Statistics and Its Interface Volume 0 (2019) 1

Markov-Switching Poisson Generalized Autoregressive Conditional Heteroscedastic Models

Jichun Liu, Yue Pan, Jiazhu Pan*, and Abdullah Almarashi

We consider a kind of regime-switching autoregressive models for nonnegative integer-valued time series when the conditional distribution given historical information is Poisson distribution. In this type of models the link between the conditional variance (i.e. the conditional mean for Poisson distribution) and its past values as well as the observed values of the Poisson process may be different when an unobservable (hidden) variable, which is a Markovian Chain, takes different states. We study the stationarity and ergodicity of Markov-switching Poisson generalized autoregressive heteroscedastic (MS-PGARCH) models, and give a condition on parameters under which a MS-PGARCH process can be approximated by a geometrically ergodic process. Under this condition we discuss maximum likelihood estimation for MS-PGARCH models. Simulation studies and application to modelling financial count time series are presented to support our methodology.

AMS 2000 subject classifications: Primary 62M10; secondary 91B84, 37M10.
Keywords and phrases: Count Time Series, Markov Regime Switching, Generalized Conditional Heteroscedasticity, Geometric Ergodicity, Poisson GARCH, Smoothing.

1. INTRODUCTION

An important topic in econometrics and statistics is to analyze the dynamic behaviour of economic and financial variables. Fortunately, there are many leading choices of linear and nonlinear models for conditional mean and conditional variance. In the past few decades, there has been growing interest in nonlinear time series models[20, 40, 18, 10, 42].

Regime-switching has been introduced in different models, including threshold models[41] and ARCH/GARCH models[22], and has various applications in economics, such as analyzing business circle[29], GNP[24], interest rate[19] and monetary policy[37]. Regime-switching models such as the Markov switching autoregressive (MSAR) models and the threshold autoregressive (TAR) have become increasingly popular for nonlinear time series and have been used to describe regime switching phenomena in various fields in science and social science. There are many studies on these kinds of models in literature[41, 27, 31, 5, 33, 34, 8, 25]. These models consist of different regimes which can describe the time series pattern in various states. The switching between states is according to an unobservable (hidden) variable or a time-delayed variable, and each state follows a linear model.

Markov switching refers to regime-switching based on a hidden Markov chain. The main purpose of the Markov switching model is to study the mean behaviour of variables. Therefore, many authors have used conditional mean models with Markov switching[20, 9, 17, 16, 36]. Since this type of models have been used satisfactorily with high success, it is informative to consider incorporating the switching into a conditional heteroscedasticity model. The best choices for conditional heteroscedasticity are the ARCH/GARCH models. So the ARCH/GARCH with Markov switching has been introduced. For example, Markov switching has been introduced to the stochastic volatility model[38], and maximum likelihood estimation for Markov switching autoregressive conditional heteroscedasticity model has been discussed[35].

Many papers have appeared in economic applications of switching conditional mean and conditional variance models, especially in applying this type of models to analysis of financial time series[22, 23, 31]. Almost all references on Markov switching are about continuous type of time series data. However, many discrete-valued data are needed to be analyzed in finance, economics, network modelling, medicine and other fields. Although there have been some papers on discrete type of data in the literature[32, 11, 14, 13, 43], the Markov-switching has not been combined to discrete-valued time series models yet. In this paper, we discuss Poisson Markov switching models for count time series with nonnegative integer values. Because the conditional mean coincides with the conditional variance for a Poisson process, we call our model a Markov switching Poisson generalized autoregressive conditional heteroscedastic (MS-PGARCH) model.

The contents of this paper are organized as follows. Section 2 introduces the model and discusses its probabilistic properties such as geometric ergodicity. We overcome mathematical difficulties to obtain a geometrically ergodic approximation to the MS-PGARCH model. We use the idea

*Corresponding author.

of the "perturbation" approach[14] but our proof process is more tricky because we have Markov-switching in the model. Section 3 gives the maximum likelihood estimation for the proposed model. Section 4 presents simulation results and a real data example to assess the modelling process by numerical evidence. Appendix A provides proofs of all the theoretical results and Appendix B gives details of the "collapsing procedure" to overcome "path dependence" problem in estimation of the model.

2. MODEL AND METHODOLOGY

2.1 Definition of the Model

Suppose that $\{Y_t\}$ is a nonnegative integer-valued time series. Consider the model defined by

$$(1) \qquad \begin{cases} Y_t|\mathcal{F}_{t-1}^{Y,\lambda,S} \sim \text{Poisson}(\lambda_t), \\ \lambda_t = c(S_t) + a(S_t)\lambda_{t-1} + b(S_t)Y_{t-1} \end{cases}$$

where $\mathcal{F}_{t-1}$ denotes the information set available at time $t-1$, which is a $\sigma-$field generated by $\{(Y_{t-1}, S_{t-1}), (Y_{t-2}, S_{t-2}), ..., (Y_0, S_0), (\lambda_0, S_0)\}$, and S_t is an irreducible and aperiodic stationary Markov chain with finite state space $S = \{1, 2, ...m\}$ and an $m \times m$ transition matrix P. A typical element of P is denoted by $p_{ij} = Pr(S_t = j|S_{t-1} = i)$, i.e. $P = [p_{ij}]$ where

$$p_{ij} = Pr(S_t = j|S_{t-1} = i), i, j = 1, 2, ..., m.$$

The stationary distribution $\{S_t\}$ is denoted by $\pi = (\pi_1, \pi_2, \pi_3, ..., \pi_m)^T$. The states of S_t represent the different regimes of the model. The coefficients $c_i = c(i)$, $a_i = a(i)$ and $b_i = b(i)$, $i = 1, 2, .., m$ are assumed to be positive. Note that the model defined by (1) is a model for the conditional variance of Y_t given the historical information. We call this model a Markov-switching Poisson generalized autoregressive heteroscedastic (MS-PGARCH) model, denoted by MS-PGARCH (m;1,1), where m is called the number of regimes of the model.

The model (1) can be rewritten in a form of Poisson process by assuming that Y_t represents the number of events $N_t(\lambda_t)$ of a Poisson process $N_t(\cdot)$ with unit intensity in time interval $(0, \lambda_t]$ as follows

$$(2) \qquad \begin{cases} Y_t = N_t(\lambda_t), \\ \lambda_t = c(S_t) + a(S_t)\lambda_{t-1} + b(S_t)Y_{t-1} \end{cases}$$

for $t \geq 1$, where the initial values Y_0 and λ_0 are assumed to be fixed, $N_t(\lambda_t)$ denotes the Poisson process with intensity λ_t, and λ_t is defined by (1).

2.2 Ergodicity of a Perturbed Model

We study the stationarity and ergodicity of $\{Y_t\}$ by proving these properties for $\{(\lambda_t, S_t)\}$, given a suitable initial distribution for (λ_0, S_0). However, establishing ψ-irreducibility and finding a small set is quite complicated because of the fact that λ_t has discrete-valued random innovations. Given the value of λ_0, the set of possible values for λ_1 is countable. In fact, the set of states that are reachable from a fixed starting state is also countable, and distinct initial values can have distinct sets of reachable locations. To avoid this issue, we consider a model with ε-perturbation defined by

$$(3) \qquad \begin{cases} Y_t^n = N_t(\lambda_t^n), \\ \lambda_t^n = c(S_t) + a(S_t)\lambda_{t-1}^n + b(S_t)Y_{t-1}^n + \varepsilon_{t,n} \end{cases}$$

where

$$(4) \qquad \varepsilon_{t,n} = c_n U_t, \quad c_n > 0, \quad c_n \to 0, \quad \text{as} \quad n \to \infty,$$

and $\{U_t\}$ is a sequence of iid uniform random variables on $(0, 1)$ and such that U_t is independent of $\mathcal{F}_{t-1}^{Y,\lambda,S}$.

We have the following results on the ergodicity of model (3). Their proofs all are postponed to Appendix A. There it is first proved that the unobserved process $\{(\lambda_t^n, S_t)\}$ is geometrically ergodic.

Before stating main results, we need some notations. Let

$$M_u := [p_{ji}(a(i) + b(i))^u], \quad i, j = 1, 2, \ldots, m; \quad u \geq 0.$$

Define a norm $\|D\| = \sum_{ij} d_{ij}$ for a nonnegative vector or matrix $D = [d_{ij}]$, and $\rho(D)$ for the spectral radius of matrix D.

Theorem 2.1. Suppose that $\rho(M_1) < 1$. Then, given an appropriate initial distribution for (λ_0^n, S_0), the process $\{(\lambda_t^n, S_t)\}$ defined in (3) is a stationary and geometrically ergodic Markov chain with finite first moment. Moreover, $\{Y_t^n\}$ defined in (3) is a stationary and ergodic process with finite first moment.

Remark 2.1. It is not difficult to verify that $\rho(M_1) = a+b$, if $a(S_t) \equiv a$ and $b(S_t) \equiv b$.

For the existence of high moments for the process $\{Y_t^n\}$, we have the following proposition.

Theorem 2.2. Suppose that $\rho(M_k) < 1$, where $k \geq 1$ is a positive integer. Given an appropriate initial distribution for (λ_0^n, S_0), the process $\{Y_t^n\}$ defined in (3) is a stationary and ergodic process with finite moment of order k.

Remark 2.2. A similar argument as Remark 2.1, we easily know that $\rho(M_k) = (a + b)^k$, if $a(S_t) \equiv a$ and $b(S_t) \equiv b$. Obviously, $a + b < 1$ if and only if $(a + b)^k < 1$. Hence, Theorems 2.1 and 2.2 above will reduce to Propositions 2.1 and 2.2 in [14], if $a(S_t) \equiv a$ and $b(S_t) \equiv b$.

The following lemma quantifies the difference between (2) and (3), as $n \to \infty$, and shows that the perturbed model can be made arbitrarily close to the unperturbed model.

Lemma 2.1. Suppose that $\rho(M_2) < 1$ and $\{(Y_t, \lambda_t)\}$ and $\{(Y_t^n, \lambda_t^n)\}$ are defined by (2) and (3) respectively, and $(\lambda_0^n, S_0) =_d (\lambda_0, S_0)$, then the following statements hold:

(1). $|E(\lambda_t^n - \lambda_t)| = |E(Y_t^n - Y_t)| \le \delta_{1,n}$,

(2). $E(\lambda_t^n - \lambda_t)^2 \le \delta_{2,n}$,

(3). $E(Y_t^n - Y_t)^2 \le \delta_{3,n}$

where $\delta_{i,n} \to 0$ as $n \to \infty$ for $i = 1, 2, 3$, and "$=_d$" denotes "equality in distribution". Furthermore, almost surely for any $\delta > 0$, with n large enough

$$|\lambda_t^n - \lambda_t| \le \delta \quad \text{and} \quad |Y_t^n - Y_t| \le \delta.$$

To make the analysis simple, we describe the model (1) with two regimes in detail next. Assume that the unobservable Markov chain S_t has two states: 1 and 2. A simple Markov switching model for the process λ_t contains two GARCH specifications: each of them follows a $GARCH(1,1)$ model. The process λ_t switches between two regimes according to the value of the state variable S_t. The limiting unconditional mean (as $t \to \infty$) of the process is

$$\frac{c_1 P(s_t = 1) + c_2 P(s_t = 2)}{1 - [(a_1 + b_1)P(s_t = 1) + (a_2 + b_2)P(s_t = 2)]}.$$

The process λ_t is governed by two regimes with special means, and the transition between regimes is determined by the value of the state variable S_t. Lemma 2.1 shows the MS-PGARCH $(2; 1, 1)$ can be approximated by the following process

$$Y_t^n = N_t(\lambda_t^n),$$
$$(5) \quad \lambda_t^n = \begin{cases} c_1 + a_1\lambda_{t-1}^n + b_1 Y_{t-1}^n + \varepsilon_{1,tn}, & \text{if} \quad S_t = 1, \\ c_2 + a_2\lambda_{t-1}^n + b_2 Y_{t-1}^n + \varepsilon_{2,tn}, & \text{if} \quad S_t = 2, \end{cases}$$

where $\varepsilon_{t,n}$ satisfies (4).

