

# Self-Exciting Threshold Binomial Autoregressive Processes\*

Tobias A. Möller<sup>†</sup>  
moellert@hsu-hh.de

Maria Eduarda Silva<sup>‡</sup>  
mesilva@fep.up.pt

Christian H. Weiß<sup>§</sup>  
weissc@hsu-hh.de

Manuel G. Scotto<sup>¶</sup>  
mscotto@ua.pt

Isabel Pereira<sup>||</sup>  
isabel.pereira@ua.pt

## Abstract

We introduce a new class of integer-valued self-exciting threshold models, which is based on the binomial autoregressive model of order one as introduced by McKenzie (1985). Basic probabilistic and statistical properties of this class of models are discussed. Moreover, parameter estimation and forecasting are addressed. Finally, the performance of these models is illustrated through a simulation study and an empirical application to a set of measles cases in Germany.

## 1 Introduction

Continuous-valued threshold autoregressive models have been extensively investigated in the literature, see the survey by Tong (2011). Some basic results on the probabilistic structure of this class of models can be found, e.g., in Chan et al (1985); Chan and Tong (1985); Cline and Pu (1999, 2004); Lanne and Saikkonen (2005); Liebscher (2005), and in the books by Tong (1990); Turkman et al (2014). Threshold models as proposed by Tong and Lim (1980); Tong (1983) have had an enormous influence in various fields of research in the past years caused by their excellent abilities to handle nonlinearity. They find usage in, e.g., actuarial science (Chan et al, 2004), biological sciences (Stenseth et al, 2006) as well as economics and finance (Chen et al, 2011; Hansen, 2011) just to mention a few.

In the field of integer-valued time series modelling (with either bounded or unbounded range of counts), limited research has been carried out so far to develop models to cope with time series of counts exhibiting piecewise-type patterns. One such approach are hidden Markov models (HMM) for counts (Zucchini and MacDonald, 2009), where a state dependence of the observed counts is introduced through an underlying (invisible) finite Markov chain (also see Section 5 below). While the HMMs are some kind of parameter-driven regime switching models, the threshold models being considered here are observation-driven regime switching

---

\*Accepted author's manuscript (AAM) to be published in [AStA Advances in Statistical Analysis ]. [DOI 10.1007/s10182-015-0264-6]

<sup>†</sup>Department of Mathematics and Statistics, Helmut Schmidt University, Hamburg, Germany

<sup>‡</sup>CIDMA and Faculdade de Economia, Universidade do Porto, Portugal

<sup>§</sup>Department of Mathematics and Statistics, Helmut Schmidt University, Hamburg, Germany

<sup>¶</sup>Department of Mathematics, IST University of Lisbon

<sup>||</sup>CIDMA and Department of Mathematics, University of Aveiro, Portugal

models. Besides the threshold regression model by Samia et al (2007), a few models being motivated by the autoregressive moving average (ARMA) approach have been proposed, see the survey by Möller and Weiß (2015). In particular, Monteiro et al (2012) introduced the class of self-exciting threshold integer-valued autoregressive (SETINAR) models of order one and with two regimes, defined by the recursive equation

$$X_t = \begin{cases} \alpha_1 \circ X_{t-1} + Z_t & \text{if } X_{t-1} \leq R, \\ \alpha_2 \circ X_{t-1} + Z_t & \text{if } X_{t-1} > R. \end{cases}$$

Here,  $(Z_t)$  constitutes a sequence of integer-valued random variables, and  $R$  represents the fixed threshold level separating the regimes. The “ $\alpha \circ$ ” is the binomial thinning operator of Steutel and van Harn (1979). It is defined as  $\alpha \circ X := \sum_{i=1}^X Y_i$ , for  $X$  with range  $\mathbb{N}_0 = \{0, 1, \dots\}$ , where the  $Y_i$ ’s are independent and identically distributed (i.i.d.) Bernoulli variables with probability  $\alpha \in (0; 1)$ . A similar SET approach related to the INAR(1) model was proposed by Thyregod et al (1999). It is important to stress here that the models by Monteiro et al (2012); Thyregod et al (1999) as well as the other recently proposed ARMA-like models by Wang et al (2014); Yu et al (2014); Zou and Yu (2014) are useful for fitting integer-valued time series exhibiting the piecewise phenomena defined over an *infinite* range of counts.

However, these SETINAR models are of little use for modelling time series taking values over a *finite* range of counts. As an illustrative example of a time series of counts exhibiting two different regimes over a bounded interval, Figure 1 shows a time series plot of the counts of districts in Germany in which at least one new case of measles was observed (weekly counts, 2004–2005); the data were downloaded from SurvStat (Robert-Koch-Institut, 2014). The time series consists of  $T = 105$  data points (53 weeks in 2004 and 52 weeks in 2005), taking values within the set  $\{0, \dots, N\}$  with  $N = 38$ , where the fixed upper bound is determined through the given number of districts in Germany. A visual inspection of the time series reveals a level shift in the graphic corresponding to the first half of 2005, so it seems appropriate to mix integer-valued and threshold models for describing this data set.

Motivated by the aforementioned example, we shall propose a simple class of threshold autoregressive models for finite counts data based on McKenzie’s binomial AR(1) (BAR(1) in short) model. For the predetermined upper limit  $N$  of the range, the latter model is defined by the recursion

$$X_t = \alpha \circ X_{t-1} + \beta \circ (N - X_{t-1}), \quad (1)$$

for  $\pi \in (0; 1)$ ,  $r \in (\max\{-\frac{\pi}{1-\pi}, -\frac{1-\pi}{\pi}\}; 1)$  and  $\beta := \pi(1-r)$ ,  $\alpha := \beta + r$ . Here, all thinnings are performed independently of each other, and the thinnings at time  $t$  are independent of  $(X_s)_{s < t}$ . Note that the representation for  $X_t$  in (1) guarantees that the range of  $X_t$  is given by the set  $\{0, 1, \dots, N\}$ . Furthermore, the condition on  $r$  guarantees that the thinning parameters  $\alpha, \beta \in (0; 1)$ . The parameter  $\alpha$  can be interpreted as a survival rate, and  $\beta$  as a recolonization rate. The process in (1) is a stationary Markov chain with  $N + 1$  states and binomial marginal distribution  $B(N, \pi)$ . The BAR(1) process shares important properties with the conventional AR(1) model, e.g., that the autocorrelation function (ACF)  $\rho(k) = r^k$  decays to 0 at an exponential rate.

The rest of the paper is organized as follows: in Section 2, diverse self-exciting threshold versions of the BAR(1) model (1) are introduced and investigated in detail. Parameter estimation

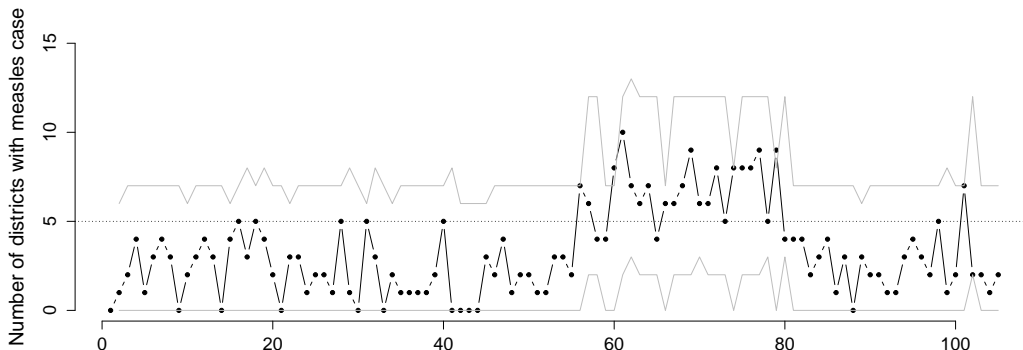


Figure 1: Weekly number of districts in Germany with at least one new case of measles (2004–2005).

Dotted line: estimated threshold dividing the range into two regimes.

Gray lines: lower and upper limit of 95% forecast intervals for fitted LSET-BAR(1) model, see Section 5.

and model specification are addressed in Section 3, while Section 4 deals with approaches for forecasting. In Section 5, we consider again the measles data shown in Figure 1 and compare our models to competing models. Section 6 presents possible extensions of our basic self-exciting threshold models, e.g., with higher delays, or for the case of bivariate counts. The article closes with some concluding remarks in Section 7.

## 2 A Basic Self-Exciting Threshold Binomial AR(1) Model

Based on the BAR(1) model (1), we build our extension by introducing a self-exciting threshold mechanism. For the moment, we restrict to a basic model, where the finite range is separated into two regimes, and where the state of the process is selected according to the previous observation (delay 1); later in Section 6, a number of possible extensions of our basic model are presented. The two regimes are determined by a specified threshold value  $0 \leq R < N$ : the lower regime consists of the states  $\{0, 1, 2, \dots, R\}$ , and the upper regime consists of  $\{R + 1, R + 2, \dots, N\}$ . In each regime, the model takes individual values for the survival rate  $\alpha_i$  and for the recolonization rate  $\beta_i$ .

### 2.1 The SET Binomial AR(1) Model

We start with the most general definition of our basic self-exciting threshold binomial AR(1) model (SET-BAR, in short), i.e., a univariate model having two regimes and delay parameter 1. Later in Section 6, we discuss possible extensions to higher-order autoregressions, to delays  $d > 1$ , to more than two regimes, and to the bivariate case.

**2.1.1 Definition** *Let  $N \in \mathbb{N}$  be the predetermined upper limit of the range, and let  $0 \leq R < N$  be the threshold value. Define  $\pi_i \in (0; 1)$ ,  $r_i \in \left(\max\left\{-\frac{\pi_i}{1-\pi_i}, -\frac{1-\pi_i}{\pi_i}\right\}; 1\right)$ , as well as  $\beta_i := \pi_i \cdot (1 - r_i) \in (0; 1)$  and  $\alpha_i := \beta_i + r_i \in (0; 1)$  for  $i \in \{1, 2\}$ . A process  $(X_t)$  is called a*

SET-BAR(1) process if  $X_t$  follows the recursion

$$X_t = \phi_t \circ X_{t-1} + \eta_t \circ (N - X_{t-1}) \quad \text{for } t \in \mathbb{Z}, \quad (2)$$

where  $\phi_t := \alpha_1 I_{t-1} + \alpha_2(1 - I_{t-1})$  and  $\eta_t := \beta_1 I_{t-1} + \beta_2(1 - I_{t-1})$  with  $I_{t-1} := \mathbf{1}_{\{X_{t-1} \leq R\}}$  as the indicator variable.

Note that, in the case  $R = 0$ , the parameter  $\alpha_1$  has no influence on the model and hence can be chosen arbitrarily, which, in turn, makes  $\alpha_1$  unidentifiable during the parameter estimation process. The same issue occurs for  $\beta_2$  in the case  $R = N - 1$ . To circumvent these problems, we set  $r_1 = r_2$  for the threshold values  $R = 0, N - 1$ , i.e., we use the LSET model as introduced in Section 2.2 below. In all remaining cases, the parameters are identifiable as long as there is a sufficient number of different observations in each regime.

Since the SET-BAR(1) model falls within the class of density-dependent binomial AR(1) (DD-BAR(1)) models as introduced by Weiß and Pollett (2014)<sup>1</sup>, it follows by expression (1) in Weiß and Pollett (2014) that the transition probabilities  $p_{k|l} := P(X_t = k | X_{t-1} = l)$  of the SET-BAR(1) process take the form

$$p_{k|l} = \sum_{m=\max\{0, k+l-N\}}^{\min\{k,l\}} \binom{l}{m} \binom{N-l}{k-m} \phi_t^m (1 - \phi_t)^{l-m} \eta_t^{k-m} (1 - \eta_t)^{N-l+m-k} > 0. \quad (3)$$

Note that the  $(N+1) \times (N+1)$ -dimensional transition matrix  $\mathbf{P} := (p_{k|l})_{k,l=0,\dots,N}$  is primitive so that the process is ergodic with uniquely determined stationary marginal distribution  $\mathbf{p}$ . Since it is hardly possible to obtain a closed-form expression for the stationary marginal distribution  $\mathbf{p}$ , we determine it numerically by solving the eigenvalue problem  $\mathbf{P}\mathbf{p} = \mathbf{p}$ .

Next, we derive marginal conditional moments. From expression (2) in Weiß and Pollett (2014), we obtain

$$E[X_t | X_{t-1}] = I_{t-1} (r_1 X_{t-1} + (1 - r_1)\pi_1 N) + (1 - I_{t-1}) (r_2 X_{t-1} + (1 - r_2)\pi_2 N), \quad (4)$$

$$\begin{aligned} V[X_t | X_{t-1}] &= I_{t-1} \left( r_1(1 - r_1)(1 - 2\pi_1)X_{t-1} + N(1 - r_1)\pi_1(1 - (1 - r_1)\pi_1) \right) \\ &\quad + (1 - I_{t-1}) \left( r_2(1 - r_2)(1 - 2\pi_2)X_{t-1} + N(1 - r_2)\pi_2(1 - (1 - r_2)\pi_2) \right). \end{aligned} \quad (5)$$

Now we are prepared to obtain the unconditional mean and the variance of the stationary process. For simplicity in notation, we define  $p := P(X_t \leq R) = E[I_{t-1}]$ ,  $\mu_X := E[X_t]$ ,  $\sigma_X^2 := V[X_t]$ , and the partial moments  $\mu_{IX} := E[I_{t-1}X_{t-1}]$ ,  $\mu_{IX,2} := E[I_{t-1}X_{t-1}^2]$ . Then,

---

<sup>1</sup>The density-dependent models by Weiß and Pollett (2014) might also be understood as special SET models with  $N + 1$  regimes.

unconditional mean and variance are given by

$$\mu_X = \frac{r_1 - r_2}{1 - r_2} \mu_{IX} + N \left( p \pi_1 \frac{1 - r_1}{1 - r_2} + (1 - p) \pi_2 \right), \quad (6)$$

$$\begin{aligned} (1 - r_2^2) \sigma_X^2 &= r_2(1 - r_2)(1 - 2\pi_2) \mu_X - 2Npr_2((1 - r_1)\pi_1 - (1 - r_2)\pi_2) \mu_X \\ &\quad - 2r_2(r_1 - r_2) \mu_X \mu_{IX} + (r_1^2 - r_2^2) \mu_{IX,2} - (r_1 - r_2)^2 \mu_{IX}^2 \\ &\quad + 2N(r_1 - p(r_1 - r_2))((1 - r_1)\pi_1 - (1 - r_2)\pi_2) \mu_{IX} \\ &\quad + (r_1(1 - r_1)(1 - 2\pi_1) - r_2(1 - r_2)(1 - 2\pi_2)) \mu_{IX} \\ &\quad + Np(1 - r_1)\pi_1(1 - (1 - r_1)\pi_1) + N(1 - p)(1 - r_2)\pi_2(1 - (1 - r_2)\pi_2) \\ &\quad + N^2p(1 - p) \left( (1 - r_1)\pi_1 - (1 - r_2)\pi_2 \right)^2. \end{aligned} \quad (7)$$

The proof of (6) and (7) can be found in Appendix A.1. Keep in mind that  $p$  strongly depends on  $\pi_1$  and  $\pi_2$ .

