3. Methods

In this chapter, methods used for performing the change point analysis in this thesis are explained. The illustration of the simple Markov switching model feature is shown here and more general model specifications namely Markov switching autoregressive model is discussed.

3.1. Markov switching model

Markov switching model is used for time series that are evolved over unobserved distinct states or regimes. This is a regime switching model where the shifting back and forth between the regimes is controlled by a latent Markov chain. The model structure consists of two stochastic processes embedded in two levels of hierarchy. One process is an underlying stochastic process that is not observable but it is possible to observe them through another stochastic process which generated the sequence of observation (Rabiner and Juang, 1986). The time of transition to different state and the duration in between is random. In addition, the state assumes to follow the Markov property that the future state depends only on the current state.

A Markov switching model is able to model more complex stochastic processes and describe changes in the dynamic behavior. A general structure of the model can be drawn in graphically as shown in Figure 3.1, where S_t and y_t denote the state sequence and observation sequence in the Markov process, respectively. The arrows from one state to another state in the diagram implied the conditional dependency.

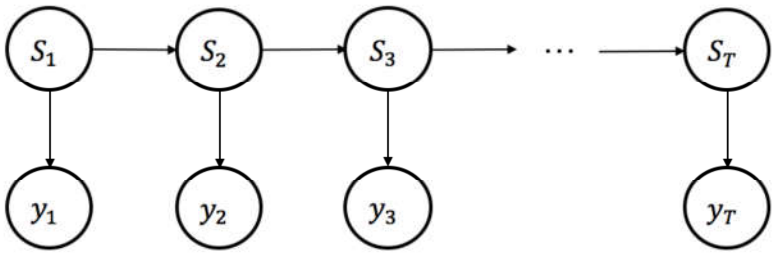

Figure 3.1.: Model structure

The process is given by (Hamilton, 1989)

$$y_t = X_t'\beta_{S_t} + \varepsilon_t \tag{3.1}$$

where y_t is the observed value of the time series at time t

X_t are the predictor variables of the time series at time t

β_{S_t} are the coefficients in state S_t, where $S_t = i$, $1 \leqslant i \leqslant k$

ε_t follows a Normal distribution with mean zero and variance given by $\sigma^2_{S_t}$ and $t = 1, 2, ..., t_0, t_0 + 1, t_0 + 2, ..., T$ where t_0 is a known change point.

The Equation 3.1 is the simplest form for the switching model. To aid understanding and simplicity, the baseline model assuming two states ($k = 2$) that is $S_t = 1$ or 2 is discussed. There are transitions between states which describe the distribution for the next state given the current state. Markov switching model assumes that S_t follows a stochastic process that satisfied Markov property. Hence, the evolution of changing from state $S_t = i$ to $S_t = j$ is defined by the transition probabilities as follow

$$p_{ij} = P(S_t = j|S_{t-1} = i, S_{t-2} = l, ..., S_1 = q) = P(S_t = j|S_{t-1} = i) \tag{3.2}$$

where $1 \leqslant i, j \leqslant k$.

The transition probabilities in a matrix notation with two-state Markov chain is given by

$$\mathbf{P} = \begin{bmatrix} P(S_t = 1|S_{t-1} = 1) & P(S_t = 1|S_{t-1} = 2) \\ P(S_t = 2|S_{t-1} = 1) & P(S_t = 2|S_{t-1} = 2) \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

It is also assumed that $\sum_{j=1}^{N} p_{ij} = 1$ and $p_{ij} \geq 0$. These two properties ensure ergodicity of Markov chain (i.e., aperiodic and irreducible) that any state will be reached eventually regardless of the initial state. They are also necessary for assessing the consistency of the model.

Since the whole process S_t is unobserved, the initial state which describes the starting distribution over states also needs to be specified.

$$\pi_j = P(S_0 = j) \tag{3.3}$$

3.2. Autoregressive (AR) model

Autoregressive model is one type of time series model that uses for describing the time-varying process. The model is flexible in handling various kinds of time series patterns. The name autoregressive comes from how the model performs a regression of the variable against its own previous outputs (Cryer and Kellet, 1986). The number of autoregressive lags is denoted by p. Hence, an autoregressive model of order p or AR(p) model can be written as

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t \tag{3.4}$$

or

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \tag{3.5}$$

where c is a constant, ϕ_i are coefficients in the autoregression and ε_t is Gaussian white noise with zero mean and variance σ^2. If p is equal to one, the model is called the first order autoregression process or AR(1).

