Stationarity of a general class of observation driven models for discrete valued processes

Stazionarietà di una classe generale di modelli observation-driven per processi a valori discreti

Mirko Armillotta, Alessandra Luati and Monia Lupparelli

Abstract A large variety of time series observation-driven models for binary and count data are currently used in different contexts. Despite the importance of stationarity and ergodicity to ensure suitable results, for many of these models stationarity is not yet proved. We specify a general class of observation-driven models for discrete valued processes, which encompasses the most frequently used models. Then, we show strict stationarity by means of Feller properties and establish easy-to-check stationarity conditions.

Abstract *Modelli observation-driven per serie storiche di dati binari e di conteggio sono correntemente utilizzati in diversi contesti. In alcuni casi, tuttavia, le proprietà di stazionarietà ed ergodicità non sono state dimostrate. In questo paper, viene specificata una classe generale di modelli observation driven per dati discreti, che comprende i modelli maggiormente utilizzati in letteratura. Tramite le proprietà di Feller, si derivano condizioni di stazionarietà semplici da verificare.*

Key words: Generalized linear ARMA, Time series of counts, Binary variables, Drift conditions

1 Introduction

Observation-driven models were originally introduced by Cox [2] and they have nowadays received new interest. There is an heterogeneous literature about such models for binary data [12, 17, 11] and for count data [3, 7, 9]; other general models

Mirko Armillotta, Alessandra Luati
Department of Statistical Sciences, University of Bologna, Via Belle Arti 41, 40126, Bologna, e-mail: `mirko.armillotta2@unibo.it`, e-mail: `alessandra.luati@unibo.it`

Monia Lupparelli
Department of Statistics, Computer Science, Applications, University of Florence, Viale Morgagni 59, 50134, Florence, e-mail: `monia.lupparelli@unifi.it`

were introduced in [1] and [19]. More recently, various attempts have been done to study the probabilistic properties of these models. Stationarity and ergodicity were proved for a very general model in [5] and in [4], but these results do not directly apply to the models mentioned above; the results of [5] and [4] provide a basis for proofs which one needs to develop from time to time, depending on different models and specific distributions. In practical applications, directly applicable stationarity condition are needed to guarantee the reliability and validity of the results obtained. Strict stationarity results have been directly derived by [13] solely for the Generalized Autoregressive Moving Average (GARMA) model of [1].

Our contribution extends the argument of [13] and provides stationarity and ergodicity conditions directly verifiable and applicable for a class of observation-driven models that encompasses the models mentioned so far and for data coming from a large family of distributions.

In Section 2 we formulate the general framework, with some examples. In Section 3 we establish stationarity and ergodicity for the model. In Section 4 we apply the results to some specific models. In Section 5, concluding remarks and future developments are highlighted.

2 The framework

Let us consider the stochastic process $\{Y_n\}_{n\in\mathbb{N}}$ and the filtration $\mathscr{F}_{n-1} = \sigma(Y_s,\ s \leq n-1, X_0 = x)$, the information set up to time $n-1$ and the starting value for X_n. An observation-driven model for Y_n has the form

$$Y_n|Y_{0:n-1} \sim f(\cdot;\mu_n) \tag{1}$$

$$\mu_n = q_{\theta,n}(Y_{0:n-1}) \tag{2}$$

where $q_{\theta,n}$ is some function parametrized by θ and $f(\cdot;\mu_n)$ is a density (or mass) function whose dynamic is captured by μ_n; usually, but not necessarily, this distribution is assumed to belong to the exponential family with μ_n as conditional expectation. We focus on models where the observation process $\{Y_n\}_{n\in\mathbb{N}}$ is integer-valued and find conditions under which there exists a stationary and ergodic version of it via Markov chain theory. However, since $\{Y_n\}_{n\in\mathbb{N}}$ is not itself a Markov chain, a classical approach is to prove the existence of a stationary ergodic process $\{Y_n\}_{n\in\mathbb{N}}$ as a function of an ergodic Markov chain $X = \{X_n\}_{n\in\mathbb{N}}$, on a state space S with σ-algebra $\mathscr{F}$ and n-step transition kernel $P(x,A) = \mathrm{P}^n(X_n \in A|X_0 = x)$ for $A \in \mathscr{F}$ and starting from $X_0 = x$.