## 2.2 The LSET Binomial AR(1) Model

Looking at Figure 1, it becomes clear that the level of the measles time series is shifted in the first half of 2005, while there is no obvious change in the serial dependence structure. This motivates to consider the model in Definition 2.1.1 but with the additional restriction  $r_1 = r_2 =: r$  with  $r \in \left( \max \left\{ -\frac{\pi_1}{1 - \pi_1}, -\frac{\pi_2}{1 - \pi_2}, -\frac{1 - \pi_1}{\pi_1}, -\frac{1 - \pi_2}{\pi_2} \right\}; 1 \right)$ . Notice that we will not have to consider the complicated restriction on the left-hand side of the interval for  $r$  if we only use positive values for the dependence parameter  $r$ .

The restriction  $r_1 = r_2$  is attractive to keep the number of model parameters low. It implies that  $\alpha_1 - \beta_1 = \alpha_2 - \beta_2$  and that  $\beta_2 - \beta_1 = (\pi_2 - \pi_1)(1 - r)$ . Since only the level of the process is shifted, we will refer to this model as the level SET-BAR(1) model, abbreviated as LSET-BAR(1) model.

**2.2.1 Definition** *A SET-BAR(1) process for which  $r_1 = r_2 =: r \neq 0$  holds is called an LSET-BAR(1) process.*

Although in this case, the transition probabilities in (3) do not change much, we get more simple expressions for the conditional moments (4) and (5):

$$E[X_t | X_{t-1}] = rX_{t-1} + N(1 - r)(I_{t-1}\pi_1 + (1 - I_{t-1})\pi_2), \quad (8)$$

$$\begin{aligned} V[X_t | X_{t-1}] &= I_{t-1} \left( r(1 - r)(1 - 2\pi_1)X_{t-1} + N(1 - r)\pi_1(1 - (1 - r)\pi_1) \right) \\ &\quad + (1 - I_{t-1}) \left( r(1 - r)(1 - 2\pi_2)X_{t-1} + N(1 - r)\pi_2(1 - (1 - r)\pi_2) \right). \end{aligned} \quad (9)$$

In particular, unconditional mean (6) and variance (7) can be simplified a lot for a stationary LSET-BAR(1) process ( $X_t$ ); see Appendix A.1:

$$\mu_X = Np\pi_1 + N(1-p)\pi_2, \quad (10)$$

$$\begin{aligned} \sigma_X^2 &= Np\pi_1(1-\pi_1) + N(1-p)\pi_2(1-\pi_2) + N^2p(1-p)(\pi_2-\pi_1)^2 \\ &\quad + \frac{2r}{1+r}(N-1)(\pi_2-\pi_1)(Np\pi_1 - \mu_{IX}). \end{aligned} \quad (11)$$

In view of the practical relevance of the parsimonious LSET model, additional stochastic properties are derived in Section 2.4 below.

### 2.3 The LSET<sup>0</sup> Binomial AR(1) Model

Relations (8), (9) highlight that the case  $r = 0$  has to be treated separately. In contrast to the usual BAR(1) model, where  $r = 0$  corresponds to serial independence, an LSET-BAR(1) model with  $r = 0$  still exhibits dependence on  $X_{t-1}$ , but only through the indicator function  $I_{t-1}$ , whereas the concrete value of  $X_{t-1}$  is without influence.

**2.3.1 Definition** *A SET-BAR(1) process with  $r_1 = r_2 = 0$  is said to be an LSET<sup>0</sup>-BAR(1) process.*

Notice that this model has only two parameters, namely  $\pi_1 = \alpha_1 = \beta_1$  and  $\pi_2 = \alpha_2 = \beta_2$ . So depending on whether  $X_{t-1} \leq R$  or  $X_{t-1} > R$ , the next count  $X_t$  is generated either from  $B(N, \pi_1)$  or  $B(N, \pi_2)$ , respectively. If, for instance,  $R = 0$ , the LSET<sup>0</sup> model allows for a simple way of causing zero-inflation or zero-deflation.

Conditional mean and variance follow from (8) and (9) as

$$E[X_t|X_{t-1}] = N(I_{t-1}\pi_1 + (1-I_{t-1})\pi_2), \quad (12)$$

$$V[X_t|X_{t-1}] = I_{t-1}(N\pi_1(1-\pi_1)) + (1-I_{t-1})(N\pi_2(1-\pi_2)). \quad (13)$$

The unconditional mean (10) remains as before, but the unconditional variance (11) simplifies to

$$\sigma_X^2 = Np\pi_1(1-\pi_1) + N(1-p)\pi_2(1-\pi_2) + N^2p(1-p)(\pi_2-\pi_1)^2. \quad (14)$$

We conclude our discussion by pointing out the analogy of the LSET<sup>0</sup> model to the ‘‘piecewise constant AR model’’ in Example 4.3 in Tong (2011) as well as to the ‘‘martingale difference model’’ (Tong, 2011, Example 4.4) (also see the more general threshold model for conditional heteroscedasticity (‘‘T-CHARM’’) in Chan et al (2014)). But while the latter models only change either the conditional mean or the conditional variance, the LSET<sup>0</sup> model changes conditional mean and conditional variance simultaneously due to the conditional binomial distribution.

### 2.4 Further Properties of the LSET-BAR(1) Model

Let us look back to the LSET-BAR(1) model according to Definition 2.2.1. The binomial index of dispersion,  $BID$ , is a useful metric when quantifying the dispersion behaviour of

count data random variables with a finite range  $\{0, \dots, N\}$ . It is defined as

$$BID \equiv BID(N, \mu, \sigma^2) = \frac{N\sigma^2}{\mu(N - \mu)} = \frac{\sigma^2}{\mu(1 - \frac{\mu}{N})} > 0. \quad (15)$$

For the binomial distribution, it holds that  $BID = 1$ . A distribution with finite range is said to have overdispersion if  $BID > 1$  (also extra-binomial variation), it is equidispersed if  $BID = 1$ , and it is underdispersed if  $BID < 1$ , each with respect to the binomial distribution.

For the LSET-BAR(1) model, the  $BID$  follows from (10) and (11) as

$$BID = 1 + \frac{N(N-1)p(1-p)(\pi_2 - \pi_1)^2 + \frac{2r}{1+r}(N-1)(\pi_2 - \pi_1)(Np\pi_1 - \mu_{IX})}{Np\pi_1(1-\pi_1) + N(1-p)\pi_2(1-\pi_2) + Np(1-p)(\pi_2 - \pi_1)^2}, \quad (16)$$

see Appendix A.2. Note that for  $r \in (0; 1)$ , underdispersion is only possible for  $\pi_2 < \pi_1$ . In contrast, for  $r \in (\max\{-\frac{\pi_1}{1-\pi_1}, -\frac{\pi_2}{1-\pi_2}, -\frac{1-\pi_1}{\pi_1}, -\frac{1-\pi_2}{\pi_2}\}; 0)$ , underdispersion is only possible for  $\pi_2 > \pi_1$ . For  $r = 0$  (LSET<sup>0</sup> model, see Definition 2.3.1), the model always shows overdispersion provided that  $\pi_1 \neq \pi_2$ .

Solving now the equation  $BID = 1$  in order to  $x := \pi_2 - \pi_1$ , it follows that

$$\begin{aligned} 0 &= N(N-1)p(1-p) \cdot x^2 + \frac{2r}{1+r}(N-1)(Np\pi_1 - \mu_{IX}) \cdot x \\ \Leftrightarrow x = 0 \quad (\Leftrightarrow \pi_1 = \pi_2) \quad \text{or} \quad x &= -\frac{\frac{2r}{1+r}(Np\pi_1 - \mu_{IX})}{Np(1-p)}. \end{aligned}$$

This result implies that for parameter values lying between these two roots, the process is underdispersed. In a nutshell, the LSET model is able to show over- and underdispersion for appropriate parameter settings.

Like for the continuous threshold models, we observed that the probability mass function (PMF) of the LSET-BAR(1) process may have multiple modes (see Figure 2 as an example). In analogy to Tong and Lim (1980), we define the skeleton of the LSET-BAR(1) model as

$$x_t = rx_{t-1} + \begin{cases} (1-r)\pi_1 N & \text{if } x_{t-1} \leq R \\ (1-r)\pi_2 N & \text{if } x_{t-1} > R \end{cases}. \quad (17)$$

We studied the limit cycles of the skeleton (Tong and Lim, 1980) aiming to find relations between the location of the modes and the elements in the limit cycle. However, it was not possible to establish a unique relation between the modes and limit cycles.

Finally, we investigate the ACF  $\rho(k)$  for the LSET-BAR(1) model. In Appendix A.3, it is proved that

$$\rho(k) = r^k + \frac{N(1-r)(\pi_1 - \pi_2)}{\sigma_X^2} \cdot \sum_{s=1}^k r^{s-1} Cov[I_{t-s}, X_{t-k}], \quad (18)$$

where  $\sigma_X^2 = V[X_t]$  and  $I_{t-s} = \mathbb{1}_{\{X_{t-s} \leq R\}}$ . The ACF in (18) reduces to the well-known AR(1)-like ACF of the BAR(1) model for  $\pi_1 = \pi_2$ . Otherwise, the ACF differs from the function

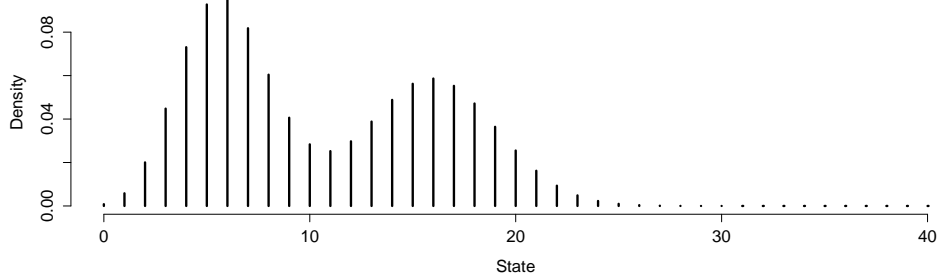


Figure 2: PMF of LSET-BAR(1) model with  $N = 40$ ,  $r = 0.3$ ,  $\pi_1 = 0.15$ ,  $\pi_2 = 0.4$ , and with threshold  $R = 10$ . For this model, we have  $p \approx 0.54$ ,  $\mu_{IX} \approx 3.21$ ,  $\mu_X \approx 10.56$  and  $BID \approx 4.14$ .

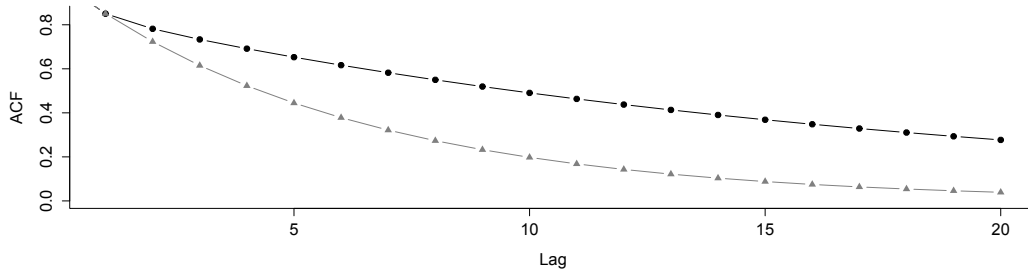


Figure 3: ACF of LSET-BAR(1) model from Figure 2, i.e., with  $N = 40$ ,  $r = 0.3$ ,  $\pi_1 = 0.15$ ,  $\pi_2 = 0.4$  and  $R = 10$  (black dots). Gray triangles show  $f(k) := (\rho(1))^k$ .

$f(k) := (\rho(1))^k$  as it would be expected from an AR(1)-like model. The model in Figure 3, for instance, has longer memory than a corresponding AR(1)-like model.

For the LSET<sup>0</sup>-BAR(1) model, the ACF takes the form

$$\rho(k) \Big|_{r=0} = \frac{N(\pi_1 - \pi_2)}{\sigma_X^2} Cov[I_{t-1}, X_{t-k}]. \quad (19)$$

### 3 Parameter Estimation and Model Specification for SET-BAR(1) Models

In this section, we consider first parameter estimation in the SET-BAR(1) model using conditional least squares (CLS) and conditional maximum likelihood (CML) methods. Later, we will take a look at a likelihood ratio test in view of uncovering nonlinearities.



### 3.1 Conditional Least Squares and Maximum Likelihood Estimation

Let  $X_0, X_1, \dots, X_T$  be the available segment from the SET-BAR(1) process with vector of unknown parameters  $\boldsymbol{\theta} := (r_1, r_2, \pi_1, \pi_2)$  taking values in

$$\Theta := \left( \max \left\{ -\frac{\pi_1}{1-\pi_1}, -\frac{1-\pi_1}{\pi_1}, \right\}; 1 \right) \times \left( \max \left\{ -\frac{\pi_2}{1-\pi_2}, -\frac{1-\pi_2}{\pi_2}, \right\}; 1 \right) \times (0; 1)^2.$$

For now, we assume a *fixed* threshold  $R$  (later we drop this assumption). Note that the alternative parametrization of the full SET model, with parameter vector  $\boldsymbol{\theta} := (\alpha_1, \alpha_2, \beta_1, \beta_2)$ , would have the practical advantage that  $(\alpha_1, \alpha_2, \beta_1, \beta_2)$  has to satisfy the box constraint  $(0; 1)^4$ .

