INFORMATICA, 2015, Vol. 26, No. 1, 67–87

DOI: http://dx.doi.org/10.15388/Informatica.2015.39

Statistical Analysis of Spatio-Temporal Data Based on Poisson Conditional Autoregressive Model

Yuriy KHARIN*, Maryna ZHURAK
Research Institute for Applied Problems of Mathematics and Informatics
Belarusian State University, Independence Av. 4, Minsk 220030, Belarus
e-mail: kharin@bsu.by, mzhurak@gmail.com

Received: July 2014; accepted: March 2015

Abstract. Poisson conditional autoregressive model of spatio-temporal data is proposed. Markov property and probabilistic characteristics of this model are presented. Algorithms for maximum likelihood estimation of the model parameters are constructed. Optimal forecasting statistic minimizing probability of forecast error is given. The "plug-in" principle based on ML-estimators is used for forecasting in the case of unknown parameters. The results of computer experiments on simulated and real medical data are presented.

Key words: spatio-temporal data, Poisson conditional autoregressive model, Markov chain, maximum likelihood estimator, optimal forecast.

1. Introduction

Statistical analysis of spatio-temporal data is an important problem for practice: it allows to model adequately the underlying stochastic phenomenon taking into account both the dependence on time and the dependence on space. Models based on spatio-temporal data become widely used for solving practical problems in meteorology (Quintian *et al.*, 2014), ecology, economics (Pragarauskaite and Dzemyda, 2013), medicine and other fields (Gelfand *et al.*, 2010). Let us present some examples of these models.

The Spatial Temporal Conditional Autoregressive model (STCAR) is considered in Mariella and Tarantino (2010) for modeling and statistical analysis of spatio-temporal data on bankruptcy of small and medium-sized enterprises in the provinces of Italy. A distinctive feature of this model is that the vector in the time slice of the process under study has the conditional Gaussian distribution.

The statistical testing for a parameter change of Poisson autoregressive models is considered and experiments are conducted on real data that describes the incidence of poliomyelitis in Kang and Lee (2014). In Fokianos and Tjøstheim (2012) the authors study statistical properties of non-linear regression models for discrete data. In particular, a non-linear Poisson autoregressive model is considered; it is a generalization of linear Poisson regression model (Fokianos *et al.*, 2009). This model is used to simulate the mortality data

*Corresponding author.

in a certain area of Portugal. Long memory autoregressive conditional Poisson (LMACP) model for modeling discrete time series with a high persistence is introduced by Grob-KlubMann and Hautsch (2013). The model is applied to forecast the bid-ask spreads. This model allows handling such properties of the bid-ask spreads as a strong autocorrelation and discreteness of observations. Homogeneous and non-homogeneous Poisson models are used by Rodrigues (2013) for statistical analysis of air pollution data. Note that the spatial dependence was not taken into account in the given above models of discrete data.

This paper is devoted to statistical analysis of the Poisson conditional autoregressive model that can be used to describe the discrete spatio-temporal data. The structure of the paper is as follows. The Poisson conditional autoregressive model of spatio-temporal data is constructed in Section 2. We study the probabilistic characteristics of this model in Section 3. The likelihood function is constructed and an algorithm for computing the maximum likelihood estimators of the model parameters is given in Section 4. Section 5 is devoted to statistical forecasting of spatio-temporal data using the Poisson conditional autoregressive model; forecasting statistic that minimizes the probability of the forecast error is built in the case of known parameters of the model. The "plug-in" principle is applied to the constructed forecasting statistic in the case of unknown parameters. Section 6 presents the results of computer experiments which were conducted on simulated and real medical data.

The paper is based on the presentation given at the X International Conference "Computer Data Analysis and Modeling" (Minsk, September 10–14, 2013).

2. Poisson Conditional Autoregressive Model

Introduce the notation: (Ω, F, P) is the probability space; N is set of positive integers, $N_0 = N \cup \{0\}$; $I\{A\}$ is the indicator function of the event A; $s \in S = \{1, 2, \dots, n\}$ is the set of indexed spatial regions or space locations (let us agree to call them sites), into which the analyzed spatial area is partitioned; n is number of sites; $t \in N$ is discrete time; T is the length of observation period; $x_{s,t} \in N_0$ is a discrete random variable at time t at site s which describes the number of incidences in the studied region; $U(s) \subseteq S$ is a subset of neighbors of site s (a subset of sites which are situated most closely to the site s, or which share common boundary with site s, or which have the greatest impact on the random variable $x_{s,t}$ in this site); $F_{\bar{s},<t} = \sigma\{x_{u,\tau} : u \neq s,\ \tau < t\} \subset F$ is the σ-algebra generated by the indicated in braces random variables; $z_{s,t} \geqslant 0$ is an observed (known) level of exogenous factors (e.g., environmental pollution) at time t at site s which influences $x_{s,t}$; $\{\varphi_k(t) : 1 \leqslant k \leqslant K\}$ is a given set of $K \in N$ basic functions which determine a trend; $L(\xi)$ means the probability distribution of the random variable ξ; $E\{\cdot\}$, $D\{\cdot\}$, $cov\{\cdot\}$ are symbols of expectation, variance, covariance of the random variables; $\Pi(l; \lambda)$ is the Poisson probability distribution of the random variable ξ with the intensity parameter $\lambda > 0$:

$$P\{\xi = l\} = \Pi(l; \lambda) = \frac{\lambda^l}{l!} e^{-\lambda}, \quad l \in N_0. \tag{1}$$

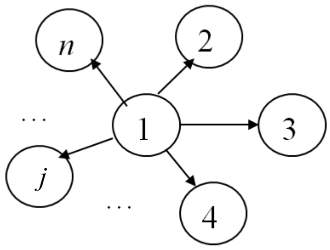

Fig. 1. Graph of sites dependence in the case $U(s) \equiv \{1\}$.

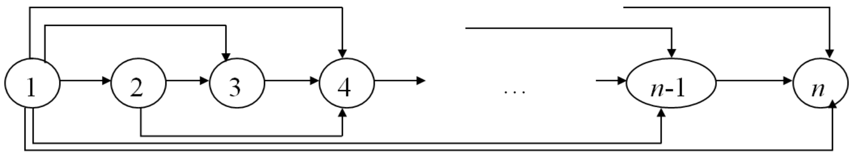

Fig. 2. Graph of sites dependence in the case $U(s) \equiv \{1, 2, \dots, s-1\}$.