In the present case, the chain is specified as $X_n = g(\mu_n)$ where g is a bijective increasing function, the *link function*; an explicit formulation for (2) is defined as follows

$$g(\mu_n) = \alpha + \sum_{j=1}^{k} \gamma_j g(\mu_{n-j}) + \sum_{j=1}^{p} \phi_j h(Y_{n-j}) + \sum_{j=1}^{q} \theta_j \left[\frac{h(Y_{n-j}) - \bar{g}(\mu_{n-j})}{\nu_{n-j}} \right] \tag{3}$$

where ν_n is some scaling sequence. The function $h(Y_n)$ is called *y-link function* because it is applied only to the observations Y_n whereas $\bar{g}(\mu_n)$ is said *mean-link function* because it is applied only to the expected value μ_n. Both link functions are monotone and could be different from the link $g(\cdot)$. In general, it is useful to choose the mean-link function as follows:

$$\bar{g}(\mu_n) = \mathrm{E}[h(Y_n)\,|\mathscr{F}_{n-1}\,] \,, \tag{4}$$

so that $\varepsilon_n = h(Y_n) - \bar{g}(\mu_n)$ is a martingale difference sequence (MDS) and can be interpreted as a *prediction error*.

For sake of clarity we focus the attention on the first order model

$$g(\mu_n) = \alpha + \gamma g(\mu_{n-1}) + \phi\, h(Y^*_{n-1}) + \theta \left[\frac{h(Y^*_{n-1}) - \bar{g}(\mu_{n-1})}{\nu_{n-1}}\right] \,, \tag{5}$$

where Y^*_n is some mapping of Y_n to the domain of $h(\cdot)$.

The general class of models (5) has a large flexibility in that it encompasses many time series models of interest. The GARMA model [1, 13] is easily obtained when $\gamma = 0$, by setting $g \equiv \bar{g} \equiv h$ and $\nu_n = 1$, such as, by equivalence of the three link functions and no scaling applied, one has

$$g(\mu_n) = \alpha + \phi\, g(Y^*_{n-1}) + \theta \left[g(Y^*_{n-1}) - g(\mu_{n-1})\right] \,. \tag{6}$$

Note that, in this case, $\varepsilon_n = g(Y^*_{n-1}) - g(\mu_{n-1})$ is a MDS only in the special case in which $g \equiv h$, the identity function.

The Binomial ARMA (BARMA) model, developed in [12, 17] is obtained by (5) when $\gamma = 0$, h is the identity ($\bar{g}(\mu_t)$ reduces to μ_t) and $Y^*_n = Y_n$. Then,

$$g(\mu_n) = \alpha + \phi\, Y_{n-1} + \theta\, [Y_{n-1} - \mu_{n-1}] \,. \tag{7}$$

Another promising branch of the literature has been developed by [16] and [3], under the name of Generalized Linear ARMA (GLARMA) models. This class is recovered here by setting $\phi = 0$ and h the identity,

$$g(\mu_n) = \alpha + \gamma g(\mu_{n-1}) + \theta \left[\frac{Y_{n-1} - \mu_{n-1}}{\nu_{n-1}}\right] \tag{8}$$

where $\bar{g}(\mu_n)$ reduces to μ_n and $Y^*_n = Y_n$.

Other models are contained in this general class, such as those in [8, 9, 11, 19].

3 Strict stationarity for the general model

In this section we present results for stationarity conditions of the chain $\{X_n\}_{n\in\mathbb{N}}$ and the process $\{Y_n\}_{n\in\mathbb{N}}$ coming from (5).

The usual practical condition for establishing stationarity and ergodicity in a Markov chain is by showing that it is positive Harris recurrent, via a drift condition in "small set". Positive Harris recurrent chains possess a unique stationary probability distribution π. However, this does not work if the chain is not φ-irreducible, as for the case of $\{Y_n\}_{n\in\mathbb{N}}$ integer-valued (for the details, see [14]). Nevertheless, as suggested by [13], one can still use the drift condition combined with the *weak Feller* property to show existence of a stationary distribution. Then, by applying the *asymptotic strong Feller* condition, one can derive uniqueness of the stationary distribution (for the definitions, see [18, 10, 13]).

Let $\mathrm{E}_x(\cdot)$ denote the expectation under the probability $\mathrm{P}_x(\cdot)$ induced on the path space of the chain when the initial state is $X_0 = x$.