The CLS estimators  $\hat{\boldsymbol{\theta}}_{CLS}$  are obtained by numerically minimizing the sum of the squared deviations

$$Q(\boldsymbol{\theta}) := \sum_{t=1}^T (X_t - g(\boldsymbol{\theta}, X_{t-1}))^2 \equiv \sum_{t=1}^T U_t^2 \quad (20)$$

with respect to  $\boldsymbol{\theta}$ , where  $g(\boldsymbol{\theta}, X_{t-1}) := E[X_t | X_{t-1}]$  is given by (4). Since the SET-BAR(1) model is stationary and ergodic, it follows from Theorems 3.1 and 3.2 in Klimko and Nelson (1978) that  $\hat{\boldsymbol{\theta}}_{CLS}$  is a consistent and asymptotically normal estimator of  $\boldsymbol{\theta}$ , i.e.,

$$\sqrt{T} (\hat{\boldsymbol{\theta}}_{CLS} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}^{-1} \mathbf{W} \mathbf{V}^{-1}), \quad (21)$$

where  $\mathbf{V}$  and  $\mathbf{W}$  are  $4 \times 4$  squared matrices with the  $ij$ th element given by

$$V_{ij} = E \left[ \frac{\partial}{\partial \theta_i} g(\boldsymbol{\theta}, X_{t-1}) \frac{\partial}{\partial \theta_j} g(\boldsymbol{\theta}, X_{t-1}) \right], \quad W_{ij} = E \left[ U_t^2 \frac{\partial}{\partial \theta_i} g(\boldsymbol{\theta}, X_{t-1}) \frac{\partial}{\partial \theta_j} g(\boldsymbol{\theta}, X_{t-1}) \right]. \quad (22)$$

Similar arguments apply to the LSET models with their reduced number of parameters.

Consider now the conditional maximum likelihood (CML) method to estimate the unknown model parameters  $\boldsymbol{\theta}$ . The CML estimators are obtained maximizing the conditional log-likelihood function

$$\ell(\boldsymbol{\theta}) := \log L(\boldsymbol{\theta}; x_0) \equiv \sum_{t=1}^T \ln P_{\boldsymbol{\theta}}(X_t = x_t | X_{t-1} = x_{t-1}), \quad (23)$$

with the transition probabilities defined in (3), i.e., they solve the following maximization problem:

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}). \quad (24)$$

Note that no closed-form expressions for the estimates can be found, so numerical procedures have to be employed. In order to prove the existence and consistency of the CML estimators, it is sufficient to show that Condition 5.1 of Billingsley (1961) holds. If Condition 5.1 holds, then Theorems 2.1 and 2.2 of Billingsley (1961) guarantee that there exists a consistent CML estimator being asymptotically normally distributed,

$$\sqrt{T} (\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_1^{-1}(\boldsymbol{\theta})), \quad (25)$$

where  $\mathbf{I}_1(\boldsymbol{\theta})$  denotes the expected Fisher information. Condition 5.1 of Billingsley (1961) is fulfilled provided that

1. the set  $D$  of  $(k, l)$  such that  $p_{kl}(\boldsymbol{\theta}) > 0$  is independent of  $\boldsymbol{\theta}$ ;

2. each  $p_{k|l}(\boldsymbol{\theta})$  has continuous partial derivatives of third order throughout  $\Theta$ ;
3. the  $d \times w$  matrix

$$\left( \frac{\partial p_{k|l}(\boldsymbol{\theta})}{\partial \theta_u} \right)_{(k,l) \in D, u=1, \dots, w}$$

has rank  $w$  throughout  $\Theta$ , where  $d := |D|$  and  $w := \dim(\Theta)$ ;

4. for each  $\boldsymbol{\theta} \in \Theta$ , there is only one ergodic set and there are no transient states.

Conditions 1 and 4 are fulfilled since all  $p_{k|l} > 0$  as stated earlier, while condition 2 holds due to the polynomial structure of the  $p_{k|l}$ . The third condition is also fulfilled if we exclude trivial cases such as  $\pi_1 = \pi_2$  (BAR(1) model) or  $r = 0$  (LSET<sup>0</sup>-BAR(1) model). Note that CML estimation for the usual BAR(1) model was already investigated by Weiß and Kim (2013). For the LSET<sup>0</sup> model, the parameter vector reduces to  $\boldsymbol{\theta} = (\pi_1, \pi_2) \in (0; 1)^2$ , and we have  $w = 2$  in this case.

For the numerical maximization of the log-likelihood (23), we use the R function `optim` with the expected Fisher information  $\mathbf{I}$  in (25) being approximated by the negative Hessian of the log-likelihood at the maximum (observed Fisher information). The initial estimates required by such numerical procedures are obtained by the CLS approach.

Next, we turn to the estimation of the threshold parameter. Note that  $R$  is a discrete-valued parameter in our case. Hence, it cannot be directly included in the parameter set  $\Theta$  of the above CLS- or CML-approaches; this is in contrast to the case of continuous-valued SET models, see, e.g., Chan et al (2014). So in order to estimate  $R$ , we considered the same approach as Wang et al (2014). Both CLS and CML estimation are performed for different values of  $R$  in a reasonable set which depends on the application. Finally, we choose the threshold value  $R$  as the value that delivers the lowest  $Q(\hat{\boldsymbol{\theta}})$  or the lowest  $-\ell_{\max}$ , respectively.

To illustrate the estimation procedures and small sample properties of the estimators, we focus on the LSET-BAR(1) models with parameters

$$\boldsymbol{\theta} = (r, \pi_1, \pi_2) \in \Theta := \left( \max \left\{ -\frac{\pi_1}{1-\pi_1}, -\frac{\pi_2}{1-\pi_2}, -\frac{1-\pi_1}{\pi_1}, -\frac{1-\pi_2}{\pi_2} \right\}; 1 \right) \times (0; 1)^2$$

and  $R$  as given in Figure 4 (a) (note that model M1 is the one from Figure 2). 1000 independent replicates of time series of length 100, 500 and 1000 are generated from each model, and the parameters are estimated by CLS and CML. Regarding the set of values for the threshold parameter  $R$ , we consider 5 values centered in the true value of  $R$ .

The results are summarized in Tables B.1–B.4 in Appendix B, while Fig. 4(b)–(d) displays boxplots of the biases for  $\hat{\boldsymbol{\theta}}$ . The tables report for each model the (mean) estimates and corresponding standard errors and, for each value of  $R$ , the percentage of series for which  $R$  leads to the minimum  $Q(\hat{\boldsymbol{\theta}})$  in the CLS or the lowest  $-\ell_{\max}$  in the CML estimation. First, note that the strategy to estimate  $R$  allows choosing the correct value most of the times already for small sample sizes, and this hit rate quickly approaches 100% for increasing  $T$ . This is similar to the case of the SETPAR models studied by Wang et al (2014), where a formal proof of the consistency of the estimation approach for  $R$  is given. Tables B.1–B.4 also illustrate the unbiasedness and consistency of the estimators  $\hat{\boldsymbol{\theta}}$ , since the bias and standard errors decrease

	Model	$N$	$r$	$\pi_1$	$\pi_2$	$R$
	M1	40	0.3	0.15	0.4	10
(a)	M2	20	0.3	0.15	0.4	4
	M3	40	0.7	0.15	0.4	10
	M4	20	0.7	0.15	0.4	5

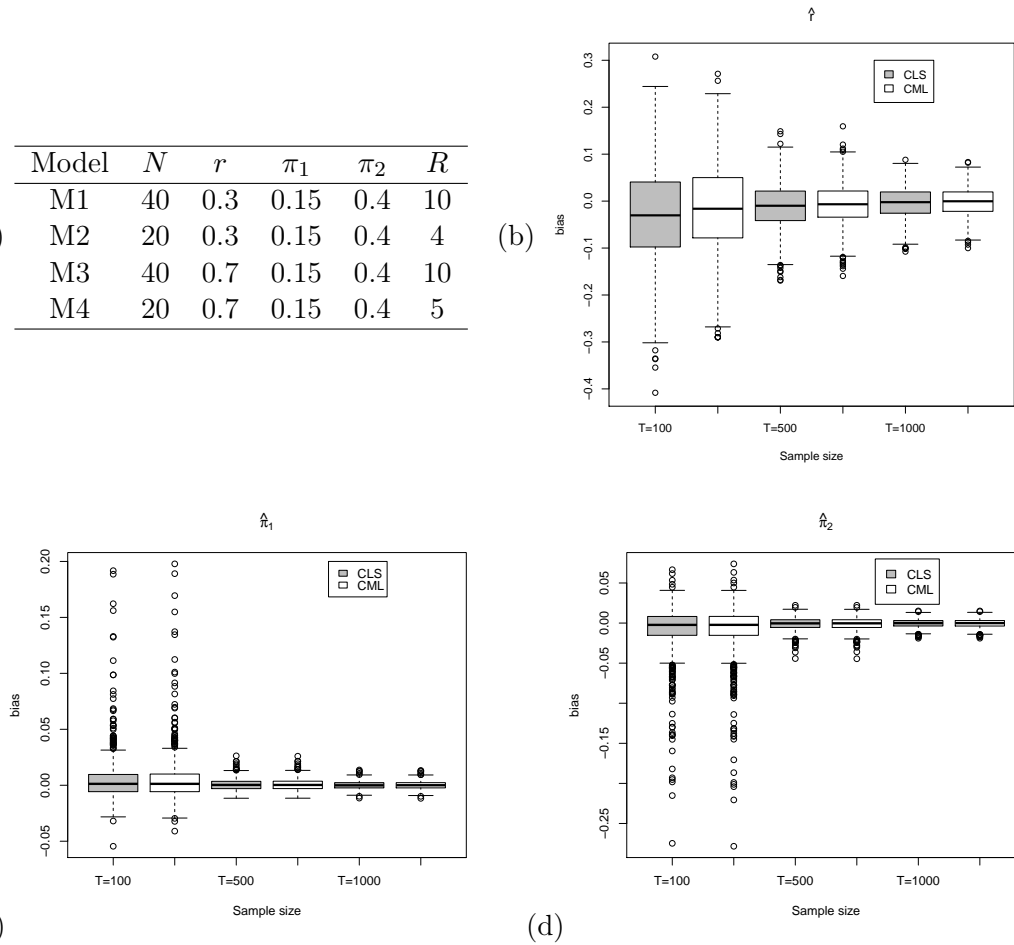


Figure 4: (a): Set of parameters for LSET-BAR(1) model.

(b)–(d): Boxplots of estimates for  $r$ ,  $\pi_1$  and  $\pi_2$  for model M1, with threshold set to its true value,  $R = 10$ .

to zero as  $T$  increases. Furthermore, Figure 4 illustrates the small sample properties of the estimators: the componentwise estimates tend to be unbiased and consistent. The results furthermore indicate that the CLS estimates present larger biases than CML when obtained under an incorrect value of the threshold.

### 3.2 Likelihood Ratio Test

Let us now tackle the issue of testing for non-linearity in the data. Petrucci (1990) investigated the performance of different tests for SETAR-type non-linearity and concluded that the likelihood ratio (LR) test was one of the best performers. Hence, we consider an LR test in the sequel, with the null hypothesis of a BAR(1) model. In order to prove the applicability of the test, we consider again the results in Billingsley (1961). Let  $\Phi$  be the parameter set of the BAR(1) model and  $\Theta$  be the one of the SET-BAR(1) model with specified threshold value  $R$ . We can define  $h : \Phi \rightarrow \Theta$  as a mapping from  $\Phi$  into  $\Theta$  in such a way that  $r_1 = r_2 = r$  and  $\pi_1 = \pi_2 = \pi$ . This mapping satisfies Condition 3.1 of Billingsley (1961). Together with

Condition 5.1, which is fulfilled as already shown before, Theorem 5.2 of Billingsley (1961) is applicable: if  $\theta^0 = h(\phi^0)$  is the true parameter value, then

$$2 \left( \max_{\Theta} \ell - \max_{\Phi} \ell \right) \xrightarrow{d} \chi_{w-c}^2, \quad (26)$$

where  $w := \dim(\Theta)$  and  $c := \dim(\Phi)$ . For the set model,  $w = 4$  and  $c = 2$ , so the LR statistic converges to a  $\chi_2^2$  distribution. If we choose the LSET model instead, we have a convergence to a  $\chi_1^2$  distribution. The finite-sample performance of the LR test is briefly considered in Section 5 below.

## 4 Forecasting for SET-BAR(1) Models

To forecast a SET-BAR(1) process, we use the forecasting distributions over all horizons  $h \in \mathbb{N}$ , i.e., the probabilistic distribution of  $X_{T+h}$  based on the observed time series up to time  $T$  (“more-than-one-step-ahead predictive distributions”, see Tong (2011)). This approach leads to forecasts being themselves counts, and therefore being coherent with the sample space. It also allows the quantification of the uncertainty associated with the future counts, which is important in a context of risk analysis. Point forecasts, if needed, are easily obtained from the median or the mode of the forecasting distribution. For the SET-BAR(1) process, the  $h$ -step-ahead conditional distribution of  $X_{T+h}$  given  $X_T$  is given by

$$P(X_{T+h} = x_{T+h} \mid X_T = x_T) = [\mathbf{P}^h]_{x_{T+h}, x_T}, \quad (27)$$

where  $\mathbf{P}$  denotes the transition matrix defined via (3). To prove (27), note that the conditional distribution of  $X_{T+h}$  given  $X_T$  satisfies the Chapman-Kolmogorov equations, since we are concerned with a homogeneous Markov chain.

As an illustration, Figures 5 and 6 represent the forecasting distributions for the horizons  $h = 1, 10, 25, 50$  steps-ahead for model M1, i.e.,  $(N, R; r, \pi_1, \pi_2) = (40, 10; 0.3, 0.15, 0.4)$ , conditioned on an observation in each of the two regimes,  $X_T = 2$  and  $X_T = 12$ , respectively. The figures show how, for growing  $h$ , these distributions converge to the stationary marginal distribution from Figure 2, as expected from the ergodicity of the process.