We construct the Poisson conditional autoregressive model for spatio-temporal data $\{x_{s,t}\}$ using the idea of the paper of Mariella and Tarantino (2010):

$$L\{x_{s,t}|F_{\bar{s},<t}\} = \Pi(\cdot; \lambda_{s,t}), \tag{2}$$

$$\begin{aligned} \ln \lambda_{s,t} &= \ln \lambda_{s,t}\big(\{x_{j,t} : j \in U(s)\}, x_{s,t-1}\big) \\ &= a_s x_{s,t-1} I\{t > 1\} + \sum_{j \in U(s)} b_{s,j} x_{j,t} + \beta_s z_{s,t} + \sum_{k=1}^{K} \gamma_{s,k} \varphi_k(t), \quad t \in N, s \in S, \end{aligned} \tag{3}$$

$$U(s) \subseteq \{1, 2, \dots, s-1\}, \quad s = 2, \dots, n,$$

$$U(1) \equiv \emptyset, \qquad |U(s)| = K_s, \qquad K_1 \equiv 0, \tag{4}$$

where $a = (a_1, a_2, \dots, a_n)' \in R^n$, $b_s = (b_{s,j_1}, \dots, b_{s,j_{K_s}})' \in R^{K_s}$, $j_k \in U(s)$: $k = 1, \dots, K_s$, $s \in S$, $\beta = (\beta_1, \beta_2, \dots, \beta_n)' \in R^n$, $\gamma_s = (\gamma_{s,1}, \dots, \gamma_{s,K})' \in R^K$, $s \in S$, are the parameters of the model. The number of parameters of the model is equal to $D = n(K+2) + \sum_{s=1}^{n} K_s$.

Depending on types of the sets of neighbors $U(s)$, $s \in S$, we propose some typical models of the sites dependence for the considered random process $x_{s,t}$:

(a) all sites depend only on the first site ($U(s) \equiv \{1\}$); graphically this type of dependence presents "star" and is shown in Fig. 1;

(b) each site depends on all previous sites ($U(s) \equiv \{1, 2, \dots, s-1\}$; see Fig. 2);

(c) each site depends on the previous one (see Fig. 3):

$$U(s) \equiv \{s-1\}. \tag{5}$$

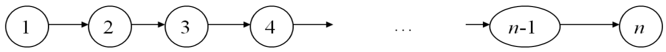

Fig. 3. Graph of sites dependence in the case $U(s) \equiv \{s-1\}$.

To reduce the number of model parameters we use an additional condition for the cardinality of sets $U(2), U(3), \dots, U(n)$:

$$K_1 = 0, \qquad K_s \leqslant 1, \quad s = 2, 3, \dots n.$$

This condition is satisfied, in particular, for types of dependence that are illustrated in Figs. 1 and 3. In this case the number of parameters of the model is equal to $D \leqslant n(K+2) + n - 1 = n(K+3) - 1$.

3. Probabilistic Properties of the Model

Let $X_t = (x_{1,t}, x_{2,t}, \dots, x_{n,t})' \in N_0^n$ be the column vector specifying the time slice of the process under consideration at time t; $L = \{l_j = (l_{1,j}, \dots, l_{n,j})' \in N_0^n : j = 0, 1, \dots\}$ be an ordered set of all admissible values of the vector X_t, e.g. the set L can take the following form:

$$L = \left\{ \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots \right\},$$

where the first vector has all zero components, then n vectors of all possible combinations of 0 and 1 follow, then – combinations of 0, 1 and 2, etc.

Theorem 1. *For the model* (2)–(4) *the observed vector process X_t is the n-dimensional nonhomogeneous Markov chain with the countable state space L, the one-step transition probability matrix $P(t) = (p_{ij}(t))$, $i, j \in N_0$, $t \geqslant 2$:*

$$p_{ij}(t) = \prod_{s=1}^{n} \frac{\exp\{-\lambda_{s,t}(\{l_{k,j} : k \in U(s)\},\ l_{s,i})\}\lambda_{s,t}^{l_{s,j}}(\{l_{k,j} : k \in U(s)\},\ l_{s,i})}{l_{s,j}!}, \tag{6}$$

and the initial probability distribution $p = (p_0, p_1, \dots)'$:

$$p_j = \prod_{s=1}^{n} \frac{\exp\{-\lambda_{s,1}(\{l_{k,j} : k \in U(s)\})\}\lambda_{s,1}^{l_{s,j}}(\{l_{k,j} : k \in U(s)\})}{l_{s,j}!}, \quad j \in N_0. \tag{7}$$

Proof. Let us show that for the model (2)–(4) vector process X_t is the n-dimensional nonhomogeneous Markov chain. Use the generalized formula for multiplying probabilities:

$$\begin{aligned} &P\{X_t = l_{i_t} | X_{t-1} = l_{i_{t-1}}, \dots, X_1 = l_{i_1}\} \\ &\quad = P\{x_{1,t} = l_{1,i_t}, \dots, x_{n,t} = l_{n,i_t} | X_{t-1} = l_{i_{t-1}}, \dots, X_1 = l_{i_1}\} \end{aligned}$$

$$= P\{x_{1,t} = l_{1,i_t} | X_{t-1} = l_{i_{t-1}}, \dots, X_1 = l_{i_1}\}$$
$$\times \prod_{s=2}^{n} P\{x_{s,t} = l_{s,i_t} | x_{1,t} = l_{1,i_t}, \dots, x_{s-1,t} = l_{s-1,i_t}, X_{t-1} = l_{i_{t-1}}, \dots, X_1 = l_{i_1}\}.$$

By condition (4) we have:

$$\{x_{k,t} = l_{k,i_t} : k \in U(s)\} \subseteq \{x_{1,t} = l_{1,i_t}, \dots, x_{s-1,t} = l_{s-1,i_t}\}, \quad s \in S.$$

