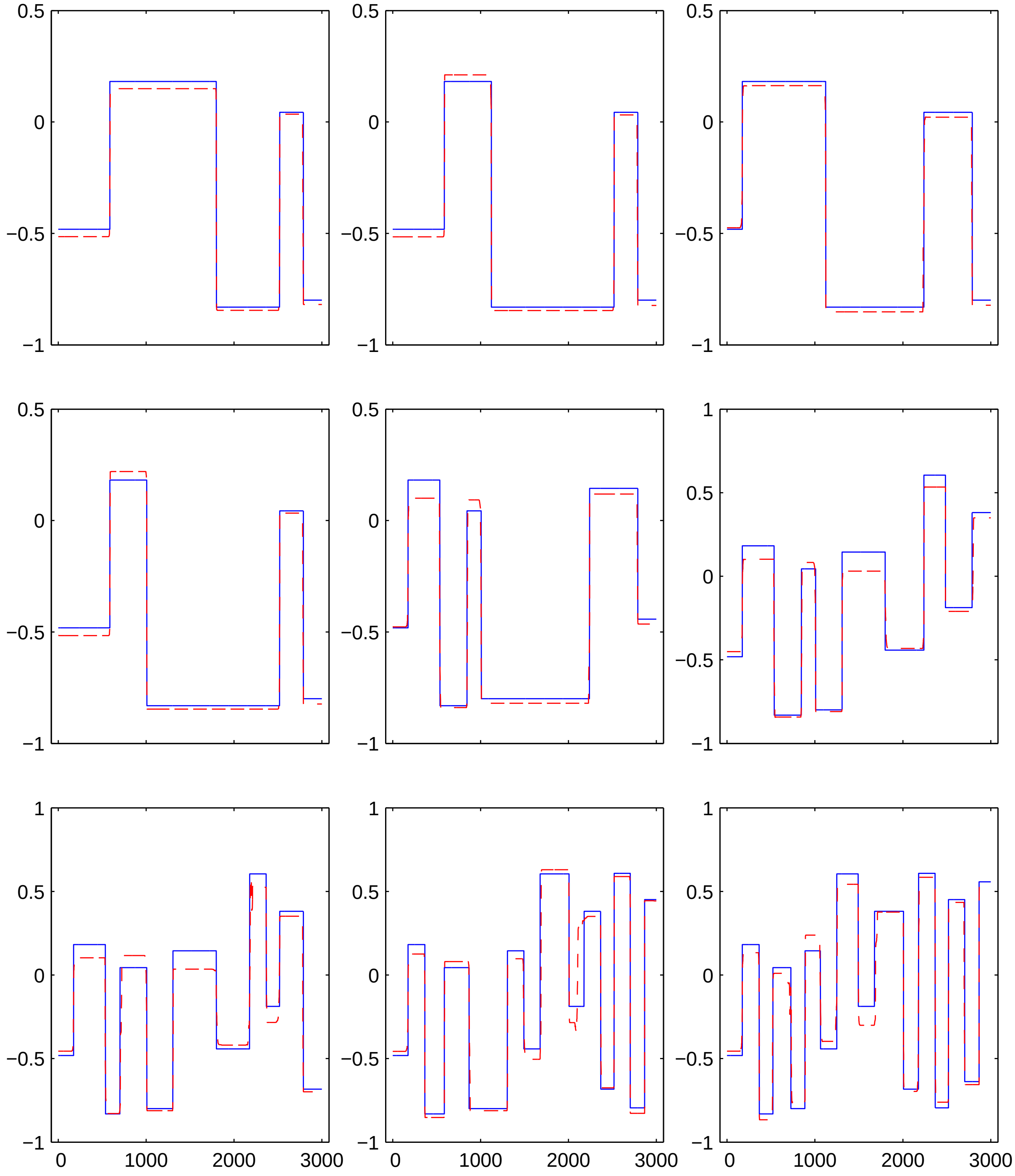
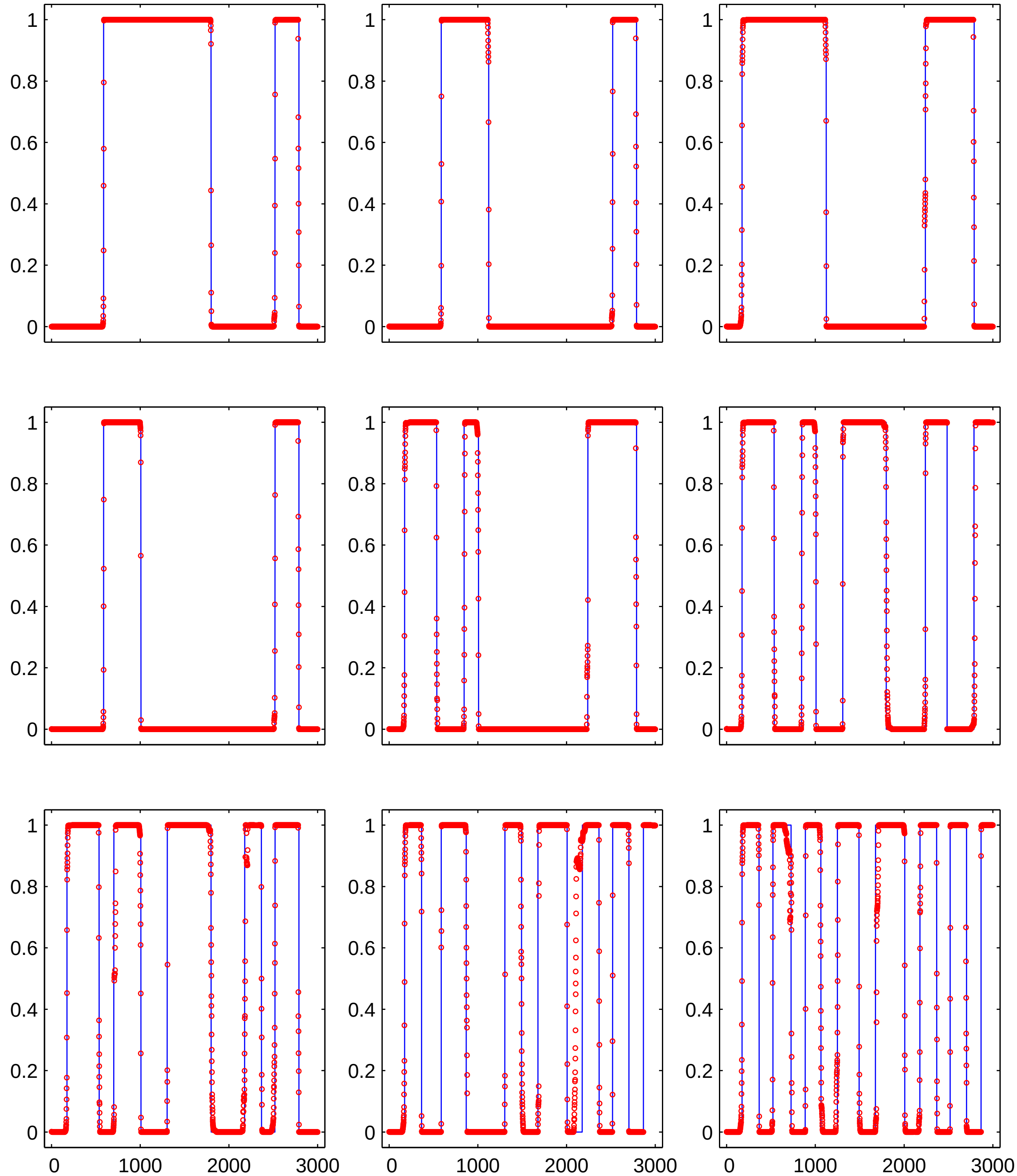


Figure 3.7: BCMIX estimates $\hat{r}_{t|T}^{(1)}$ (dashed line) and true $P(s_t = 1)$ (solid line) of the selected series in Scenarios 1-9 (from left to right and top to bottom).



after $t = 1000$ as the magnitude of y_t decreases. Both fitted values and forecasts capture the transitions. Three series in the figure are very similar.

Figure 3.9 compares the filtering estimates $\hat{\alpha}_{t|t}$ and $\hat{\beta}_{t|t}$ with the corresponding true values. Compared with the smoothing estimates shown in Figure 3.5 and Figure 3.6, the filtering estimates have more errors and fluctuations, especially at the beginning of the series and around each transition. The filtering estimate $\hat{\boldsymbol{\beta}}_{t|t} = E(\boldsymbol{\beta}_t|\mathcal{F}_t)$. At the beginning of the series, $\mathcal{F}_t$ contains very little information, while the smoothing estimate benefits a lot from the information between time t and T. Thus the filtering estimate is much worse. Before each transition, $E(\boldsymbol{\beta}_t|\mathcal{F}_t)$ does not contain any information about the forthcoming regime. So the adjustment after the transition might be slow and gradual as more and more information about the new regime comes to $\mathcal{F}_t$). Sometimes there is an over adjustment, as shown around the second transition in both $\hat{\alpha}_{t|t}$ and $\hat{\beta}_{t|t}$. Overall, the jump of the filtering estimate around each transition is not as sharp as the smoothing estimate.

Figure 3.10 compares the filtering estimate $\xi_t^{(1)}$, which is the estimated probability that $P(s_t = 1|\mathcal{F}_t)$, to the true values. Similar to the estimates in Figure 3.9, this filtering estimates show more fluctuations than the smoothing estimates shown in Figure 3.7. But two transitions in the series are identified very accurately.

Figure 3.11 compares the forecasts $\hat{\alpha}_{t+1|t}$ and $\hat{\beta}_{t+1|t}$ with the corresponding true values. Comparing the formula to calculate $\hat{\boldsymbol{\beta}}_{t|t}$ given in (2.2.8) and $\hat{\boldsymbol{\beta}}_{t+1|t}$ given in (2.6.2), you will notice that they are very similar to each other if the probability of staying in the current regime is much larger than that of switching to another regime. That explains the similarity of Figure 3.9 and Figure 3.11. Although there are errors in the values of the forecasts, the forecast jumps right after the true transition. So using the forecasting algorithm shown in Section 2.6, we can detect the transitions very quickly.

Figure 3.8: Top panel: A selected series y_t in Scenario 1. Middle panel: Fitted values $\hat{y}_t$ using the filtering estimates $\hat{\boldsymbol{\beta}}_{t|t}$. Bottom panel: Forecasts of y_{t+1} using the forecast $\hat{\boldsymbol{\beta}}_{t+1|t}$.

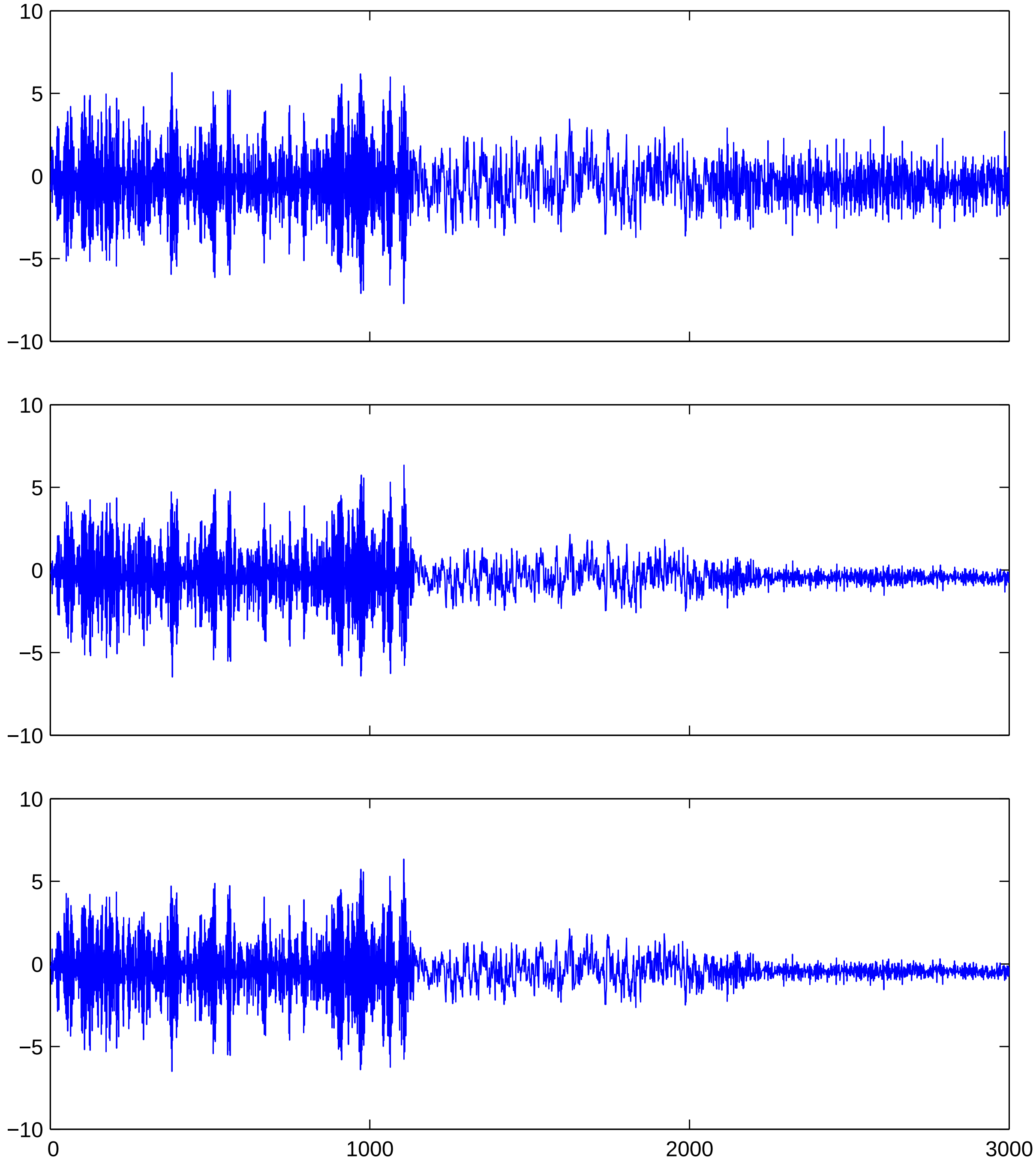


Figure 3.9: Top panel: filtering estimates $\hat{\alpha}_{t|t}$ (dashed line) and true α_t (solid line) of the selected series. Bottom panel: filtering estimates $\hat{\beta}_{t|t}$ (dashed line) and true β_t (solid line) of the selected series.