We handle three separate cases:

1. $f(\cdot;\mu)$ is defined for any $\mu \in \mathbb{R}$. In this case the domain of g and h is $\mathbb{R}$ and $Y_n^* = Y_n$ is taken;
2. $f(\cdot;\mu)$ is defined for only $\mu \in \mathbb{R}^+$(or μ on any one-sided open interval by analogy). In this case the domain of g and h is $\mathbb{R}^+$ and $Y_n^* = \max\{Y_n, c\}$ for some $c \geq 0$ is taken;
3. $f(\cdot;\mu)$ is defined for only $\mu \in (0,a)$ where $a > 0$ (or any bounded interval by analogy). In this case the domain of g and h is $(0,a)$ and for some $c \in [0, a/2)$ $Y_t^* = \min\{\max(Y_n, c), (a-c)\}$ is taken.

Let $Y_0(x)$ denote the random variable Y_0 conditional on $\mu_0 = x$.

Definition 1. The Lipschitz condition

$$|\tilde{g}(z) - \tilde{g}(w)| \leq L|z - w| \tag{9}$$

with $L \leq 1$, is satisfied in the following different scenarios:

1. $\bar{g} \equiv h \neq g$
2. $\bar{g} \neq g$ and h : *identity*
3. $\mathrm{E}[h(Y_t^*)|\mathscr{F}_{t-1}] = \bar{g}(\mu_t) \neq g(\mu_t)$

under the assumption that the link functions g^{-1}, h, $\bar{g}$ are Lipschitz with constant smaller or equal than 1.

Theorem 1. *The process $\{\mu_n\}_{n\in\mathbb{N}}$ specified by the model* (5) *has a stationary distribution, and thus is stationary for an appropriate initial distribution for μ_0 (then, $\{Y_n\}_{n\in\mathbb{N}}$ is stationary), under the conditions below.*

1. $Y_0(x) \Rightarrow Y_0(x')$ *as* $x \to x'$.
2. $\mathrm{E}(Y_n \mid \mu_n) = \mu_n$.
3. *There exist $\delta > 0$, $r \in [0, 1+\delta)$ and nonnegative constants d_1, d_2 such that*

$$\mathrm{E}(|Y_n - \mu_n|^{2+\delta} \mid \mu_n) \leq d_1 |\mu_n|^r + d_2.$$

4. *g and h are bijective and increasing, and*
   - *If $\bar{g}(\mu_t) = g(\mu_t)$,*

a. $h : \mathbb{R} \mapsto \mathbb{R}$ *concave on* $\mathbb{R}^+$ *and convex on* $\mathbb{R}^-$, $g : \mathbb{R} \mapsto \mathbb{R}$ *concave on* $\mathbb{R}^+$ *and convex on* $\mathbb{R}^-$, *and* $|\phi| + |\gamma| < 1$
b. $h : \mathbb{R}^+ \mapsto \mathbb{R}$ *concave on* $\mathbb{R}^+$, $g : \mathbb{R}^+ \mapsto \mathbb{R}$ *concave on* $\mathbb{R}^+$, *and* $(|\gamma| + |\phi|) \vee |\theta + \gamma| < 1$
c. $|\theta + \gamma| < 1$*; no additional conditions on* $h : (0,a) \mapsto \mathbb{R}$ *and* $g : (0,a) \mapsto \mathbb{R}$.

- *If* $\bar{g}(\mu_t) \neq g(\mu_t)$ *and* $\tilde{g}$ *satisfies the Lipschitz condition* (9),
  a. $h : \mathbb{R} \mapsto \mathbb{R}$ *concave on* $\mathbb{R}^+$ *and convex on* $\mathbb{R}^-$, $g : \mathbb{R} \mapsto \mathbb{R}$ *concave on* $\mathbb{R}^+$ *and convex on* $\mathbb{R}^-$, *and* $|\phi| + |\gamma| < 1$
  b. $h : \mathbb{R}^+ \mapsto \mathbb{R}$ *concave on* $\mathbb{R}^+$, $g : \mathbb{R}^+ \mapsto \mathbb{R}$ *concave on* $\mathbb{R}^+$, *and* $|\gamma| + (|\phi| \vee |\theta|) < 1$
  c. $|\theta| + |\gamma| < 1$*; no additional conditions on* $h : (0,a) \mapsto \mathbb{R}$ *and* $g : (0,a) \mapsto \mathbb{R}$.

Let $X_n = g(\mu_n)$. For $X_0 = x$ and $g(\mu) = x$ we have that

$$X_1(x) = \alpha + \phi h(Y_0^*(g^{-1}(x))) + \theta[h(Y_0^*(g^{-1}(x))) - \tilde{g}(x)] + \gamma x$$

where $\tilde{g}(x) = (\bar{g} \circ g^{-1})(x) = \bar{g}(g^{-1}(x)) = \bar{g}(\mu)$.