According to model assumptions (2)–(4) we obtain:

$$\begin{aligned}
&P\{X_t = l_{i_t} | X_{t-1} = l_{i_{t-1}}, \dots, X_1 = l_{i_1}\} \\
&\quad = P\{x_{1,t} = l_{1,i_t} | x_{1,t-1} = l_{1,i_{t-1}}\} \\
&\qquad \times \prod_{s=2}^{n} P\big\{x_{s,t} = l_{s,i_t} \big| \{x_{k,t} = l_{k,i_t} : k \in U(s)\}, x_{s,t-1} l_{s,i_{t-1}}\big\} \\
&\quad = P\{X_t = l_t | X_{t-1} = l_{i_{t-1}}\}, \quad t \geqslant 2.
\end{aligned} \tag{8}$$

Thus, the Markov property is satisfied, and X_t is the n-dimensional nonhomogeneous Markov chain. The one-step transition probability matrix $P(t) = (p_{ij}(t))$ is determined by taking into account (8) and (1), (3):

$$\begin{aligned}
p_{ij}(t) &= P\{X_t = l_j | X_{t-1} = l_i\} \\
&= P\{x_{1,t} = l_{1,j} | x_{1,t-1} = l_{1,i}\} \\
&\quad \times \prod_{s=2}^{n} P\left\{x_{s,t} = l_{s,j} \Big| \bigcap_{k \in U(s)} \{x_{k,t} = l_{k,j}\}, x_{s,t-1} = l_{s,i}\right\} \\
&= \Pi\big(l_{1,j}; \lambda_{1,t}(l_{1,i})\big) \prod_{s=2}^{n} \Pi\big(l_{s,j}; \lambda_{s,t}\big(\{l_{k,j} : k \in U(s)\}, l_{s,i}\big)\big) \\
&= \prod_{s=1}^{n} \frac{\exp\{-\lambda_{s,t}(\{l_{k,j} : k \in U(s)\}, l_{s,i})\} \lambda_{s,t}^{l_{s,j}}(\{l_{k,j} : k \in U(s)\}, l_{s,i})}{l_{s,j}!},
\end{aligned}$$
$$i, j \in N_0, \ t \geqslant 2,$$

that coincides with (6).

Calculate the initial probability distribution $p = (p_0, p_1, \dots)'$:

$$\begin{aligned}
p_j &= P\{X_1 = l_j\} \\
&= P\{x_{1,1} = l_{1,j}\} \prod_{s=2}^{n} P\left\{x_{s,1} = l_{s,j} \Big| \bigcap_{k \in U(s)} \{x_{k,1} = l_{k,j}\}\right\}
\end{aligned}$$

$$= \Pi(l_{1,j};\lambda_{1,1})\prod_{s=2}^{n}\Pi\big(l_{s,j};\lambda_{s,1}\big(\{l_{k,j}:k\in U(s)\}\big)\big)$$

$$= \prod_{s=1}^{n}\frac{\exp\{-\lambda_{s,1}(\{l_{k,j}:k\in U(s)\})\}\lambda_{s,1}^{l_{s,j}}(\{l_{k,j}:k\in U(s)\})}{l_{s,j}!},\quad j\in N_0,$$

that coincides with (7). □

Corollary 1. *Under conditions of Theorem* 1*, the matrix of transition probabilities* $H(t_1,t_2)=(h_{i,j}(t_1,t_2))$*,* $h_{i,j}(t_1,t_2)=P\{X_{t_2}=l_j|X_{t_1}=l_i\}$ *,* $i,j\in N_0$*, for* t_2-t_1 *steps from time point* t_1 *to time point* t_2 *(*$t_1<t_2$*,* $t_1,t_2\in N$*) is:*

$$H(t_1,t_2)=P(t_1+1)P(t_1+2)\dots P(t_2). \tag{9}$$

Proof. Proof follows from the Kolmogorov–Chapman formula (Kemeny and Snell, 1976). □

Corollary 2. *Under conditions of Theorem* 1*, if* $\beta_s\equiv 0$*,* $K=1$*,* $\varphi_1\equiv 1$*, then the one-step transition probability matrix does not depend on t, and Markov chain is homogeneous*:

$$P(t)=P=(p_{ij}),$$

$$p_{ij}=\prod_{s=1}^{n}\frac{\exp\{-\lambda_{s,t}(\{l_{k,j}:k\in U(s)\},l_{s,i})\}\lambda_{s,t}^{l_{s,j}}(\{l_{k,j}:k\in U(s)\},l_{s,i})}{l_{s,j}!},$$

$$H(t_1,t_2)=P^{t_2-t_1},$$

where $\ln\lambda_{s,t}=a_sl_{s,i}+\sum_{k\in U(s)}b_{s,k}l_{k,j}+\gamma_{s,1}$ *does not depend on t.*

Corollary 3. *The current probability distribution* $p(t)=(p_0(t),p_1(t),\dots)'$*,* $p_j(t)=P\{X_t=l_j\}$*,* $j=0,1,\dots$*, is determined by formula*:

$$p_j(t)=\sum_{i=0}^{+\infty}p_ih_{i,j}(1,t),\quad j\in N_0,\ t\in N, \tag{10}$$

or in matrix form:

$$p(t)=H'(1,t)p,\quad t\in N,$$

where $H(1,1)=I_\infty$ *is the infinite dimensional identity matrix.*

Proof. Proof follows from the total probability formula and Corollary 1. □

Present now some probabilistic properties of the model (2)–(4) for the case $U(s)=\{s-1\}$ that will be used for statistical estimation of the model parameters.

Introduce the notation: $\lfloor a/b \rfloor$, $a \bmod b$ are the integer part and the residual of a modulo b respectively.

Lemma 1. *For the model* (2)–(5) *the* 2*-dimensional probability distribution* $P\{x_{s-1,t} = i_{st-1}, x_{s,t-1} = i_{st-s}\}$, $s \geqslant 2$, $t \in N$, *is*

$$P\{x_{s-1,t} = i_{st-1}, x_{s,t-1} = i_{st-s}\} = \sum_{i_1=0}^{+\infty} \cdots \sum_{i_{st-s-1}=0}^{+\infty} \sum_{i_{st-s+1}=0}^{+\infty} \cdots \sum_{i_{st-2}=0}^{+\infty} \exp\left\{ \sum_{k=1}^{st-1} (i_k \ln \lambda_{s',t'} - \lambda_{s',t'} - \ln i_k!) \right\},$$

where $k = (t'-1)s + s'$, $s' = (k-1) \bmod s + 1$, $t' = \lfloor (k-1)/s \rfloor + 1$.