Assume S_t in follows a first-order Markov chain with the transition matrix

$$P_2 \quad =: \quad \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{bmatrix}$$

where $p_{ij}, (i, j = 1, 2)$ denote the transition probabilities of the state variable $S_t = j$ given $S_{t-1} = i$. The transition probabilities satisfy $\sum_{j=1}^2 p_{ij} = 1$, and $0 \le p_{ij} \le 1, \quad i = 1, 2$. A small value of the transition probabilities p_{12} and p_{21} means that the process tends to stay longer in state 1 and 2 respectively. In general, the expected duration of the model to stay in the state 1 is $\sum_{k=1}^\infty k p_{11}^{k-1}(1 - p_{11}) = \frac{1}{1-p_{11}}$, and the expected duration of the other state is $\frac{1}{1-p_{22}}$[20]. The transition matrix P_2 above contains only two parameters (p_{11}, p_{22}). The stationary distribution of the Markov chain $\{S_t\}$ is

$$\pi = \begin{bmatrix} Pr(S_t = 1) \\ Pr(S_t = 2) \end{bmatrix} = \begin{bmatrix} \frac{1-p_{22}}{2-p_{11}-p_{22}} \\ \frac{1-p_{11}}{2-p_{11}-p_{22}} \end{bmatrix}.$$

According to (1), the MS-PGARCH$(2; 1, 1)$ model uses an unobservable Markov chain to govern the transition from one conditional mean (i.e. variance) to another. Note that the state variable S_t is independent of Y_{t-1} (over time) and this consequently may make it difficult for people to apply this model to analyzing time series data.

3. MAXIMUM LIKELIHOOD ESTIMATION OF MS-PGARCH MODEL

In the last few decades, many authors have discussed the different methods to estimate the Markov switching model[20, 21, 23, 28, 29, 15, 35]. Here, we concentrate on the MS-PGARCH(2;1,1) model of order (1,1). However, the procedure for high order MS-PGARCH model would be straightforward.

Because the states of the hidden Markov chain are not directly observable, it is quite difficult to estimate a MS-GARCH model. The EM algorithm[20] and the Gibbs sampler[1] could be applied by treating the hidden Markov process as parameters. The first method is more difficult to implement in the presence of AR lags[12], while the second approach requires heavy computation. Nevertheless, in this section, we show that it is possible to numerically approximate the maximum likelihood estimates of our model by adopting the collapsing procedure.

Denote the parameter vector as

$$\theta' = (\gamma', \eta')$$

where

$$\begin{cases} \gamma' = (a_1, a_2, b_1, b_2, c_1, c_2), \\ \eta' = (p_{11}, p_{22}). \end{cases}$$

The intensity λ_t could be written in an alternative form as $\lambda_t(S_{1:t})$ to demonstrate its dependence on past trajectory of the Markov process $\{S_t\}$, where $S_{1:t} := \{S_1, S_2, ..., S_t\}$. [23] proposed a filtering algorithm to evaluate the likelihood function for certain types of Markov switching models. Given T samples $Y_{1:T} := \{Y_1, Y_2, ..., Y_T\}$, the log-likelihood function of $(Y_{1:T}, S_{1:T})$ is

$$(6) \qquad L_T(\theta) = \sum_{t=1}^T \ell_{t-1}(\theta) := \sum_{t=1}^T \log\{f(Y_t|Y_{1:t-1}, \theta)\}.$$

In the remaining part of this section, the symbol θ is omitted from the likelihood functions for simplicity.

Since $S_{1:T}$ are not observable and the distribution of Y_t depends on the trajectory of $S_{1:t}$, the likelihood function of Y_t could only be evaluated by integrating over all possible realizations of $S_{1:t}$:

(7)
$$\begin{aligned} & f(Y_t|Y_{1:t-1}) \\ = & \textstyle\sum_{S_{1:t}} f(Y_t, S_{1:t}|Y_{1:t-1}) \\ = & \textstyle\sum_{S_{1:t}} f(Y_t|Y_{1:t-1}, S_{1:t}) Pr\{S_t|S_{t-1}\} g(S_{1:t-1}|Y_{1:t-1}) \end{aligned}$$

where

$$(8)\qquad f(Y_t|Y_{1:t-1},S_{1:t})=\frac{\{\lambda_t(S_{1:t})\}^{Y_t}exp\{-\lambda_t(S_{1:t})\}}{Y_t!}$$

and $g(S_{1:t}|Y_{1:t})$ is called the filtering probability of $S_{1:t}$ given $Y_{1:t}$, which could easily be derived as

$$(9)\qquad g(S_{1:t}|Y_{1:t})=\frac{f(Y_t,S_{1:t}|Y_{1:t-1})}{f(Y_t|Y_{1:t-1})}.$$

Given an initial distribution of S_0, the likelihood function (6) could be evaluated iteratively according to (7) − (9). However, note that the number of potential trajectories is 2^t in the evaluation of (7), hence such filtering algorithm would become numerically infeasible as t becomes large. To address this computational problem caused by the path dependence of λ_t on the hidden Markov process, a widely used technique in the literature is to approximate model (1) by a collapsed version where the intensity process only depends on a segment instead of the entire history of $\{S_t\}$. Different such approximations of Markov switching GARCH has been proposed[19, 7, 30], and then been generalized as the General Collapsing Procedure [2], which could also be applied to the ML estimation of our MSPGARCH model.

Set the length of segment as q, at $t=1,2,...,q$. The Hamilton filtering procedure (7) − (9) are adopted to calculate the exact likelihood function of model (1). Starting from $t=q+1$, the likelhood function is evaluated using the collapsed model

$$(10)\qquad \begin{cases}\tilde\lambda_t(S_{t-q+1:t})=c(S_t)+a(S_t)E_{t-1}+b(S_t)Y_{t-1},\\ E_t=\mathbb{E}\left\{\tilde\lambda_t(S_{t-q+1:t})|Y_{1:t},S_{t-q+2:t+1}\right\}.\end{cases}$$

Note that at each t, the dependence of $\tilde\lambda_t(S_{t-q+1:t})$ on S_{t-q+1} is removed by taking expectation conditioning on $Y_{1:t}$ and one-step-ahead moving segment of $S_{t-q+2:t+1}$, such that the approximated intensity process only depends on a window of lenghth q of its past trajectory. Similar to (7) − (9) we could approximate the likelihood function starting from $t=q+1$ by iterating (11) to (13) as follows:

$$(11)\qquad \tilde f(Y_t|Y_{1:t-1})=\sum_{S_{t-q+1:t}}\tilde f(Y_t|Y_{1:t-1},S_{t-q+1:t})\times Pr\{S_t|S_{t-1}\}\tilde g(S_{t-q:t-1}|Y_{1:t-1}),$$

where

$$(12)\qquad \tilde f(Y_t|Y_{1:t-1},S_{t-q+1:t})=\{\tilde\lambda_t(S_{t-q+1:t})\}^{Y_t}\times\frac{exp\{-\tilde\lambda_t(S_{t-q+1:t})\}}{Y_t!}$$

and

$$(13)\qquad \tilde g(S_{t-q+1:t}|Y_{1:t})=\frac{\tilde f(Y_t,S_{t-q+1:t}|Y_{1:t-1})}{\tilde f(Y_t|Y_{1:t-1})}.$$

The calculation of this conditional expectaion in (10) is actually a by-product of the filtering probabilities in the iteration of (11) to (13) since

$$(14)\qquad \begin{aligned}&\mathbb{E}\left\{\tilde\lambda_t(S_{t-q+1:t})|Y_{1:t},S_{t-q+2:t+1}\right\}\\ =&\sum_{S_{t-q+1}}\tilde\lambda_t(S_{t-q+1:t})\tilde g(S_{t-q+1}|S_{t-q+2:t+1},Y_{1:t})\end{aligned}$$

where

$$(15)\qquad \begin{aligned}&\tilde g(S_{t-q+1}|S_{t-q+2:t+1},Y_{1:t})\\ =&\begin{cases}\frac{Pr\{S_{t+1}|S_t\}\tilde g(S_t|Y_{1:t})}{\sum_{S_t}Pr\{S_{t+1}|S_t\}\tilde g(S_t|Y_{1:t})}, & if\quad q=1,\\ \frac{\tilde g(S_{t-q+1:t}|Y_{1:t})}{\sum_{S_{t-q+1}}\tilde g(S_{t-q+1:t}|Y_{1:t})}, & if\quad q\geq 2.\end{cases}\end{aligned}$$

Since the summation in (11) is over 2^q of segments, the computational complexity remain fixed for $t=q+1,q+2,\cdots$. According to [2], there is a deterministic bias in the estimation results produced in this approach due to the collapsing procedure, such bias is proved to be neglectable with sufficiently large q nevertheless. Thus there is a trade-off between accuracy and efficiency when it comes to the choice of q.

The collapsing procedure is discussed more in detail in Appendix B of this paper.

4. NUMERICAL RESULTS: SIMULATION STUDY AND A REAL DATA EXAMPLE

4.1 Simulations

In this section, we carry out simulation studies to show the finite sample properties of the MLEs for the MSPGARCH model. Datasets with sizes of $T=500,1500,5000$ are generated by following data generating process:

$$(16)\qquad \lambda_t=\begin{cases}0.3+0.2\lambda_{t-1}+0.1Y_{t-1}, & if\quad S_t=1,\\ 2+0.4\lambda_{t-1}+0.3Y_{t-1}, & if\quad S_t=2,\end{cases}$$

$$(17)\qquad P=\begin{bmatrix}0.98 & 0.02\\ 0.04 & 0.96\end{bmatrix}.$$

The MLEs are obtained by optimizing the log likelihood function derived based on the filtering algorithm with collapsing procedure, which we introduced in section 3 and the details of such algorithm are presented in Appendix B. The simulation is repeated for 1000 times with $q=8$ and different initial values of θ, and the results are summarized in Table 1 and Table 2.

The mean values of estimates are presented in Table 1, the root-mean-square errors (RMSE) are given in the parentheses to evaluate the performance of the results. As it is indicated by the decreasing trend of RMSE as the sample size becomes larger, the MLEs based on the collapsed model approximate the true parameters of model (16) quite well, hence the deterministic bias caused by the collapsing procedure is neglectable as expected with sufficiently large sample

Table 1. Maximum likelihood estimates with their RMSEs (in parentheses) for model (16) for different sample sizes.