To assess the accuracy of the probabilistic forecasting in the case of estimated parameters, we use an approach suggested by Corradi and Swanson (2006), which measures accuracy using a distributional analog of mean squared error. Focussing on the one-step-ahead forecast distribution, i.e.,  $h = 1$ , we denote the conditional distribution  $P(X_{T+1} = i \mid X_T = j)$  by  $f_{i|j}$  with  $i, j \in 0, 1, \dots, N$ . Then, the mean squared error of the estimator  $\hat{f}_{\cdot|j}$  for the predictive distribution  $f_{\cdot|j}$  is the average over the support  $i \in \{0, 1, \dots, N\}$  of  $E \left[ (\hat{f}_{i|j} - f_{i|j})^2 \right]$  (note that the bias  $\frac{1}{N+1} \sum_{i=0}^N E(\hat{f}_{i|j} - f_{i|j})$  across the full support is equal to 0 since both the estimated and true forecast distribution sum up to 1). We computed the accuracy of the predictive distributions based on finite sample paths for the models M1 to M4 from Section 3.1. It turned out that the distributional ‘MSE’, as estimated from 1000 sample paths for each of the models, is negligible. To further assess the accuracy of the forecasting distribution, ‘Bias’ and ‘MSE’ of the median of the forecasting distributions  $\hat{f}_{\cdot|j}$  are averaged over all  $X_T = j$ , see Table 1. There, we also show the results for two particular cases concerning

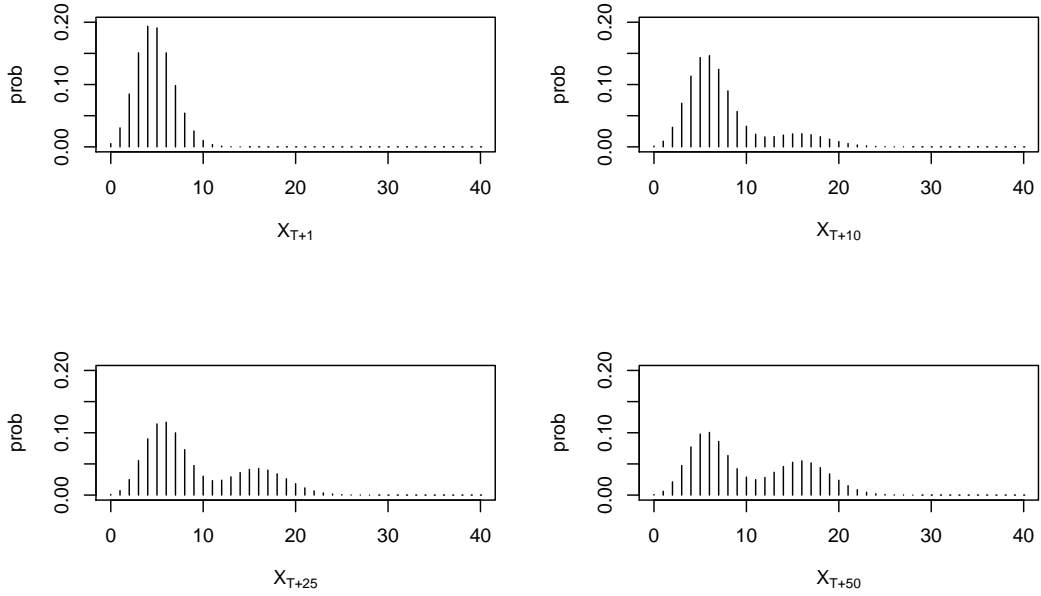


Figure 5: Forecasting distributions for model M1 conditional on  $X_T = 2$ .

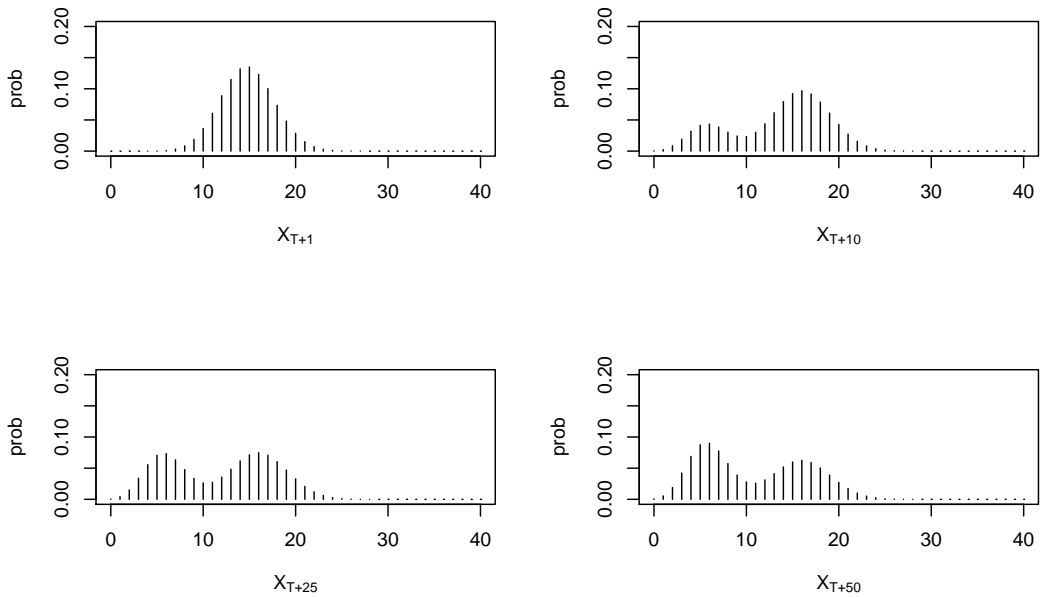


Figure 6: Forecasting distributions for model M1 conditional on  $X_T = 12$ .

the distributions conditioned on  $X_T = 2$  (observation in lower regime) and  $X_T = 12$  (upper regime), respectively. The results presented in Table 1 indicate that the estimated median is slightly negatively biased, but both the bias and the MSE generally decrease when the sample size increases.

		... over all conditional distributions			... conditional on					
					$X_T = 2$			$X_T = 12$		
$T$		100	500	1000	100	500	1000	100	500	1000
M1	Bias	-0.2448	-0.0645	-0.0235	-0.1198	-0.1680	-0.0960	-0.3925	-0.1690	-0.0940
	MSE	2.1327	0.4273	0.2493	0.5348	0.1700	0.0960	1.1754	0.1750	0.0940
M2	Bias	-0.0348	-0.0001	0.0117	-0.2457	-0.2690	-0.1760	0.0819	0.0580	0.0140
	MSE	0.6619	0.1942	0.1239	0.4681	0.2690	0.1760	0.3428	0.0620	0.0140
M3	Bias	-0.2711	-0.0556	-0.0327	-0.2592	-0.0060	0	-0.1704	-0.0120	-0.0040
	MSE	1.3139	0.1998	0.1346	0.4840	0.0060	0	1.0533	0.0260	0.0040
M4	Bias	-0.1667	-0.0401	-0.0297	-0.0866	0	0	-0.3454	-0.0380	-0.0040
	MSE	0.4939	0.1423	0.0994	0.0866	0	0	0.4814	0.0380	0.0040

Table 1: Estimated bias and MSEs for median forecasts in models M1–M4.

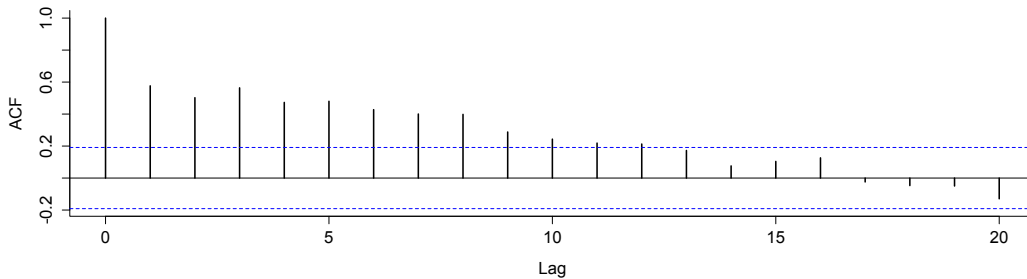


Figure 7: Plot of SACF for measles data (2004–2005).

## 5 Data Example: Measles in Germany in 2004–2005

We consider again the time series of measles infections in Germany in 2004 and 2005, see Section 1. We have  $T = 105$  data points and  $N = 38$  districts. A plot of the data was already shown in Figure 1. The observed data have an empirical mean of  $\bar{x} \approx 3.305$ , an empirical variance of  $s^2 \approx 6.137$  and a binomial index of dispersion of  $\widehat{BID} \approx 2.034$ . We have tested for overdispersion by using the approach of Weiß and Kim (2014) and found that the data are significantly overdispersed on a 5%-level. The plot of the sample ACF (SACF) in Figure 7 shows a slowly decaying extend of serial dependence.

In view of the level shift being visible in Figure 1, we start with fitting an LSET model to the data (later, we also consider the more general SET model and the more special LSET<sup>0</sup> model). We estimate the model parameters for threshold values of  $R \in \{3, \dots, 7\}$ , which is a reasonable range when we take a look at the plot of the data in Figure 1. The estimates given different threshold values  $R$  are compared in Table 2. Looking at  $-\ell_{\max}$ , we decide to consider a model with a threshold value of  $R = 5$ , also see Figure 1. The initial values for the CML estimation procedure were obtained from the CLS estimates computed for threshold values of  $R \in \{3, \dots, 7\}$ . The threshold value which minimizes  $Q(\boldsymbol{\theta})$  from equation (20) is also  $R = 5$ , and the corresponding CLS estimates are  $\hat{r} = 0.28$ ,  $\hat{\pi}_1 = 0.06$  and  $\hat{\pi}_2 = 0.15$ .

$R$	$r$	$\pi_1$	$\pi_2$	$-\ell_{\max}$
3	0.2546 (0.0907)	0.0681 (0.0074)	0.1174 (0.0114)	217.62
4	0.2245 (0.0906)	0.0697 (0.0063)	0.1376 (0.0142)	215.05
5	0.1947 (0.0876)	0.0707 (0.0057)	0.1604 (0.0169)	212.28
6	0.3083 (0.0744)	0.0787 (0.0065)	0.1482 (0.0229)	218.24
7	0.3381 (0.0679)	0.0813 (0.0066)	0.1594 (0.0299)	218.81

Table 2: Comparison of the LSET's CML estimates for different threshold values. Standard errors are given in parentheses.

	$T = 100$	$T = 500$	$T = 1000$
SET-BAR(1)	0.053	0.049	0.049
LSET-BAR(1)	0.045	0.043	0.048

Table 3: Simulated sizes concerning the critical values  $\chi_{2,0.95}^2$  (SET-BAR(1) model) and  $\chi_{1,0.95}^2$  (LSET-BAR(1) model), respectively.

We also applied the LR test of Section 3.2 to check if such a nonlinear model is appropriate for the data. For the LSET-BAR(1) model against the BAR(1) model by McKenzie (1985) (also see Table 4 below), we obtain a value about 19.6 for the LR test statistic, while our critical value on a 5%-level is given by  $\chi_{1,0.95}^2 = 3.841$ . So we have to reject the null hypothesis of a BAR(1) model. In order to verify the applicability of the LR test from Section 3.2 for this data example, we simulated  $n = 1000$  paths of the BAR(1) model with the estimated parameters from Table 4 for different time series length  $T = 100, 500, 1000$  (with  $T = 100$  being close to our data). We calculated the test statistic for the SET- and LSET-BAR(1) model and studied, among others, the size if using a critical value from the asymptotic  $\chi^2$ -distribution (again nominal level 5%). The results in Table 3 show a good agreement already for time series length  $T = 100$ . Furthermore, none of the simulated test statistic values reached the value of the test statistic from the measles data example, which also confirms the rejection of the null hypothesis.

Next, we compare different integer-valued models with a finite range. Besides the threshold models introduced in this paper as well as the simple BAR(1) model (also remember the above LR test), we consider the following models: the above-mentioned DD-BAR(1) model as well as its boundary case, the binomial INARCH(1) model (Weiß and Pollett, 2014, Section 4). Table 4 shows the comparison of the fitted models for the given data example. All estimates of the LSET-BAR(1) model are significant regarding the approximated standard errors, while for the full SET model,  $r_2$  is not significant. Similarly, for the full DD model,  $r$  is not significant, which implies to prefer its boundary case INARCH(1). Overall, however, the LSET-BAR(1) model has the lowest AIC and BIC of all models and is the model of choice considering these information criteria.

	Par. 1	Par. 2	Par. 3	Par. 4	AIC	BIC
BAR(1)	0.0882	0.4158	-	-	448.2	453.5
$(\pi, r)$	(0.0070)	(0.0550)				
DD-BAR(1)	0.0419	0.5270	0	-	436.9	444.9
$(a, b, r)$	(0.0095)	(0.1077)	(0.1765)			
Bin. INARCH(1)	0.0419	0.5270	-	-	434.9	440.2
$(a, b)$	(0.0060)	(0.0682)				
SET-BAR(1)	0.0706	0.1558	0.1916	0.2904	432.5	443.1
$(\pi_1, \pi_2, r_1, r_2)$	(0.0056)	(0.0269)	(0.0884)	(0.375)		
LSET-BAR(1)	0.0707	0.1604	0.1947	-	430.6	438.5
$(\pi_1, \pi_2, r)$	(0.0057)	(0.0169)	(0.0876)			
LSET <sup>0</sup> -BAR(1)	0.0689	0.1671	-	-	433.4	438.7
$(\pi_1, \pi_2)$	(0.0045)	(0.0135)				

Table 4: Comparison of estimated parameters (standard errors in parentheses) for different models for measles data. All threshold models include a threshold value of  $R = 5$ .

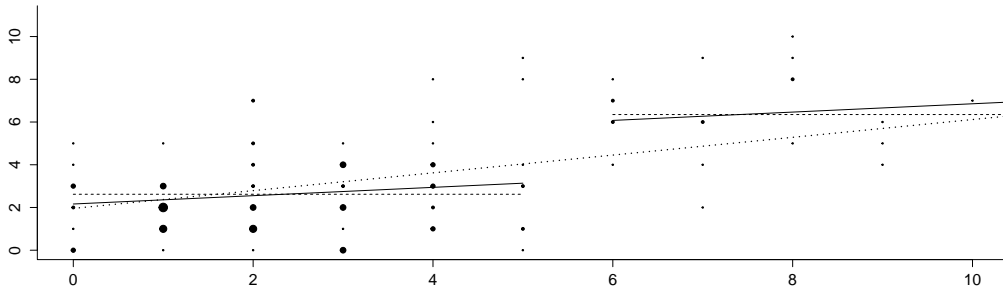


Figure 8: Dots represent observed frequencies for measles data with coordinates  $(x_{t-1}, x_t)$ . Lines represent conditional means for BAR(1) (dotted), LSET-BAR(1) (solid) and LSET<sup>0</sup>-BAR(1) (dashed).

The difference between the models becomes clear through Figure 8. The dotted straight line represents the conditional mean of the BAR(1) model against the previous observation  $x_{t-1}$ , which is a linear function. The dashed straight lines represent the conditional mean of the LSET<sup>0</sup>-BAR(1) model, which is piecewise constant in the regimes, see (12). The remaining solid straight lines refer to the LSET-BAR(1) model and show a piecewise linear behaviour with a unique slope in both regimes. The conditional means of the DD-BAR(1) and the binomial INARCH(1) model as well as that of the SET-BAR(1) model are not shown in Figure 8, since they are very similar to the ones of the BAR(1) or LSET-BAR(1) model, respectively. If comparing the plotted conditional means with the observed bivariate frequencies for  $(x_{t-1}, x_t)$ , the piecewise approaches seem more appropriate since the frequencies concentrate either in the lower left or upper right quarter.