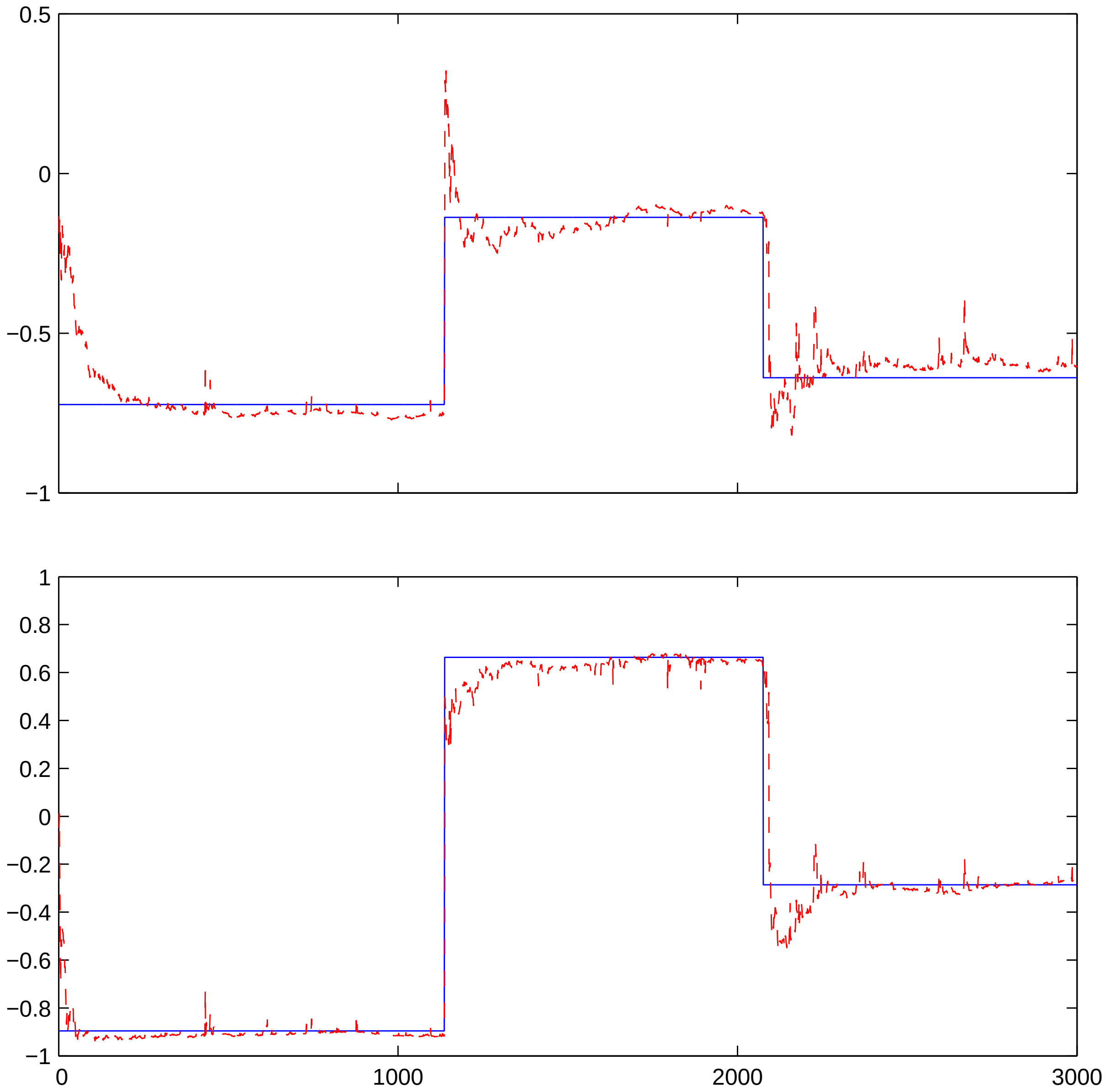


Figure 3.10: filtering estimates (dashed line) and true values of $P(s_t|\mathcal{F}_t)$.

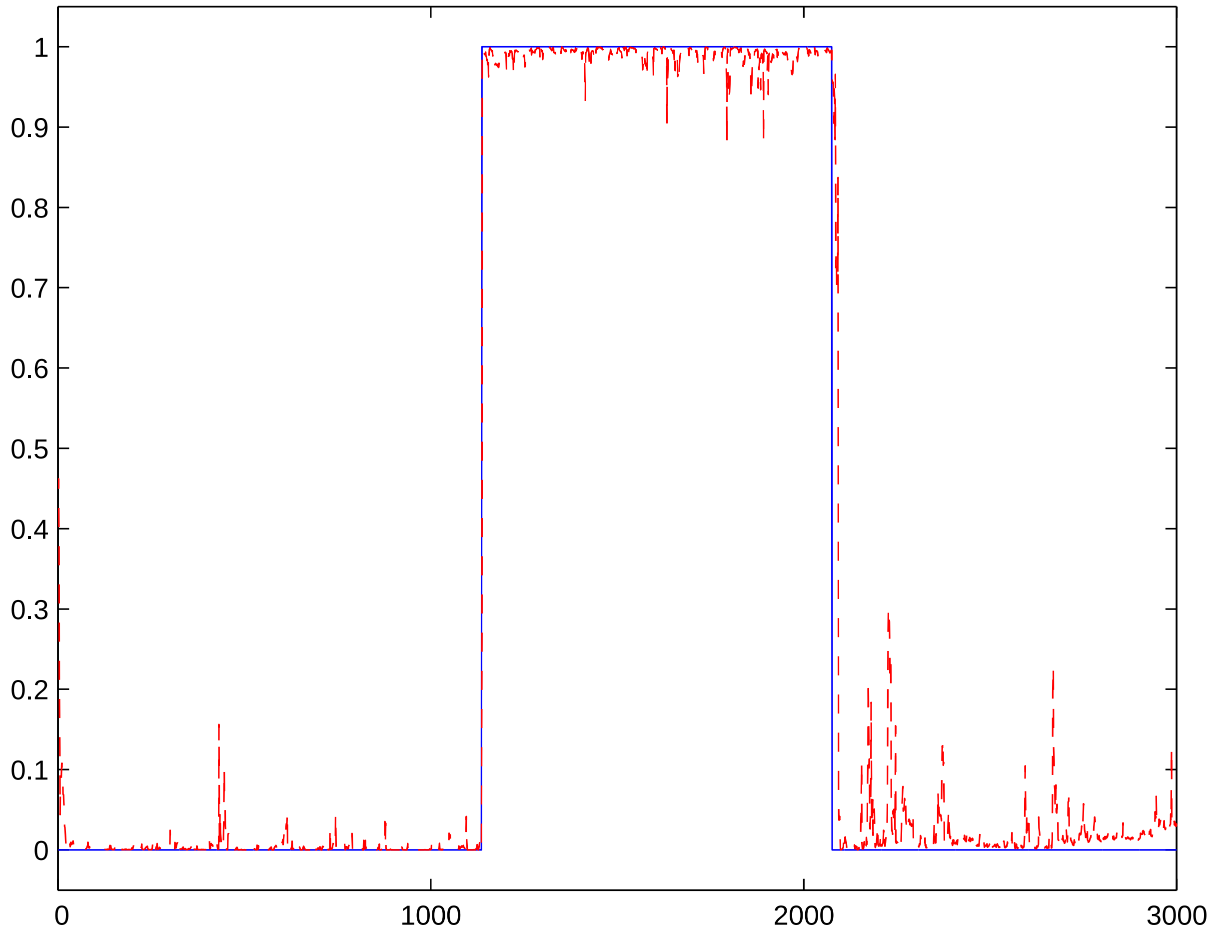
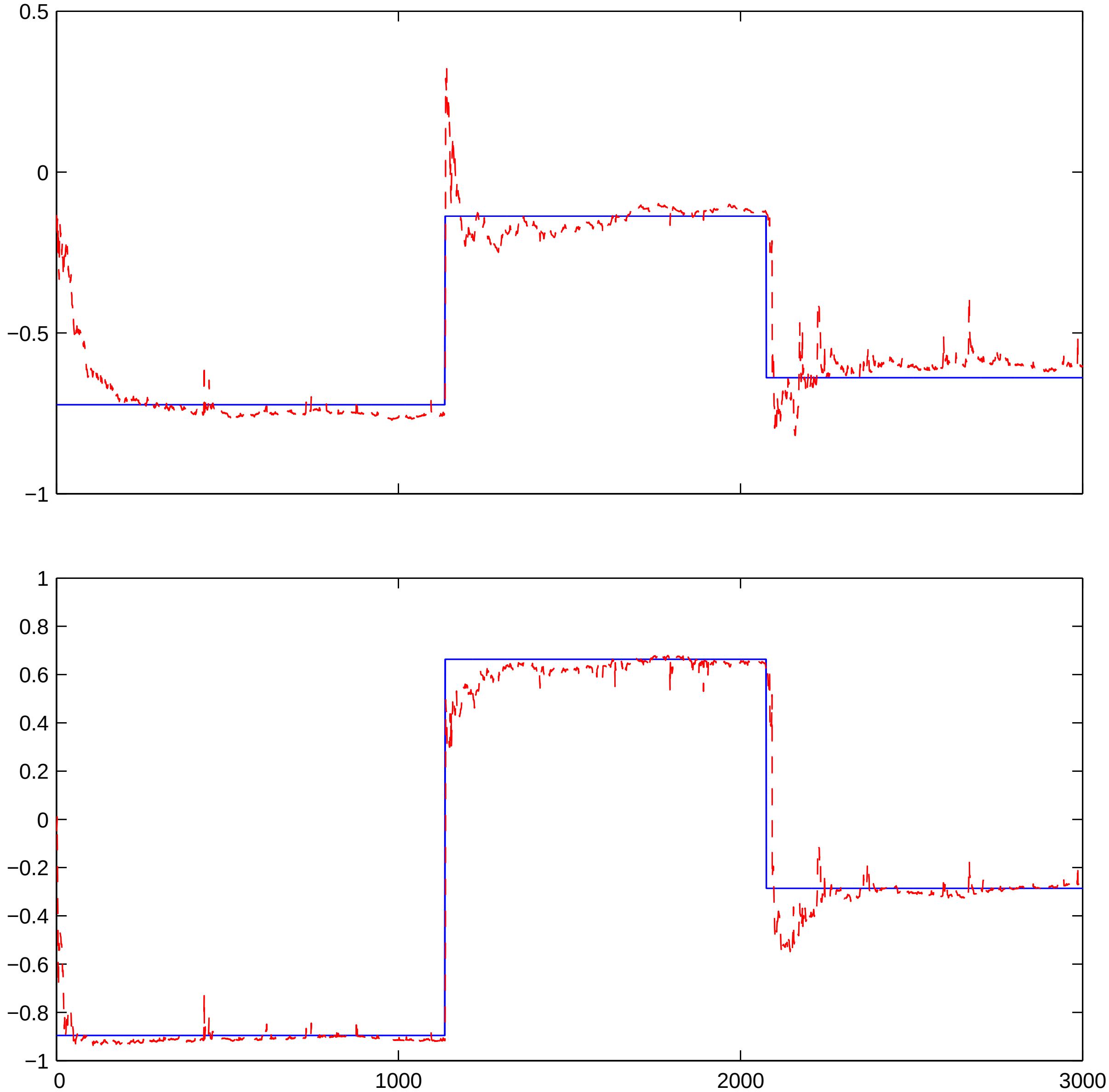


Figure 3.11: Top panel: Forecasts $\hat{\alpha}_{t+1|t}$ (dashed line) and true α_{t+1} (solid line) of the selected series. Bottom panel: Forecasts $\hat{\beta}_{t+1|t}$ (dashed line) and true β_{t+1} (solid line) of the selected series.