3.3. Markov switching autoregressive model

This model is an extension of the basic Markov switching model which has dependencies at the observation level. Markov switching autoregressive model relaxes the conditional independent assumption by allowing an observation to also depends on a past observation and a current state (Shannon and Byrne, 2009). The observations are drawn from autoregression process and the first order Markov switching autoregressive model is

$$y_t = X_t' \beta_{S_t} + \phi_{1,S_t} y_{t-1} + \varepsilon_t \tag{3.6}$$

where ϕ_{1,S_t} is the autoregression coefficient of the observed value at time $t-1$ in state S_t and ε_t follows a Normal distribution with mean zero and variance given by $\sigma^2_{S_t}$.

The structure of the model is shown in Figure 3.2. It can be clearly seen that observation is, now, not independent from one another.

Assuming two states $S_t = 1$ or 2, the set of parameters for the Markov switching autoregressive model that are necessary for describing the law of probability governs y_t are $\theta = \{\beta_1, \beta_2, \phi_{1,1}, \phi_{1,2}, \sigma_1^2, \sigma_2^2, \pi_1, \pi_2, p_{11}, p_{22}\}$.

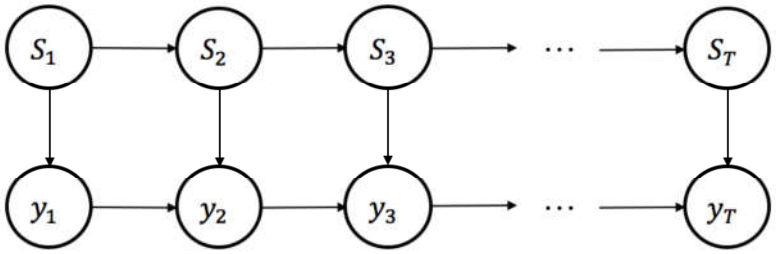

Figure 3.2.: Model structure of Markov switching AR(1)

3.4. Parameter Estimation

There are various ways to estimate parameters of Markov switching model. Methods which have been widely use are as follow: E-M algorithm (Hamilton, 1990; Kim, 1994) used the maximum likelihood criterion, Segmental K-mean (Juang and Rabiner, 1990) used K-means algorithm and maximized the state-optimized likelihood criterion, and Gibbs sampling (Kim et al., 1999) used a Markov chain Monte Carlo simulation method based on the Bayesian inference. In this thesis framework, E-M algorithm is used in estimating parameters and is briefly described a general procedure.

3.4.1. The Expectation-Maximization algorithm

E-M algorithm is originally designed to deal with the incomplete or missing values in data (Dempster et al., 1977). Nevertheless, it can potentially implement in Markov switching model since the unobserved state S_t can be viewed as the missing values. The set of parameters is estimated by iterative two-step procedure. It starts with an arbitrary initial parameters and finds the expected values of the state process given the observations. Next, the new maximum likelihood from the derived parameters in previous step is calculated. These two steps are repeated until the maximum value of the likelihood function is reached (Janczura and Weron, 2012).

3.4.1.1. E-step

Assume that $\theta^{(n)}$ is the derived set of parameters in M-step from the previous iteration and the available observations of time $t-1$ is denoted as $\Omega_{t-1} = (y_1, y_2, ..., y_{t-1})$. The general idea of this step is to calculate the expectation of S_t under the current estimation of the parameters. The obtained result is called smoothed inferences probability and is denoted by$P(S_t = j|\Omega_T; \theta)$. The E-step consists of filtering and smoothing algorithm and the process is described as follow (Kim, 1994):

Filtering Filtered probability is the probability of the non-observable Markov chain being in a given state j at time t, conditional on information up to time t. The algorithm starts from $t = 1$ to $t = T$. The starting point for the first iteration where $t = 1$ is chosen as arbitrary values. The probabilities of each state given that the available observation is up to time $t - 1$ is calculated.

$$P(S_t = j|\Omega_{t-1}; \theta^{(n)}) = \sum_{i=1}^{N} p_{ij}^{(n)} P(S_{t-1} = i|\Omega_{t-1}; \theta^{(n)}) \tag{3.7}$$

and the conditional densities of y_t given Ω_{t-1} are

$$f(y_t|\Omega_{t-1}; \theta^{(n)}) = \sum_{j=1}^{N} f(y_t|S_t = j, \Omega_{t-1}; \theta^{(n)}) P(S_t = j|\Omega_{t-1}; \theta^{(n)}) \tag{3.8}$$

where $f(y_t|S_t = j, \Omega_{t-1}; \theta) = \frac{1}{\sqrt{2\pi\sigma_{S_t}^2}} exp\left\{-\frac{(y_t-\beta_{S_t})^2}{2\sigma_{S_t}^2}\right\}$ is the likelihood function in each state for time t.