An important application of the fitted LSET model would be forecasting. The forecast distri-



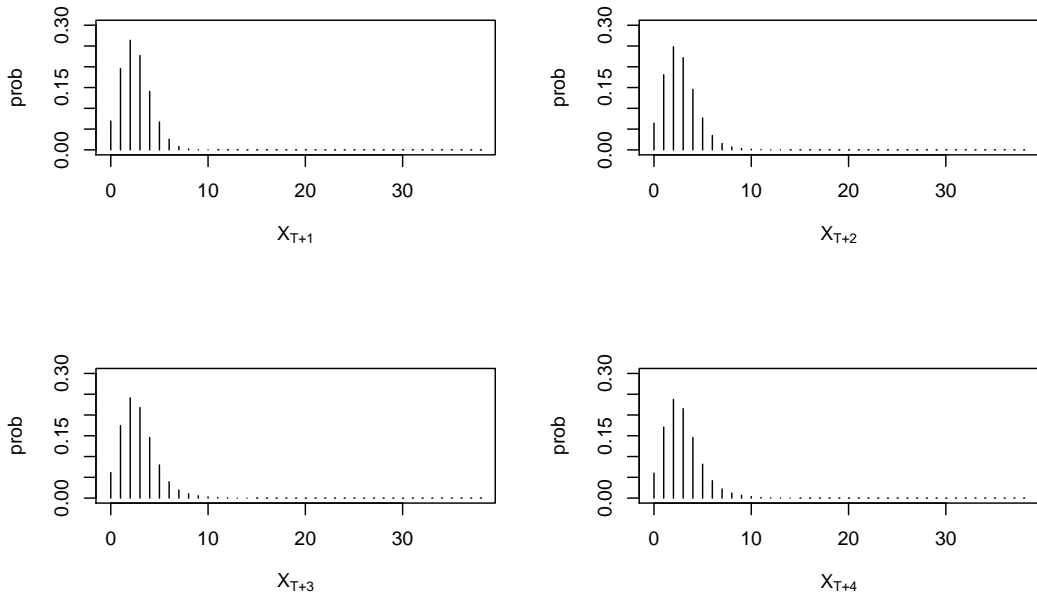


Figure 9: Forecast distribution for 1, 2, 3 and 4 weeks ahead, conditioned on  $x_{105} = 2$ , for the number of districts with new measles infections.

butions up to four weeks ahead, based on the fitted LSET model, are represented in Figure 9 and convey information that is important for public health monitoring. For example, we can say that, conditional on the last observed value  $x_{105} = 2$  (infections observed in two districts), we observe infections in  $\geq 6$  districts (upper regime) in the following week with a probability around 3.7%. With increasing forecast horizon, this conditional probability increases up to about 11.0%. We also applied the one-step-ahead forecast distributions to the measles data itself, see Figure 1, where the limits of the 95% prediction intervals are shown. Only two observations lie outside these intervals, which adapt nicely to the level shift in the data. For the BAR(1)-based forecasts instead (not shown), the 95% bands look clearly worse, which again confirms preferring the LSET model for our measles data. Generally, if looking at the coverage rates of the one-step-ahead prediction intervals for different levels, the BAR(1) model always performs worst, and the LSET<sup>0</sup> model is second worst, while all remaining models do comparably well in terms of this retrospective forecasting.

At this point, it is also interesting to look at a completely different approach towards modeling the piecewise behaviour. As already pointed out in Section 1, a parameter-driven alternative would be a (two-state) binomial HMM (Zucchini and MacDonald, 2009). Although the estimates for the HMM's transition matrix are not significant, there are well-interpretable analogies between the fitted HMM and the LSET-BAR(1) model. The estimated binomial parameters of the HMM are 0.0590 in the lower and 0.1740 in the upper regime, which are fairly close to the estimates for  $\pi_1, \pi_2$  in the LSET-BAR(1) model, the latter being 0.0707, 0.1604. The stationary distribution in the HMM gives us a probability of 0.8132 to be in the lower regime, while the LSET-BAR(1) model results in  $p = 0.8905$ . So both types of regime switching model lead to similar conclusions with respect to the measles data.

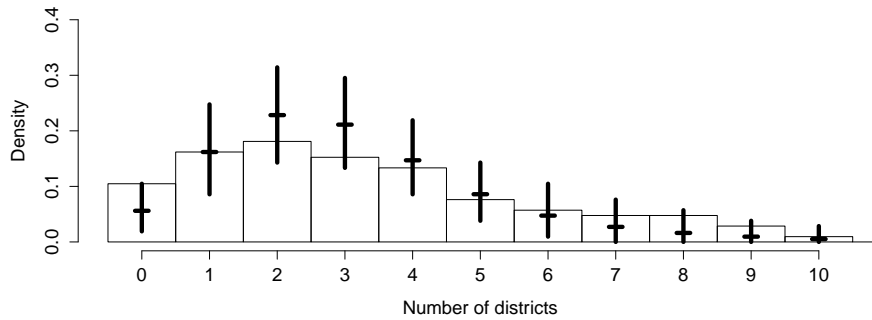


Figure 10: Comparison of histogram of measles data (bars) and marginal distribution of fitted LSET model (ticks). 95 % confidence bands from parametric bootstrap (black lines).

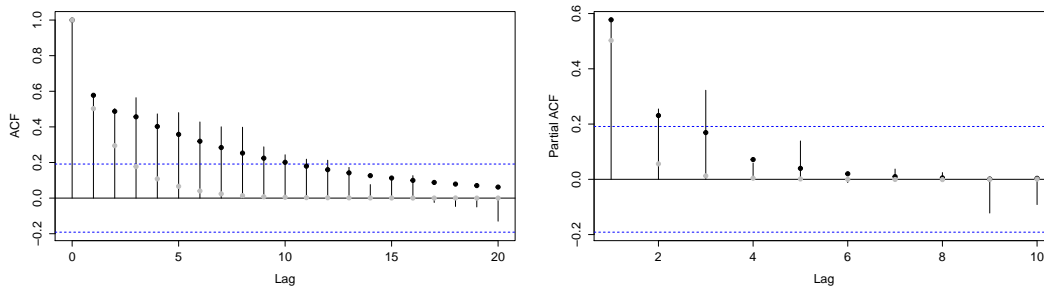


Figure 11: SACF and SPACF of measles data (bars) compared to estimated higher delay model (black dots) and to estimated LSET-BAR(1) model from Table 4 (gray dots).

Although the LSET model performs best among all candidate models, it does not perfectly describe the measles data. While the mean within the fitted LSET model is close to the empirically observed one (3.060 vs. 3.305), its *BID* is visibly smaller than the empirically observed one (1.440 vs. 2.034). So the fitted LSET model is not able to reflect the full extent of observed overdispersion, which also becomes clear by looking at Figure 10, where the LSET's marginal distribution (together with 95 % confidence bands from a parametric bootstrap) is plotted against a histogram of the data. Although confidence bands and histogram always overlap, the LSET's distribution is stronger centered around values from 1 to 4. Another weakness of the fitted LSET model is that its ACF does not fit the SACF of the measles data particularly well, see Figure 11. This issue motivates to discuss possible extensions of our basic SET models in the following section, where we shall see that a higher delay parameter solves the ACF problem.

## 6 Possible Extensions of the SET-BAR(1) Model

### 6.1 Multiple Regimes or Higher-Order Dependence

The SET-BAR(1) model according to Definition 2.1.1 is a basic first-order model with two regimes and with delay parameter 1. Depending on the concrete application, it can be extended in several ways. First, more than two regimes are easily possible, e.g., by dividing the range into  $k > 2$  partitions of the form  $\{0, \dots, R_1\}$ ,  $\{R_1 + 1, \dots, R_2\}$ ,  $\dots$ ,  $\{R_{k-1} + 1, \dots, N\}$  with  $0 \leq R_1 < \dots < R_{k-1} < N$ , with an individual set of parameters  $(\pi_1, r_1), \dots, (\pi_k, r_k)$  for each regime. But since the finite range will be quite small in many real data scenarios, the choice of only two regimes will often be reasonable. Secondly, our approaches could be extended to higher-order autoregressions by adapting the binomial AR( $p$ ) model of Weiß (2009), thus leading to a  $p$ th order Markov process (at the price of an increased number of parameters). Here, we shall discuss a third possibility in some more detail, an extension of the simple SET-BAR(1) model with a higher delay  $d > 1$  by considering  $\mathbb{1}_{\{X_{t-d} \leq R\}}$  as the condition for switching between the regimes, also see the analogous discussion in Tong (2011). This extension comes along with the definition of more complex threshold conditions, and it does not require additional parameters. Also here, the Markov property is preserved (now  $d$ th order Markov process), which is helpful for parameter estimation and forecasting.

For delays  $d > 1$ , one may simply shift the threshold condition in time, leading to  $X_{t-d} \leq R$ . But a larger delay can also be combined with some function on the last  $d$  observations, i.e., the threshold condition becomes  $f(X_{t-d}, \dots, X_{t-1}) \leq \tilde{R}$ . For example, the function  $f$  could be the mean of the past  $d$  observations, in analogy to Thyregod et al (1999). Of course, it may happen that  $f(X_{t-d}, \dots, X_{t-1})$  attains values outside the range  $\{0, \dots, N\}$ , so we have to choose the threshold value  $\tilde{R}$  in an appropriate way. If, as another example,  $f$  returns the minimum of  $X_{t-d}, \dots, X_{t-1}$ , then at least one of the observations has to be not larger than  $\tilde{R}$ .

For the measles data, we finally decided to discriminate the regimes in the following way: The process is in the lower regime if  $\max\{X_{t-3}, X_{t-2}, X_{t-1}\} \leq 5$ , and in the upper regime otherwise. Note that the max-condition is equivalent to requiring that *all* values  $X_{t-3}, X_{t-2}, X_{t-1} \leq 5$ . The results for the ACF and partial ACF (PACF) of this extended model, compared to the LSET-BAR(1) model, are shown in Figure 11. Obviously, we achieve a much better fit to both SACF and SPACF by this threshold condition, the new model shows a much longer memory. Furthermore, mean (3.49), variance (5.83) and BID (1.84) of this model better agree with the empirically observed values for the measles data.

### 6.2 A Bivariate Extension

The SET-BAR(1) process can also be generalized to the case of multivariate observations. A possible extension is to induce piecewise-type patterns to the class of bivariate binomial autoregressive models introduced by Scotto et al (2014), which are based on the bivariate binomial thinning operator “ $\otimes$ ”, defined via

$$(\alpha_1, \alpha_2, \varphi_\alpha) \otimes \mathbf{X} \mid \mathbf{X} \sim \text{BVB}_{\text{II}}(X_1, X_2, \min\{X_1, X_2\}; \alpha_1, \alpha_2, \varphi_\alpha). \quad (28)$$

The definition of the SET-BVB<sub>II</sub>-AR(1) is given below.

**6.2.1 Definition** Let  $\mathbf{N} := [N_1 \ N_2]' \in \mathbb{N}^2$  be the vector of upper limits for the bivariate range, let  $0 \leq R_j < N_j$  for  $j \in \{1, 2\}$  be the threshold values.

For  $\mathbf{i} \in \{0, 1\}^2$  and  $j \in \{1, 2\}$ , let  $\pi_j^{(\mathbf{i})} \in (0; 1)$ , and

$$r_j^{(\mathbf{i})} \in \left( \max \left\{ -\frac{\pi_j^{(\mathbf{i})}}{1 - \pi_j^{(\mathbf{i})}}, -\frac{1 - \pi_j^{(\mathbf{i})}}{\pi_j^{(\mathbf{i})}} \right\}; 1 \right).$$

In addition, define  $\beta_j^{(\mathbf{i})} := \pi_j^{(\mathbf{i})} \cdot (1 - r_j^{(\mathbf{i})}) \in (0; 1)$  and  $\alpha_j^{(\mathbf{i})} := \beta_j^{(\mathbf{i})} + r_j^{(\mathbf{i})} \in (0; 1)$ . Let  $\boldsymbol{\alpha}^{(\mathbf{i})} := (\alpha_1^{(\mathbf{i})}, \alpha_2^{(\mathbf{i})}, \varphi_\alpha^{(\mathbf{i})})$  and  $\boldsymbol{\beta}^{(\mathbf{i})} := (\beta_1^{(\mathbf{i})}, \beta_2^{(\mathbf{i})}, \varphi_\beta^{(\mathbf{i})})$ .

The process  $(\mathbf{X}_t)$  of bivariate random variables  $\mathbf{X}_t := [X_{t,1} \ X_{t,2}]'$  is called SET-BVB<sub>II</sub>-AR(1) if  $\mathbf{X}_t$  satisfies the recursion

$$\mathbf{X}_t = \boldsymbol{\phi}_t \otimes \mathbf{X}_{t-1} + \boldsymbol{\eta}_t \otimes (\mathbf{N} - \mathbf{X}_{t-1}) \quad \text{for } t \in \mathbb{Z}, \quad (29)$$

where  $\boldsymbol{\phi}_t := \boldsymbol{\alpha}^{(\mathbf{I}_{t-1})}$ ,  $\boldsymbol{\eta}_t := \boldsymbol{\beta}^{(\mathbf{I}_{t-1})}$ , and  $\mathbf{I}_{t-1} := [\mathbb{1}_{\{X_{t-1,1} > R_1\}} \ \mathbb{1}_{\{X_{t-1,2} > R_2\}}]'$ .

Furthermore, it is assumed that the thinnings are performed independently of each other.