3.1.3 Choice of Hyperparameters

In this section, I will present some results to discuss the effects of hyperparameters on the estimation results. More specifically, we will discuss the choice of $\boldsymbol{z}(k)$ and $\boldsymbol{V}(k)$. In the first two simulation studies, we assume two regimes with distinctable hyperparameters. For example, in the first simulation study, we assume the means of β_t in regime 1 and 2 are $z(1) = 0.5$ and $z(2) = -0.5$, respectively. The values are far from each other considering the constraint that $|\beta_t| < 1$. That is why we see the clear transitions in Figure 3.2. We will further investigate the case when the regimes are very close to each other. We adopt the same hyperparameters as in the first simulation study shown in Section 3.1.1, where $\epsilon_t \sim N(0, \sigma^2)$. There are two regimes, $K = 2$, and the values of the parameter $\boldsymbol{\beta}_t$ depend on the hidden state s_t. The hyperparameters are $V(1) = V(2) = 0.16$, $P = \begin{pmatrix} 0.999 & 0.001 \\ 0.001 & 0.999 \end{pmatrix}$, and $\sigma^2 = 1$, sample size $T = 1000$. We keep $z(1) = 0.5$, but change $z(2)$ to 0.4. We use the setting of Scenario 4, where there are three transitions between regime 1 and regime 2. $s_t = 1$ for $1 \leq t \leq 200$; $s_t = 2$ for $201 \leq t \leq 500$; $s_t = 1$ for $501 \leq t \leq 600$; $s_t = 2$ for $601 \leq t \leq 1000$.

We try to use initial values of hyperparameters which are very close to the true hyperparameters and give an example shown in Figure 3.12. The EM algorithm is used to estimate the hyperparameters. The first plot shows a series y_t which is generated using the above setting. The middle plot compares the estimates $\hat{\beta}_{t|T}$ with the true values of β_t. The last plot compares the estimated probability of regime 1 with the true value. We cannot see clear patterns in the original series. Although $z(1)$ and $z(2)$ are far from each other, the variance $V(1) = V(2) = 0.16$ are large enough. So the realized β_t still show transitions with significant size. The estimates $\hat{\beta}_{t|T}$ show small errors in the first half of the sample period. Around the close transitions at $t = 500$ and 600, the errors are more significant. As shown in the third plot, the regimes cannot be distinguished clearly as the estimated probabilities are 0.8 or 0.2 in the first half of the sample period, and get even worth during the frequent transitions. This is due to the close distance between $z(1)$ and $z(2)$.

We further try a case using the above setting to generate a series but starting the estimation with initial values of hyperparameters which are far from the true hyperparameters. More specifically, we specify $z(1) = 0.2$, $z(2) = 0.6$, $V(1) = V(2) = 0.2$ as the initial values of the EM algorithm. The estimation results are shown in Figure 3.13. Compared with the second plot in Figure 3.12, the estimates do not work well around the frequent transitions at $t = 500$ and 600. This is more clearly shown in the third plot. The estimated probabilities of $P(s_t) = 1$ are changing gradually over time. It seems that the EM algorithm cannot converge to the true values of the hyperparameters when the initials are too far away. The algorithm is confused about the regimes and cannot separate them efficiently. The algorithm identifies the first transition successfully, and ignores the second transition which is too close to the third transition. After that, the information between $t = 500$ and 600 is realized and the estimated probabilities change from 0.4 to almost 1 to identify the change. However, the correct regime is 2.

We further try a case using the above setting:$z(1) = 0.4$, $z(2) = 0.5$, $V(1) = V(2) = 0.01$. In this case, the variation of β_t within each regime is smaller compared with the first case. We still use initial values of hyperparameters which are very close to the true hyperparameters and use the EM algorithm to estimate the hyperparameters. The example is shown in Figure 3.14. The magnitude of the series as shown in the first plot is much smaller than the one in Figure 3.12. By observing the second plot, we cannot tell how many regimes in the series as the values are not separate enough between regimes. Compared with the second plot in Figure 3.12, the estimates do not work efficiently to identify the first transition, and around the frequent transitions at $t = 500$ and 600 the estimated values are fluctuating. From the

Figure 3.12: Top panel: A selected series y_t. Middle panel: BCMIX estimates $\hat{\beta}_{t|T}$ (dashed line) and true β_t (solid line) of the selected series. Bottom panel: BCMIX estimated probability of $P(s_t) = 1$ (dashed line) and true values (solid line).

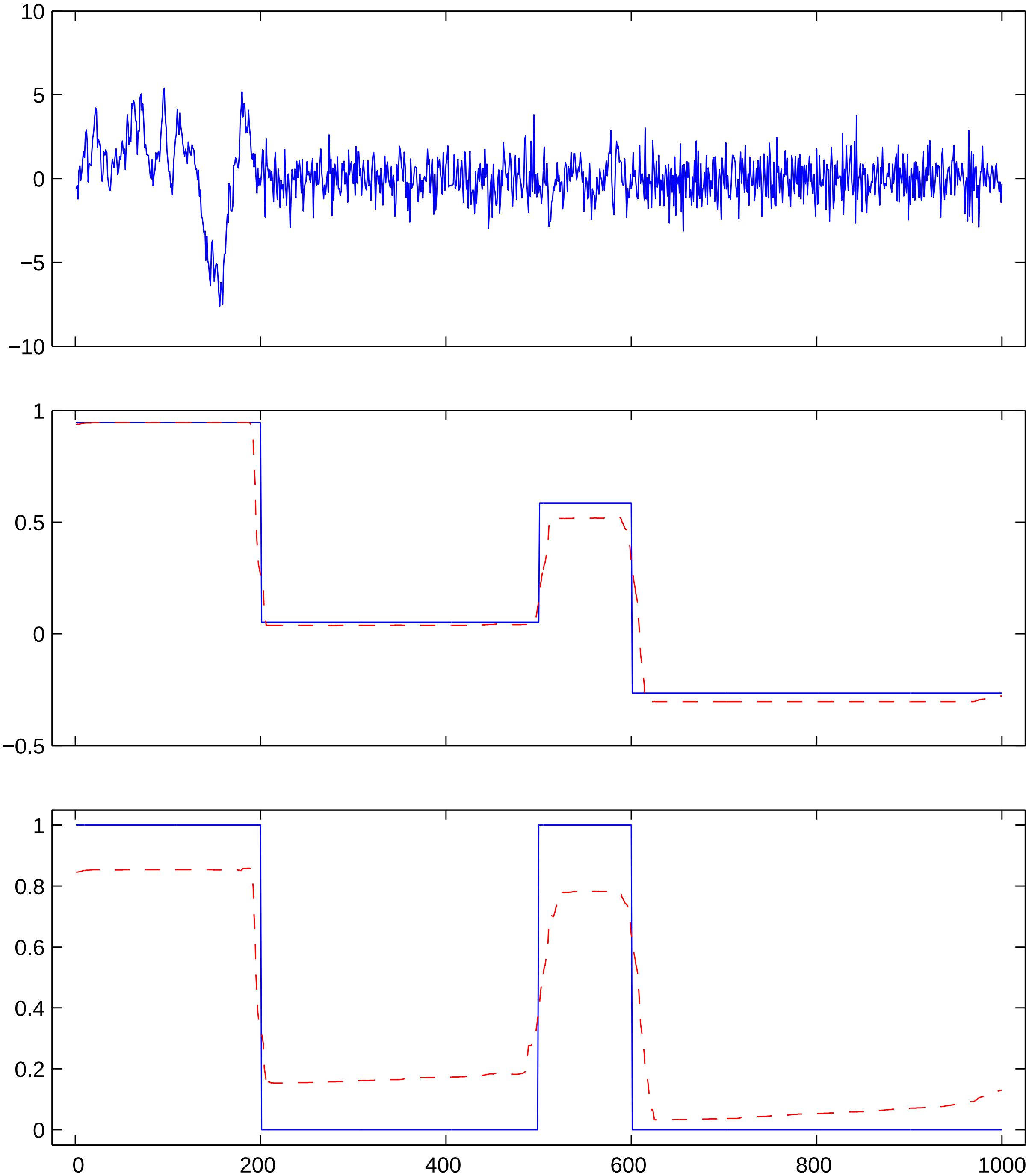
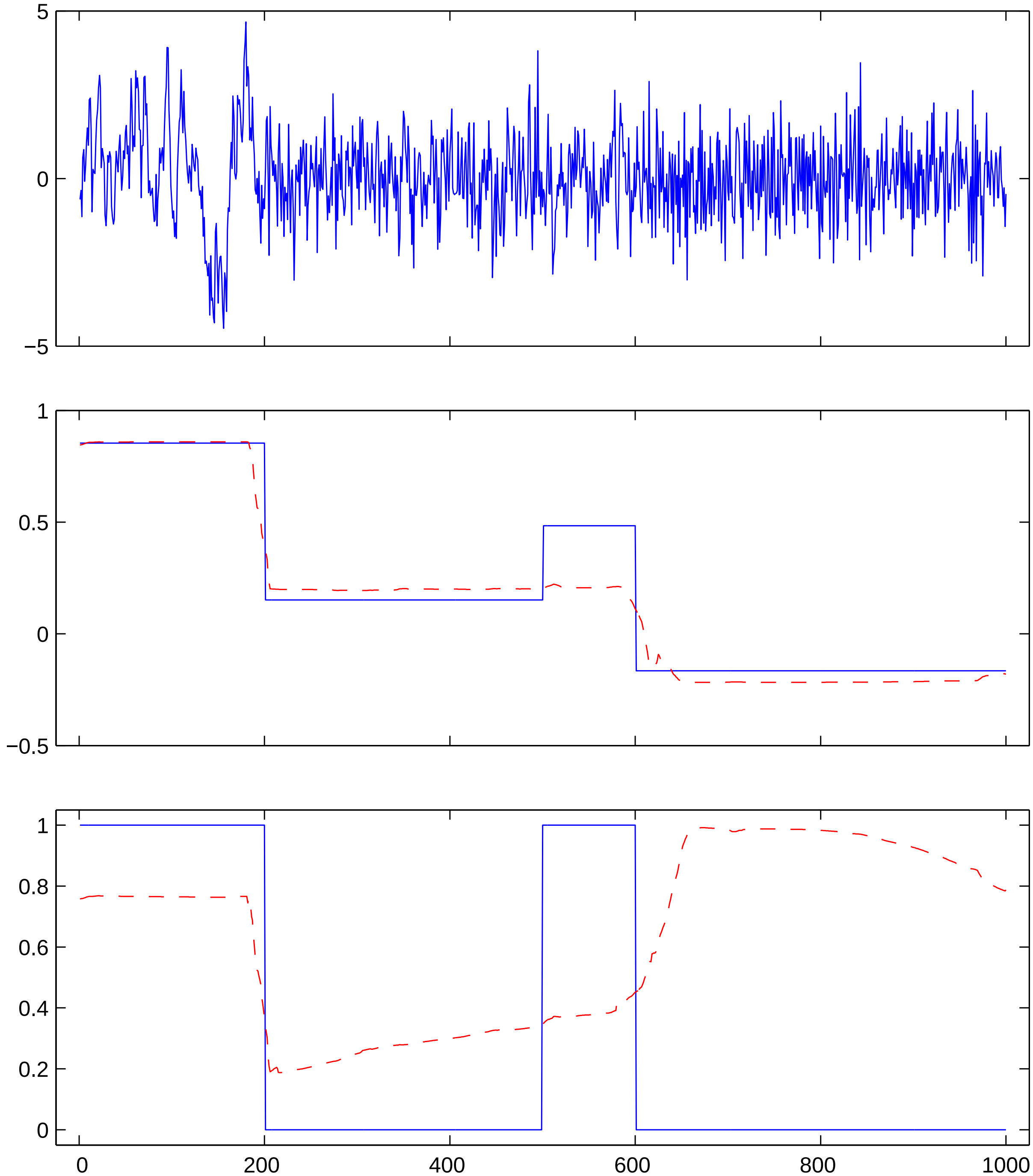


Figure 3.13: Top panel: A selected series y_t. Middle panel: BCMIX estimates $\hat{\beta}_{t|T}$ (dashed line) and true β_t (solid line) of the selected series. Bottom panel: BCMIX estimated probability of $P(s_t) = 1$ (dashed line) and true values (solid line).



third plot, we can see that actually the algorithm only identifies one transition with the probabilities changing gradually from almost 1 to 0.1. So when the difference between two regimes is too small, the algorithm cannot identify the difference at all.

In the last example, we follow the exact setting of Scenario 4 in Section 3.1.1 to generate the series, and use initial values which are far away from the true values for the EM algorithm. We want to see when the regimes are separate enough whether we need "accurate" initial values to conduct the algorithm correctly. In this case, we specify $z(1) = 0.2$, $z(2) = -0.8$, $V(1) = V(2) = 0.1$. The result in shown in Figure 3.15. We can see that although as in case 2, our information about the hyperparameters is highly inaccurate, the estimation results are very close to the true values. There are only small errors around the close transitions at $t = 500$ and 600.