Sample size	MLEs							
T	a_1	a_2	b_1	b_2	c_1	c_2	p_{11}	p_{22}
500	0.2024	0.4169	0.0970	0.2761	0.2990	2.0655	0.9787	0.9564
	(0.1241)	(0.1374)	(0.0588)	(0.0922)	(0.0660)	(0.5587)	(0.0100)	(0.0248)
1500	0.1979	0.3971	0.0961	0.2973	0.3028	2.0371	0.9786	0.9585
	(0.0826)	(0.0731)	(0.0360)	(0.0480)	(0.0370)	(0.2949)	(0.0057)	(0.0091)
5000	0.2049	0.3993	0.0986	0.3016	0.2973	1.9909	0.9804	0.9598
	(0.0426)	(0.0399)	(0.0230)	(0.0236)	(0.0186)	(0.1658)	(0.0025)	(0.0055)
θ_0	0.2	0.4	0.1	0.3	0.3	2	0.98	0.96

Table 2. The simulated standard error for model (16) for different sample sizes.

Sample size	Simulated Standard Error							
T	a_1	a_2	b_1	b_2	c_1	c_2	p_{11}	p_{22}
500	0.0125	0.0137	0.0059	0.0090	0.0066	0.0558	0.0010	0.0025
1500	0.0083	0.0073	0.0036	0.0048	0.0037	0.0294	0.0006	0.0009
5000	0.0043	0.0040	0.0023	0.0024	0.0019	0.0166	0.0002	0.0006

size. On the other hand, the estimation approach we use become more efficient as the standard error of estimates from 100 replications is significantly lower with larger sample size according to Table 2.

4.2 A Real Data Example

As illustration of our model and methodology, we fit a MS-PGARCH(2;1,1) to a real data set consisting of 650 observations of the number of transactions per minute for the stock Ericsson B which conveys around 11 hours of transactions in the period 2-3 July, 2002. The sample mean of this particular realization was 10.08, with a standard deviation of 5.6207. Figure 1 presents the observation numbers, histogram and sample autocorrelation function. From Figure 1, it can be seen that this time series does not exhibit any trending behaviour. The plot of sample autocorrelation function of the number of transactions suggests significant serial correlation between transactions.

Table 3 below shows the estimates of parameters in the model defined by (1) for $m = 2$.

The steady-state distribution of the Markov chain S_t is also estimated as

$$\pi = (0.721, 0.279)^\tau.$$

The results in Table 3 can be expressed as

$$\hat{\lambda}_t = \begin{cases} 0.053 + 0.870\hat{\lambda}_{t-1} + 0.099Y_{t-1}, & if \quad S_t = 1, \\ 3.484 + 0.915\hat{\lambda}_{t-1} + 0.084Y_{t-1}, & if \quad S_t = 2 \end{cases}$$

with transition matrix

$$P_2 = \begin{bmatrix} 0.716 & 0.284 \\ 0.820 & 0.180 \end{bmatrix}.$$

From Table 3, it can be seen that there are several interesting findings. First, the limiting unconditional mean (as $t \to \infty$) of the number of transactions per minute for the stock Ericsson B in the period July 2-22,2002 is

$$\frac{c_1 P(S_t = 1) + c_2 P(S_t = 2)}{1 - [(a_1 + b_1)P(S_t = 1) + (a_2 + b_2)P(S_t = 2)]} = 40.136.$$

Second, the transition probabilities appear to be different for different states. The process tends to stay longer in state 1 than of the state 2 . Third, there are two regimes of this transactions: one is called “contraction period” in which the process stays in the state 1, the other one is called “expansion period” in which the process stays in the state 2. Note that the expected duration of state 1 is $\frac{1}{1-p_{11}} = 3.3501$, while the expected duration of the other state is $\frac{1}{1-p_{22}} = 1.3005$. That is, the expected duration for a contraction period and an expansion period are approximately 3.3501 and 1.3005 minutes respectively.

5. CONCLUSION

This paper incorporates Markov regime-switching into Poisson process and proposes the Markov-switching Poisson generalized autoregressive conditional heteroscedastic (MS-PGARCH) model. This model can be used to describe the dynamics of an integer-valued time series whose conditional variance/mean varies over time switching between different patterns. In theory, we mainly studied the fundamental properties such as stationarity and ergodicity of MS-PGARCH models. It is proved that the proposed model can be approximated by a series of geometrically ergodic processes under mild condition on parameters. This enables the statistical inference about this model. For practical empirical implementation, the maximum likelihood estimate

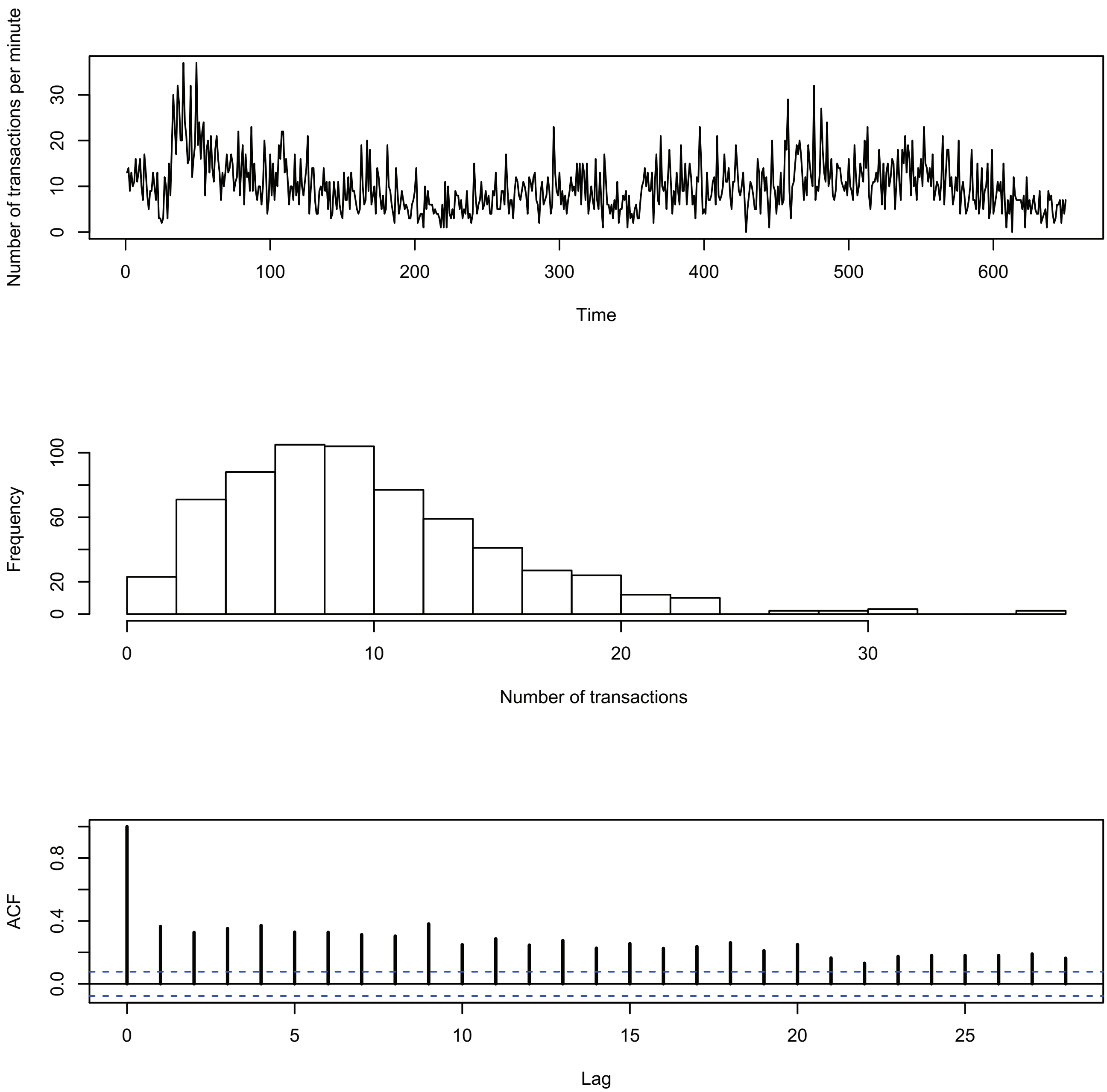

Figure 1. From top to bottom: Time plot of the number of transactions per minute for the stock Ericsson B in the period 2-3 July,2002, histogram of the observation number and the sample autocorrelation function.

Table 3. The estimates of parameters for MS-PGARCH(2;1,1).

Parameters	a_1	a_2	b_1	b_2	c_1	c_2	p_{11}	p_{22}
Estimated value	0.870	0.915	0.099	0.084	0.053	3.484	0.716	0.180

Note: The estimates are obtained by using starting values from the uniform distribution. The results are based on 1000 simulations.

for this model was approximated by collapsing procedure to overcome so called "path dependence" problem. The accuracy of estimation is evaluated by simulation studies, and the methodology is illustrated by a real data example.

APPENDIX A. PROOFS OF THEORETICAL RESULTS

Let $A := \{i : 0 < a(i) < 1, 1 \le i \le m\}$ and $\lambda^* := \min_{i \in A}\{c(i)/(1-a(i))\}$, and ϕ and v denote the Lebesgue and the counting measure of a set, respectively. If $A \neq \emptyset$, without loss of generality, write $\lambda^* := c(1)/(1-a(1))$. Moreover, for convenience, write

$$\begin{aligned} &A_t := [1(S_t = i, S_{t-1} = j)a(i)],\\ &B_t := [1(S_t = i, S_{t-1} = j)b(i)],\\ &C_{t,kl} := [1(S_t = i, S_{t-1} = j)a^k(i)b^l(i)],\\ &M_{kl} := [p_{ji}a^k(i)b^l(i)],\\ &i, j = 1, \ldots, m, \end{aligned}$$

where $1(\cdot)$ is the indicator function. Again, let $\mathbf{1}_{S_t} := (1(S_t = 1), \ldots, 1(S_t = m))^\tau$, $|J|$ denote the length of the interval J, and $Z_t^n := (\lambda_t^n, S_t)$.

We are now ready to state the ergodicity of the process $\{Y_t^n\}$ defined by (3). But, for the proof of ergodicity, we need the following lemma.

Lemma A-1 Let $\{Z_t^n\}$ be a Markov chain defined by (3). If $A \neq \emptyset$, then every point in $D = \{(\lambda, s) \in \mathbb{R}_+ \times \mathbb{R}_+ : (\lambda, s) \in \bigcup_{i=1}^m D_i\}$ is reachable, where $D_i = \{(\lambda, s) \in \mathbb{R}_+ \times \mathbb{R}_+ : \lambda \ge c(i) + a(i)\lambda^*, s \in T_i\}$ with $T_i = \{j \in S : c(j) + a(j)\lambda^* \le c(i) + a(i)\lambda^*\}$, $i = 1, 2, \ldots, m$.