Proof. Use the total probability formula:

$$P\{x_{s-1,t} = i_{st-1}, x_{s,t-1} = i_{st-s}\} = \sum_{i_{st-2}, i_{st-s-1}, i_{st-2s}=0}^{+\infty} P\{x_{s-1,t} = i_{st-1}, x_{s,t-1} = i_{st-s} | \mathcal{A}\} P\{\mathcal{A}\},$$

where $\mathcal{A} = \{x_{s-2,t} = i_{st-2}, x_{s-1,t-1} = i_{st-s-1}, x_{s,t-2} = i_{st-2s}\}$.

Since under fixed values $x_{s-2,t}$, $x_{s-1,t-1}$, $x_{s,t-2}$ the random variables $x_{s,t-1}$, $x_{s-1,t}$ are independent and have conditional Poisson distribution (2)–(4), then

$$\begin{aligned} P\{x_{s-1,t} &= i_{st-1}, x_{s,t-1} = i_{st-s}\} \\ = &\sum_{i_{st-2}, i_{st-s-1}, i_{st-2s}=0}^{+\infty} P\{x_{s-1,t} = i_{st-1} | x_{s-2,t} = i_{st-2}, x_{s-1,t-1} = i_{st-s-1}\} \\ &\times P\{x_{s,t-1} = i_{st-s} | x_{s-1,t-1} = i_{st-s-1}, x_{s,t-2} = i_{st-2s}\} P\{\mathcal{A}\} \\ = &\sum_{i_{st-2}, i_{st-s-1}, i_{st-2s}=0}^{+\infty} \Pi\big(i_{st-1}; \lambda_{s-1,t}(i_{st-2}, i_{st-s-1})\big) \\ &\times \Pi\big(i_{st-s}; \lambda_{s,t-1}(i_{st-s-1}, i_{st-2s})\big) P\{\mathcal{A}\}. \end{aligned}$$

Probability $P\{\mathcal{A}\}$ can be found analogously. Doing similar calculations, we obtain the following result:

$$P\{x_{s-1,t} = i_{st-1}, x_{s,t-1} = i_{st-s}\} = \sum_{i_1=0}^{+\infty} \cdots \sum_{i_{st-s-1}=0}^{+\infty} \sum_{i_{st-s+1}=0}^{+\infty} \cdots \sum_{i_{st-2}=0}^{+\infty} \prod_{k=3,\, k \neq s+1}^{st-1} \Pi(i_k; \lambda_{s',t'})$$

$$\times P\{x_{1,2}=i_{s+1}, x_{2,1}=i_2\}$$
$$=\sum_{i_1=0}^{+\infty}\cdots\sum_{i_{st-s-1}=0}^{+\infty}\sum_{i_{st-s+1}=0}^{+\infty}\cdots\sum_{i_{st-2}=0}^{+\infty}\exp\left\{\sum_{k=1}^{st-1}(i_k\ln\lambda_{s',t'}-\lambda_{s',t'}-\ln i_k!)\right\},$$

where $k=(t'-1)s+s'$, $s'=(k-1) \bmod s+1$, $t'=\lfloor(k-1)/s\rfloor+1$. □

Corollary 4. *Under conditions of Lemma* 1 *the* 1*-dimensional probability distribution of the discrete random variable $x_{s,t}$ in time point $t\in N$ at site $s\in S$ has the following form:*

$$P\{x_{s,t}=i_{st}\}=\sum_{i_1,\dots,i_{st-1}=0}^{+\infty}\exp\left\{\sum_{k=1}^{st}(i_k\ln\lambda_{s',t'}-\lambda_{s',t'}-\ln i_k!)\right\}. \tag{11}$$

where $k=(t'-1)s+s'$, $s'=(k-1) \bmod s+1$, $t'=\lfloor(k-1)/s\rfloor+1$.

Proof. Use the total probability formula and model (2)–(5) definition:

$$P\{x_{s,t}=i_{st}\}=\sum_{i_{st-1},i_{st-s}=0}^{+\infty}P\{x_{s,t}=i_{st}|x_{s-1,t}=i_{st-1},x_{s,t-1}=i_{st-s}\}$$
$$\times P\{x_{s-1,t}=i_{st-1},x_{s,t-1}=i_{st-s}\}.$$

Considering the result of Lemma 1, we obtain

$$P\{x_{s,t}=i_{st}\}$$
$$=\sum_{i_{st-1},i_{st-s}=0}^{+\infty}\Pi\big(i_{st};\lambda_{s,t}(i_{st-1},i_{st-s})\big)$$
$$\times\sum_{i_1=0}^{+\infty}\cdots\sum_{i_{st-s-1}=0}^{+\infty}\sum_{i_{st-s+1}=0}^{+\infty}\cdots\sum_{i_{st-2}=0}^{+\infty}\exp\left\{\sum_{k=1}^{st-1}(i_k\ln\lambda_{s',t'}-\lambda_{s',t'}-\ln i_k!)\right\}$$
$$=\sum_{i_1,\dots,i_{st-1}=0}^{+\infty}\exp\left(\sum_{k=1}^{st}(i_k\ln\lambda_{s',t'}-\lambda_{s',t'}-\ln i_k!)\right),$$

that coincides with (11). □

Lemma 2. *For the model* (2)–(5) *expectation and variance of the random variable $x_{s,t}$ have the following form:*

$$E\{x_{s,t}\}$$
$$=c_{s,t}\sum_{i_1,\dots,i_{st-1}=0}^{+\infty}\exp\left\{a_s i_{st-1}+b_{s,s-1}i_{st-s}+\sum_{k=1}^{st-1}(i_k\ln\lambda_{s',t'}-\lambda_{s',t'}-\ln i_k!)\right\}, \tag{12}$$

$$D\{x_{s,t}\} = c_{s,t}^2 \sum_{i_1,\ldots,i_{st-1}=0}^{+\infty} \exp\left\{2a_s i_{st-1} + 2b_{s,s-1} i_{st-s} + \sum_{k=1}^{st-1} (i_k \ln \lambda_{s',t'} - \lambda_{s',t'} - \ln i_k!)\right\} + E\{x_{s,t}\}\big(1 - E\{x_{s,t}\}\big), \quad (13)$$

where $c_{s,t} = \exp\{\beta_s z_{s,t} + \sum_{k=1}^{K} \gamma_{s,k}\varphi_k(t)\}$.