Based on the results of the cases shown here, the algorithm works well in two cases. When the distance between regimes is not too small, even if the information of the hyperparameters is not accurate, the EM algorithm is efficient and the estimation results are good. When the means $\boldsymbol{z}(k)$ are close, as long as the variance is large enough and there is enough variation inside each regime, the regimes can be distinguished correctly. Furthermore, when the initial values of the hyperparameters are close to the true values, the algorithm works efficiently even if the regimes do not spread out enough. So in a real data analysis, some historical information and experience are very important for identifying the priors.

3.2 Real Data Analysis

In this section, we will apply the stochastic regime switching autoregressive model (3.1.2)

$$y_t = \alpha_t + \beta_t y_{t-1} + \epsilon_t,$$

Figure 3.14: Top panel: A selected series y_t. Middle panel: BCMIX estimates $\hat{\beta}_{t|T}$ (dashed line) and true β_t (solid line) of the selected series. Bottom panel: BCMIX estimated probability of $P(s_t) = 1$ (dashed line) and true values (solid line).

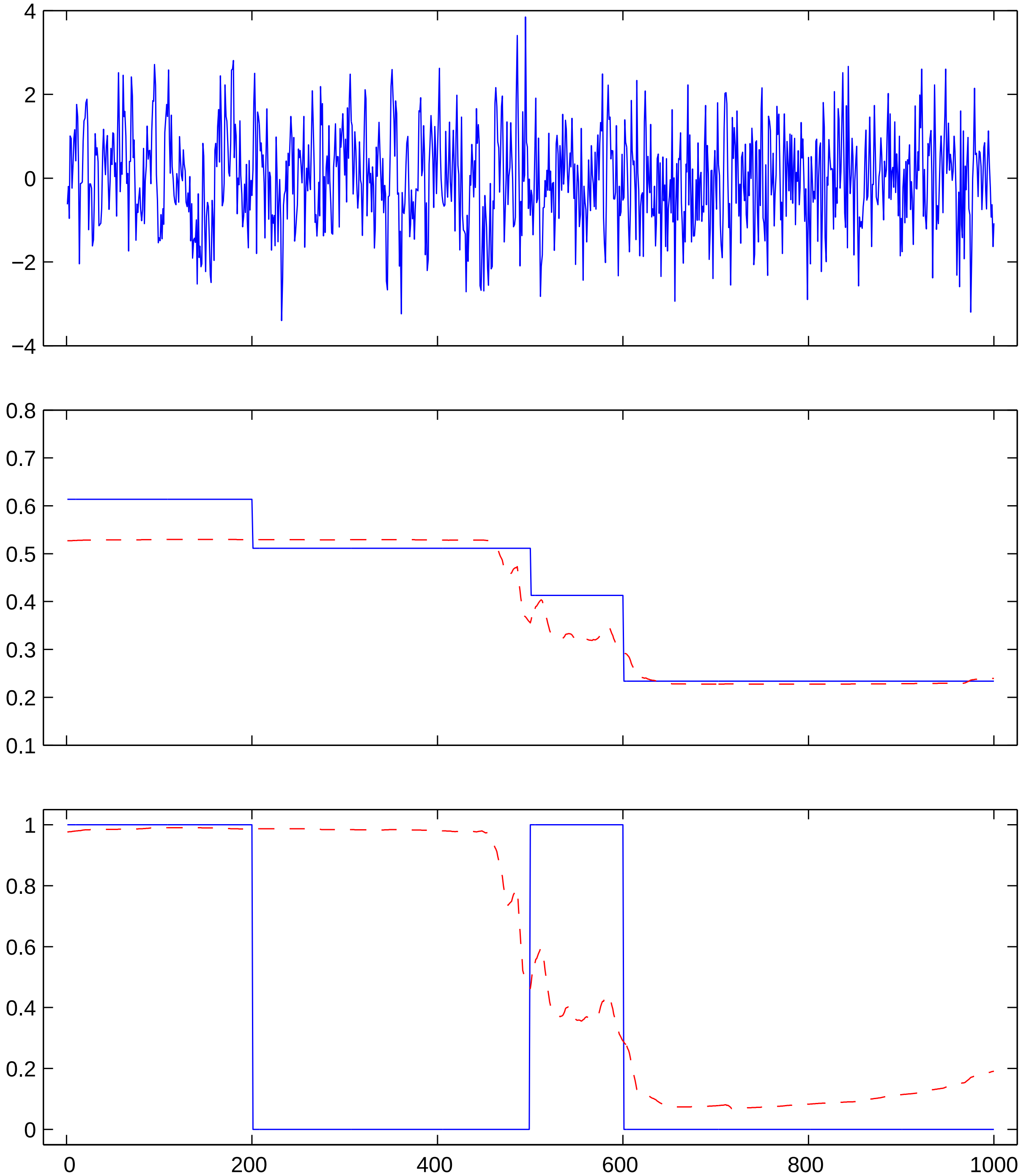
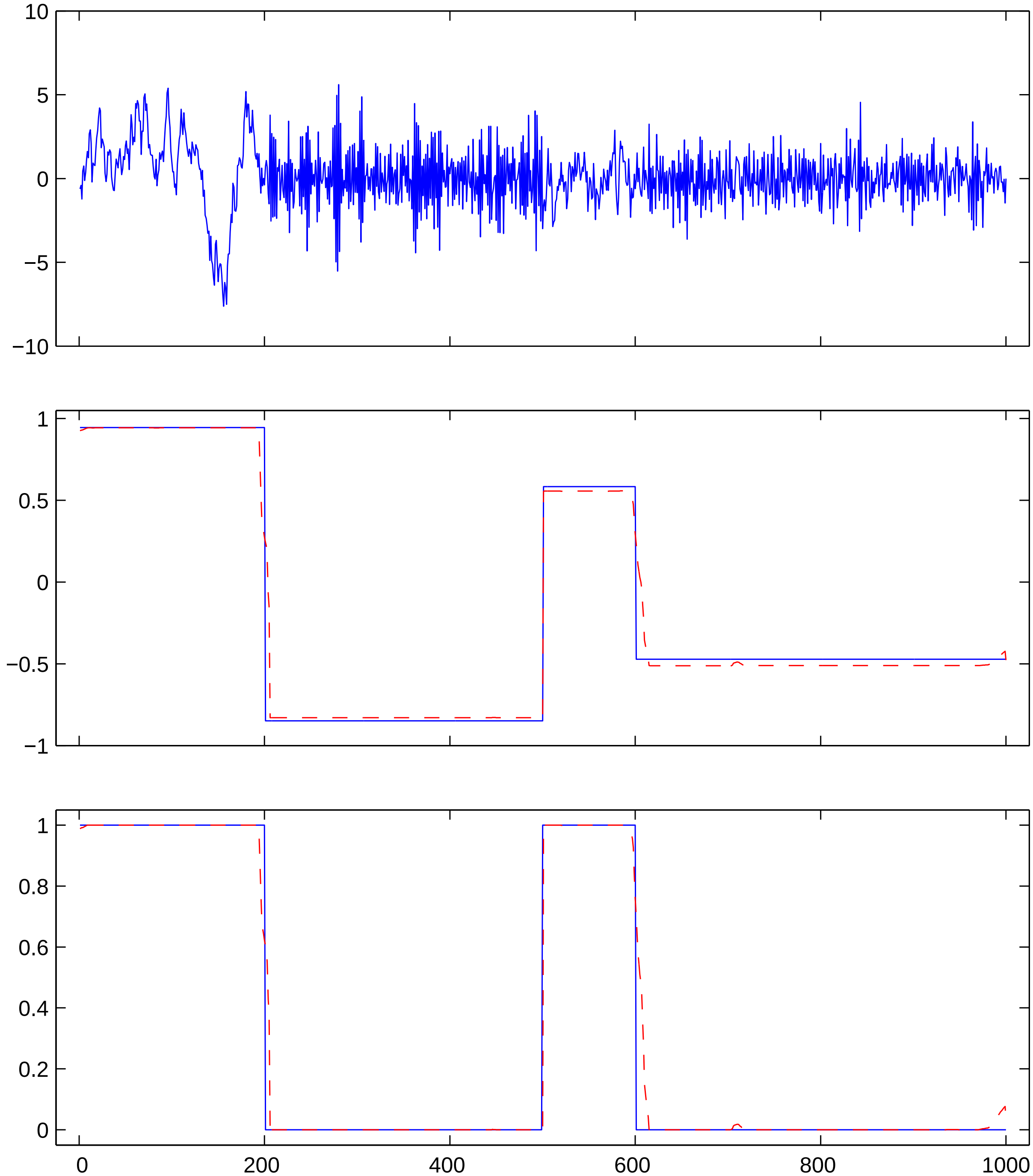


Figure 3.15: Top panel: A selected series y_t. Middle panel: BCMIX estimates $\hat{\beta}_{t|T}$ (dashed line) and true β_t (solid line) of the selected series. Bottom panel: BCMIX estimated probability of $P(s_t) = 1$ (dashed line) and true values (solid line).



to analyze some real econometric time series.

3.2.1 Total Financial Assets - Assets - Balance Sheet of Nonfarm Nonfinancial Corporate Business

The first series we use is "Total Financial Assets - Assets - Balance Sheet of Nonfarm Nonfinancial Corporate Business (TFAABSNNCB)" available from the website of Federal Reserve Bank of St. Louis at `http://research.stlouisfed.org/fred2/series/TFAABSNNCB?rid=52`. This data comes from the Z.1 Flow of Funds release of the Board of Governors of the Federal Reserve System.

The nonfarm nonfinancial corporate business sector includes all private domestic corporations with the exception of corporate farms and financial institutions. We use the quarterly data (in billions of dollars) from the fourth quarter of 1951 to the first quarter of 2011. Figure 3.16 shows the original series. It is clear that the series shows an increasing trend over time and is therefore not stationary.

Instead of using the original series y'_t, we use the continuously compounded rate of change which is calculated as $y_t = 100 \cdot \ln \frac{y'_t}{y'_{t-1}}$. The series is shown in Figure 3.17. The p-value is less than 0.01 in the augmented Dickey-Fuller test with the null hypothesis that the series is non-stationary. Hence this series is stationary.

We use $K = 2$ regimes. Model (3.1.2) is fitted to the adjusted series. The estimated hyperparameters by the EM algorithm are: $\hat{\boldsymbol{z}}(1) = (2.61, 0.14)'$, $\hat{\boldsymbol{z}}(2) = (1.19, 0.08)'$, $\hat{\boldsymbol{V}}(1) = \begin{pmatrix} 0.12 & 0.03 \\ 0.03 & 0.12 \end{pmatrix}$, $\hat{\boldsymbol{V}}(2) = \begin{pmatrix} 0.13 & 0.04 \\ 0.04 & 0.08 \end{pmatrix}$ and $\hat{\sigma}^2 = 1.18^2$. The estimated transition matrix is $\hat{P} = \begin{pmatrix} 0.59 & 0.41 \\ 0.34 & 0.66 \end{pmatrix}$. The smoothing estimates of $\hat{\alpha}_{t|T}$ and $\hat{\beta}_{t|T}$ are shown in Figure 3.18. The estimated probability of staying in regime 1, $P(s_t = 1)$, is also shown in the figure. The

Figure 3.16: TFAABSNNCB series: In billions of dollars.

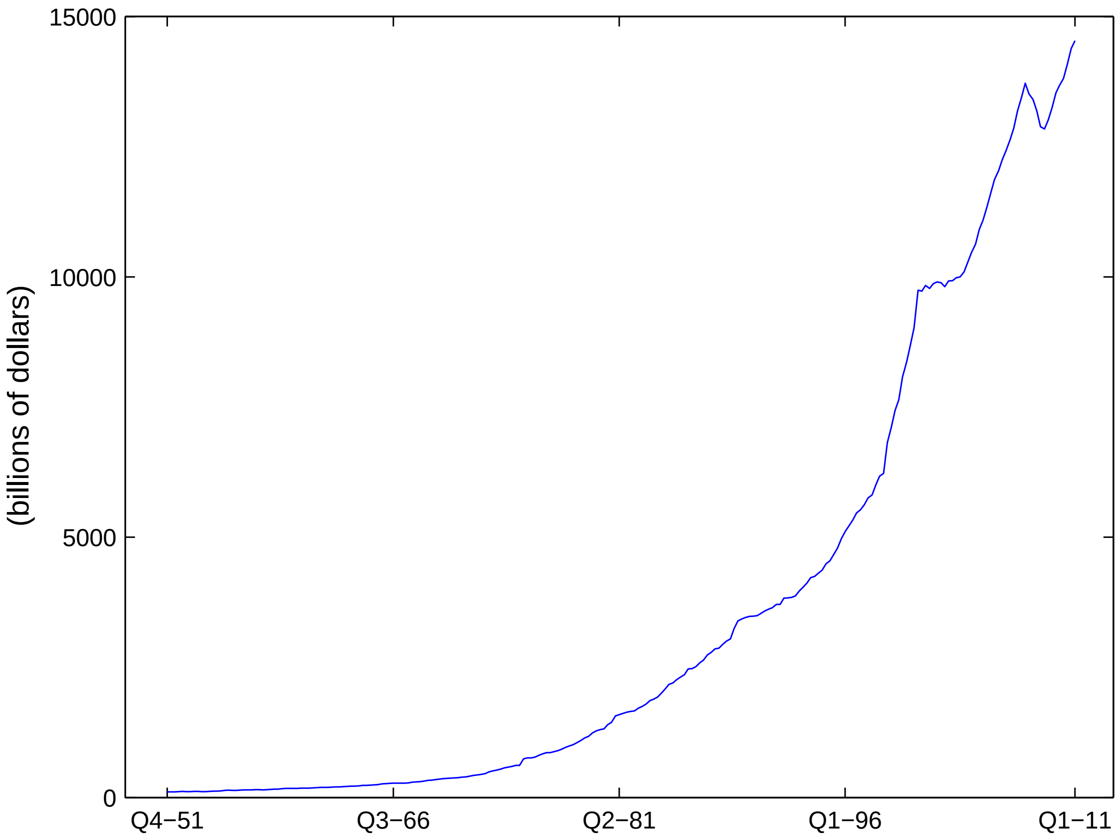


Figure 3.17: TFAABSNNCB series: Continuously compounded rate of change.