Proof. Notice that if $\lambda_{t-1}^n = \lambda \ge \lambda^*$, then

$$\begin{aligned} &\lambda_t^n\\ =&c(S_t) + a(S_t)\lambda + b(S_t)N_{t-1}(\lambda) + \varepsilon_{t,n}\\ \ge&c(S_t) + a(S_t)\lambda\\ =&(c(S_t) + a(S_t)\lambda)(1(a(S_t) \ge 1) + 1(a(S_t) < 1))\\ =&(c(S_t) + a(S_t)\lambda)1(a(S_t) \ge 1)\\ &+ (c(S_t) + a(S_t)\lambda)1(a(S_t < 1))\\ \ge&\lambda^* 1(a(S_t) \ge 1) + (c(S_t) + a(S_t)\lambda^*)1(a(S_t < 1))\\ =&\lambda^* 1(a(S_t) \ge 1)\\ &+ (1 - a(S_t))\left(\frac{c(S_t)}{1 - a(S_t)} + \frac{a(S_t)}{1 - a(S_t)}\lambda^*\right)\\ &\times 1(a(S_t < 1))\\ \ge&\lambda^* 1(a(S_t) \ge 1)\\ &+ (1 - a(S_t))\left(\lambda^* + \frac{a(S_t)}{1 - a(S_t)}\lambda^*\right)1(a(S_t < 1))\\ =&\lambda^*. \end{aligned}$$

Hence, if $\lambda_0^n = \lambda \ge \lambda^*$, we know that $\lambda_t^n \ge \lambda^*$ for all $t \ge 1$. Now consider a point $(d, s_i) \in D_i \subset D$. We will show that there exists a $j \ge 1$ and an interval J_d such that $d \in J_d$, $|J_d| \le \varepsilon$ for sufficiently small $\varepsilon > 0$, and $P^j(z_0, G) > 0$, where $z_0 = (\lambda, s_0) \in D$ and $G = \{(\lambda, s) \in D : \lambda \in J_d, s = s_i\}$.

First assume that $d = c(i) + a(i)\lambda^*$ and $s_i = i$. In this case, define j to be the smallest positive integer such that $a(1)^j a(i)(\lambda - \lambda^*) < \varepsilon$, since $a(1) < 1$. Consider a path such that $\lambda_0^n = \lambda$, $S_0 = s_0$, $Y_0^n = Y_1^n = \ldots = Y_j^n = 0$, $S_1 = S_2 = \ldots = S_j = 1$, and $S_{j+1} = i$. This implies that

$$\text{(A.1)} \qquad \begin{aligned} \lambda_{j+1}^n = \;& c(i) + a(i)\lambda^* + a(1)^j a(i)(\lambda - \lambda^*)\\ &+ a(i)\textstyle\sum_{i=0}^{j-1} a(1)^i \varepsilon_{j-i,n} + \varepsilon_{j+1,n}. \end{aligned}$$

From (A.1)together with

$$\begin{aligned} &P(S_1 = S_2 = \ldots = S_j = 1, S_{j+1} = i | Z_0^n = z_0) > 0,\\ &P(Y_0^n = Y_1^n = \ldots = Y_j^n = 0 | Z_0^n = z_0, W_{ij}) > 0,\\ &P(a(i)\sum_{i=0}^{j-1} a(1)^i \varepsilon_{j-i,n} + \varepsilon_{j+1,n} < \varepsilon - a(1)^j a(i)(\lambda - \lambda^*)) > 0, \end{aligned}$$

we conclude that for $G = \{(\lambda, s) \in D : \lambda \in J_d, s = i\}$ with $J_d = [d, d + \varepsilon)$,

$$\begin{aligned} &P^{j+1}(z_0, G)\\ \ge&P(Y_0^n = Y_1^n = \ldots = Y_j^n = 0,\\ &S_1 = S_2 = \ldots = S_j = 1, S_{j+1} = i,\\ &a(i)\sum_{i=0}^{j-1} a(1)^i \varepsilon_{j-i,n} + \varepsilon_{j+1,n} < \varepsilon - a(1)^j a(i)(\lambda - \lambda^*) | Z_0^n = z_0)\\ =&P(Y_0^n = \ldots = Y_j^n = 0 | Z_0^n = z_0, W_{ij})\\ &\times P(S_1 = \ldots = S_j = 1, S_{j+1} = i | Z_0^n = z_0)\\ &\times P(a(i)\sum_{i=0}^{j-1} a(1)^i \varepsilon_{j-i,n} + \varepsilon_{j+1,n} < \varepsilon - a(1)^j a(i)(\lambda - \lambda^*))\\ >&0, \end{aligned}$$

where $W_{ij} = \{\omega : S_1 = \ldots = S_j = 1, S_{j+1} = i, a(i)\sum_{i=0}^{j-1} a(1)^i \varepsilon_{j-i,n} + \varepsilon_{j+1,n} < \varepsilon - a(1)^j a(i)(\lambda - \lambda^*)\}$.

Next assume that $d = c(i) + a(i)\lambda^*$ and $s_i = k \neq i$. Obviously, $c(k) + a(k)\lambda^* \le c(i) + a(i)\lambda^*$. Hence $\delta := c(i) + a(i)\lambda^* - c(k) - a(k)\lambda^* \ge 0$. If $\delta = 0$, replace i with k in (A.1), the assertion is also true. If $\delta > 0$, define j to be the smallest positive integer such that $a(1)^{j-1}a(k)b(1) < \varepsilon$ and $a(1)^j a(k)(\lambda - \lambda^*) < \delta/2$, since $a(1) < 1$. Consider a path such that $\lambda_0^n = \lambda$, $S_0 = s_0$, $Y_0^n = N$, $Y_1^n = \ldots = Y_j^n = 0$, $S_1 = S_2 = \ldots = S_j = 1$, and $S_{j+1} = k$. This shows that

$$\begin{aligned} \lambda_{j+1}^n =& c(k) + a(k)(\lambda^* + a(1)^j(\lambda - \lambda^*)\\ &+ a(1)^{j-1}b(1)N + \sum_{i=0}^{j-1} a(1)^i \varepsilon_{j-i,n}) + \varepsilon_{j+1,n} \end{aligned}$$

$$=\lambda^n_{k,j+1}(N)+a(k)\sum_{i=0}^{j-1}a(1)^i\varepsilon_{j-i,n}+\varepsilon_{j+1,n},$$

where $\lambda^n_{k,j+1}(N) := c(k)+a(k)(\lambda^*+a(1)^j(\lambda-\lambda^*)+a(1)^{j-1}b(1)N)$. If there exists an N such that $\lambda^n_{k,j+1}(N)=d$, in an analogous way as above $s=i$ case, one can show that $P^{j+1}(z_0,G)>0$, where $G=\{(\lambda,s)\in D:\lambda\in J_d, s=k\}$ with $J_d=[d,d+\varepsilon)$. Otherwise, let $N:=N_{j+1}$ be the least integer such that $\lambda^n_{k,j+1}(N-1)<d<\lambda^n_{k,j+1}(N)$ ($\lambda^n_{k,j+1}(0)<d$). Taking $J_d=[\lambda^n_{k,j+1}(N-1),\lambda^n_{k,j+1}(N))$, we easily obtain that $P^{j+1}(z_0,G)>0$ and $|J_d|=a(1)^{j-1}a(k)b(1)<\varepsilon$, where $G=\{(\lambda,s):\lambda\in J_d, s=k\}$.

Now, we consider the case that $d>c(i)+a(i)\lambda^*$. If $s_i=i$, define j to be the smallest positive integer such that $a(1)^{j-1}b(1)a(i)<\varepsilon$ and $a(1)^j a(i)(\lambda-\lambda^*)<(d-c(i)-a(i)\lambda^*)/2$, and consider a path such that $Z^n_0=z_0$, $Y^n_0=N$, $Y^n_1=Y^n_2=\ldots=Y^n_j=0$, $S_1=S_2=\ldots=S_j=1$, and $S_{j+1}=i$. We can also obtain that there exists a $j\geq 1$ and $|J_d|\leq\varepsilon$ such that $P^{j+1}(z_0,G)>0$, where $G=\{(\lambda,s)\in D:\lambda\in J_d, s=i\}$ with $J_d=[d,d+\varepsilon)$ if there exists an N such that $\lambda^n_{i,j+1}(N)=d$, otherwise $J_d=[\lambda^n_{i,j+1}(N-1),\lambda^n_{i,j+1}(N))$ with $N:=N_{j+1}$ being the least integer such that $\lambda^n_{i,j+1}(N-1)<d<\lambda^n_{i,j+1}(N)$ ($\lambda^n_{i,j+1}(0)<d$). If $s=k\neq i$ and $c(k)+a(k)\lambda^*<c(i)+a(i)\lambda^*$, define j to be the smallest positive integer such that $a(1)^{j-1}b(1)a(k)<\varepsilon$ and $a(1)^j a(k)(\lambda-\lambda^*)<(d-c(k)-a(k)\lambda^*)/2$, and consider a path such that $Z^0_0=z_0$, $Y^n_0=N$, $Y^n_1=Y^n_2=\ldots=Y^n_j=0$, $S_1=S_2=\ldots=S_j=1$, and $S_{j+1}=k$. If there exists an N such that $\lambda^n_{k,j+1}(N)=d$, in an analogous way as above, one can show that $P^j(z_0,G)>0$, where $G=\{(\lambda,s)\in D:\lambda\in[d,d+\varepsilon), s=k\}$. Otherwise, let $N:=N_{j+1}$ be the least integer such that $\lambda^n_{k,j+1}(N-1)<d<\lambda^n_{k,j+1}(N)$ ($\lambda^n_{k,j+1}(0)<d$). Taking $G=\{(\lambda,s)\in D:\lambda\in[\lambda^n_{k,j+1}(N-1),\lambda^n_{k,j+1}(N)), s=k\}$, we easily obtain that $P^j(z_0,G)>0$ and $|[\lambda^n_{k,j+1}(N-1),\lambda^n_{k,j+1}(N))|=a(1)^{j-1}a(k)b(1)<\varepsilon$. Similarly, we may get that $P^j(z_0,G)>0$ if $s=k\neq i$ and $c(k)+a(k)\lambda^*=c(i)+a(i)\lambda^*$, where G is defined as above in the same manner. Since ε is arbitrary, (d,s_i) can be approximated arbitrarily closely, and (d,s_i) is reachable if $(d,s_i)\in D$. This completes the proof of Lemma A-1.

A.1 Proof of Theorem 2.1.

We will first prove that the Markov chain $\{Z^n_t\}$ is ψ-irreducible, aperiodic, and positive Harris recurrent. These properties will imply that $\{Z^n_t\}$ has a unique stationary distribution $\pi(\lambda,S)$, and that if $Z^n_0\sim\pi(\lambda,S)$, then $\{Z^n_t\}$ is stationary and geometrically ergodic.