So the bivariate indicator  $\mathbf{I}_{t-1}$  distinguishes between the following events:

$$\begin{aligned} \mathbf{I}_{t-1} = [0 \ 0]' & \quad \text{iff } \{X_{t-1,1} \leq R_1, X_{t-1,2} \leq R_2\}; \\ \mathbf{I}_{t-1} = [0 \ 1]' & \quad \text{iff } \{X_{t-1,1} \leq R_1, X_{t-1,2} > R_2\}; \\ \mathbf{I}_{t-1} = [1 \ 0]' & \quad \text{iff } \{X_{t-1,1} > R_1, X_{t-1,2} \leq R_2\}; \\ \mathbf{I}_{t-1} = [1 \ 1]' & \quad \text{iff } \{X_{t-1,1} > R_1, X_{t-1,2} > R_2\}. \end{aligned}$$

The transition probabilities at lag 1 of the SET-BVB<sub>II</sub>-AR(1) model are computed through the expression

$$\begin{aligned} p_{\mathbf{k}|\mathbf{l}} & := P(\mathbf{X}_t = \mathbf{k} \mid \mathbf{X}_{t-1} = \mathbf{l}) \equiv P(\boldsymbol{\phi}_t \otimes \mathbf{X}_{t-1} + \boldsymbol{\eta}_t \otimes (\mathbf{N} - \mathbf{X}_{t-1}) = \mathbf{k} \mid \mathbf{X}_{t-1} = \mathbf{l}) \\ & = \sum_{a_1=0}^{\min\{k_1, l_1\}} \sum_{a_2=0}^{\min\{k_2, l_2\}} p_{(l_1, l_2; \boldsymbol{\alpha}^{(\mathbf{i})})(a_1, a_2)} p_{(N_1 - l_1, N_2 - l_2; \boldsymbol{\beta}^{(\mathbf{i})})(k_1 - a_1, k_2 - a_2)}, \end{aligned}$$

where  $\mathbf{k} := [k_1 \ k_2]'$ ,  $\mathbf{l} := [l_1 \ l_2]'$  as well as  $\mathbf{i} := [\mathbb{1}_{\{l_1 > R_1\}} \ \mathbb{1}_{\{l_2 > R_2\}}]'$ , and where the bivariate probability mass functions  $p_{(\cdot)}(\cdot)$  defined as in equation (13) in Scotto et al (2014, p. 236). Note that since these transition probabilities are truly positive, the SET-BVB<sub>II</sub>-AR(1) process is a primitive and finite-state Markov chain, which, in turn, implies irreducibility and aperiodicity. Hence, a uniquely determined stationary marginal distribution exists. Denoting the transition matrix by  $\mathbf{P} := (p_{\mathbf{k}|\mathbf{l}})$ , the unique stationary marginal distribution, expressed as a vector  $\mathbf{p}$ , is obtained as the solution of the linear equation  $\mathbf{P}\mathbf{p} = \mathbf{p}$ .

Now we are prepared to obtain the mean, the variance and the autocovariance function of the process. For simplicity in notation, we define  $q_{\mathbf{i}} := P(\mathbf{I}_{t-1} = \mathbf{i})$  and  $u_{j,\mathbf{i}} := E[X_{t-1,j} \mid \mathbf{I}_{t-1} = \mathbf{i}]$ ,  $\sigma_{j,\mathbf{i}}^2 := V[X_{t-1,j} \mid \mathbf{I}_{t-1} = \mathbf{i}]$  for  $\mathbf{i} \in \{0, 1\}^2$  and  $j \in \{1, 2\}$ . The mean of the process is given by

$$E[\mathbf{X}_t] = \sum_{\mathbf{i}} q_{\mathbf{i}} \begin{bmatrix} \alpha_1^{(\mathbf{i})} u_{1,\mathbf{i}} + \beta_1^{(\mathbf{i})} \nu_{1,\mathbf{i}} \\ \alpha_2^{(\mathbf{i})} u_{2,\mathbf{i}} + \beta_2^{(\mathbf{i})} \nu_{2,\mathbf{i}} \end{bmatrix}.$$

In order to calculate the variance (componentwise), note first that

$$\begin{aligned} V[\mathbf{X}_t] &= V[\boldsymbol{\phi}_t \otimes \mathbf{X}_{t-1} + \boldsymbol{\eta}_t \otimes (\mathbf{N} - \mathbf{X}_{t-1})] \\ &= V[\boldsymbol{\phi}_t \otimes \mathbf{X}_{t-1}] + V[\boldsymbol{\eta}_t \otimes (\mathbf{N} - \mathbf{X}_{t-1})] + 2 \text{Cov}[\boldsymbol{\phi}_t \otimes \mathbf{X}_{t-1}, \boldsymbol{\eta}_t \otimes (\mathbf{N} - \mathbf{X}_{t-1})] \\ &=: \text{I} + \text{II} + \text{III}. \end{aligned}$$

The term I can be obtained through the expression

$$\text{I} = V[E(\boldsymbol{\phi}_t \otimes \mathbf{X}_{t-1} | \mathbf{X}_{t-1})] + E[V(\boldsymbol{\phi}_t \otimes \mathbf{X}_{t-1} | \mathbf{X}_{t-1})] = \sum_{\mathbf{i}} q_{\mathbf{i}} \begin{bmatrix} (\alpha_1^{(\mathbf{i})})^2 \sigma_{1,\mathbf{i}}^2 + \alpha_1^{(\mathbf{i})} (1 - \alpha_1^{(\mathbf{i})}) u_{1,\mathbf{i}} \\ (\alpha_2^{(\mathbf{i})})^2 \sigma_{2,\mathbf{i}}^2 + \alpha_2^{(\mathbf{i})} (1 - \alpha_2^{(\mathbf{i})}) u_{2,\mathbf{i}} \end{bmatrix}.$$

By similar arguments, it follows that

$$\text{II} = \sum_{\mathbf{i}} q_{\mathbf{i}} \begin{bmatrix} \beta_1^{(\mathbf{i})^2} \sigma_{1,\mathbf{i}}^2 + \beta_1^{(\mathbf{i})} (1 - \beta_1^{(\mathbf{i})}) (N_1 - u_{1,\mathbf{i}}) \\ \beta_2^{(\mathbf{i})^2} \sigma_{2,\mathbf{i}}^2 + \beta_2^{(\mathbf{i})} (1 - \beta_2^{(\mathbf{i})}) (N_2 - u_{2,\mathbf{i}}) \end{bmatrix}.$$

Finally, to obtain III, note that

$$\begin{aligned} \text{Cov}[(\boldsymbol{\phi}_t \otimes \mathbf{X}_{t-1})_j, (\boldsymbol{\eta}_t \otimes (\mathbf{N} - \mathbf{X}_{t-1}))_j] &= \sum_{\mathbf{i}} q_{\mathbf{i}} \text{Cov}[(\alpha^{(\mathbf{i})} \otimes \mathbf{X}_{t-1})_j, (\beta^{(\mathbf{i})} \otimes (\mathbf{N} - \mathbf{X}_{t-1}))_j | \mathbf{I}_{t-1} = \mathbf{i}] \\ &= - \sum_{\mathbf{i}} q_{\mathbf{i}} \alpha_j^{(\mathbf{i})} \beta_j^{(\mathbf{i})} \text{Cov}[X_{t-1,j}, X_{t-1,j} | \mathbf{I}_{t-1} = \mathbf{i}]. \end{aligned}$$

To calculate the cross-covariance function, we proceed as follows:

$$\begin{aligned} \text{Cov}[X_{t,1}, X_{t,2}] &= \text{Cov}[(\boldsymbol{\phi}_t \otimes \mathbf{X}_{t-1})_1, (\boldsymbol{\phi}_t \otimes \mathbf{X}_{t-1})_2] \\ &\quad + \text{Cov}[(\boldsymbol{\eta}_t \otimes (\mathbf{N} - \mathbf{X}_{t-1}))_1, (\boldsymbol{\eta}_t \otimes (\mathbf{N} - \mathbf{X}_{t-1}))_2] \\ &\quad + \text{Cov}[(\boldsymbol{\phi}_t \otimes \mathbf{X}_{t-1})_1, (\boldsymbol{\eta}_t \otimes (\mathbf{N} - \mathbf{X}_{t-1}))_2] \\ &\quad + \text{Cov}[(\boldsymbol{\eta}_t \otimes (\mathbf{N} - \mathbf{X}_{t-1}))_1, (\boldsymbol{\phi}_t \otimes \mathbf{X}_{t-1})_2] \quad =: \text{I} + \text{II} + \text{III} + \text{IV}. \end{aligned}$$

Straightforward (although tedious) algebraic calculations lead to

$$\begin{aligned} \text{I} + \text{II} &= \sum_{\mathbf{i}} q_{\mathbf{i}} \cdot \left[ (\alpha_1^{(\mathbf{i})} \alpha_2^{(\mathbf{i})} + \beta_1^{(\mathbf{i})} \beta_2^{(\mathbf{i})}) \cdot \text{Cov}[X_{t-1,1}, X_{t-1,2} | \mathbf{I}_{t-1} = \mathbf{i}] \right. \\ &\quad + \varphi_{\alpha}^{(\mathbf{i})} \cdot \sqrt{\alpha_1^{(\mathbf{i})} \alpha_2^{(\mathbf{i})} (1 - \alpha_1^{(\mathbf{i})}) (1 - \alpha_2^{(\mathbf{i})})} \cdot E[\min\{X_{t-1,1}, X_{t-1,2}\} | \mathbf{I}_{t-1} = \mathbf{i}] \\ &\quad \left. + \varphi_{\beta}^{(\mathbf{i})} \cdot \sqrt{\beta_1^{(\mathbf{i})} \beta_2^{(\mathbf{i})} (1 - \beta_1^{(\mathbf{i})}) (1 - \beta_2^{(\mathbf{i})})} \cdot E[\min\{N_1 - X_{t-1,1}, N_2 - X_{t-1,2}\} | \mathbf{I}_{t-1} = \mathbf{i}] \right]. \end{aligned}$$

Similarly,

$$\text{III} + \text{IV} = - \sum_{\mathbf{i}} q_{\mathbf{i}} (\alpha_1^{(\mathbf{i})} \beta_2^{(\mathbf{i})} + \alpha_2^{(\mathbf{i})} \beta_1^{(\mathbf{i})}) \cdot \text{Cov}[X_{t-1,1}, X_{t-1,2} | \mathbf{I}_{t-1} = \mathbf{i}],$$

and one concludes in analogy to Theorem 5.2 in Scotto et al (2014).

## 7 Concluding Remarks

In this paper, we proposed types of self-exciting threshold models for integer-valued time series with a finite range, which are based on the BAR(1) model by McKenzie (1985). We analyzed their marginal means and variances as well as further stochastic properties, and we considered the topic of forecasting as an application. For estimation purposes, we considered the conditional least squares and the maximum likelihood approach, and we investigated both their asymptotic and finite-sample behaviour. We successfully applied the novel self-exciting threshold models to a case study of measles infections in Germany. Finally, we exemplified the potential for further generalizing our models by proposing a model with a higher delay as well as a bivariate extension.

### Acknowledgement

The authors thank the referees for carefully reading the article and for their comments, which greatly improved the article. This work was supported by Portuguese funds through the CIDMA-Center for Research and Development in Mathematics and Applications, and the Portuguese Foundation for Science and Technology (FCT-Fundação para a Ciência e a Tecnologia), within project UID/MAT/04106/2013.

### References

- Billingsley P (1961) *Statistical Inference for Markov Processes*. Statistical research monographs, University of Chicago Press
- Chan KS, Tong H (1985) On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Advances in Applied Probability* 17:666–678
- Chan KS, Petrucci JD, Tong H, Woolford SW (1985) A multiple-threshold AR(1) model. *Journal of Applied Probability* 22:267–279
- Chan KS, Li D, Ling S, Tong H (2014) On conditionally heteroscedastic AR models with thresholds. *Statistica Sinica* 24(2):625–652
- Chan WS, Wong ACS, Tong H (2004) Some nonlinear threshold autoregressive time series models for actuarial use. *North American Actuarial Journal* 8:37–61, DOI 10.1080/10920277.2004.10596170, <http://dx.doi.org/10.1080/10920277.2004.10596170>
- Chen CWS, So MKP, Liu FC (2011) A review of threshold time series models in finance. *Statistics and Its Interface* 4:167–182
- Cline DBH, Pu HH (1999) Stability of nonlinear AR(1) time series with delay. *Stochastic Processes and their Applications* 82:307–333
- Cline DBH, Pu HH (2004) Stability and the Lyapounov exponent of threshold AR-ARCH models. *Annals of Applied Probability* 14:1920–1949
- Corradi V, Swanson NR (2006) Predictive density and conditional confidence interval accuracy tests. *Journal of Econometrics* 135(1 - 2):187 – 228

- Hansen BE (2011) Threshold autoregression in economics. *Statistics and its Interface* 4:123–127
- Klimko LA, Nelson PI (1978) On conditional least squares estimation for stochastic processes. *The Annals of Statistics* 6:629–642
- Lanne M, Saikkonen P (2005) Non-linear GARCH models for highly persistent volatility. *Econometrics Journal* 8:251–276
- Liebscher E (2005) Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *Journal of Time Series Analysis* 26:669–689
- McKenzie E (1985) Some simple models for discrete variate time series. *Water Resources Bulletin* 21:645–650, DOI 10.1111/j.1752-1688.1985.tb05379.x
- Möller T, Weiß CH (2015) Threshold models for integer-valued time series with infinite or finite range. In: Steland A, Rafajłowicz E, Szajowski K (eds) *Stochastic Models, Statistics and Their Applications*, Springer Proceedings in Mathematics & Statistics, vol 122, Springer International Publishing, pp 327–334, DOI 10.1007/978-3-319-13881-7\_36
- Monteiro M, Scotto MG, Pereira I (2012) Integer-valued self-exciting threshold autoregressive processes. *Communications in Statistics–Theory and Methods* 41:2717–2737
- Petrucelli JD (1990) A comparison of tests for setar-type non-linearity in time series. *Journal of Forecasting* 9(1):25–36, DOI 10.1002/for.3980090104, URL <http://dx.doi.org/10.1002/for.3980090104>
- Robert-Koch-Institut (2014) Robert-Koch-Institut: SurvStat@RKI. <http://www3.rki.de/SurvStat>, accessed: 2014-07-02
- Samia NI, Chan KS, Stenseth NC (2007) A generalized threshold mixed model for analyzing nonnormal nonlinear time series, with application to plague in Kazakhstan. *Biometrika* 94(1):101–118, DOI 10.1093/biomet/asm006
- Scotto M, Weiß CH, Silva ME, Pereira I (2014) Bivariate binomial autoregressive models. *Journal of Multivariate Analysis* 125:233–251
- Stenseth NC, Samia NI, Viljugrein H, Kausrud KL, Begon M, Davis S, Leirs H, Dubyan-skiy VM, Esper J, Ageyev VS, Klassovskiy NL, Pole SB, Chan KS (2006) Plague dynamics are driven by climate variation. *Proceedings of the National Academy of Sciences* 103:13,110–13,115, DOI 10.1073/pnas.0602447103, <http://www.pnas.org/content/103/35/13110.full.pdf+html>
- Steutel FW, van Harn K (1979) Discrete analogues of self-decomposability and stability. *Annals of Probability* 7:893–899
- Thyregod P, Carstensen J, Madsen H, Arnbjerg-Nielsen K (1999) Integer valued autoregressive models for tipping bucket rainfall measurements. *Environmetrics* 10(4):395–411, DOI 10.1002/(SICI)1099-095X(199907/08)10:4<395::AID-ENV364>3.0.CO;2-M
- Tong H (1983) *Threshold Models in Non-linear Time Series Analysis*. Lecture Notes in Statistics, No. 21. Springer-Verlag