Proof. By (2)–(5) and the total expectation formula we have

$$E\{x_{s,t}\} = E\big\{E\{x_{s,t}|x_{s,t-1}, x_{s-1,t}\}\big\} = E\big\{\lambda_{s,t}(x_{s,t-1}, x_{s-1,t})\big\}.$$

Further, by Lemma 1, we obtain:

$$\begin{aligned} &E\big\{\lambda_{s,t}(x_{s,t-1}, x_{s-1,t})\big\} \\ &= \sum_{i_{st-1}, i_{st-s}=0}^{+\infty} \lambda_{s,t}(i_{st-1}, i_{st-s}) P\{x_{s-1,t} = i_{st-1}, x_{s,t-1} = i_{st-s}\} \\ &= \exp\left\{\beta_s z_{s,t} + \sum_{k=1}^{K} \gamma_{s,k}\varphi_k(t)\right\} \sum_{i_{st-1}, i_{st-s}=0}^{+\infty} \exp\{a_s i_{st-1} + b_{s,s-1} i_{st-s}\} \\ &\quad \times \sum_{i_1=0}^{+\infty} \cdots \sum_{i_{st-s-1}=0}^{+\infty} \sum_{i_{st-s+1}=0}^{+\infty} \cdots \sum_{i_{st-2}=0}^{+\infty} \exp\left\{\sum_{k=1}^{st-1} (i_k \ln \lambda_{s',t'} - \lambda_{s',t'} - \ln i_k!)\right\} \\ &= c_{s,t} \sum_{i_1,\ldots,i_{st-1}=0}^{+\infty} \exp\left\{a_s i_{st-1} + b_{s,s-1} i_{st-s} + \sum_{k=1}^{st-1} (i_k \ln \lambda_{s',t'} - \lambda_{s',t'} - \ln i_k!)\right\}, \end{aligned}$$

which leads to (12).

Using the total expectation formula, variance property of the Poisson distribution and the previous result, we find

$$\begin{aligned} D\{x_{s,t}\} &= E\big\{x_{s,t}^2\big\} - E^2\{x_{s,t}\} \\ &= E\big\{E\big\{x_{s,t}^2|x_{s,t-1}, x_{s-1,t}\big\}\big\} - E^2\big\{E\{x_{s,t}|x_{s,t-1}, x_{s-1,t}\}\big\} \\ &= E\big\{\lambda_{s,t}^2 + \lambda_{s,t}\big\} - E^2\{\lambda_{s,t}\} = E\big\{\lambda_{s,t}^2\big\} + E\{\lambda_{s,t}\} - E^2\{\lambda_{s,t}\} \\ &= E\big\{\lambda_{s,t}^2\big\} + E\{x_{s,t}\}\big(1 - E\{x_{s,t}\}\big), \end{aligned}$$

where

$$E\big\{\lambda_{s,t}^2\big\} = c_{s,t}^2 \sum_{i_1,\ldots,i_{st-1}=0}^{+\infty} \exp\left\{2a_s i_{st-1} + 2b_{s,s-1} i_{st-s} + \sum_{k=1}^{st-1} (i_k \ln \lambda_{s',t'} - \lambda_{s',t'} - \ln i_k!)\right\},$$

and it leads to (13). □

Lemma 3. *If the parameters of the model* (2)–(5) *satisfy the conditions* $a_s < 0$, $b_{s,s-1} < 0$, $s \in S$, $\varphi_k(t) \to a_k \neq 0$, $a_k \in R$, $k = 1, \dots, K$ *if* $t \to +\infty$ *and* $f(x_{s,t-1}, x_{s-1,t}) > 0$ *is some polynomial function of variables* $x_{s,t-1}$, $x_{s-1,t}$, *then*

$$0 < E\left\{\lambda_{s,t} f(x_{s,t-1}, x_{s-1,t})\right\} < +\infty.$$

Proof. Calculate $E\{\lambda_{s,t} f(x_{s,t-1}, x_{s-1,t})\}$ using the results of Corollary 4:

$$\begin{aligned} &E\left\{\lambda_{s,t} f(x_{s,t-1}, x_{s-1,t})\right\} \\ &\quad = c_{s,t} E\left\{\exp\{a_s x_{s,t-1} + b_{s,s-1} x_{s-1,t}\} f(x_{s,t-1}, x_{s-1,t})\right\} \\ &\quad = c_{s,t} \sum_{i_1,\dots,i_{st-1}=0}^{+\infty} \exp\{a_s i_{st-1} + b_{s,s-1} i_{st-s}\} f(i_{st-1}, i_{st-s}) \\ &\qquad \times \exp\left\{\sum_{k=1}^{st-1} (i_k \ln \lambda_{s',t'} - \lambda_{s',t'} - \ln i_k!)\right\}, \end{aligned}$$

where $c_{s,t} = \exp\{\beta_s z_{s,t} + \sum_{k=1}^{K} \gamma_{s,k} \varphi_k(t)\}$. Since under the condition of this lemma $a_s < 0$, $b_{s,s-1} < 0$, $s \in S$, then $\lambda_{s',t'} \to 0$ if $i_k \to +\infty$, $k = 1, 2, \dots, st-1$, and the series $\sum_{i_{st-1}, i_{st-s}=0}^{+\infty} \exp\{a_s i_{st-1} + b_{s,s-1} i_{st-s}\} f(i_{st-1}, i_{st-s})$ converges uniformly. Therefore, $\exp\{\sum_{k=1}^{st-1} (i_k \ln \lambda_{s',t'} - \lambda_{s',t'} - \ln i_k!)\} \to 0$ if $i_k \to +\infty$, $k = 1, 2, \dots, st-1$, and the sequence of partial sums $\sum_{i_{st-1}, i_{st-s}=0}^{N} \exp\{a_s i_{st-1} + b_{s,s-1} i_{st-s}\} f(i_{st-1}, i_{st-s})$ is bounded. Then by Dirichlet's feature for the convergence of a series we obtain that the infinite series converges, that is $E\{\lambda_{s,t} f(x_{s,t-1}, x_{s-1,t})\} < +\infty$. Since $\varphi_k(t) \to a_k \neq 0$, $a_k \in R$ if $t \to +\infty$, $k = 1, \dots, K$ and $f(x_{s,t-1}, x_{s-1,t}) > 0$, then

$$\begin{aligned} &E\left\{\lambda_{s,t} f(x_{s,t-1}, x_{s-1,t})\right\} \\ &\quad = c_{s,t} \sum_{i_1,\dots,i_{st-1}=0}^{+\infty} \exp\{a_s i_{st-1} + b_{s,s-1} i_{st-s}\} f(i_{st-1}, i_{st-s}) \\ &\qquad \times \exp\left\{\sum_{k=1}^{st-1} (i_k \ln \lambda_{s',t'} - \lambda_{s',t'} - \ln i_k!)\right\} > 0. \end{aligned}$$