First note that $A\neq\emptyset$ if $\rho(M_1)<1$. The state space is equipped with $\mathcal{F}_D$, the Borel σ-algebra on $\mathbb{R}_+\times\mathbb{R}_+$ restricted to D. The measure ψ is the product measure $\phi\otimes v$ on $(D,\mathcal{F}_D)$. Let $G\in\mathcal{F}_D$ such that $\psi(G)>0$. Then there exists a point $(d,k)\in G\subset D_i$ such that $\psi(D^+_\delta)>0$, where $D^+_\delta=G\cap B^+_\delta$ with $B^+_\delta=\{(\lambda,s):\lambda\in[d,d+\delta/2), s=k\}$ for any $\delta>0$. Write $D^{d+}_\delta=\{(x,s):x=y-d, y\in D^+_\delta, s=k\}$. Since $d\geq c(i)+a(i)\lambda^*\geq c(k)+a(k)\lambda^*$, then $(d-c(k))/a(k)\geq\lambda^*$. Thus using the technique of proof of Lemma A-1 for some j, (λ^n_j,S_j) will be arbitrarily close to $((d-c(k))/a(k),1)$ by choosing j large enough and $S_j=1$. In particular, j can be chosen so that $|c(k)+a(k)\lambda^n_j-d|<\varepsilon/2$, where $\varepsilon<\delta$. Consider a path of the next step such that $Y^n_j=0$ and $S_{j+1}=k$, we have

$$\begin{aligned}&P((\lambda',1),G)\geq P((\lambda',1),D^+_\delta)\\ \geq&P(\varepsilon_{j+1,n}\in D^{d+}_{\delta-\varepsilon}(1))P(S_{j+1}=k|S_j=1)P(Y^n_j=0|\lambda^n_j=\lambda')\\ =&\frac{1}{c_n}\phi(D^{d+}_{\delta-\varepsilon}(1)\cap(0,c_n))P(S_{j+1}=k|S_j=1)P(Y^n_j=0|\lambda^n_j=\lambda')\\ >&0,\end{aligned}$$

where $D^{d+}_{\delta-\varepsilon}(1)=\{x:(x,s)\in D^{d+}_{\delta-\varepsilon}\}$. Hence $P^{j+1}(z_0,G)>0$, which implies ψ-irreducibility. It remains to prove the existence of a small set, aperiodicity and positive Harris recurrent.

Let $C=\{(\lambda,s)\in D:\lambda\leq M^*\}$ for any $M^*>\lambda^*$, and define j to be the smallest positive integer such that $a(1)^{j-1}(M^*-\lambda^*)<\varepsilon/2$ for sufficiently small ε. Then

$$\begin{aligned}&\inf_{\lambda\in C}P(S_1=S_2=\ldots=S_{j-1}=1|Z^n_0=z_0)>0,\\ &\inf_{\lambda\in C}P(Y^n_0=Y^n_1=\ldots=Y^n_{j-2}=0|Z^n_0=z_0,W_{j-1})>0,\\ &P(\varepsilon_{1,n}+\varepsilon_{2,n}+\ldots+\varepsilon_{j-1,n}<\frac{\varepsilon}{2}-a(1)^{j-1}(M^*-\lambda^*))>0,\end{aligned}$$

where $W_{j-1}=\{\omega:S_1=\ldots=S_{j-1}=1,\varepsilon_{1,n}+\varepsilon_{2,n}+\ldots+\varepsilon_{j-1,n}<\varepsilon/2-a(1)^{j-1}(M^*-\lambda^*)\}$. Thus $\inf_{z\in C}P^{j-1}(z,H)>0$, where $H=\{(\lambda,s)\in D:\lambda\in[\lambda^*,\lambda^*+\varepsilon/2]\}$. Taking $\mu=\mathrm{Unif}(\lambda^*+\varepsilon/2,\lambda^*+\varepsilon)$ and

$$\begin{aligned}\gamma=&\frac{1}{c_n}\inf_{z\in C}P^{j-1}(z,H)\\ &\times\inf_{z\in H}\{P(Y^n_{j-1}=0|Z^n_{j-1}=z)P(S_j=1|Z^n_{j-1}=z)\}\\ >&0,\end{aligned}$$

then, for all $G\in\mathcal{B}(\mathbb{R}_+\times\mathbb{R}_+)$,

$$\begin{aligned}&P^j(z,G)\\ \geq&P^j(z,G\cap H_\varepsilon)\\ =&\int_{\mathbb{R}_+\times\mathbb{R}_+}P(y,G\cap H_\varepsilon)P^{j-1}(z,dy)\\ \geq&\int_H P(y,G\cap H_\varepsilon)P^{j-1}(z,dy)\\ \geq&\inf_{z\in H}\{P(Y^n_{j-1}=0|Z^n_{j-1}=z)P(S_j=1|Z^n_{j-1}=z)\}\\ &\times\int_H P(\varepsilon_{j,n}\in G_y\cap(\varepsilon/2-a(1)(y-\lambda^*),\varepsilon-a(1)(y-\lambda^*)))\\ &\times P^{j-1}(z,dy)\\ \geq&\gamma P(U_j\in G_{\lambda^*}\cap(\varepsilon/2,\varepsilon))\\ =&\gamma\mu(G),\end{aligned}$$

which establishes C as a small set, where $H_\varepsilon = \{(\lambda, s) \in D : \lambda \in (\lambda^* + \varepsilon/2, \lambda^* + \varepsilon)\}$ and $G_u = \{x : x = z_1 - \lambda^* - a(1)(u - \lambda^*), (z_1, z_2) \in G\}$.

We now show that $\{Z_t^n\}$ is aperiodic. Consider the small set $C = \{(\lambda, s) \in D : \lambda \le M^*\}$ with $M^* > \lambda^*$. Note that $\psi(C) > 0$. If $Z_{t-1}^n = (\lambda, s) \in C$, consider a path such that $Y_{t-1}^n = 0$, $S_t = 1$, and $\varepsilon_{t,n} \le (1 - a(1))(M^* - \lambda^*)$, then

$$\begin{aligned}\lambda^* \le& \lambda_t^n \\ =& c(1) + a(1)\lambda + \varepsilon_{t,n} \\ \le& c(1) + a(1)M^* + (1 - a(1))(M^* - \lambda^*) \\ =& M^*.\end{aligned}$$

This implies

$$\begin{aligned}& P(Z_t^n \in C | Z_{t-1}^n = (\lambda, s) \in C) \\ \ge& P\left\{Y_{t-1}^n = 0, S_t = 1, \right. \\ & \left. \varepsilon_{t,n} \le (1 - a(1))(M^* - \lambda^*) | Z_{t-1}^n = (\lambda, s) \in C\right\} \\ =& P(N_{t-1}(\lambda) = 0) P(S_t = 1 | S_{t-1} = s) \\ & \times P\left\{\varepsilon_{t,n} \le (1 - a(1))(M^* - \lambda^*)\right\} \\ >& 0.\end{aligned}$$

Similarly,

$$\begin{aligned}& P(Z_{t+1}^n \in C | Z_{t-1}^n = (\lambda, s) \in C) \\ \ge& P\left\{Y_t^n = Y_{t-1}^n = 0, S_{t+1} = S_t = 1, \right. \\ & \varepsilon_{t+1,n} \le (1 - a(1))(M^* - \lambda^*), \\ & \varepsilon_{t,n} \le (1 - a(1))(M^* - \lambda^*) \\ & \left. | Z_{t-1}^n = (\lambda, s) \in C\right\} \\ >& 0.\end{aligned}$$

It follows that $\{Z_t^n\}$ is aperiodic by Proposition A1.1 in [6].

Finally, we prove that $\{Z_t^n\}$ is positive Harris recurrent. It is well-known that $M_1^k \to 0$ as $k \to \infty$, if $\rho(M_1) < 1$. Thus, for some positive constant $\alpha < 1$, there exists a positive integer N such that $\|M_1^k\| \le \alpha$ for all $k \ge N$. Consider $V(x) = 1 + x_1$, where $x = (x_1, x_2)$ with $x_1 > 0$. Define the small set $C = \{(\lambda, s) \in D : \lambda \le M^*, s \in S\}$, where $M^* = K/(1 - 2\varepsilon - \alpha)$ with $\varepsilon > 0$ such that $1 - 2\varepsilon - \alpha > 0$, $M^* = K/(1 - 2\varepsilon - \alpha) > \max\{1, \lambda^*\}$, and constant $K := (c_n/2 + c)\sum_{i=1}^N \|M^{i-1}\|$ with $c = \max\{c(1), \dots, c(m)\}$. First note that S_t and Y_{t-1}^n are conditionally independent under given $(S_{t-1}, \lambda_{t-1}^n)$, which implies that

$$\begin{aligned}& E(\|B_t \mathbf{1}_{S_{t-1}} Y_{t-1}^n\| \lambda_{t-1}^n, S_{t-1}) \\ =& E(\|B_t \mathbf{1}_{S_{t-1}}\| | \lambda_{t-1}^n, S_{t-1}) E(Y_{t-1}^n | \lambda_{t-1}^n, S_{t-1}) \\ =& \lambda_{t-1}^n E(\|B_t \mathbf{1}_{S_{t-1}}\| | \lambda_{t-1}^n, S_{t-1}) \\ =& \lambda_{t-1}^n \|M_{01} \mathbf{1}_{S_{t-1}}\|.\end{aligned}$$

In addition,

$$E(\|A_t \mathbf{1}_{S_{t-1}} \lambda_{t-1}^n\| | \lambda_{t-1}^n, S_{t-1}) = \lambda_{t-1}^n \|M_{10} \mathbf{1}_{S_{t-1}}\|.$$

It follows that

$$\begin{aligned}& E(\|A_t \mathbf{1}_{S_{t-1}} \lambda_{t-1}^n + B_t \mathbf{1}_{S_{t-1}} Y_{t-1}^n\| | \lambda_{t-1}^n, S_{t-1}) \\ =& \lambda_{t-1}^n \|M_1 \mathbf{1}_{S_{t-1}}\|.\end{aligned}$$

Therefore, for any $\lambda > M^*$,

$$\begin{aligned}& E[V(Z_{Nt}^n) | Z_{N(t-1)}^n = z = (\lambda, s)] \\ =& E[1 + \lambda_{Nt}^n | Z_{N(t-1)}^n = z] \\ =& 1 + E[\|\mathbf{1}_{S_{Nt}} \lambda_{Nt}^n\| | Z_{N(t-1)}^n = z] \\ =& 1 + E[\|\mathbf{1}_{S_{Nt}} c(S_{Nt}) + A_{Nt} \mathbf{1}_{S_{Nt-1}} \lambda_{Nt-1}^n \\ & + B_{Nt} \mathbf{1}_{S_{Nt-1}} Y_{Nt-1}^n + \mathbf{1}_{S_{Nt}} \varepsilon_{Nt,n}\| | Z_{N(t-1)}^n = z] \\ =& 1 + E\varepsilon_{Nt,n} + E[c(S_{Nt}) | Z_{N(t-1)}^n = z] \\ & + E[E(\|A_{Nt} \mathbf{1}_{S_{Nt-1}} \lambda_{Nt-1}^n + B_{Nt} \mathbf{1}_{S_{Nt-1}} Y_{Nt-1}^n\| | \mathcal{F}_{N(t-1),Nt-1}^{\lambda^n,S}) \\ & | Z_{N(t-1)}^n = z)] \\ =& 1 + E\varepsilon_{Nt,n} + E[c(S_{Nt}) | Z_{N(t-1)}^n = z] \\ & + E[\|M_1 \mathbf{1}_{S_{Nt-1}} \lambda_{Nt-1}^n\| | Z_{N(t-1)}^n = z] \\ =& \dots \\ =& 1 + \sum_{i=1}^N \|M_1^{i-1} E[\mathbf{1}_{S_{Nt-i+1}} \varepsilon_{Nt-i+1,n} | Z_{N(t-1)}^n = z]\| \\ & + \sum_{i=1}^N \|M_1^{i-1} E[\mathbf{1}_{S_{Nt-i+1}} c(S_{Nt-i+1}) | Z_{N(t-1)}^n = z]\| \\ & + \|M_1^N E[\mathbf{1}_{S_{N(t-1)}} \lambda | Z_{N(t-1)}^n = z]\| \\ \le& 1 + \frac{c_n}{2} \sum_{i=1}^N \|M_1^{i-1}\| + c \sum_{i=1}^N \|M_1^{i-1}\| + \|M_1^N\| \lambda \\ =& 1 + K + \|M_1^N\| \lambda \\ \le& 1 + \lambda \left(\alpha + \frac{K}{\lambda}\right) \\ \le& 1 + \lambda (\alpha + K) \\ \le& 1 + \lambda (1 - 2\varepsilon) \\ =& 1 - \lambda\varepsilon + \lambda(1 - \varepsilon) \\ \le& 1 - \varepsilon + \lambda(1 - \varepsilon) \\ =& (1 - \varepsilon) V(\lambda, s),\end{aligned}$$