Since g^{-1} is continuous, $Y_0(g^{-1}(x)) \Rightarrow Y_0(g^{-1}(x'))$ as $x \to x'$. Since the * that maps Y_0 to the domain of h is continuous, it follows that $Y_0^*(g^{-1}(x)) \Rightarrow Y_0^*(g^{-1}(x'))$ as $x \to x'$. Since h is continuous, we have that $h(Y_0^*(g^{-1}(x))) \Rightarrow h(Y_0^*(g^{-1}(x')))$. Since $\tilde{g}(x)$ is continuous, we have that $\tilde{g}(x) \Rightarrow \tilde{g}(x')$. So $X_1(x) \Rightarrow X_1(x')$ as $x \to x'$, showing the weak Feller property. So by combining this fact and Theorem 1, Theorem in [18] is satisfied and, then, a stationary distribution for $\{\mu_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ exists.

Assume that the distribution $\pi_z(\cdot)$ of $g(Y_n^*)$ conditional on $g(\mu_t) = z$ has the Lipschitz property

$$\sup_{w,z\in\mathbb{R}:w\neq z} \frac{\|\pi_w(\cdot) - \pi_z(\cdot)\|_{TV}}{|w - z|} < B < \infty \tag{10}$$

where $\|\cdot\|_{TV}$ is the total variation norm (see [14], pag. 315).

Theorem 2. *Suppose that the conditions of Theorem 1 and the Lipschitz condition* (10) *hold, and that there is some* $x \in \mathbb{R}$ *that is in the support of* Y_0 *for all values of* μ_0*. Then, there is a unique stationary distribution for* $\{\mu_n\}_{n\in\mathbb{N}}$.

Proof. A sketch of the proofs of Theorems 1 and 2 is postponed to the Appendix.

4 Strict stationarity of specific models

In this section, the results obtained in Theorems 1 and 2 are applied to specific models of potential interest.

We first remark that for the GARMA model (6), Theorems 1 and 2 reduce exactly to the results of [13].

For the BARMA model, Theorem 1 and 2 reduce to the following proposition.

Proposition 1. *Suppose that conditional on μ_n, Y_n is Binomial(n,μ_n), with fixed number of trials n, the link function $g:(0,a)\mapsto\mathbb{R}$ is bijective and increasing, g^{-1} is Lipschitz and $|\theta|<1$. Then, the process $\{\mu_n\}_{n\in\mathbb{N}}$ defined in* (7) *has a unique stationary distribution π. Hence, when μ_0 is initialized according to π, the process $\{Y_n\}_{n\in\mathbb{N}}$ is strictly stationary.*

In [13] (pag. 820-821) is proved that, for Poisson and Binomial distributions, the Lipschitz conditions (10) holds when g^{-1} is Lipschitz. We proved that the same holds for the Negative Binomial distribution. Note that the conditions on g and g^{-1} are clearly satisfied for the usual link, like logit or probit.

For GLARMA models, no stationarity results are available, apart form the simplest case when $k=0$, $q=1$ (see [3], [7], [4]). Our Theorems 1 and 2 imply the following proposition.

Proposition 2. *The process $\{\mu_n\}_{n\in\mathbb{N}}$ specified by the model* (8)*, has a unique stationary distribution π, and thus is stationary when μ_0 is initialized according to π, under the conditions below. This implies that $\{Y_n\}_{n\in\mathbb{N}}$ is strictly stationary when μ_0 is initialized according to π. The conditions are:*

1. $\mathrm{E}(Y_n\mid\mu_n)=\mu_n$.
2. *($2+\delta$ moment condition): There exist $\delta>0$, $r\in[0,1+\delta)$ and nonnegative constants d_1,d_2 such that*

$$\mathrm{E}(|Y_n-\mu_n|^{2+\delta}\mid\mu_n)\le d_1\,|\mu_n|^r+d_2.$$

3. *g is bijective and increasing, and*
   a. *$g:\mathbb{R}\mapsto\mathbb{R}$ concave on $\mathbb{R}^+$ and convex on $\mathbb{R}^-$, and $|\gamma|<1$*
   b. *$g:\mathbb{R}^+\mapsto\mathbb{R}$ concave on $\mathbb{R}^+$, and $|\gamma|+|\theta|<1$*
   c. *$|\gamma|+|\theta|<1$; no additional conditions on $g:(0,a)\mapsto\mathbb{R}$.*
4. *g^{-1} is Lipschitz with constant not greater than 1.*
5. *If $\{Y_t\}_{t\in\mathbb{N}}$ is discrete-valued, then* (10) *need to hold.*