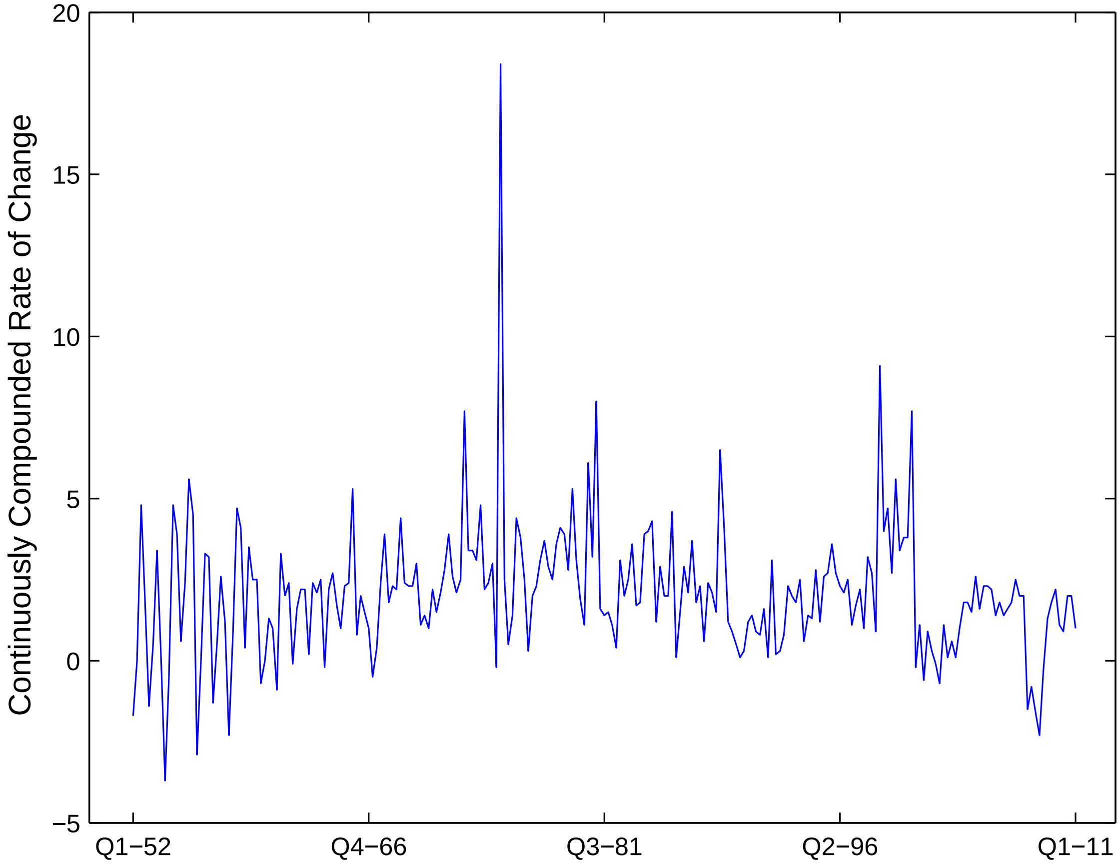


Table 3.9: Summary of $\hat{\alpha}_{t|T}$ and $\hat{\beta}_{t|T}$ in periods between pairs of transitions.

	Subperiods		*Estimates of Coefficients*		
	Regime	Starting	Ending	$\hat{\alpha}_{t\|T}$	$\hat{\beta}_{t\|T}$
Period 1	2	Q2 1952	Q3 1971	1.42	0.16
Period 2	1	Q4 1971	Q2 1981	3.37	-0.05
Period 3	2	Q3 1981	Q3 1998	1.50	0.22
Period 4	1	Q4 1998	Q4 2000	3.18	0.22
Period 5	2	Q1 2001	Q1 2011	0.90	0.18

NOTE: The first NBER recession list here started from December 1969, before the beginning of our sample period which was January 1970.

figure shows four transitions between two regimes, occurring in the third quarter of 1971, the second quarter of 1981, the third quarter of 1998 and the fourth quarter of 2000. Because of the fuzziness, the jumps are gradual over a period. For example, the probabilities of staying in regime 1 during the four quarters in 1971 are 0.23, 0.47, 0.66 and 0.72, respectively. The estimates $\hat{\alpha}_{t|T}$ and $\hat{\beta}_{t|T}$ for all the subperiods between two consecutive transitions are shown in Table 3.9. Consistent with our intuition, the estimated parameters within each regime are not the same. For example, the hidden regime is 1 for the period between the first and second transitions, with an $\hat{\alpha}_{t|T}$ of about 3.37. The regime is also 1 for the period between the third and fourth transitions, with $\hat{\alpha}_{t|T}$ of about 3.18. The hidden regime is 2 for the period before the first transition, with an estimate $\hat{\beta}_{t|T}$ of about 0.16. The regime is also 2 for the period between the second and third transitions, with an estimate $\hat{\beta}_{t|T}$ of about 0.22. The results justify the assumption of our model.

The fitted values $\hat{y}_t$ and the errors between y_t and $\hat{y}_t$ are shown in Figure 3.19. The switching trend can be found in the fitted series. The errors are distributed evenly around zero without any pattern.

Figure 3.18: Top panel: Smoothed probability $\hat{r}^{(1)}_{t|T}$ of TFAABSNNCB series. Middle panel: BCMIX estimates $\hat{\alpha}_{t|T}$ of TFAABSNNCB series. Bottom panel: BCMIX estimates $\hat{\beta}_{t|T}$ of TFAABSNNCB series.

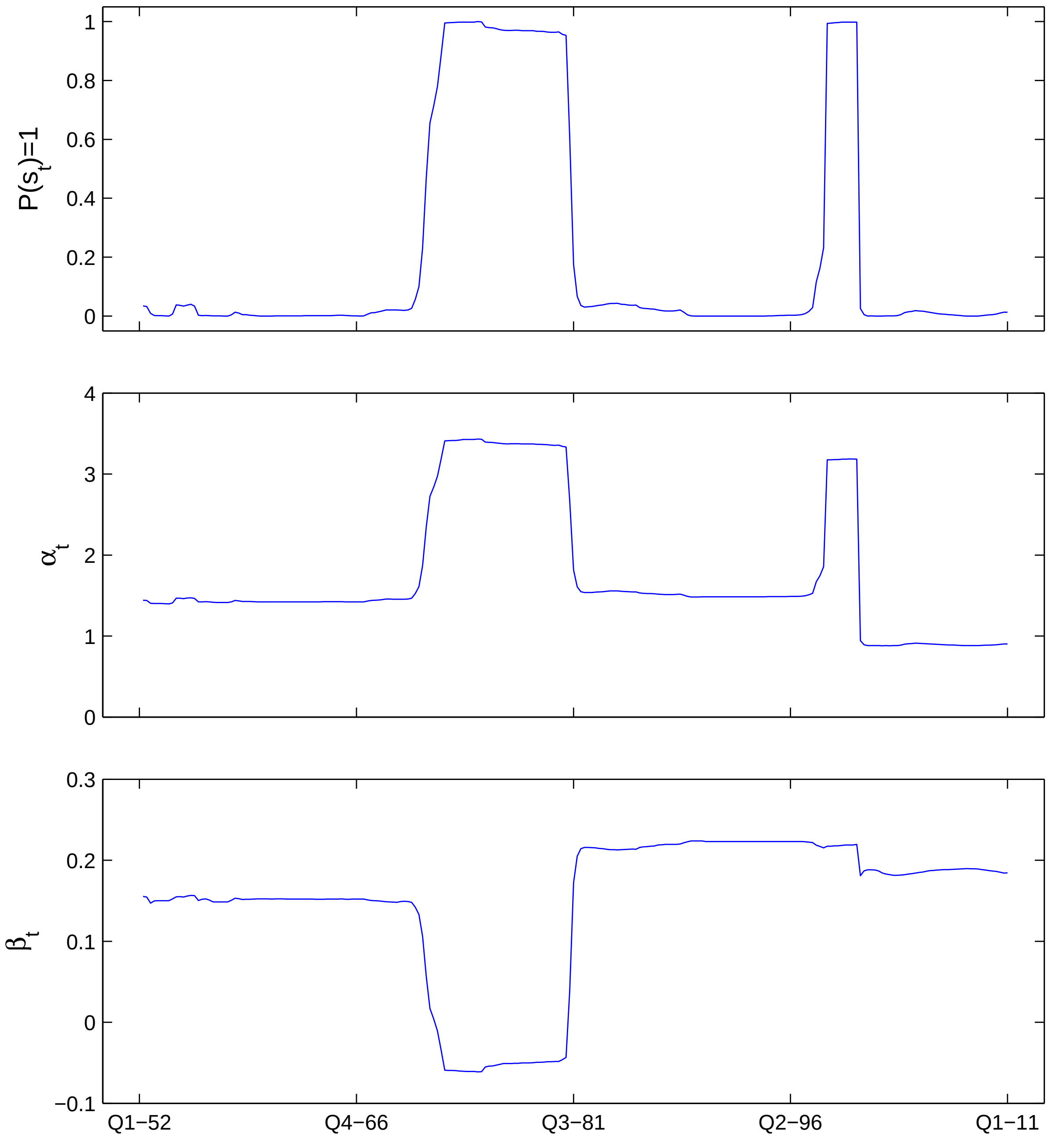
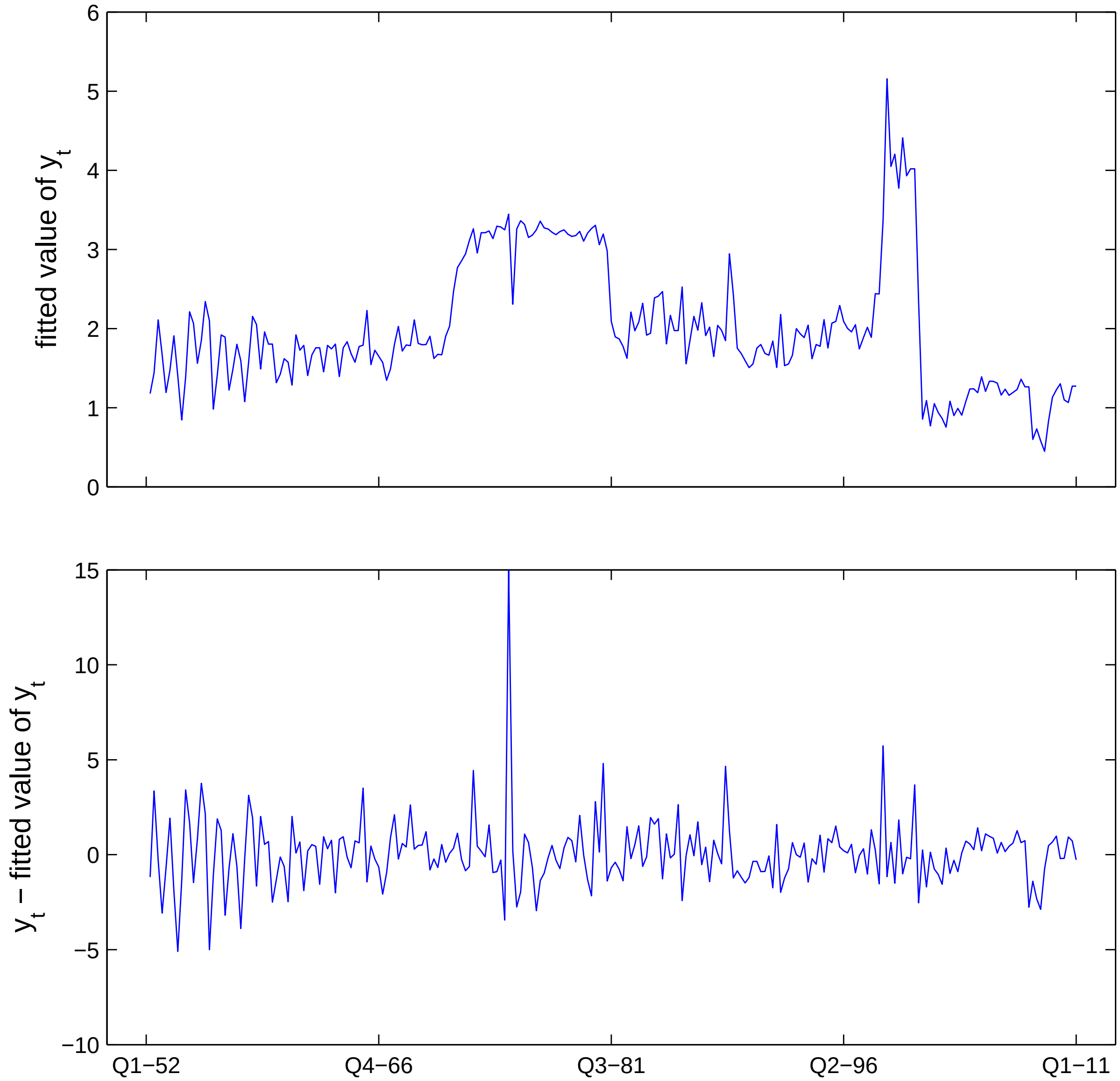


Figure 3.19: Top panel: Fitted series $\hat{y}_t$ of TFAABSNNCB series. Bottom panel: Error series $y_t - \hat{y}_t$ of TFAABSNNCB series.