- Tong H (1990) Non-linear Time Series. A Dynamical System Approach. Oxford Clarendon Press
- Tong H (2011) Threshold models in time series analysis—30 years on. *Statistics and its Interface* 4:107–118
- Tong H, Lim KS (1980) Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society Series B* 42:245–292
- Turkman KF, Scotto MG, de Zea Bermudez P (2014) Non-Linear Time Series: Extreme Events and Integer Value Problems. Springer-Verlag, Switzerland
- Wang C, Liu H, Yao JF, Davis RA, Li WK (2014) Self-excited threshold Poisson autoregression. *Journal of the American Statistical Association* 109:777–787, DOI 10.1080/01621459.2013.872994
- Weiß CH (2009) A new class of autoregressive models for time series of binomial counts. *Communications in Statistics–Theory and Methods* 38(4):447–460, DOI 10.1080/03610920802233937
- Weiß CH, Kim HY (2013) Parameter estimation for binomial AR(1) models with applications in finance and industry. *Statistical Papers* 54:563–590
- Weiß CH, Kim HY (2014) Diagnosing and modeling extra-binomial variation for time-dependent counts. *Applied Stochastic Models in Business and Industry* 30:588–608, DOI 10.1002/asmb.2005
- Weiß CH, Pollett PK (2014) Binomial autoregressive processes with density-dependent thinning. *Journal of Time Series Analysis* 35:115–132
- Yu K, Zou H, Shi D (2014) Integer-valued moving average models with structural changes. *Mathematical Problems in Engineering* 2014(Article ID 231592):5 pages
- Zou H, Yu K (2014) First order threshold integer-valued moving average processes. *DCDIS: Dynamics of Continuous, Discrete and Impulsive Systems; Series B: Applications & Algorithms* 21:197–205
- Zucchini W, MacDonald IL (2009) Hidden Markov Models for Time Series: An Introduction Using R. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, CRC Press

## A Proofs

### A.1 Unconditional Mean and Variance

The unconditional mean (6) is a direct consequence of (4):

$$\begin{aligned}\mu_X &= E[E[X_t|X_{t-1}]] = r_1 \mu_{IX} + (1 - r_1)\pi_1 N p + r_2 (\mu_X - \mu_{IX}) + (1 - r_2)\pi_2 N (1 - p) \\ &= r_2 \mu_X + (r_1 - r_2) \mu_{IX} + N p \pi_1 (1 - r_1) + N(1 - p) \pi_2 (1 - r_2).\end{aligned}$$



For the unconditional variance (7), consider first

$$\begin{aligned}
E[V[X_t|X_{t-1}]] &\stackrel{(5)}{=} E\left[I_{t-1}\left(r_1(1-r_1)(1-2\pi_1)X_{t-1} + N(1-r_1)\pi_1(1-(1-r_1)\pi_1)\right)\right] \\
&\quad + E\left[(1-I_{t-1})\left(r_2(1-r_2)(1-2\pi_2)X_{t-1} + N(1-r_2)\pi_2(1-(1-r_2)\pi_2)\right)\right] \\
&= r_1(1-r_1)(1-2\pi_1)\mu_{IX} + pN(1-r_1)\pi_1(1-(1-r_1)\pi_1) \quad (\text{A.1}) \\
&\quad + r_2(1-r_2)(1-2\pi_2)(\mu_X - \mu_{IX}) + (1-p)N(1-r_2)\pi_2(1-(1-r_2)\pi_2),
\end{aligned}$$

as well as (note that  $E[I_{t-1}(1-I_{t-1}) \cdot Y] = 0$ )

$$\begin{aligned}
V[E[X_t|X_{t-1}]] &\stackrel{(4)}{=} V\left[I_{t-1}(r_1X_{t-1} + (1-r_1)\pi_1N) + (1-I_{t-1})(r_2X_{t-1} + (1-r_2)\pi_2N)\right] \\
&= V\left[I_{t-1}(r_1X_{t-1} + (1-r_1)\pi_1N)\right] + V\left[(1-I_{t-1})(r_2X_{t-1} + (1-r_2)\pi_2N)\right] \\
&\quad + 2\text{Cov}\left[I_{t-1}(r_1X_{t-1} + (1-r_1)\pi_1N), (1-I_{t-1})(r_2X_{t-1} + (1-r_2)\pi_2N)\right] \\
&= V\left[I_{t-1}r_1X_{t-1}\right] + V\left[I_{t-1}(1-r_1)\pi_1N\right] + 2\text{Cov}\left[I_{t-1}r_1X_{t-1}, I_{t-1}(1-r_1)\pi_1N\right] \\
&\quad + V\left[(1-I_{t-1})r_2X_{t-1}\right] + V\left[(1-I_{t-1})(1-r_2)\pi_2N\right] \\
&\quad + 2\text{Cov}\left[(1-I_{t-1})r_2X_{t-1}, (1-I_{t-1})(1-r_2)\pi_2N\right] \\
&\quad + 0 - 2E\left[I_{t-1}(r_1X_{t-1} + (1-r_1)\pi_1N)\right] \cdot E\left[(1-I_{t-1})(r_2X_{t-1} + (1-r_2)\pi_2N)\right] \\
&= r_1^2V\left[I_{t-1}X_{t-1}\right] + (1-r_1)^2\pi_1^2N^2p(1-p) + 2r_1(1-r_1)\pi_1N\text{Cov}\left[I_{t-1}X_{t-1}, I_{t-1}\right] \\
&\quad + r_2^2V\left[(1-I_{t-1})X_{t-1}\right] + (1-r_2)^2\pi_2^2N^2p(1-p) \\
&\quad + 2r_2(1-r_2)\pi_2N\text{Cov}\left[(1-I_{t-1})X_{t-1}, (1-I_{t-1})\right] \\
&\quad - 2r_1r_2\mu_{IX}(\mu_X - \mu_{IX}) - 2r_1(1-r_2)\pi_2(1-p)N\mu_{IX} \\
&\quad - 2r_2(1-r_1)\pi_1pN(\mu_X - \mu_{IX}) - 2p(1-p)(1-r_1)\pi_1(1-r_2)\pi_2N^2 \\
&= r_1^2(\mu_{IX,2} - \mu_{IX}^2) + (1-r_1)^2\pi_1^2N^2p(1-p) + 2r_1(1-r_1)\pi_1N(1-p)\mu_{IX} \\
&\quad + r_2^2(\sigma_X^2 + 2\mu_X\mu_{IX} - \mu_{IX,2} - \mu_{IX}^2) + (1-r_2)^2\pi_2^2N^2p(1-p) \\
&\quad + 2r_2(1-r_2)\pi_2Np(\mu_X - \mu_{IX}) \quad (\text{A.2}) \\
&\quad - 2r_1r_2\mu_{IX}(\mu_X - \mu_{IX}) - 2r_1(1-r_2)\pi_2(1-p)N\mu_{IX} \\
&\quad - 2r_2(1-r_1)\pi_1pN(\mu_X - \mu_{IX}) - 2p(1-p)(1-r_1)\pi_1(1-r_2)\pi_2N^2.
\end{aligned}$$

Insertion of (A.1) and (A.2) into  $\sigma_X^2 = E[V[X_t|X_{t-1}]] + V[E[X_t|X_{t-1}]]$  and reordering gives

$$\begin{aligned}
(1 - r_2^2)\sigma_X^2 &= r_1(1 - r_1)(1 - 2\pi_1)\mu_{IX} + r_2(1 - r_2)(1 - 2\pi_2)(\mu_X - \mu_{IX}) \\
&\quad + Np(1 - r_1)\pi_1(1 - (1 - r_1)\pi_1) + N(1 - p)(1 - r_2)\pi_2(1 - (1 - r_2)\pi_2) \\
&\quad + r_1^2(\mu_{IX,2} - \mu_{IX}^2) + 2r_2^2\mu_X\mu_{IX} - r_2^2(\mu_{IX,2} + \mu_{IX}^2) - 2r_1r_2\mu_{IX}(\mu_X - \mu_{IX}) \\
&\quad + (1 - r_1)^2\pi_1^2N^2p(1 - p) + (1 - r_2)^2\pi_2^2N^2p(1 - p) \\
&\quad - 2(1 - r_1)\pi_1(1 - r_2)\pi_2N^2p(1 - p) \\
&\quad + 2r_1(1 - r_1)\pi_1N(1 - p)\mu_{IX} - 2r_1(1 - r_2)\pi_2N(1 - p)\mu_{IX} \\
&\quad + 2r_2(1 - r_2)\pi_2Np(\mu_X - \mu_{IX}) - 2r_2(1 - r_1)\pi_1Np(\mu_X - \mu_{IX}) \\
&= (r_1(1 - r_1)(1 - 2\pi_1) - r_2(1 - r_2)(1 - 2\pi_2))\mu_{IX} + r_2(1 - r_2)(1 - 2\pi_2)\mu_X \\
&\quad + Np(1 - r_1)\pi_1(1 - (1 - r_1)\pi_1) + N(1 - p)(1 - r_2)\pi_2(1 - (1 - r_2)\pi_2) \\
&\quad + (r_1^2 - r_2^2)\mu_{IX,2} - (r_1 - r_2)^2\mu_{IX}^2 - 2r_2(r_1 - r_2)\mu_X\mu_{IX} \\
&\quad + N^2p(1 - p)((1 - r_1)\pi_1 - (1 - r_2)\pi_2)^2 \\
&\quad + 2N(1 - p)r_1((1 - r_1)\pi_1 - (1 - r_2)\pi_2)\mu_{IX} \\
&\quad - 2Npr_2((1 - r_1)\pi_1 - (1 - r_2)\pi_2)(\mu_X - \mu_{IX}) \\
&= r_2(1 - r_2)(1 - 2\pi_2)\mu_X - 2Npr_2((1 - r_1)\pi_1 - (1 - r_2)\pi_2)\mu_X \\
&\quad - 2r_2(r_1 - r_2)\mu_X\mu_{IX} + (r_1^2 - r_2^2)\mu_{IX,2} - (r_1 - r_2)^2\mu_{IX}^2 \\
&\quad + 2N(r_1 - p(r_1 - r_2))((1 - r_1)\pi_1 - (1 - r_2)\pi_2)\mu_{IX} \\
&\quad + (r_1(1 - r_1)(1 - 2\pi_1) - r_2(1 - r_2)(1 - 2\pi_2))\mu_{IX} \\
&\quad + Np(1 - r_1)\pi_1(1 - (1 - r_1)\pi_1) + N(1 - p)(1 - r_2)\pi_2(1 - (1 - r_2)\pi_2) \\
&\quad + N^2p(1 - p)((1 - r_1)\pi_1 - (1 - r_2)\pi_2)^2.
\end{aligned}$$

This completes the proof of the variance formula (7).

To get the properties of the LSET-BAR(1) model, we have to insert  $r := r_1 = r_2$  into the equations (6) and (7). We start with the mean (10):

$$(1 - r)\mu_X = 0 + Np\pi_1(1 - r) + N(1 - p)\pi_2(1 - r).$$

The derivation of the variance (11) is more tedious:

$$\begin{aligned}
(1-r^2)\sigma_X^2 &= r(1-r)(1-2\pi_2)\mu_X + 2Npr(1-r)(\pi_2-\pi_1)\mu_X \\
&\quad - 2Nr(1-r)(\pi_2-\pi_1)\mu_{IX} + r(1-r)((1-2\pi_1)-(1-2\pi_2))\mu_{IX} \\
&\quad + Np(1-r)\pi_1(1-(1-r)\pi_1) + N(1-p)(1-r)\pi_2(1-(1-r)\pi_2) \\
&\quad + N^2p(1-p)(1-r)^2(\pi_2-\pi_1)^2 \\
&= r(1-r)(1-2\pi_2)N(p\pi_1 + (1-p)\pi_2) + 2Npr(1-r)(\pi_2-\pi_1)N(\pi_2 - p(\pi_2-\pi_1)) \\
&\quad - 2(N-1)r(1-r)(\pi_2-\pi_1)\mu_{IX} + Np(1-r^2)\pi_1(1-\pi_1) - Npr(1-r)\pi_1(1-2\pi_1) \\
&\quad + N(1-p)(1-r^2)\pi_2(1-\pi_2) - N(1-p)r(1-r)\pi_2(1-2\pi_2) \\
&\quad + N^2p(1-p)(1-r^2)(\pi_2-\pi_1)^2 - 2N^2p(1-p)r(1-r)(\pi_2-\pi_1)^2 \\
&= Npr(1-r)\pi_1((1-2\pi_2) - (1-2\pi_1)) \\
&\quad + 2N^2pr(1-r)\pi_2(\pi_2-\pi_1) - 2N^2p^2r(1-r)(\pi_2-\pi_1)^2 \\
&\quad - 2N^2p(1-p)r(1-r)(\pi_2-\pi_1)^2 - 2(N-1)r(1-r)(\pi_2-\pi_1)\mu_{IX} \\
&\quad + Np(1-r^2)\pi_1(1-\pi_1) + N(1-p)(1-r^2)\pi_2(1-\pi_2) + N^2p(1-p)(1-r^2)(\pi_2-\pi_1)^2 \\
&= -2Npr(1-r)\pi_1(\pi_2-\pi_1) + 2N^2pr(1-r)\pi_2(\pi_2-\pi_1) \\
&\quad - 2N^2pr(1-r)(\pi_2-\pi_1)^2 - 2(N-1)r(1-r)(\pi_2-\pi_1)\mu_{IX} \\
&\quad + Np(1-r^2)\pi_1(1-\pi_1) + N(1-p)(1-r^2)\pi_2(1-\pi_2) + N^2p(1-p)(1-r^2)(\pi_2-\pi_1)^2 \\
&= 2(N-1)r(1-r)(\pi_2-\pi_1)Np\pi_1 - 2(N-1)r(1-r)(\pi_2-\pi_1)\mu_{IX} \\
&\quad + Np(1-r^2)\pi_1(1-\pi_1) + N(1-p)(1-r^2)\pi_2(1-\pi_2) + N^2p(1-p)(1-r^2)(\pi_2-\pi_1)^2.
\end{aligned}$$

This completes the proof.