Note that, in the GLARMA model, the conditional distribution of $\{Y_t\}_{t\in\mathbb{N}}$ belongs to the exponential family, thus the first two moment conditions are satisfied. As mentioned above, for usual choices of discrete distributions (Poisson, Binomial, or Negative Binomial) the Lipschitz conditions (10) holds when g^{-1} is Lipschitz.

Finally, the conditions on g and g^{-1} clearly hold for the usual link functions.

In practical applications, one just needs to verify the condition on the coefficients to establish the stationarity of the model.

5 Concluding remarks and further developments

This paper provides a framework for proving the existence of stationarity end ergodic solutions for a wide class of observation-driven time series models. For many models in the class, no such results were available in the literature.

The Lipschitz assumption (10) is not satisfied when the y-link function h is the logarithmic function. In [5], different assumptions are considered to weaken the Lipschitz condition. We shall investigate them in the future.

All the models encompassed are often used with covariates. Extending our results to accomplish for covariates would be a further aspect to investigate.

Finally, our stationarity results could be used for proving consistency and asymptotic normality of estimators in discrete-valued models. The results of [5] and [6] could be used in future work with the aim to develop the asymptotic theory for the class of models considered in this paper.

Appendix

The proof for Theorem 1 and 2 follows the line of Theorems 5 and 15 in [13]. We provide a sketch the proof here in the following.

Having showed that the set $A = [-M;M]$, $M > 0$, is a small set, it is possible to prove a drift condition by taking the energy function $V(x) = |x|$ for the model (5):

$$\mathrm{E}_x V(X_1) = \mathrm{E}_x|\alpha + \gamma x + \phi h(Y_0^*) + \theta[h(Y_0^*) - \bar{g}(\mu)]| \tag{11}$$

and find that it is bounded under certain conditions on the coefficients, in the same fashion of Theorem 5 of [13]. For sake of brevity, we omit the details.

The last step required for completing the proof it to show that the Markov chain $\{X_t\}_{t\in\mathbb{N}}$ is asymptotically strong Feller. This is accomplished by a modification of the proof for Theorem 15 of [13]: set $g \equiv \bar{g}$, then, the random variables $g(Y_0^*(z))$ and $g(Y_0^*(w))$ have marginal distributions π_z and π_w, and $\mathrm{P}(g(Y_0^*(w)) = g(Y_0^*(z))) = 1 - \|\pi_w(\cdot) - \pi_z(\cdot)\|_{TV} > 1 - B|z-w|$.

If $h(Y_0^*(w)) = h(Y_0^*(z))$ then $|Z_1(w) - Z_1(z)| = |-\theta(\tilde{g}(w) - \tilde{g}(z)) + \gamma(z-w)| = |\theta+\gamma||z-w|$ and so $\left\|\pi_{Z_1(z)}(\cdot) - \pi_{Z_1(w)}(\cdot)\right\|_{TV} < B|Z_1(z) - Z_1(w)| < B|\theta+\gamma||z-w|$. Then we can construct $g(Y_1^*(z))$ and $g(Y_1^*(w))$ so that they have the correct marginal distributions and that $\mathrm{P}(g(Y_1^*(w)) = g(Y_1^*(z))|g(Y_0^*(w)) = g(Y_0^*(z))) = \mathrm{P}(g(Y_1^*(w)) = g(Y_1^*(z))|h(Y_0^*(w)) = h(Y_0^*(z))) > 1 - \left\|\pi_{Z_1(z)}(\cdot) - \pi_{Z_1(w)}(\cdot)\right\|_{TV} > 1 - B|\theta+\gamma||z-w|$ where the first equality works because g and h are one-to-one functions
$g(Y_0^*(w)) = g(Y_0^*(z)) \iff Y_0^*(w) = Y_0^*(z) \iff h(Y_0^*(w)) = h(Y_0^*(z))$.
If $h(Y_1^*(z)) = h(Y_1^*(w))$ then we can continue to "couple" the chains as above.
Notice that the probability that the chains couple for all times $0, 1, \ldots$ is at least

$$1 - B|z-w|\sum_{t=0}^{\infty}(|\theta+\gamma|)^t = 1 - \frac{|z-w|B}{1-|\theta+\gamma|},$$

where the inequality applies by imposing $|\theta+\gamma| < 1$. Thus, we combine this coefficient condition with those obtained from the drift condition (11) into Theorem 1. The remaining follows as in [13].