where $\mathcal{F}_{u,t}^{\lambda^n,S}$ is the σ-algebra generated by $\{Z_v^n, u \le v \le t\}$. Furthermore, it is easy to see that, for $\lambda \in C$,

$$E[V(Z_{Nt}^n) | Z_{N(t-1)}^n = z] \le 1 + K + \|M_1^N\| \lambda \le L$$

for some constant L. This implies that the chain $\{Z_{Nt}^n\}$ has a unique stationary distribution $\pi(\lambda, s)$, and that if $Z_0^n \sim \pi(\lambda, s)$ then $\{Z_{Nt}^n\}$ is stationary and geometrically ergodic, and that $E\lambda_{Nt}^n < \infty$. Furthermore, by Lemma 3.1 in [39], $\{Z_t^n\}$ is also stationary and geometrically ergodic. Finally, note that

$$Y_t^n = b^{-1}(S_{t+1})(\lambda_{t+1}^n - c(S_{t+1}) - a(S_{t+1})\lambda_t^n - \varepsilon_{t+1,n})$$

$$:= g(Z_{t+1}^n, Z_t^n, \varepsilon_{t+1,n}),$$

which shows that Y_t^n depends on $\{Z_t^n\}$ only through Z_t^n and Z_{t+1}^n. Since $\{(Z_t^n, \varepsilon_{t,n})\}$ is stationary and ergodic, Theorem 36.4 of [4] gives rise to that $\{Y_t^n\}$ is stationary and ergodic. It is not difficult to see that $EY_t^n < \infty$. This completes the proof of Proposition 2.1.

In order to prove Theorem 2.2, we need the following lemma.

Lemma A-2. If $\rho(M_{u_2}) < 1$, then $\rho(M_{u_1}) < 1$ for $0 < u_1 < u_2$.

Proof. Firstly recall some properties of nonnegative matrices and nonnegative irreducible matrices[3, 26]. For a $m\times m$ nonnegative matrix $R = [r_{ij}]$,

$$\min_j \left\{\sum_{i=1}^m r_{ij}\right\} \le \rho(R) \le \max_j \left\{\sum_{i=1}^m r_{ij}\right\}. \tag{A.2}$$

In particular, for two nonnegative irreducible matrices $Q = [q_{ij}]$ and $R = [r_{ij}]$, if $q_{ij} \ge r_{ij}$, $i,j = 1,2,\dots,m$, and $Q \ne R$, then $\rho(Q) > \rho(R)$. This immediately implies that if R is a nonnegative irreducible matrix and $\min_j \{\sum_{i=1}^m r_{ij}\} < \max_j \{\sum_{i=1}^m r_{ij}\}$, then, by (A.2),

$$\min_j \left\{\sum_{i=1}^m r_{ij}\right\} < \rho(R) < \max_j \left\{\sum_{i=1}^m r_{ij}\right\}. \tag{A.3}$$

Next notice that the transition matrix P is nonnegative irreducible and $a(S_t) + b(S_t) > 0$. Hence it is easy to prove that M_u is nonnegative irreducible for any $u \ge 0$. From (A.3) and $\rho(M_{u_2}) < 1$, we conclude that there exists some $1 \le i_1 \le m$ such that $a(i_1)+b(i_1) < 1$. If $\max_i\{a(i)+b(i)\} \le 1$, by (A.2) and (A.3), we know that Lemma A-2 is true. Otherwise suppose that there exists some $1 \le i_2 \le m$ such that $a(i_2)+b(i_2) > 1$. For convenience, now define a function

$$f(u) = \rho(M_u), u \ge 0.$$

We will show that $\log f(u)$ is a convex function on $[0,\infty)$.

Since the transition matrix P is irreducible and aperiodic, and $a(i) + b(i) > 0$, $i = 1,2,\dots,m$, this implies that M_u is also an irreducible and aperiodic $m\times m$ matrix. Therefore, by Theorem 8.5.1 in [26], we may get that

$$\lim_{n\to\infty} f^{-n}(u) M_u^n = L_u,$$

where L_u is a constant positive matrix. It follows that

$$\lim_{n\to\infty} [\mathbf{1}^\tau M_u^n \pi]^{1/n} = f(u),$$

where $\mathbf{1} = (1,1,\dots,1)^\tau$.

Again write $M_u^{(n)}(\delta) = [((M_u^n)_{ij})^\delta]$, where $(M_u^n)_{ij}$ denotes the (i,j)-element of M_u^n. Then, we have

$$\begin{aligned}
&\log f(\gamma v_1 + (1-\gamma) v_2)\\
&= \lim_{n\to\infty} \frac{1}{n}\log(\mathbf{1}^\tau M^n_{\gamma v_1 + (1-\gamma)v_2}\pi)\\
&\le \lim_{n\to\infty} \frac{1}{n}\log[\mathbf{1}^\tau (M_{v_1}^{(n)}(\gamma)\circ M_{v_2}^{(n)}(1-\gamma))\pi)]\\
&\le \lim_{n\to\infty} \frac{1}{n}\log[(\mathbf{1}^\tau M_{v_1}^n\pi)^\gamma(\mathbf{1}^\tau M_{v_2}^n\pi)^{1-\gamma}]\\
&= \gamma \lim_{n\to\infty} \frac{1}{n}\log(\mathbf{1}^\tau M_{v_1}^n\pi) + (1-\gamma)\lim_{n\to\infty} \frac{1}{n}\log(\mathbf{1}^\tau M_{v_2}^n\pi)\\
&= \gamma \log f(v_1) + (1-\gamma)\log f(v_2),
\end{aligned}$$

where $0 < \gamma < 1$ and $v_1, v_2 \ge 0, v_1 \ne v_2$, and $\circ$ denotes the Hadamard product of matrices, the two inequalities are applications of the Hölder's inequality. This shows that $\log f(u)$ is a convex function. Hence, together with $f(0) = 1$ and $f(u_2) < 1$,

$$\begin{aligned}
\rho(M_{u_1}) &= f(u_1)\\
&= \exp\{\log f(u_1)\}\\
&= \exp\{\log f(\gamma\times 0 + (1-\gamma)u_2)\}\\
&\le \exp\{\gamma \log f(0) + (1-\gamma)\log f(u_2)\}\\
&< 1,
\end{aligned}$$

where $\gamma = 1 - u_1/u_2$.

A.2 Proof of Theorem 2.2

By Lemma A-2 and Theorem 2.1, we can obtain the stationarity and ergodicity of the process $\{Y_t^n\}$ defined as in (3). A similar argument of the proof of Proposition 2.1 can be used to show the process $\{Y_t^n\}$ has finite moment of order k.

By $\rho(M_k) < 1$, we know that there exists a positive integer N such that $\|M_k^l\| \le \alpha_k$ with some $\alpha_k < 1$ for all $l \ge N$. Consider $V(x) = 1 + x_1^k$, where $x = (x_1, x_2)$ with $x_1 > 0$. Define the small set $C = \{(\lambda, s) \in D : \lambda \le M^*, s \in S\}$, where $M^* = \beta/(1-2\varepsilon-\alpha_k)$ with $\varepsilon > 0$ such that $1-2\varepsilon-\alpha_k > 0$, $\beta/(1-2\varepsilon-\alpha_k) > \max\{1,\lambda^*\}$, and constant $\beta = \sum_{j=0}^{k-1} d_j$ is defined as follows. Note that

$$\begin{aligned}
&E[\|A_{Nt}\mathbf{1}_{S_{Nt-1}}\lambda^n_{Nt-1}\|^i \|B_{Nt}\mathbf{1}_{S_{Nt-1}}Y^n_{Nt-1}\|^{k-i}|\lambda^n_{Nt-1}, S_{Nt-1}]\\
=&(\lambda^n_{Nt-1})^i E[(\mathbf{1}^\tau A_{Nt}\mathbf{1}_{S_{Nt-1}})^i(\mathbf{1}^\tau B_{Nt}\mathbf{1}_{S_{Nt-1}})^{k-i})|\lambda^n_{Nt-1}, S_{Nt-1}]\\
&\times E[(Y^n_{Nt-1})^{k-i}|\lambda^n_{Nt-1}, S_{Nt-1}]\\
=&(\lambda^n_{Nt-1})^k E[\mathbf{1}^\tau C_{Nt,i,k-i}\mathbf{1}_{S_{Nt-1}}|\lambda^n_{Nt-1}, S_{Nt-1}]\\
&+\sum_{l=1}^{k-1}(\lambda^n_{Nt-1})^l f_l(a(S_{Nt-1}), b(S_{Nt-1}))\\
=&\|M_{i,k-i}\mathbf{1}_{S_{Nt-1}}(\lambda^n_{Nt-1})^k\|\\
&+\sum_{l=1}^{k-1}(\lambda^n_{Nt-1})^l f_l(a(S_{Nt-1}), b(S_{Nt-1}))
\end{aligned}$$