3.2.2 All Employees: Total Nonfarm

In Section 1.3, we use the “All Employees: Total nonfarm (PAYEMS)” data to illustrate the motivation of this dissertation study. The estimated probability of $P(s_t = 1)$ is shown in Figure 1.3. In this section, we will use $K = 2$ regimes, and fit model (3.1.2) to the same series to compare our results with the one in Section 1.3, and to see if the transitions between regimes match the recession history. The estimated hyperparameters by the EM algorithm are: $\hat{\boldsymbol{z}}(1) = (187.27, 0.15)'$, $\hat{\boldsymbol{z}}(2) = (-64.34, 0.24)'$, $\hat{\boldsymbol{V}}(1) = \begin{pmatrix} 2811.69 & 0.53 \\ 0.53 & 0.06 \end{pmatrix}$, $\hat{\boldsymbol{V}}(2) = \begin{pmatrix} 1249.40 & -9.10 \\ -9.10 & 0.22 \end{pmatrix}$ and $\hat{\sigma}^2 = 66.73^2$. The estimated transition matrix is $\hat{P} = \begin{pmatrix} 0.87 & 0.123 \\ 0.35 & 0.65 \end{pmatrix}$. Figure 3.20 shows the smoothing estimates $\hat{\alpha}_{t|T}$ and $\hat{\beta}_{t|T}$, and the probability of staying in regime 1, $P(s_t = 1)$. Different from the first example, this case shows much more transitions between two regimes. The estimated $P(s_t = 1)$ do not fit the results shown in 1.3. There are periods in which $P(s_t = 1)$ is close to zero and the identified regime is 2. More interestingly, there are 7 subperiods identified as in regime 2, each of which is at least partially overlapped by an NBER recession. So we can think of regime 1 as “good” period, and regime 2 as “bad” period. The $\hat{\alpha}_{t|T}$ in the second plot of Figure 3.20 further prove this finding: The values $\hat{\alpha}_{t|T}$ in regime 2 are negative and much lower than those in regime 1, indicating that the bad employment situation will become worse during a recession. The values of $\hat{\beta}_{t|T}$ in the third plot show high and positive autocorrelations in regime 2, indicating that the bad employment situation will last for a while during a recession. Moreover, the estimates $\hat{\alpha}_{t|T}$ and $\hat{\beta}_{t|T}$ are different within each regime. The estimates $\hat{\alpha}_{t|T}$ and $\hat{\beta}_{t|T}$ for the periods identified as regime 2 are shown in Table 3.10. For example, during the period which is identified as regime 2 and overlapped by the third NBER recession (July

Figure 3.20: Top panel: Smoothed probability $\hat{r}^{(1)}_{t|T}$ of PAYEMS series. Middle panel: BCMIX estimates $\hat{\alpha}_{t|T}$ of PAYEMS series. Bottom panel: BCMIX estimates $\hat{\beta}_{t|T}$ of PAYEMS series. NBER recessions are shown as shaded areas.

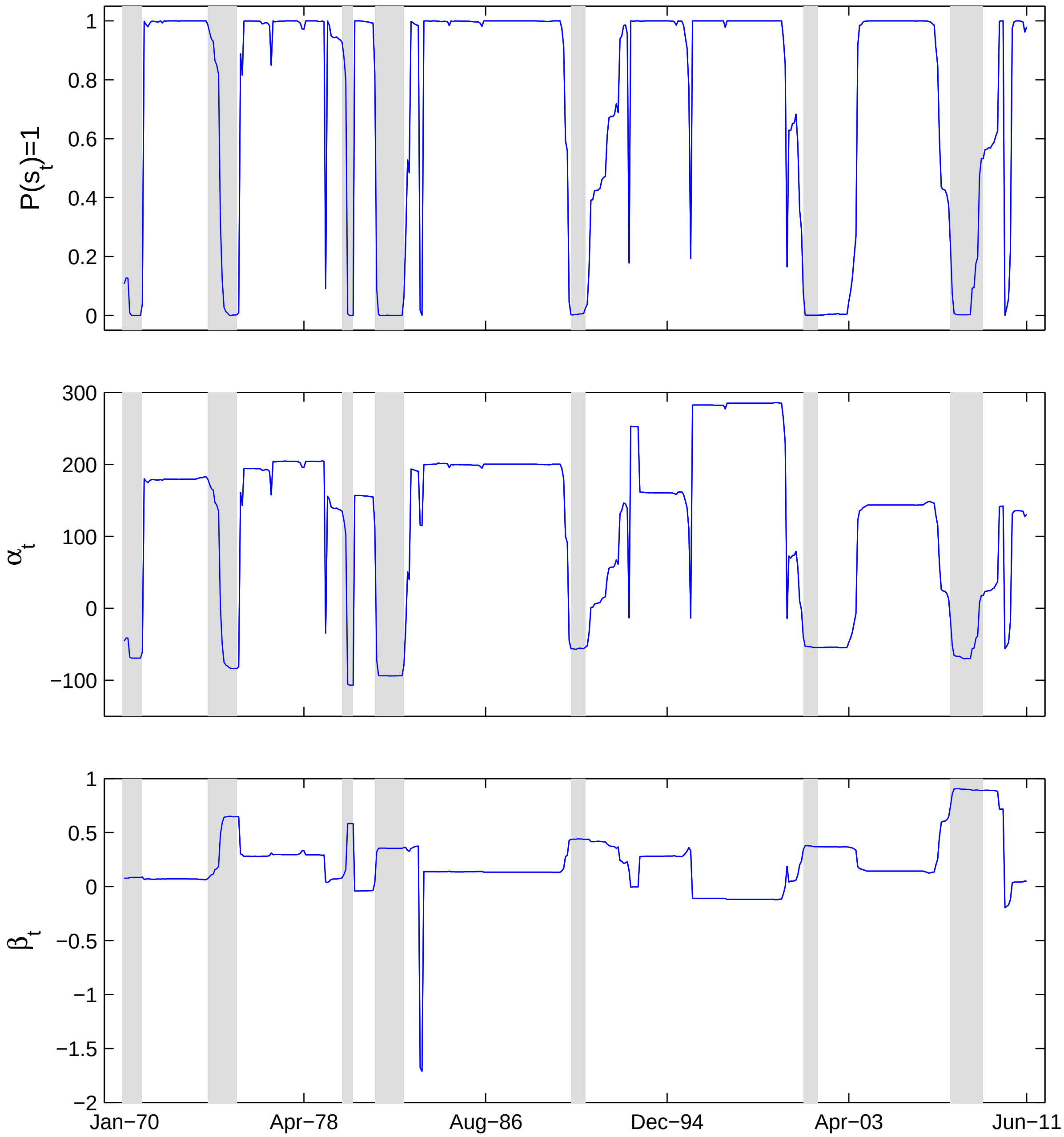


Figure 3.21: Top panel: Fitted series $\hat{y}_t$ of PAYEMS series. Bottom panel: Error series $y_t - \hat{y}_t$ of PAYEMS series.

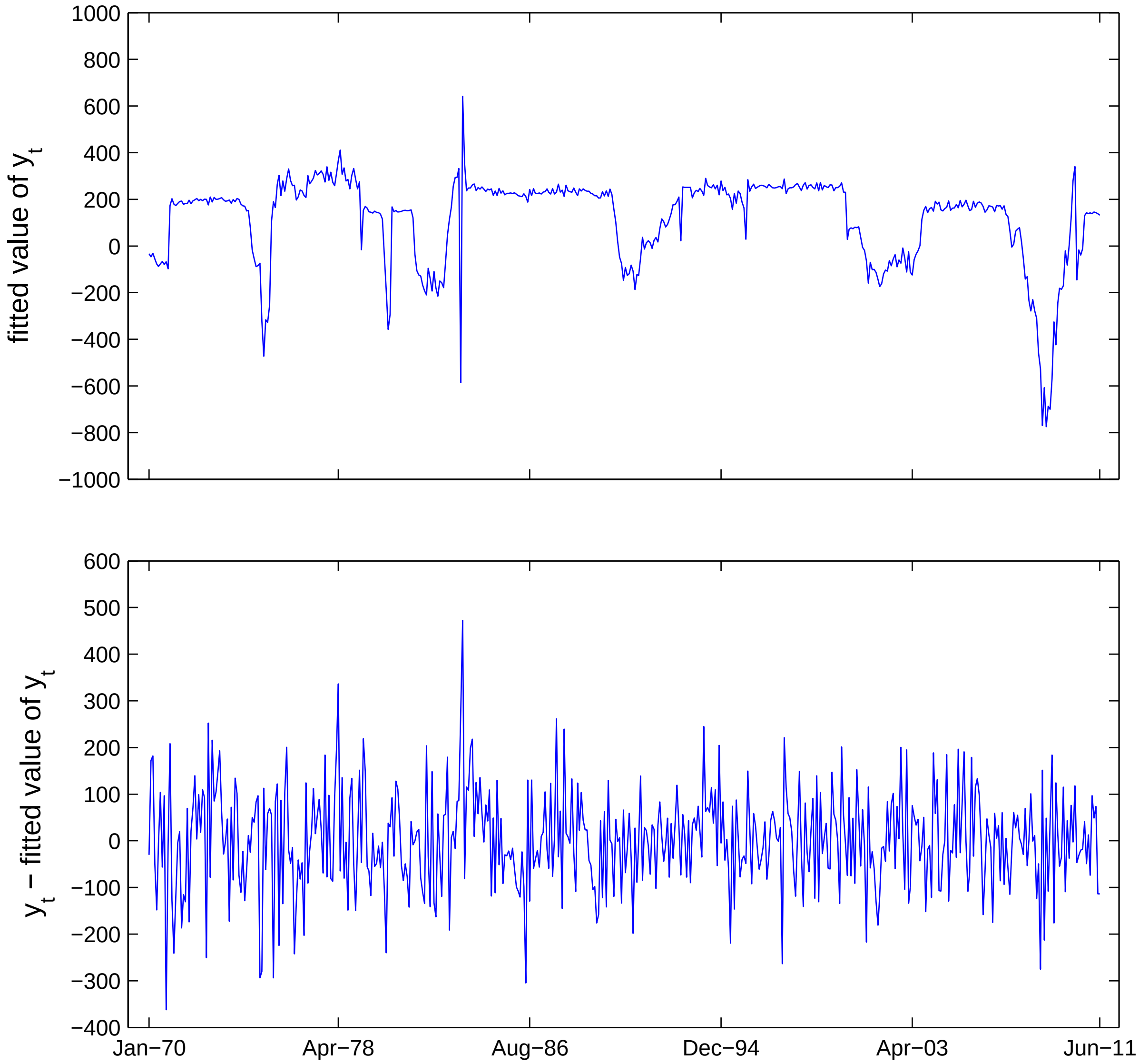


Table 3.10: Summary of the estimated parameters in periods identified as in regime 2 and the corresponding NBER recessions.