□

Corollary 5. *Under conditions of Lemma* 3 *expectation for* $x_{s,t}$, $s \in S$, $t \in N$, *is finite*: $0 < E\{x_{s,t}\} < +\infty$.

Proof. Proof follows from Lemmas 2 and 3. □

4. Statistical Estimation of Parameters for the Poisson Conditional Autoregressive Model

Introduce the notation: $\theta_s = (a_s, b_s', \beta_s, \gamma_s')' \in R^{2+K+K_s}$, $\theta = (\theta_1', \ldots, \theta_n')' \in R^D$ is the composite vector of D unknown parameters.

Theorem 2. *The log-likelihood function for the model* (2)–(4) *under the observed spatio-temporal data* $\{X_t : t = 1, 2, \ldots, T\}$ *takes the separable additive form:*

$$l(\theta) = \sum_{s=1}^{n} l_s(\theta_s), \qquad l_s(\theta_s) = \sum_{t=1}^{T} (-\lambda_{s,t} + x_{s,t} \ln \lambda_{s,t} - \ln x_{s,t}!) \tag{14}$$

where $\lambda_{s,t}$ *is defined by* (3).

Proof. Use the generalized formula for multiplying probabilities and properties of the Markov chain defined in Theorem 1:

$$L(\theta) = P\{X_1, \ldots, X_T\} = P\{X_1\} \prod_{t=2}^{T} P\{X_t | X_{t-1}\},$$

$$P\{X_1\} = \prod_{s=1}^{n} \frac{\exp\{-\lambda_{s,1}\}\lambda_{s,1}^{x_{s,1}}}{x_{s,1}!},$$

$$P\{X_t | X_{t-1}\} = p_{ij}(t) = \prod_{s=1}^{n} \frac{\exp\{-\lambda_{s,t}\}\lambda_{s,t}^{x_{s,t}}}{x_{s,t}!}, \quad t \geqslant 2,$$

where $\lambda_{s,t}$ is defined by formula (3). Then we have the log-likelihood function

$$l(\theta) = \ln L(\theta) = \sum_{s=1}^{n} \sum_{t=1}^{T} (-\lambda_{s,t} + x_{s,t} \ln \lambda_{s,t} - \ln x_{s,t}!),$$

that coincides with (14). □

To find the maximum likelihood estimators (MLE) $\hat{\theta} = (\hat{\theta}_1', \ldots, \hat{\theta}_n')'$ for the parameters of the model we need to maximize the log-likelihood function (14):

$$l(\theta) \to \max_{\theta}. \tag{15}$$

Based on the model (2)–(4) and Theorem 2 the problem (15) splits into n maximization problems:

$$l_s(\theta_s) \to \max_{\theta_s}, \quad s \in S. \tag{16}$$

A necessary condition for a local maximum in (16) is

$$\nabla_{\theta_s} l_s(\theta_s) = \mathrm{O}_{2+K+K_s}, \tag{17}$$

where O_{2+K+K_s} is a vector with $2 + K + K_s$ zero elements.

To construct a numerical algorithm for computation of MLE $\hat{\theta}$ we give some auxiliary results.

Lemma 4. *For the model* (2)–(4) *the column vector of first-order derivatives and the matrix of second-order derivatives* (*with respect to the parameter* θ_s) *of the function* $l_s(\theta_s)$, $s \in S$, *have the form:*

$$\nabla_{\theta_s} l_s(\theta_s) = \sum_{t=1}^{T} (-\lambda_{s,t} Y_{s,t} + x_{s,t} Y_{s,t}), \qquad \nabla^2_{\theta_s} l_s(\theta_s) = -\sum_{t=1}^{T} \lambda_{s,t} Y_{s,t} Y'_{s,t},$$

where

$$\begin{aligned} &Y_{s,t} = \big(x_{s,t-1} I\{t > 1\}, x_{j_1,t}, \dots, x_{j_{K_s},t}, z_{s,t}, \varphi_1(t), \dots, \varphi_K(t)\big)', \\ &Y_{s,t} \in R^{2+K+K_s}, \quad s \in S,\ t \in N. \end{aligned} \tag{18}$$

If $|\sum_{t=1}^{T} \lambda_{s,t} Y_{s,t} Y'_{s,t}| \neq 0$, *the matrix* $\nabla^2_{\theta_s} l_s(\theta_s)$ *is negative defined.*

Proof. By (14), (18) and properties of matrix differentiation we have:

$$\begin{aligned} &\lambda_{s,t} = \exp\{\theta'_s Y_{s,t}\}, \\ &\nabla_{\theta_s} l_s(\theta_s) = \sum_{t=1}^{T} (-\nabla_{\theta_s} \lambda_{s,t} + x_{s,t} \nabla_{\theta_s} \ln \lambda_{s,t}) = \sum_{t=1}^{T} (-\lambda_{s,t} Y_{s,t} + x_{s,t} Y_{s,t}), \\ &\nabla^2_{\theta_s} l_s(\theta_s) = \nabla_{\theta_s}\big(\nabla_{\theta_s} l_s(\theta_s)\big) = -\sum_{t=1}^{T} \nabla_{\theta_s} \exp\{\theta'_s Y_{s,t}\} Y_{s,t} = -\sum_{t=1}^{T} \lambda_{s,t} Y_{s,t} Y'_{s,t}. \end{aligned} \tag{19}$$