If $g \neq \bar{g}$ is it possible to adapt the previous proof in the following way $|Z_1(w)-Z_1(z)| = |-\theta(\tilde{g}(w)-\tilde{g}(z))+\gamma(z-w)| \leq |\theta||\tilde{g}(w)-\tilde{g}(z)|+|\gamma||z-w|$ and, under the Lipschitz condition (9) we obtain $|Z_1(w)-Z_1(z)| \leq |\theta||\tilde{g}(w)-\tilde{g}(z)| + |\gamma||z-w| \leq (|\theta|+|\gamma|)|z-w|$. Hence, the proof for the former case $\bar{g} \equiv g$ works also for other shapes of $\bar{g}$, by substituting the condition $|\theta+\gamma|$ with $|\theta|+|\gamma|$ and by combining it in Theorem 1. Clearly, the condition (9) depends on the shape of $\tilde{g}$. However, it is easy to show that it works for all the specific shapes of the link functions considered in Definition 1. We omit the details.

References

1. Benjamin, M.A., Rigby, R.A., Stasinopoulos, D.M.: Generalized autoregressive moving average models. J. Amer. Stat. Assoc. **98**, (461), 214–223 (2003)
2. Cox, D.: Statistical analysis of time-series: some recent developments. Scand. J. Stat. **8**, (2), 93–115 (1981)
3. Davis, R., Dunsmuir, W., Streett, S.: Observation-driven models for Poisson counts. Biometrika **90**, (4), 777–790 (2003)
4. Davis, R.,Liu, H.: Theory and inference for a class of observation-driven models with application to time series of counts. Stat. Sin. **26**, (4), 1673–1707 (2016)
5. Douc, R., Doukhan, P., Moulines, E.: Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. Stoch. Proc. Appl. **123**, (7), 2620–2647 (2013)
6. Douc, R., Fokianos, K., Moulines,E.: Asymptotic properties of quasi-maximum likelihood estimators in observation-driven time series models. Electron. J. Stat. **11**, (2), 2707–2740 (2017)
7. Dunsmuir, W.T., Scott, D.J.: The glarma package for observation driven time series regression of counts. J. Stat. Soft. **67**, (7), 1–36 (2015)
8. Ferland, R., Latour,A., Oraichi, D.: Integer-valued GARCH process. J. Time Ser. Anal. textbf27, (6), 923–942 (2006)
9. Fokianos, K., Tjstheim, D.: Log-linear Poisson autoregression. J. Multivar. Anal. **102**, (3), 563–578 (2011)
10. Hairer, M., Mattingly, J.C.: Ergodicity of the 2D Navier-Stokes equations with degenerate stochastic forcing. Ann. Math. **164** 993–1032 (2006)
11. Kauppi, H., Saikkonen, P.: Predicting US recessions with dynamic binary response models. Rev. Econ. Stat. **90**, (4), 777–791 (2008)
12. Li, W.K.: Time series models based on generalized linear models: some further results. Biometrics **50**, (2), 506–511 (1994)
13. Matteson, D.S., Woodard, D.B., Henderson, S.G.: Stationarity of generalized autoregressive moving average models. Electron. J. Stat. **5**, 800–828 (2011)
14. Meyn, S., Tweedie, R.L., Glynn, P. W.: Markov Chains and Stochastic Stability. Cambridge University Press, London (2009)
15. Roberts, G.O., Rosenthal, J.S.: General state space markov chains and mcmc algorithms. Probab. Surv. **1**, 20–71 (2004)
16. Rydberg, T.H., Shephard N.: Dynamics of trade-by-trade price movements: Decomposition and models. J. Financ. Economet. **1**, (1), 2–25 (2003)
17. Startz, R.: Binomial autoregressive moving average models with an application to U.S. recessions. J. Bus. Econ. Stat. **26**, (1), 1–8 (2008)
18. Tweedie, R.L.: Invariant measures for Markov chains with no irreducibility assumptions. J. Appl. Probab., **25**, (A), 275–285 (1988)
19. Zheng, T., Xiao, H., Chen, R.: Generalized ARMA models with martingale difference errors. J. Economet. **189**, (2), 492–506 (2015)