	NBER Recessions		*"Regime 2" Periods*		*Estimates of Coefficients*	
Subperiods	Starting	Ending	Starting	Ending	$\hat{\alpha}_{t\|T}$ ($\times 10^2$)	$\hat{\beta}_{t\|T}$
Period 1	Dec 1969	Nov 1970	Jan 1970	Nov 1970	-0.69	0.08
Period 2	Nov 1973	Mar 1975	Jun 1974	Apr 1975	-0.81	0.65
Period 3	Jan 1980	Jul 1980	Apr 1980	Jul 1980	-1.07	0.58
Period 4	Jul 1981	Nov 1982	Aug 1981	Dec 1982	-0.94	0.35
Period 5	Jul 1990	Mar 1991	Jun 1990	Feb 1992	-0.56	0.44
Period 6	Mar 2001	Nov 2001	Jan 2001	Aug 2003	-0.54	0.37
Period 7	Dec 2007	Jun 2009	Jul 2007	Apr 2009	-0.68	0.90

NOTE: The first NBER recession list here started from December 1969, before the beginning of our sample period which was January 1970.

1981 to November 1982), $\hat{\alpha}_{t|T}$ and $\hat{\beta}_{t|T}$ are -93.82 and 0.35, respectively. During the period which is also identified as regime 2 and overlapped by the sixth NBER recession (March 2001 to November 2001), $\hat{\alpha}_{t|T}$ and $\hat{\beta}_{t|T}$ are -54.10 and 0.37, respectively. The results further document the effectiveness of our method.

The estimates $\hat{\alpha}_{t|T}$ and $\hat{\beta}_{t|T}$ within regime 1 are not as stable as the estimates within regime 2. For example, during the period between the fifth and sixth recessions, the regime is identified as 1 since $P(s_t) = 1$ is greater than 0.5. But the estimates $\hat{\alpha}_{t|T}$ and $\hat{\beta}_{t|T}$ fluctuate a lot as several step functions. In Table 3.11, we list all the values of those step functions within regime 1. Some unstable periods, for example, the period right after the fifth recession and the one after the seventh recession, are not included.

Compare the results with the one shown in Figure 1.3, in which most subperiods are identified as regime 1, fewer transitions are recognized and most of the transitions are not coincident with the NBER recessions, our results are more convincing and consistent with the empirical experience that employment situation varies directly and simultaneously with the business cycle. Furthermore, consistent with our model assumptions, the behaviors of

Table 3.11: Summary of the estimated parameters in periods identified as in regime 1.

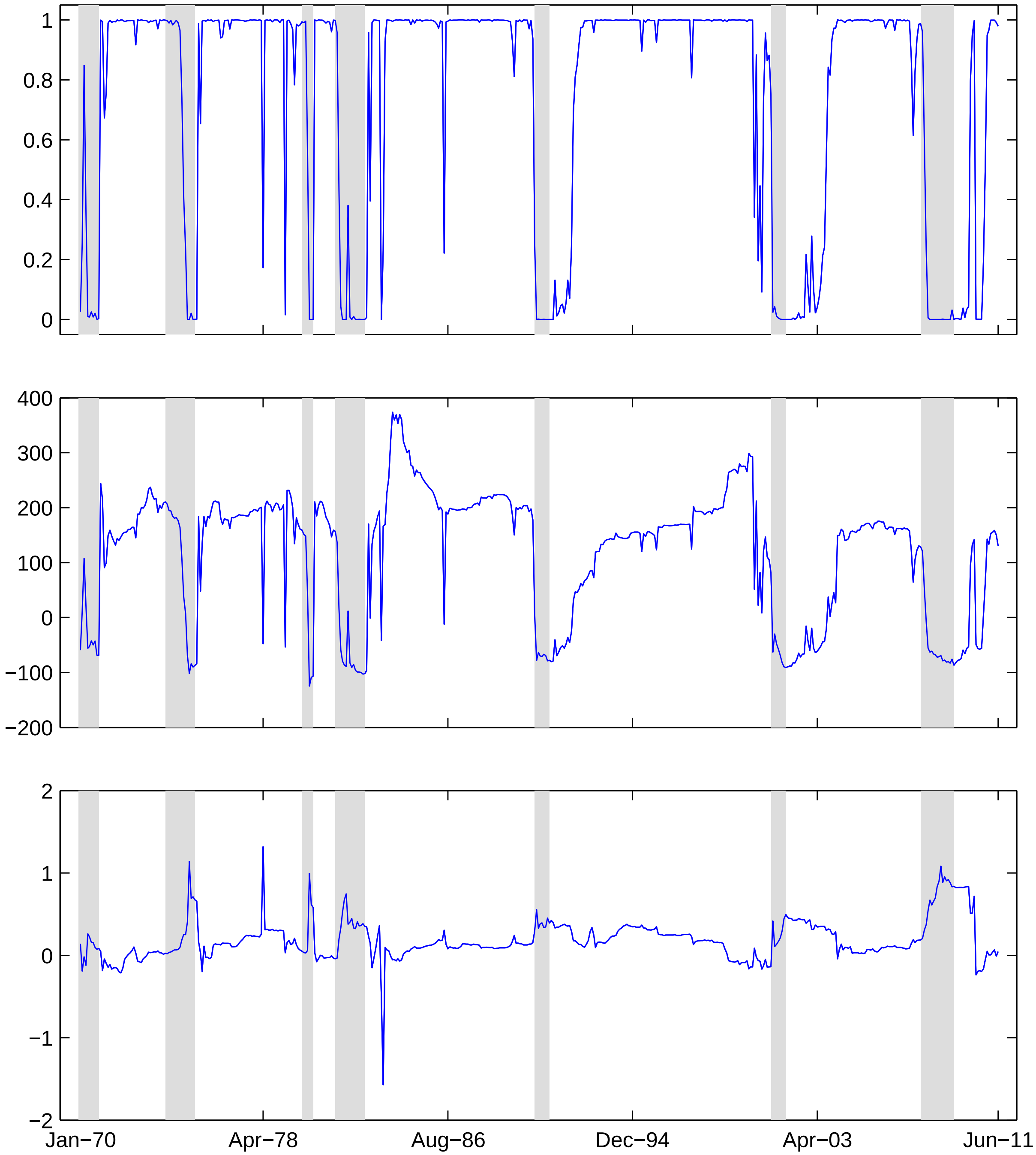
	"Regime 1" Periods		*Estimates of Coefficients*	
Subperiods	Starting	Ending	α_t ($\times 10^2$)	β_t
Period 1	Dec 1970	May 1974	1.80	0.07
Period 2	May 1975	Mar 1979	2.04	0.30
Period 3	May 1979	Mar 1980	1.39	0.07
Period 4	Aug 1980	Jul 1981	1.56	-0.04
Period 5	Jan 1983	Jul 1983	1.91	0.37
Period 6	Oct 1983	May 1990	2.00	0.13
Period 7	Sept 1993	Dec 1995	1.60	0.28
Period 8	Feb 1996	May 2000	2.85	-0.12
Period 9	Sept 2003	Jun 2007	1.43	0.14

employment figures during different recessions are different. In this sense, our model provides a framework to obtain more information from the economic time series than the classic regime switching model. The fitted values $\hat{y}_t$ and the errors between y_t and $\hat{y}_t$ are shown in Figure 3.21. The errors are distributed evenly around zero without any pattern.

We further calculate the filtering estimates $\hat{\boldsymbol{\beta}}_{t|t}$ and forecasts $\hat{\boldsymbol{\beta}}_{t+1|t}$ of the parameters which are important for the ongoing estimation and prediction. Figure 3.22 shows the filtering estimates. Comparing the figure with Figure 3.20, we can see more fluctuations in the estimated probability of $P(s_t) = 1$. But most "regime 2" periods are captured accurately. The filtering estimates of parameters do not look like step functions as in Figure 3.20. The spikes are more significant. Moreover, the jumps in the values are gradual around transitions. So the transitions are not very clear. But overall the transitions are realized quickly.

Figure 3.23 shows the forecasts $\hat{\alpha}_{t+1|t}$ and $\hat{\beta}_{t+1|t}$. Compared with the filtering estimates shown in Figure 3.22, the predicted values are more unstable, especially the $\hat{\beta}_{t+1|t}$. Since there are too many fluctuations in $\hat{\beta}_{t+1|t}$, we cannot identify the transitions efficiently. The intercept terms, $\hat{\alpha}_{t+1|t}$, show a much better shape. A "regime 2" period can be identified

Figure 3.22: Top panel: Filtering probability $\hat{r}^{(1)}_{t|t}$ of PAYEMS series. Middle panel: Filtering estimates $\hat{\alpha}_{t|t}$ of PAYEMS series. Bottom panel: Filtering estimates $\hat{\beta}_{t|t}$ of PAYEMS series. NBER recessions are shown as shaded areas.



when a local minimum is reached.

Figure 3.24 compares the predicted values of y_{t+1} with the true values. Since the predicted parameters are volatile, the predicted values of y_{t+1} show a lot of fluctuations. But the forecasts capture the feature that the decreases in the series are overlapped by NBER recessions.

Figure 3.23: Top panel: Forecasts $\hat{\alpha}_{t+1|t}$ of PAYEMS series. Bottom panel: Forecasts $\hat{\beta}_{t+1|t}$ of PAYEMS series. NBER recessions are shown as shaded areas.

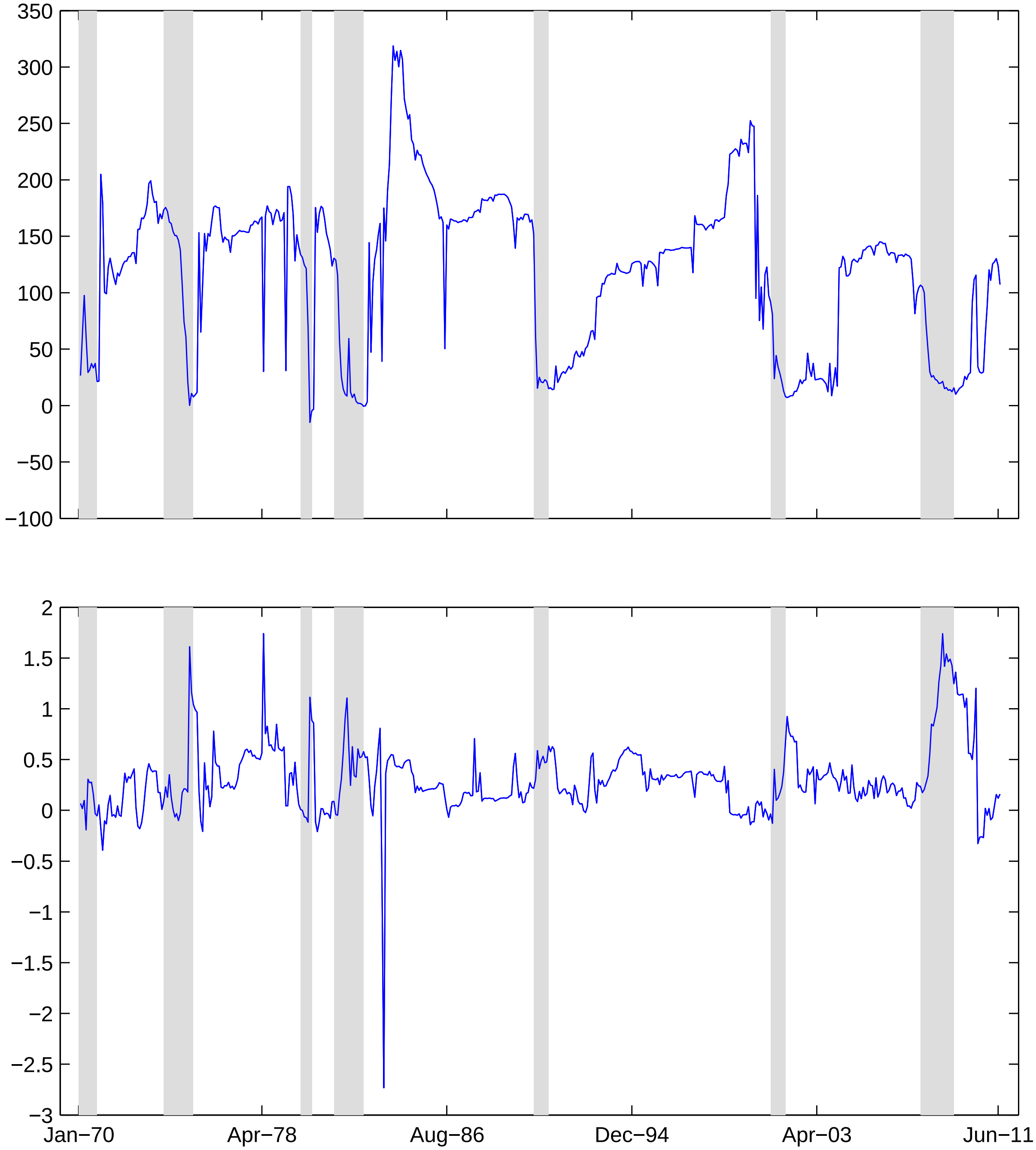
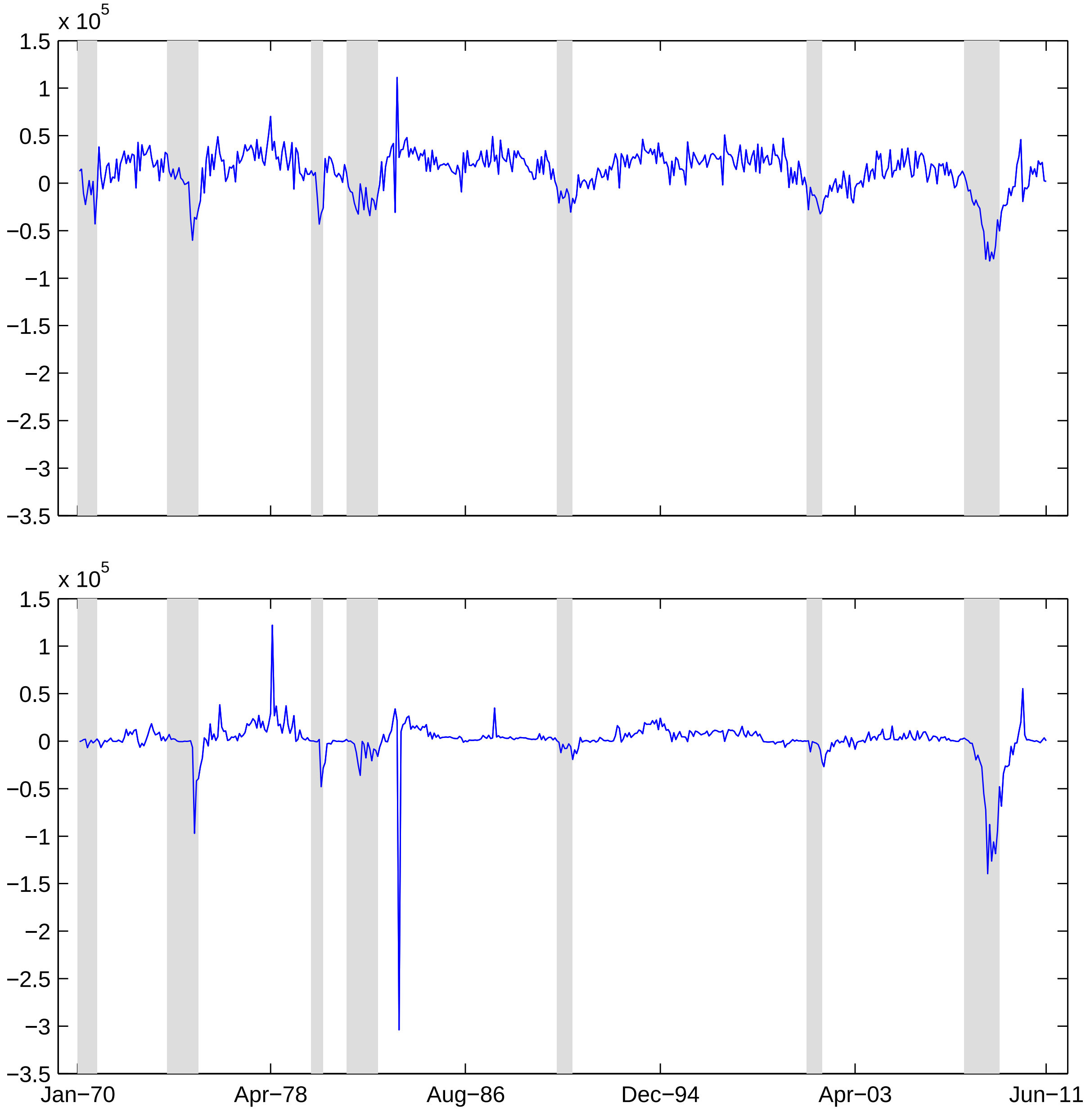