Since $|\sum_{t=1}^{T} \lambda_{s,t} Y_{s,t} Y'_{s,t}| \neq 0$, and for arbitrary vector $z \in R^{2+K+K_s}$

$$\begin{aligned} z' \nabla^2_{\theta_s} l_s(\theta_s) z &= -z' \left(\sum_{t=1}^{T} \lambda_{s,t} Y_{s,t} Y'_{s,t} \right) z = -\sum_{t=1}^{T} \lambda_{s,t} z' Y_{s,t} Y'_{s,t} z \\ &= -\sum_{t=1}^{T} \lambda_{s,t} (z' Y_{s,t})(z' Y_{s,t})' \leqslant 0, \end{aligned}$$

where the equality is attained only if z is the zero vector, then the property of negative definiteness of matrix is satisfied: $\nabla^2_{\theta_s} l_s(\theta_s) \prec 0$. □

Let θ_s^* be some solution of Eq. (17), then the sufficient condition for a local maximum (16) at the point θ_s^* is the condition of negative definiteness of the matrix of second derivatives at this point, which is satisfied because of Lemma 4 results.

We solve (17) numerically using the Newton iterative method which has the quadratic convergence rate. For this method the $(k+1)$th iteration is:

$$\theta_s^{(k+1)} = \theta_s^{(k)} - \left(\nabla^2_{\theta_s} l_s\left(\theta_s^{(k)}\right)\right)^{-1} \cdot \nabla_{\theta_s} l_s\left(\theta_s^{(k)}\right), \quad k = 0, 1, 2, \ldots, \tag{20}$$

where $\theta_s^{(k)}$ is an approximation of the MLE $\hat{\theta}_s$ on the kth step. The iterative process stops if $\|\nabla_{\theta_s} l_s(\theta_s^{(k)})\| < \varepsilon$, where $\varepsilon \geqslant 0$ is a given small quantity which determines the computation accuracy for the MLE; in this case we take the statistic $\hat{\theta}_s = \theta_s^{(k+1)}$ as the MLE.

Problem (16) can define several local maxima, so to find the global maximum of $l_s(\theta_s)$ we apply (20) several times with different initial values, and then we choose the solution of (16) with the greatest value of the likelihood function as the estimate $\hat{\theta}_s$. As one of the vectors of the initial values $\theta_s^{(0)}$ for the iterative algorithm (20) we propose to take

$$\theta_s^{(0)} = \left(\sum_{t=1}^{T} Y_{s,t} Y'_{s,t}\right)^{-1} \left(\sum_{t=1}^{T} v_{s,t} Y_{s,t}\right),$$

where $Y_{s,t}$ is defined in (18), $v_{s,t} = \ln x_{s,t}$. Other vectors of initial values we propose to generate randomly in the set $\{\theta : \|\theta_s^{(0)} - \theta\| < M\}$, where M is a given radius of the ball.

Theorem 3. *If parameters of the model* (2)–(5) *satisfies the conditions* $a_s < 0$, $b_{s,s-1} < 0$, $s \in S$, *and* $\varphi_k(t) \to a_k \neq 0$ $(k = 1, \ldots, K)$ *if* $t \to +\infty$, *then the constructed maximum likelihood estimators* $\{\hat{\theta}_s\}$ *are consistent, effective, asymptotically normal and asymptotically unbiased under the asymptotics with* $T \to +\infty$.

Proof. Check regularity conditions R_1–R_4 (Borovkov, 1998). Since the function $p_s(X_t; \theta_s) = \Pi(x_{s,t}; \lambda_{s,t}) = \frac{\exp\{-e^{\theta'_s Y_{s,t}} + x_{s,t}\theta'_s Y_{s,t}\}}{x_{s,t}!}$ is twice differentiable with respect to θ_s for all $X_t \in N_0^n$, then regularity condition R_1 is satisfied.

By Lemma 4 we have

$$\frac{\partial \ln p_s(X_t; \theta_s)}{\partial \theta_{sj}} = -\lambda_{s,t} y^j_{s,t} + x_{s,t} y^j_{s,t},$$

$$\frac{\partial^2 \ln p_s(X_t; \theta_s)}{\partial \theta_{sj} \partial \theta_{sk}} = -\lambda_{s,t} y^j_{s,t} y^k_{s,t}, \quad j, k = 1, \ldots, 2 + K + K_s, \ t \in N,$$

where θ_{sj} is the jth component of the vector θ_s, $y^j_{s,t}$ is the jth component of the vector $Y_{s,t}$. Using the total expectation formula and properties of the Poisson distribution, we perform the following transformation:

$$E\left\{\left(\frac{\partial \ln p_s(X_t;\theta_s)}{\partial\theta_{sj}}\right)^2\right\} = E\left\{(-\lambda_{s,t}+x_{s,t})^2\left(y_{s,t}^j\right)^2\right\}$$
$$= E\left\{E\left\{(-\lambda_{s,t}+x_{s,t})^2\left(y_{s,t}^j\right)^2\middle|F_{\bar{s},<t}\right\}\right\}$$
$$= E\left\{E\left\{\left(-E\{x_{s,t}|F_{\bar{s},<t}\}+x_{s,t}\right)^2\middle|F_{\bar{s},<t}\right\}\left(y_{s,t}^j\right)^2\right\}$$
$$= E\left\{D\{x_{s,t}|F_{\bar{s},<t}\}\left(y_{s,t}^j\right)^2\right\} = E\left\{\lambda_{s,t}\left(y_{s,t}^j\right)^2\right\}.$$

So we have

$$E\left\{\left(\frac{\partial \ln p_s(X_t;\theta_s)}{\partial\theta_{sj}}\right)^2\right\} = E\left\{\lambda_{s,t}\left(y_{s,t}^j\right)^2\right\},$$
$$E\left\{\left|\frac{\partial^2 \ln p_s(X_t;\theta_s)}{\partial\theta_{sj}\partial\theta_{sk}}\right|\right\} = E\left\{\lambda_{s,t}y_{s,t}^j y_{s,t}^k\right\}, \quad j,k = 1,\dots,2+K+K_s,\ t\in N.$$

Evaluate the value of $E\{\lambda_{s,t}y_{s,t}^j y_{s,t}^k\} = E\{\lambda_{s,t}f_{jk}(x_{s,t-1},x_{s-1,t})\}$, where $f_{jk}(x_{s,t-1}, x_{s-1,t}) = y_{s,t}^j y_{s,t}^k$. Since under conditions of this theorem $a_s < 0$, $b_{s,s-1} < 0$, $s \in S$, then by Lemma 3 we get $E\{\lambda_{s,t}y_{s,t}^j y_{s,t}^k\} < +\infty$, therefore, regularity condition R_2 is satisfied. Regularity condition R_3 is satisfied due to the fact that the range of the vector X_t does not depend on the parameters θ_s of the model (2)–(5).