and

$$\sum_{i=0}^{k} C_k^i M_{i,k-i} = M_k.$$

Therefore, for any $\lambda > M^*$,

$$\begin{aligned}
&E[V(Z_{Nt}^n)|Z_{N(t-1)}^n = z = (\lambda, s)] \\
=&E[1 + (\lambda_{Nt}^n)^k | Z_{N(t-1)}^n = z] \\
=&1 + E[\|\mathbf{1}_{S_{Nt}} \lambda_{Nt}^n\|^k | Z_{N(t-1)}^n = z] \\
=&1 + E[\|\mathbf{1}_{S_{Nt}} c(S_{Nt}) + A_{Nt}\mathbf{1}_{S_{Nt-1}} \lambda_{Nt-1}^n \\
&+ B_{Nt}\mathbf{1}_{S_{Nt-1}} Y_{Nt-1}^n + \mathbf{1}_{S_{Nt}} \varepsilon_{Nt,n}\|^k | Z_{N(t-1)}^n = z] \\
=&1 + E[\|A_{Nt}\mathbf{1}_{S_{Nt-1}} \lambda_{Nt-1}^n \\
&+ B_{Nt}\mathbf{1}_{S_{Nt-1}} Y_{Nt-1}^n\|^k | Z_{N(t-1)}^n = z] \\
&+ \sum_{j=1}^{k} C_k^j E[\|\mathbf{1}_{S_{Nt}} c(S_{Nt}) + \mathbf{1}_{S_{Nt}} \varepsilon_{Nt,n}\|^j \\
&\times \|A_{Nt}\mathbf{1}_{S_{Nt-1}} \lambda_{Nt-1}^n + B_{Nt}\mathbf{1}_{S_{Nt-1}} Y_{Nt-1}^n\|^{k-j} \\
&\quad |Z_{N(t-1)}^n = z] \\
=&1 + \sum_{i=0}^{k} C_k^i E[E(\|A_{Nt}\mathbf{1}_{S_{Nt-1}} \lambda_{Nt-1}^n\|^i \\
&\times \|B_{Nt}\mathbf{1}_{S_{Nt-1}} Y_{Nt-1}^n\|^{k-i} | \mathcal{F}_{N(t-1),Nt-1}^{\lambda^n,S}) | Z_{N(t-1)}^n = z] \\
&+ \sum_{j=1}^{k} C_k^j E[E(\|\mathbf{1}_{S_{Nt}} c(S_{Nt}) + \mathbf{1}_{S_{Nt}} \varepsilon_{Nt,n}\|^j \\
&\times \|A_{Nt}\mathbf{1}_{S_{Nt-1}} \lambda_{Nt-1}^n + B_{Nt}\mathbf{1}_{S_{Nt-1}} Y_{Nt-1}^n\|^{k-j} \\
&\quad |\mathcal{F}_{N(t-1),Nt-1}^{\lambda^n,S}) | Z_{N(t-1)}^n = z] \\
=&1 + \sum_{i=0}^{k} C_k^i E[\|M_{i,k-i}\mathbf{1}_{S_{Nt-1}} (\lambda_{Nt-1}^n)^k\| | Z_{N(t-1)}^n = z] \\
&+ \sum_{l=1}^{k-1} E[(\lambda_{Nt-1}^n)^l f_l(a(S_{Nt-1}), b(S_{Nt-1})) | Z_{N(t-1)}^n = z] \\
&+ \sum_{j=1}^{k} C_k^j E[E(\|\mathbf{1}_{S_{Nt}} c(S_{Nt}) + \mathbf{1}_{S_{Nt}} \varepsilon_{Nt,n}\|^j \\
&\times \|A_{Nt}\mathbf{1}_{S_{Nt-1}} \lambda_{Nt-1}^n + B_{Nt}\mathbf{1}_{S_{Nt-1}} Y_{Nt-1}^n\|^{k-j} \\
&\quad |\mathcal{F}_{N(t-1),Nt-1}^{\lambda^n,S}) | Z_{N(t-1)}^n = z] \\
=&\ldots \\
=&1 + \|M_k^N E[\mathbf{1}_{S_{N(t-1)}} \lambda^k | Z_{N(t-1)}^n = z\| + \sum_{j=0}^{k-1} d_j \lambda^j \\
\leq&1 + \lambda^k \left(\alpha_k + \sum_{j=0}^{k-1} \frac{d_j}{\lambda^{k-j}}\right) \\
\leq&1 + \lambda^k (\alpha_k + \beta) \\
\leq&1 + \lambda^k (1 - 2\varepsilon) \\
=&(1-\varepsilon)V(\lambda, s).
\end{aligned}$$

Furthermore, it is easy to see that, for $\lambda \leq M^*$,

$$E[V(Z_{Nt}^n)|Z_{N(t-1)}^n = z] \leq L_1$$

for some constant L_1. This implies that there exists finite moment of order k for $\{Z_{Nt}^n, t \geq 0\}$. It follows that there exists finite moment of order k for $\{Y_t^n, t \geq 0\}$.

A.3 Proof of Lemma 2.1

Since $\rho(M_2) < 1$, by Lemma A-2, we have that $\rho(M_1) < 1$ and

$$\sum_{k=0}^{\infty} M_1^k = (I - M_1)^{-1}, \qquad \sum_{k=0}^{\infty} M_2^k = (I - M_2)^{-1},$$

where I denotes the identity matrix. By the equations (2) and (3),

(A.4)
$$\lambda_t^n - \lambda_t = a(S_t)(\lambda_{t-1}^n - \lambda_{t-1}) + b(S_t)(Y_{t-1}^n - Y_{t-1}) + \varepsilon_{t,n}.$$

Therefore, by using (A.4),

$$\begin{aligned}
&E(\lambda_t^n - \lambda_t) \\
=&E[E\|A_t\mathbf{1}_{S_{t-1}}(\lambda_{t-1}^n - \lambda_{t-1}) + B_t\mathbf{1}_{S_{t-1}}(Y_{t-1}^n - Y_{t-1})\| \\
&\quad |S_{t-1}, \lambda_{t-1}^n, \lambda_{t-1}] + E(\varepsilon_{t,n}) \\
=&E[\|M_1\mathbf{1}_{S_{t-1}}(\lambda_{t-1}^n - \lambda_{t-1})\|] + E(\varepsilon_{t,n}) \\
=&\|M_1 E[\mathbf{1}_{S_{t-1}}(\lambda_{t-1}^n - \lambda_{t-1})]\| + E(\varepsilon_{t,n}) \\
=&\|M_1 E[\mathbf{1}_{S_{t-1}}\|A_{t-1}\mathbf{1}_{S_{t-2}}(\lambda_{t-2}^n - \lambda_{t-2}) \\
&+ B_{t-1}\mathbf{1}_{S_{t-2}}(Y_{t-2}^n - Y_{t-2}) + \varepsilon_{t-1,n}\|]\| + E(\varepsilon_{t,n}) \\
=&\|M_1 E[A_{t-1}\mathbf{1}_{S_{t-2}}(\lambda_{t-2}^n - \lambda_{t-2}) + B_{t-1}\mathbf{1}_{S_{t-2}}(Y_{t-2}^n - Y_{t-2})\|]\| \\
&+ E(\varepsilon_{t,n}) + \|M_1 E[\mathbf{1}_{S_{t-1}}\varepsilon_{t-1,n}]\| \\
=&\ldots \\
=&E(\varepsilon_{t,n}) \left\|(I + M_1 + \ldots + M_1^{t-1})\right\| + \|M_1^t E[\mathbf{1}_{S_0}(\lambda_0^n - \lambda_0)]\| \\
=&\frac{c_n}{2} \left\|(I + M_1 + \ldots + M_1^{t-1})\right\| \\
\leq&\frac{c_n}{2} \left\|(I + M_1 + \ldots + M_1^{t-1} + \ldots)\right\| \\
=&\frac{c_n}{2} \left\|(I - M_1)^{-1}\right\| \leq \frac{c_n}{2} \left\|(I - M_1)^{-1}\right\| \\
:=&\delta_{1,n} \to 0
\end{aligned}$$

as $n \to \infty$, which implies the first assertion.

Next consider the second statement. By using (A.4) again,

$$\begin{aligned}
&E(\lambda_t^n - \lambda_t)^2 \\
=&E[a^2(S_t)(\lambda_{t-1}^n - \lambda_{t-1})^2] + E[b^2(S_t)(Y_{t-1}^n - Y_{t-1})^2] \\
&+ 2E[a(S_t)b(S_t)(\lambda_{t-1}^n - \lambda_{t-1})(Y_{t-1}^n - Y_{t-1})] + E(\varepsilon_{t,n}^2)
\end{aligned}$$

$$+ 2E[a(S_t)(\lambda_{t-1}^n - \lambda_{t-1})\varepsilon_{t,n}]$$
$$+ 2E[b(S_t)(Y_{t-1}^n - Y_{t-1})\varepsilon_{t,n}].$$

We first calculate the second term of the above expression, by using the properties of the Poisson process and the conditional expectation, assuming that $\lambda_{t-1}^n \geq \lambda_{t-1}$ and $Y_{t-1}^n - Y_{t-1} \sim \text{Poisson}(\lambda_{t-1}^n - \lambda_{t-1})$ independently of $Y_{t-1} \sim \text{Poisson}(\lambda_{t-1})$ (i.e $\lambda_1^n > \lambda_1$, say $\lambda_1^n = \lambda_1 + \mu_1$, then $Y_1^n = Y_1 + Z_1$, $Z_1 \sim \text{Poisson}(\mu_1)$ and $Y_1^n > Y_1$; $\lambda_2^n > \lambda_2$, say $\lambda_2^n = \lambda_2 + \mu_2$, then $Y_2^n = Y_2 + Z_2$ and $Z_2 \sim \text{Poisson}(\mu_2)$; and so on). Now we have that

$$\begin{aligned}
& E[b^2(S_t)(Y_{t-1}^n - Y_{t-1})^2] \\
= \; & E[E(b^2(S_t)(Y_{t-1}^n - Y_{t-1})^2 | S_{t-1}, \lambda_{t-1}^n, \lambda_{t-1})] \\
= \; & E[E(b^2(S_t)|S_{t-1}, \lambda_{t-1}^n, \lambda_{t-1}) E(Y_{t-1}^n - Y_{t-1})^2 \\
& \quad |S_{t-1}, \lambda_{t-1}^n, \lambda_{t-1})] \\
= \; & E[\|M_{02}\mathbf{1}_{S_{t-1}}\| E(N_{t-1}(\lambda_{t-1}^n) - N_{t-1}(\lambda_{t-1}))^2] \\
= \; & E[\|M_{02}\mathbf{1}_{S_{t-1}}\|(\lambda_{t-1}^n - \lambda_{t-1})^2] \\
& + E[\|M_{02}\mathbf{1}_{S_{t-1}}\|(\lambda_{t-1}^n - \lambda_{t-1})] \\
\leq \; & \|M_{02}\|[E(\lambda_{t-1}^n - \lambda_{t-1})^2 + E(\lambda_{t-1}^n - \lambda_{t-1})].
\end{aligned}$$

Similarly,

$$E[a^2(S_t)(\lambda_{t-1}^n - \lambda_{t-1})^2] \leq \|M_{20}\| E(\lambda_{t-1}^n - \lambda_{t-1})^2$$

and

$$\begin{aligned}
& 2E[a(S_t)b(S_t)(\lambda_{t-1}^n - \lambda_{t-1})(Y_{t-1}^n - Y_{t-1})] \\
\leq \; & 2\|M_{11}\| E(\lambda_{t-1}^n - \lambda_{t-1})^2.
\end{aligned}$$

Finally, with a positive constant K,

$$\begin{aligned}
& E(\varepsilon_{t,n}^2) + 2E[a(S_t)(\lambda_{t-1}^n - \lambda_{t-1})\varepsilon_{t,n}] \\
& + 2E[b(S_t)(Y_{t-1}^n - Y_{t-1})\varepsilon_{t,n}] \\
\leq & c_n^2/3 + c_n^2\|M_{10}\| \left\|(I - M_1)^{-1}\right\| \\
& + c_n^2\|M_{01}\| \left\|(I - M_1)^{-1}\right\| \\
= & \left(\frac{1}{3} + \|M_1\| \left\|(I - M_1)^{-1}\right\|\right) c_n^2 := Kc_n^2.
\end{aligned}$$

Thus, by simple recursion,

$$\begin{aligned}
& E(\lambda_t^n - \lambda_t)^2 \\
\leq \; & \|M_2\| E(\lambda_{t-1}^n - \lambda_{t-1})^2 + \|M_{02}\|\delta_{1,n} + Kc_n^2 \\
\leq \; & \|(I - M_2)^{-1}\|(\|M_{02}\|\delta_{1,n} + Kc_n^2) \\
:= \; & \delta_{2,n} \to 0
\end{aligned}$$

as $n \to \infty$, which implies the second assertion. The third assertion follows that

$$\begin{aligned}
& E(Y_{t-1}^n - Y_{t-1})^2 \\
= \; & E(N_{t-1}(\lambda_{t-1}^n) - N_{t-1}(\lambda_{t-1}))^2 \\
= \; & E(\lambda_{t-1}^n - \lambda_{t-1})^2 + E(\lambda_{t-1}^n - \lambda_{t-1}) \\
\leq \; & \delta_{2,n} + \delta_{1,n} := \delta_{3,n}.
\end{aligned}$$

The last statement of this lemma follows from the second and the third assertions.