Figure 3.24: PAYEMS: Predicted y_{t+1} and the true values. NBER recessions are shown as shaded areas.



Chapter 4

Conclusions

For the analysis of econometric time series, we proposed a class of stochastic regime switching models and an associated inference framework that has attractive statistical and computational properties. The stochastic regime switching model in Chapter 2 assumes that $y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_t + \epsilon_t$, in which ϵ_t are independent normal random variables with mean zero and variance σ^2, and $\boldsymbol{\beta}_t$ is an unknown step function whose prior distribution depends on a finite state hidden Markov chain s_t. After the hidden state shifts from one regime to another regime, the model parameters jump to another set of values, which are generated by regime-dependent prior distributions and hence are not necessarily same as those within the same regime during the past.

A forward filtering procedure shows the posterior distribution of the parameter as a mixture distribution with explicit weights which can be calculated recursively. Furthermore, based on the reversibility of the hidden Markov chain, a backward filtering procedure can be conducted in a similar way. Based on Bayes' theorem, both the smoothing estimate of parameter and probability of regimes can be calculated explicitly to save a time-consuming numerical filtering procedure. The hyperparameters in the model can be estimated by the Expectation-Maximum (EM) algorithm. Furthermore, a Bounded Complexity Mixture Approximation (BCMIX) is shown to have much lower computational complexity yet comparable to the Bayes estimates in statistical efficiency. Simulation studies evaluate the Bayes and BCMIX estimates in terms of the sum of squared errors (SSE), the Kullback-Leibler divergence (κ) and L_2 errors. Moreover, the accuracy of identifying the transitions is evaluated by an Identification Ratio (IR). Applying this model to the historical data of "all employees: Total Nonfarm" shows that the fluctuations of the AR parameter can be recovered via our model. They are closely related to the recession history identified by NBER.

An important benefit of our Bayesian model is that we can derive analytical filtering and smoothing formulas for the posterior distributions of model parameters and make inference on regimes. The BCMIX estimate has much lower computational complexity yet comparable to the Bayes estimate in statistical efficiency.

Reference

Abel, A. (1994). Exact solutions for expected rates of return under Markov regime switching: Implications for the equity premium. *Journal of Money, Credit and Banking*, 26, 345-361.

Abel, A.(1999). Risk premia and term premia in general equilibrium. *Journal of Monetary Economics*, 43, 3-33.

Albert, J.H. and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*, 88, 669-679.

Ang, A. and Bekaert, G. (2002). International asset allocation with regime shifts. *Review of Financial Studies*, 15, 1137-1187.

Bekaert, G. and Hodrick, R. (1993). Characterizing predictable components in excess returns on equity and foreign exchange markets. *Journal of Finance*, 47, 467-509.

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

Box, G.E. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley.

Brown, R.L. and Durbin, J. (1968). Methods of investigating whether a regression relationship is constant over time. Paper presented at European Statistical Meeting.

Cai, J. (1994). A Markov model of unconditional variance in ARCH. *Journal of Business and Economic Statistics*, 12, 309-316.

Cerra, V. and Saxena, S.C. (2005). Did Output Recover from the Asian Crisis? IMF Staff Papers, 52, 1-23.

Cecchetti, S.G., Lam, P. and Mark, N.C. (1990). Mean reversion in equilibrium asset prices. *American Economic Review*, 80, 398-418.

Cecchetti, S.G., Lam, P. and Mark, N.C. (1993). The equity premium and the risk-free rate: Matching the moments. *Journal of Monetary Economics*, 31, 21-45.

Cho, J.S. and White, H. (2007). Testing for Regime Switching. *Econometrica*, 75, 1671-1720.

Cosslett, S.R. and Lee, L. (1985). Serial correlation in discrete variable models. *Journal of Econometrics* 27, 79-97.

Dai, Q., Singleton, K.J., and Wang, W. (2007). Regime shifts in a dynamic term structure model of U.S. treasury bond yields. *The Review of Financial Studies*, 20, 1669-1706.

Davig, T. (2004). Regime-switching debt and taxation. *Journal of Monetary Economics*, 51, 837-859.

Diebold, F.X., Lee, J. and Weinbach, G.C. (1994). Regime switching with time-varying transition probabilities. in Hargreaves, C. (ed.) *Nonstationary Time Series Analysis and Cointegration*, Oxford University Press.

Engel, C. (1994). Can the Markov Switching model forecast exchanges rates? *Journal of International Economics*, 36, 151-165.

Engel, C. and Hamilton, J.D. (1990). Long Swings in the Dollar: Are They in the Data and Do Markets Know It? *The American Economic Review*, 80, 689-713.

Fair, R.C. and Jaffee, D.M. (1972). Methods of estimation for markets in disequilibrium. *Econometrica*, 40, 497-514.

Farley, J.U. and Hinich, M.J. (1970). A test of a shifting slope coefficient in a linear model. *Journal of the American Statistical Association*, 65, 1320-1329.

Fiess, N. and Shankar, R. (2009). Determinants of exchange rate regime switching. *Journal of International Money and Finance*, 28, 68-98.

Filardo, A.J. (1994). Business cycle phases and their transitional dynamics. *Journal of Business and Economic Statistics*, 12, 299-308.

Frömmel M., MacDonald, R. and Menkhoff, L. (2005). Markov switching regimes in a monetary exchange rate model. *Economic Modelling*, 22, 485-502.

Garcia, R. and Perron, P. (1996). An analysis of the real interest rate under regime shifts. *The Review of Economics and Statistics*, 78, 111-125.

Garcia, R. (1998). Asymptotic null distribution of the likelihood ratio test in Markov switching models. *International Economic Review*, 39, 763-788.

Goldfeld, S.M. and Quandt, R.E. (1973). A Markov model for switching regressions. *Journal of Econometrics*, 1, 3-15.

Gray, S.F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42, 27-62.

Guidolin, M. and Timmermann, A. (2006). Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control*, 31, 3503-3544.

Guidolin, M. and Timmermann, A. (2008). International asset allocation under regime switching, skew and kurtosis preference. *Review of Financial Studies*, 21, 889-893.

Guo, F., Chen, C.R. and Huang, Y.S. (2011). Markets contagion during financial crisis: A regime-switching approach. *International Review of Economics and Finance*, 20, 95-109.

Hamilton, J.D. (1988). Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control*, 12, 385-423.

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

Hamilton, J.D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45, 39-70.

Hamilton, J.D. (1994). *Time Series Analysis*, Princeton University Press.

Hamilton, J.D. (1996). Specification testing in Markov-switching time-series models. *Journal of Econometrics*, 70, 127-157.

Hamilton, J.D. (2005). What's real about the business cycle? *Federal Reserve Bank of St. Louis Review*, 87, 435-452.

Hamilton, J.D. and Susmel, R. (1994). Autoregressive conditional heteroscedasticity and changes in regime. *Journal of Econometrics*, 64, 307-333.

Hansen, B. (1992). The likelihood ratio test under non-standard conditions: Testing the Markov switching model of GNP. *Journal of Applied Econometrics*, 7, S61-S82.

Hauzenberger, K. (2010). Flexible regime-switching projections to estimate the dynamic effects of a government spending stimulus, working paper.

Jeanne, O. and Masson,P. (2000). Currency crises, sunspots, and Markov-Switching regimes. *Journal of International Economics*, 50, 327-350.

Kim, C. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60, 1-22.

Kim, C. and Nelson, C.R. (1999). *State-space models with regime switching*, MIT Press.

Lai, T.L., Liu, H., and Xing, H. (2005). Autoregressive models with piecewise constant volatility and regression parameters. *Statistica Sinica*, 15, 279-301.

Lai, T.L., Xing, H., and Zhang, N. (2008). Stochastic segmentation models for array-based comparative genomic hybridization data analysis. B*iostatistics*, 9, 290-307.

Lai, T.L. and Xing, H. (2011). A simple Bayesian approach to multiple change-points. *Statistica Sinica*, 21, 539-569.

Lettau, M., Ludvigson, S. and Wachter, J. (2003). The declining equity premium: What role does macroeconomic risk play? *The Review of Financial Studies*, 21, 1653-1687.

Perlin, M. (2010). MS Regress - The MATLAB Package for Markov Regime Switching Models, Working Paper. `http://ssrn.com/abstract=1714016`.

Peria, M.S.M. (2002). A regime-switching approach to the study of speculative attacks: A focus on EMS crises. in Hamilton, J.D. and Raj, B. (ed.), *Advances in Markov-Switching Models: Applications in Business Cycle Research and Finance*, Physica-Verlag HD.

Quandt, R.E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, 53, 873-880.

Quandt, R.E. (1960). Test of the hypothesis that a linear regression obeys two separate regimes. *Journal of the American Statistical Association*, 55, 324-330.

Quandt R.E. (1972). A new approach to estimating switching regressions. *Journal of the American Statistical Association*, 67, 306-310.

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

Sims, C. and Zha, T. (2006). Were there regime switches in U.S. monetary policy? *The American Economic Review*, 1, 54-81.

So, M.K.P., Lam, K. and Li, W.K. (1998). A stochastic volatility model with Markov switching. *Journal of Business and Economic Statistics*, 16, 244-253.

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

Turner, C., Startz, R. and Nelson, C. (1989). A Markov model of heteroskedasticity, risk and learning in the stock market. *Journal of Financial Economics*, 25, 3-22.

Veronesi, P. (1999). Stock market overreactions to bad news in good times: a rational expectations equilibrium model.*Review of Financial Studies*, 12, 975-1007.

Whitelaw, R.F. (2000). Stock market risk and return: An equilibrium approach. *Review of Financial Studies*, 13, 521-547.