Then, with the new observation at time t, the probabilities of each state are updated by using Bayes' rule

$$P(S_t = j|\Omega_t; \theta^{(n)}) = \frac{f(y_t|S_t = j, \Omega_{t-1}; \theta^{(n)}) P(S_t = j|\Omega_{t-1}; \theta^{(n)})}{f(y_t|\Omega_{t-1}; \theta^{(n)})} \tag{3.9}$$

It is computing iteratively until all the observation is reached i.e., $t = T$.

Smoothing Smoothed probability is the probability of the non-observable Markov chain being in state j at time t, conditional on all available information. The algorithm iterates over $t = T - 1, T - 2, ..., 1$. The starting value is obtained from the final iteration of the filtered probabilities.

By nothing that

$$\begin{aligned} P(S_t = j|S_{t+1} = i, \Omega_T; \theta^{(n)}) &\approx P(S_t = j|S_{t+1} = i, \Omega_t; \theta^{(n)}) \\ &= \frac{P(S_t = j, S_{t+1} = i|\Omega_t; \theta^{(n)})}{P(S_{t+1} = i|\Omega_t; \theta^{(n)})} \\ &= \frac{P(S_t = j|\Omega_t; \theta^{(n)}) p_{ij}^{(n)}}{P(S_{t+1} = i|\Omega_t; \theta^{(n)})} \end{aligned} \tag{3.10}$$

and

$$P(S_t = j|\Omega_T; \theta^{(n)}) = \sum_{i=1}^{N} P(S_t = j, S_{t+1} = i|\Omega_T; \theta^{(n)}) \tag{3.11}$$

Then, the smoothed probabilities can be expressed as

$$P(S_t = j|\Omega_T; \theta^{(n)}) = \sum_{i=1}^{N} \frac{P(S_{t+1} = i|\Omega_T; \theta^{(n)}) P(S_t = j|\Omega_t; \theta^{(n)}) p_{ij}^{(n)}}{P(S_{t+1} = i|\Omega_t; \theta^{(n)})} \tag{3.12}$$

Once the filtered probabilities are estimated, there is necessarily enough information to calculate the full log-likelihood function.

$$\ln L(\theta) = \sum_{t=1}^{T} \ln(f(y_t|\Omega_{t-1}; \theta^{(n)}) = \sum_{t=1}^{T} \ln \sum_{j=1}^{N} ((f(y_t|S_t = j, \Omega_{t-1}; \theta^{(n)}) P(S_t = j|\Omega_{t-1})) \tag{3.13}$$

This is simply a weighted average of the likelihood function in each state. The probabilities of states are considered as weights.

3.4.1.2. M-step

The new estimated model parameters $\theta^{(n+1)}$ is obtained by finding the set of parameters that maximizes the Equation 3.13. This new set of parameters is more exact value of the maximum likelihood estimates than the previous one. It serves as the set of parameters in the next iteration of the E-step. The estimated parameters are derived by taking the partial derivative of the log-likelihood function with respect to the specific parameter and then setting it to zero. Generally, this process is similar to the standard maximum likelihood estimation except that it has to be weighted by the smoothed probabilities since each observation y_t carries probability of coming from any of the k state.

3.5. Model selection

Model selection is a task of selecting the most suitable model for a given set of data based on the quality of model. In this thesis framework, the Bayesian Information

Criterion (BIC) is widely employed in the applied literature and proved to be useful in selecting the model among a finite set of models. BIC is defined as

$$-2\ln(L(\hat{\theta})) + m \cdot \ln(T)$$

where $L(\hat{\theta})$ represents the maximized value of the likelihood function, T is the number of observations and m is the number of parameters to be estimated in the model.

3.6. ecp

ecp is an extension package in R which mainly focus on computing the non-parametric test for multiple change point analysis. This method is applicable to both univariate and multivariate time series. A fundamental idea for building algorithms that can identify the change point is based on either divisive or agglomerative procedure in hierarchical clustering approach (James and Matteson, 2013).

3.6.1. E-divisive

E-divisive algorithm recursively partitions a time series. An estimation of the change point is computed at each iteration. Permutation test is performed to find the statistical significance of an estimated change point.

3.6.2. E-agglomerative

E-agglomerative algorithm tries to maximize a goodness of fit test after merging segments in the iteration. The estimated change point is defined by the iteration that maximized a goodness of fit statistic. This algorithm allows user to input an initial segmentation for the time series or a prior knowledge of the possible change point in order to reduce the computation time.