## A.2 Binomial Index of Dispersion

First, we consider the denominator of the *BID* (15) for the case of the LSET model ( $r_1 = r_2 = r$ ). Using (10), we obtain

$$\begin{aligned}
\frac{\mu_X}{N} \left(1 - \frac{\mu_X}{N}\right) &= (p\pi_1 + (1-p)\pi_2) \left(1 - (p\pi_1 + (1-p)\pi_2)\right) \\
&= p\pi_1 + (1-p)\pi_2 - p^2\pi_1^2 - 2p(1-p)\pi_1\pi_2 - (1-p)^2\pi_2^2 \\
&= p\pi_1 + (1-p)\pi_2 - p\pi_1^2 + p(1-p)\pi_1^2 - (1-p)\pi_2^2 + p(1-p)\pi_2^2 - 2p(1-p)\pi_1\pi_2 \\
&= p\pi_1(1-\pi_1) + (1-p)\pi_2(1-\pi_2) + p(1-p)(\pi_2-\pi_1)^2. \tag{A.3}
\end{aligned}$$

Then we look for matching terms in the numerator. First note that

$$N^2p(1-p)(\pi_2-\pi_1)^2 = Np(1-p)(\pi_2-\pi_1)^2 + N(N-1)p(1-p)(\pi_2-\pi_1)^2.$$

Using this together with (11) and (A.3), we find

$$\sigma_X^2 = \mu_X(1 - \mu_X/N) + N(N-1)p(1-p)(\pi_2-\pi_1)^2 + \frac{2r}{1+r}(N-1)(\pi_2-\pi_1)(Np\pi_1 - \mu_{IX}).$$

Bringing the results together leads to (16) for the *BID*. Note that only the last term,  $\frac{2r}{1+r} \dots$ , might become negative.

If  $r = 0$ , (16) reduces to

$$BID = 1 + \frac{p(1-p)N(N-1)(\pi_2 - \pi_1)^2}{p\pi_1(1-\pi_1) + (1-p)\pi_2(1-\pi_2) + p(1-p)(\pi_2 - \pi_1)^2} \geq 1.$$

### A.3 Autocovariance Function

By the law of total covariance, we obtain

$$\begin{aligned} \gamma(k) &:= Cov[X_t, X_{t-k}] = Cov[E[X_t|X_{t-1}, \dots], E[X_{t-k}|X_{t-1}, \dots]] + 0 \\ &\stackrel{(8)}{=} Cov[rX_{t-1} + N(1-r)(\pi_2 + I_{t-1}(\pi_1 - \pi_2)), X_{t-k}] \\ &= r \cdot Cov[X_{t-1}, X_{t-k}] + N(1-r) \cdot (\pi_1 - \pi_2) \cdot Cov[I_{t-1}, X_{t-k}] \\ &= \dots = r^k V[X_{t-k}] + N(1-r) \cdot (\pi_1 - \pi_2) \cdot \sum_{s=1}^k r^{s-1} Cov[I_{t-s}, X_{t-k}], \end{aligned}$$

which proves (18).

## B Tables

Model M1:  $(N, R; r, \pi_1, \pi_2) = (40, 10; 0.3, 0.15, 0.4)$ .

$T$	$R$	CLS				CML			
		$r$	$\pi_1$	$\pi_2$	%	$r$	$\pi_1$	$\pi_2$	%
100	8	0.57	0.17	0.31	3.2	0.39	0.15	0.32	4.1
		(0.18)	(0.16)	(0.27)		(0.11)	(0.03)	(0.09)	
	9	0.42	0.16	0.35	8.5	0.33	0.15	0.35	7.3
		(0.14)	(0.02)	(0.07)		(0.11)	(0.03)	(0.09)	
	10	0.27	0.15	0.39	83.7	0.26	0.14	0.37	84.9
		(0.08)	(0.02)	(0.03)		(0.12)	(0.04)	(0.10)	
11	0.39	0.18	0.39	2.9	0.32	0.17	0.37	3.1	
	(0.13)	(0.05)	(0.03)		(0.11)	(0.06)	(0.09)		
12	0.53	0.21	0.37	1.7	0.37	0.18	0.37	0.6	
	(0.16)	(0.07)	(0.06)		(0.12)	(0.07)	(0.09)		
500	8	0.64	0.17	0.35	0	0.42	0.16	0.36	0
		(0.08)	(0.01)	(0.03)		(0.03)	(0.01)	(0.02)	
	9	0.46	0.15	0.38	0	0.36	0.15	0.38	0
		(0.01)	(0.01)	(0.00)		(0.04)	(0.01)	(0.01)	
	10	0.29	0.15	0.40	99.9	0.29	0.15	0.40	100
		(0.05)	(0.00)	(0.00)		(0.04)	(0.01)	(0.01)	
11	0.43	0.16	0.40	0.1	0.35	0.16	0.40	0	
	(0.06)	(0.01)	(0.01)		(0.04)	(0.01)	(0.01)		
12	0.58	0.19	0.38	0	0.41	0.18	0.39	0	
	(0.07)	(0.02)	(0.01)		(0.03)	(0.02)	(0.01)		
1000	8	0.64	0.17	0.35	0	0.43	0.15	0.36	0
		(0.06)	(0.01)	(0.02)		(0.02)	(0.00)	(0.01)	
	9	0.46	0.15	0.38	0	0.36	0.15	0.38	0
		(0.05)	(0.00)	(0.01)		(0.02)	(0.00)	(0.00)	
	10	0.30	0.15	0.40	100	0.30	0.15	0.40	100
		(0.03)	(0.00)	(0.00)		(0.03)	(0.00)	(0.00)	
11	0.44	0.16	0.40	0	0.36	0.16	0.40	0	
	(0.04)	(0.01)	(0.01)		(0.02)	(0.01)	(0.01)		
12	0.59	0.18	0.39	0	0.41	0.17	0.40	0	
	(0.05)	(0.01)	(0.01)		(0.02)	(0.01)	(0.01)		

Table B.1: Conditional Least Squares and Conditional Maximum Likelihood estimates for  $(r, \pi_1, \pi_2)$  and  $R$  in the LSET-BAR(1) model. Standard errors in parentheses.

Model M2:  $(N, R; r, \pi_1, \pi_2) = (20, 4; 0.3, 0.15, 0.4)$ .

$T$	$R$	CLS				CML			
		$r$	$\pi_1$	$\pi_2$	%	$r$	$\pi_1$	$\pi_2$	%
100	2	0.61	0.25	0.34	4.9	0.48	0.20	0.35	5.1
		(0.13)	(0.10)	(0.05)		(0.08)	(0.06)	(0.04)	
	3	0.49	0.19	0.37	1.5	0.40	0.17	0.37	1.7
		(0.13)	(0.05)	(0.03)		(0.09)	(0.05)	(0.03)	
	4	0.27	0.16	0.40	91.9	0.28	0.16	0.39	92.0
		(0.11)	(0.03)	(0.02)		(0.11)	(0.04)	(0.04)	
5	0.42	0.25	0.39	1.7	0.36	0.23	0.39	1.1	
	(0.14)	(0.06)	(0.03)		(0.09)	(0.06)	(0.02)		
6	0.62	0.37	0.30	0	0.43	0.28	0.38	0.1	
	(0.17)	(0.41)	(0.52)		(0.08)	(0.07)	(0.02)		
500	2	0.63	0.22	0.35	0	0.48	0.18	0.36	0
		(0.06)	(0.03)	(0.02)		(0.03)	(0.02)	(0.02)	
	3	0.50	0.17	0.37	0	0.41	0.16	0.38	0
		(0.06)	(0.02)	(0.01)		(0.04)	(0.01)	(0.01)	
	4	0.29	0.15	0.40	100	0.30	0.15	0.40	100
		(0.05)	(0.01)	(0.01)		(0.05)	(0.01)	(0.01)	
5	0.46	0.23	0.39	0	0.38	0.22	0.40	0	
	(0.06)	(0.02)	(0.01)		(0.04)	(0.02)	(0.01)		
6	0.67	0.33	0.34	0	0.45	0.27	0.39	0	
	(0.07)	(0.04)	(0.04)		(0.03)	(0.03)	(0.01)		
1000	2	0.63	0.22	0.35	0	0.48	0.18	0.36	0
		(0.04)	(0.02)	(0.01)		(0.02)	(0.01)	(0.01)	
	3	0.51	0.17	0.37	0	0.41	0.16	0.38	0
		(0.04)	(0.01)	(0.01)		(0.02)	(0.01)	(0.01)	
	4	0.30	0.15	0.40	100	0.30	0.15	0.40	100
		(0.03)	(0.01)	(0.01)		(0.03)	(0.01)	(0.01)	
5	0.46	0.23	0.39	0	0.38	0.22	0.40	0	
	(0.04)	(0.02)	(0.01)		(0.03)	(0.02)	(0.01)		
6	0.68	0.33	0.34	0	0.46	0.27	0.39	0	
	(0.05)	(0.03)	(0.02)		(0.02)	(0.02)	(0.01)		

Table B.2: Conditional Least Squares and Conditional Maximum Likelihood estimates for  $(r, \pi_1, \pi_2)$  and  $R$  in the LSET-BAR(1) model. Standard errors in parentheses.

Model M3:  $(N, R; r, \pi_1, \pi_2) = (40, 10; 0.7, 0.15, 0.4)$ .

$T$	$R$	CLS				CML			
		$r$	$\pi_1$	$\pi_2$	%	$r$	$\pi_1$	$\pi_2$	%
100	8	0.77	0.18	0.28	2.8	0.71	0.16	0.30	3.6
		(0.11)	(0.07)	(0.13)		(0.04)	(0.06)	(0.11)	
	9	0.71	0.17	0.32	12	0.70	0.16	0.32	12.2
		(0.09)	(0.05)	(0.11)		(0.04)	(0.05)	(0.12)	
	10	0.66	0.17	0.37	61.5	0.69	0.17	0.37	61.5
		(0.08)	(0.05)	(0.07)		(0.05)	(0.05)	(0.08)	
11	0.71	0.21	0.38	8.9	0.70	0.21	0.38	8.9	
	(0.09)	(0.10)	(0.06)		(0.04)	(0.11)	(0.07)		
12	0.76	0.21	0.37	3.8	0.71	0.22	0.38	3.8	
	(0.10)	(0.10)	(0.08)		(0.04)	(0.10)	(0.07)		
500	8	0.83	0.17	0.34	0	0.72	0.15	0.36	0
		(0.05)	(0.02)	(0.05)		(0.02)	(0.01)	(0.03)	
	9	0.76	0.15	0.37	0.5	0.70	0.15	0.38	1
		(0.04)	(0.01)	(0.03)		(0.02)	(0.01)	(0.02)	
	10	0.69	0.15	0.40	97.4	0.70	0.15	0.40	97.2
		(0.04)	(0.01)	(0.02)		(0.02)	(0.01)	(0.02)	
11	0.74	0.16	0.39	1.9	0.70	0.16	0.40	1.6	
	(0.04)	(0.02)	(0.02)		(0.02)	(0.02)	(0.02)		
12	0.80	0.19	0.38	0.2	0.71	0.17	0.40	0.2	
	(0.05)	(0.03)	(0.02)		(0.02)	(0.02)	(0.02)		
1000	8	0.84	0.16	0.34	0	0.72	0.15	0.36	0
		(0.03)	(0.03)	(0.03)		(0.01)	(0.01)	(0.02)	
	9	0.76	0.15	0.38	0	0.71	0.15	0.38	0
		(0.03)	(0.01)	(0.02)		(0.01)	(0.01)	(0.01)	
	10	0.70	0.15	0.40	99.9	0.70	0.15	0.40	99.9
		(0.02)	(0.01)	(0.01)		(0.01)	(0.00)	(0.02)	
11	0.75	0.16	0.39	0.1	0.71	0.16	0.39	0.1	
	(0.03)	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)		
12	0.81	0.18	0.38	0	0.71	0.17	0.40	0	
	(0.03)	(0.02)	(0.02)		(0.01)	(0.01)	(0.01)		

Table B.3: Conditional Least Squares and Conditional Maximum Likelihood estimates for  $(r, \pi_1, \pi_2)$  and  $R$  in the LSET-BAR(1) model. Standard errors in parentheses.

Model M4:  $(N, R; r, \pi_1, \pi_2) = (20, 5; 0.7, 0.15, 0.4)$ .

$T$	$R$	CLS				CML			
		$r$	$\pi_1$	$\pi_2$	%	$r$	$\pi_1$	$\pi_2$	%
100	3	0.82	0.23	0.24	6.5	0.73	0.17	0.27	3.1
		(0.10)	(0.14)	(0.10)		(0.04)	(0.05)	(0.09)	
	4	0.75	0.17	0.29	1.1	0.71	0.16	0.31	7.8
		(0.10)	(0.05)	(0.10)		(0.05)	(0.03)	(0.10)	
	5	0.65	0.16	0.37	68.9	0.69	0.16	0.37	74.8
		(0.09)	(0.03)	(0.07)		(0.04)	(0.05)	(0.09)	
	6	0.76	0.21	0.34	9.9	0.71	0.19	0.36	9.8
(0.09)		(0.08)	(0.10)		(0.04)	(0.05)	(0.09)		
7	0.82	0.24	0.29	3.9	0.73	0.21	0.35	4.5	
	(0.08)	(0.10)	(0.12)		(0.04)	(0.06)	(0.10)		
500	3	0.90	0.28	0.19	0	0.73	0.16	0.29	0
		(0.03)	(0.14)	(0.16)		(0.02)	(0.01)	(0.04)	
	4	0.81	0.17	0.32	0	0.71	0.15	0.04	0
		(0.04)	(0.02)	(0.05)		(0.02)	(0.01)	(0.03)	
	5	0.69	0.15	0.40	99.5	0.70	0.15	0.40	99.9
		(0.04)	(0.01)	(0.02)		(0.02)	(0.01)	(0.02)	
	6	0.80	0.19	0.37	0.4	0.71	0.17	0.39	0.1
(0.04)		(0.03)	(0.04)		(0.02)	(0.02)	(0.02)		
7	0.88	0.24	0.26	0.1	0.73	0.19	0.39	0	
	(0.04)	(0.13)	(0.34)		(0.02)	(0.02)	(0.03)		
1000	3	0.90	0.27	0.20	0	0.74	0.16	0.30	0
		(0.02)	(0.05)	(0.08)		(0.01)	(0.01)	(0.03)	
	4	0.82	0.17	0.32	0	0.71	0.15	0.3	0
		(0.03)	(0.01)	(0.03)		(0.01)	(0.01)	(0.02)	
	5	0.69	0.15	0.40	99.9	0.70	0.15	0.40	99.9
		(0.02)	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)	
	6	0.80	0.18	0.37	0.1	0.71	0.17	0.40	0.1
(0.03)		(0.02)	(0.02)		(0.01)	(0.01)	(0.01)		
7	0.88	0.23	0.26	0	0.73	0.19	0.39	0	
	(0.03)	(0.04)	(0.08)		(0.01)	(0.02)	(0.02)		

Table B.4: Conditional Least Squares and Conditional Maximum Likelihood estimates for  $(r, \pi_1, \pi_2)$  and  $R$  in the LSET-BAR(1) model. Standard errors in parentheses.