□

APPENDIX B. ALGORITHM OF MAXIMUM LIKELIHOOD ESTIMATION BASED ON COLLAPSING PROCEDURE

B.1 Exact filtering stage

For $t = 1, 2, ..., q$:

1. When $t = 1$, for $k = 1, 2$, calculate

$$\lambda_1^{(k)} = c_k + a_k\lambda_0 + b_kY_0,$$
$$g_1^{(k)} = f(Y_1|S_1 = k)Pr\{S_1 = k\},$$
$$w_1^{(k)} = \frac{g_1^{(k)}}{g_1^{(1)} + g_1^{(2)}}.$$

   An initial distribution of S_0 should be specified to calculate $Pr\{S_1 = k\} = p_{1k}Pr\{S_0 = 1\} + p_{2k}Pr\{S_0 = 2\}$.

2. When $t = 2, 3, ..., q$, for $k = 1, 2, ..., 2^{t-1}$ and $j = 1, 2$, set $S_{1:t}^{(k,j)} := (S_{1:t-1}^{(k)}, S_t = j)$ and calculate

$$\lambda_t^{(k,j)} = c_j + a_j\lambda_{t-1}^{(k)} + b_jY_{t-1},$$
$$g_t^{(k,j)} = f_{t-1}(Y_t|S_t = j, S_{1:t-1}^{(k)})Pr\{S_t = j|S_{t-1}^{(k)}\}w_{t-1}^{(k)},$$
$$w_t^{(k,j)} = \frac{g_t^{(k,j)}}{\sum_{k=1}^{2^{t-1}}\sum_{j=1}^{2} g_t^{(k,j)}}.$$

   Combining indices (k, j) into a single index k, we have $\left\{\lambda_t^{(k)}\right\}_{k=1}^{2^t}$ and $\left\{S_{1:t}^{(k)}, w_t^{(k)}\right\}_{k=1}^{2^t}$. The latter is an exact representation of filtering probabilities.

B.2 Approximated filtering stage

Let $\tilde{\lambda}_q^{(k)} = \lambda_q^{(k)}$ for $k = 1, 2, ..., 2^q$. Then, for $t = q+1, q+2, ..., T$:

3. Let $S_{t-q:t-1}^{(k)} = (S_{t-q} = j, S_{t-q+1:t-1}^{(m)})$, i.e. re-index $k \in \{1, 2, ..., 2^q\}$ as $(j, m) \in \{1, 2\} \times \{1, 2, ..., 2^{q-1}\}$. Then re-index the corresponding $\tilde{\lambda}_{t-1}^{(k)}$ and $w_{t-1}^{(k)}$ as $\tilde{\lambda}_{t-1}^{(j,m)}$ and $w_{t-1}^{(j,m)}$.
4. Integrate out S_{t-q} from $\tilde{\lambda}_{t-1}^{(j,m)}$ based on (15):

$$E_{t-1}^{(m)} = \sum_{j=1}^{K}\left\{\tilde{\lambda}_{t-1}^{(j,m)} \frac{w_{t-1}^{(j,m)}}{w_{t-1}^{(1,m)} + w_{t-1}^{(2,m)}}\right\}.$$

5. Let $S^{(m,j)}_{t-q+1:t} = (S^{(m)}_{t-q+1:t-1}, S_t = j)$. Calculate

$$\begin{aligned}\tilde{\lambda}^{(m,j)}_t =& c_j + a_j E^{(m)}_{t-1} + b_j Y_{t-1},\\ g^{(m,j)}_t =& \tilde{f}_{t-1}(Y_t|S_t = j, S^{(m)}_{t-q+1:t-1})\\ & \times Pr\{S_t = j|S^{(m)}_{t-1}\} w^{(m)}_{t-1},\\ w^{(m,j)}_t =& \frac{g^{(m,j)}_t}{\sum_{m=1}^{2^{t-1}} \sum_{j=1}^{2} g^{(m,j)}_t}.\end{aligned}$$

Combining indices (m, j) into a single index k, we have $\left\{\tilde{\lambda}^{(k)}_{it}\right\}_{k=1}^{2^t}$ and $\left\{S^{(k)}_{t-q+1:t},\ w^{(k)}_t\right\}_{k=1}^{2^t}$.

After iterating 1-5 for $t = 1, 2, ..., T$, the log likelihood function could be evaluated as

$$\sum_{t=1}^{T} \log(\sum_{m=1}^{2^{t-1}} \sum_{j=1}^{2} g^{(m,j)}_t).$$

REFERENCES

[1] Albert, J. H. and Chib, S. (1993). Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business & Economic Statistics* **11** 1–15.

[2] Augustyniak, M., Boudreault, M. and Morales, M. (2018). Maximum likelihood estimation of the Markov-switching GARCH model based on a general collapsing procedure. *Methodology and Computing in Applied Probability* **20** 165–188.

[3] Berman, A. and Plemmons, R. J. (1994). *Nonnegative matrices in the mathematical sciences*. SIAM.

[4] Billingsley, P. (1995). *Probability and Measure*. Wiley, New York.

[5] Caner, M. and Hansen, B. E. (2001). Threshold autoregression with a unit root. *Econometrica* **69** 1555–1596.

[6] Chan, K. S. (1990). Deterministic stability, stochastic stability, and ergodicity. In *Non-linear Time Series: A Dynamical Systems Approach*, **6** Clarendon Press.

[7] Dueker, M. J. (1997). Markov switching in GARCH processes and mean-reverting stock-market volatility. *Journal of Business & Economic Statistics* **15** 26–34.

[8] Enders, W. (2015). *Applied econometric time series*. Wiley.

[9] Engel, C. and Hamilton, J. D. (1989). Long swings in the exchange rate: are they in the data and do markets know it?

[10] Fan, J. and Yao, Q. (2008). *Nonlinear time series: nonparametric and parametric methods*. Springer Science & Business Media.

[11] Ferland, R., Latour, A. and Oraichi, D. (2006). Integer-valued GARCH process. *Journal of time series analysis* **27** 923–942.

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

[13] Fokianos, K. and Fried, R. (2010). Interventions in INGARCH processes. *Journal of Time Series Analysis* **31** 210–225.

[14] Fokianos, K., Rahbek, A. and Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association* **104** 1430–1439.

[15] Francq, C. and Zakoian, J.-M. (2001). Stationarity of multivariate Markov–switching ARMA models. *Journal of Econometrics* **102** 339–364.

[16] Ghysels, E. (1994). On the periodic structure of the business cycle. *Journal of Business & Economic Statistics* **12** 289–298.

[17] Goodwin, T. H. (1993). Business-cycle analysis with a Markov-switching model. *Journal of Business & Economic Statistics* **11** 331–339.

[18] Granger, C. W., Terasvirta, T. et al. (1993). Modelling nonlinear economic relationships. *OUP Catalogue*.

[19] Gray, S. F. (1996). Regime-switching in Australian short-term interest rates. *Accounting & Finance* **36** 65–88.

[20] Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the econometric society* 357–384.

[21] Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of econometrics* **45** 39–70.

[22] Hamilton, J. D. and Lin, G. (1996). Stock market volatility and the business cycle. *Journal of applied econometrics* **11** 573–593.

[23] Hamilton, J. D. and Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of econometrics* **64** 307–333.

[24] Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica: Journal of the econometric society* 413–430.

[25] Hansen, B. E. (2011). Threshold autoregression in economics. *Statistics and its Interface* **4** 123–127.

[26] Horn, R. A. and Johnson, C. R. (2012). *Matrix analysis*. Cambridge university press.

[27] Huang, C. H. (1999). Phases and characteristics of Taiwan business cycles: A Markov switching analysis. *Taiwan Economic Review* **27** 185–214.

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

[29] Kim, C.-J. and Nelson, C. R. (2017). *State-space models with REGIME Switching: Classical and gibbs-sampling approaches with applications*. MIT press.

[30] Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. In *Advances in Markov-switching models* 223–254. Springer.

[31] Kuan, C.-M. (2002). Lecture on the Markov switching model. *Institute of Economics Academia Sinica* **8** 1–30.

[32] McKenzie, E. (2003). Discrete variate time series. In *Handbook of statistics*, **21** 573–606. Elsevier.

[33] Medeiros, M. C., Veiga, A. and Resende, M. G. C. (2002). A combinatorial approach to piecewise linear time series analysis. *Journal of Computational and graphical Statistics* **11** 236–258.

[34] Pan, J., Wang, H. and Tong, H. (2008). Estimation and tests for power-transformed and threshold GARCH models. *Journal of Econometrics* **142** 352–378.

[35] Safaei, M. (2012). Maximum Likelihood Estimator for Markov Switching Autoregressive Conditional Heteroskedasticity Model. *Journal of Basic and Applied Scientific Research* **2** 5652–5655.

[36] Schaller, H. and Norden, S. V. (1997). Regime switching in stock market returns. *Applied Financial Economics* **7** 177–191.

[37] Sims, C. A. and Zha, T. (2006). Were there regime switches in US monetary policy? *American Economic Review* **96** 54–81.

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

[39] Tjøstheim, D. (1990). Non-linear time series and Markov chains. *Advances in applied probability* **22** 587–611.

[40] Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*. Oxford University Press.

[41] Tong, H. and Lim, K. S. (1980). Threshold Autoregression, Limit Cycles and Cyclical Data. *Journal of the Royal Statistical Society. Series B (Methodological)* **42** 245–292.

[42] Tsay, R. S. and Chen, R. (2018). *Nonlinear time series analysis* **891**. John Wiley & Sons.

[43] Wang, C., Liu, H., Yao, J.-F., Davis, R. A. and Li, W. K. (2014). Self-excited threshold Poisson autoregression. *Journal of the American Statistical Association* **109** 777–787.

Jichun Liu,
School of Mathematical Science, Xiamen University, Xiamen 361005,
China.
E-mail address: liujichun65@126.com

Yue Pan,
Department of Mathematics and Statistics, University of Strathclyde, Glasgow, G1 1XH,
United Kingdom.
E-mail address: yue.pan@strath.ac.uk

Jiazhu Pan,
Department of Mathematics and Statistics, University of Strathclyde, Glasgow, G1 1XH,
United Kingdom;
School of Mathematics and Statistics, Yangtze Normal University, Chongqing 408100,
China.
E-mail address: jiazhu.pan@strath.ac.uk

Abdullah Almarashi
Department of Mathematics and Statistics, University of Strathclyde, Glasgow, G1 1XH,
United Kingdom.
E-mail address: abdullah.almarashi@strath.ac.uk