Compute now the Fisher information matrix $I(s) = (i_{jk}(s))_{j,k=1}^{2+K+K_s}$ for the sample value X_t, $t = 1,\dots,T$:

$$i_{jk}(s) = E\left\{-\frac{\partial^2 \ln p_s(X_t;\theta_s)}{\partial\theta_{sj}\partial\theta_{sk}}\right\} = E\left\{\lambda_{s,t}y_{s,t}^j y_{s,t}^k\right\}, \quad j,k = 1,\dots,2+K+K_s.$$

Since $\varphi_k(t) \to a_k \neq 0$ $(k = 1,\dots,K)$ if $t \to +\infty$, then $f_{jk}(x_{s,t-1},x_{s-1,t}) > 0$; by Lemma 3 we have $i_{jk}(s) = E\{\lambda_{s,t}y_{s,t}^j y_{s,t}^k\} > 0$, therefore, $|I(s)| \neq 0$, and regularity condition R_4 also is satisfied.

Thus, according to Borovkov (1998), constructed estimators $\{\hat{\theta}_s\}$ if $T \to +\infty$ are consistent, effective, asymptotically normal and asymptotically unbiased. □

5. Statistical Forecasting

5.1. *Optimal Forecasting Statistic*

Consider now the problem of forecasting of the future state $X_{T+\tau}$ in $\tau \geqslant 1$ steps ahead based on observations until the time $t = T$ inclusively: $X_1,\dots,X_T$. Denote some forecasting statistic $\hat{X}_{T+\tau} = g_\tau(X_1,\dots,X_T;\theta)$, where θ is the vector of true values of the model parameters. Since X_t is the Markov chain, as in Kharin (2013b), we use the probability of the forecast error as a measure of forecast performance:

$$r(\tau) = P\{\hat{X}_{T+\tau} \neq X_{T+\tau}\}.$$

Theorem 4. *Under the model* (2)–(4) *the optimal forecasting statistic minimizing the probability of the forecast error* $r(\tau)$ *is*

$$\hat{X}_{T+\tau} = \arg\max_{l_j \in L} h_{i,j}(T, T+\tau), \tag{21}$$

where i *is determined by the equation* $l_i = X_T$. *In this case the minimum of the forecast error*

$$r(\tau) = 1 - \sum_{i=0}^{+\infty} \sum_{k=0}^{+\infty} \max_{l_j \in L} h_{i,j}(T, T+\tau) p_k h_{k,i}(1, T). \tag{22}$$

Proof. The optimal forecasting statistic that minimizes the probability of the forecast error is defined by the following condition (Kharin, 2013b):

$$\hat{X}_{T+\tau} = \arg\max_{l_j \in L} P\{X_{T+\tau} = l_j | X_1, X_2, \ldots, X_T\}. \tag{23}$$

Since dependence on the time in model (2)–(4) is determined by Markov chain the forecasting statistic (23) depends only on X_T:

$$\hat{X}_{T+\tau} = \arg\max_{l_j \in L} P\{X_{T+\tau} = l_j | X_T\} = \arg\max_{l_j \in L} h_{i,j}(T, T+\tau),$$

that coincides with (21).

Calculate the probability of the forecast error for the forecasting statistic (21), using the total probability formula:

$$\begin{aligned} r(\tau) &= P\{\hat{X}_{T+\tau} \neq X_{T+\tau}\} = 1 - P\{\hat{X}_{T+\tau} = X_{T+\tau}\} \\ &= 1 - \sum_{i=0}^{+\infty} P\{\hat{X}_{T+\tau} = X_{T+\tau} | X_T = l_i\} P\{X_T = l_i\}. \end{aligned} \tag{24}$$

By (21) and the Markov property we have

$$\begin{aligned} &P\{\hat{X}_{T+\tau} = X_{T+\tau} | X_T = l_i\} \\ &\quad = P\left\{\arg\max_{l_j \in L} h_{i,j}(T, T+\tau) = X_{T+\tau} | X_T = l_i\right\} \\ &\quad = \sum_{k=0}^{+\infty} h_{i,k}(T, T+\tau) I\left\{\arg\max_{l_j \in L} h_{i,j}(T, T+\tau) = l_k\right\} = \max_{l_j \in L} h_{i,j}(T, T+\tau). \end{aligned} \tag{25}$$

Substituting (25) into (24), we obtain:

$$\begin{aligned} r(\tau) &= 1 - \sum_{i=0}^{+\infty} \max_{l_j \in L} h_{i,j}(T, T+\tau) \sum_{k=0}^{+\infty} p_k h_{k,i}(1, T) \\ &= 1 - \sum_{i=0}^{+\infty} \sum_{k=0}^{+\infty} \max_{l_j \in L} h_{i,j}(T, T+\tau) p_k h_{k,i}(1, T), \end{aligned}$$

that coincides with (22). □

Note that if there is more than one maximum element in (21), then we have equivalent forecasts with the same error probability.

5.2. *Forecasting in Case of Unknown Parameters*

In case of parametric prior uncertainty we construct the forecasting statistic using "the plug-in" principle (Kharin, 2013b, 2011):

$$\tilde{X}_{T+\tau} = g_\tau(X_1, \ldots, X_T; \hat{\theta}),$$

where $\hat{\theta}$ is the MLE for the parameters of the model (2)–(4) based on observations $\{X_1, \ldots, X_T\}$. The "plug-in" forecasting statistic based on (21)

$$\tilde{X}_{T+\tau} = \arg\max_{l_j \in L} \tilde{h}_{ij}(T, T+\tau), \tag{26}$$

where i is determined by the equation $l_i = X_T$, $\tilde{h}_{i,j}(T, T+\tau)$ depends on $\hat{\theta}$ according to Theorem 1 and Corollary 1.

Theorem 5. *The minimal probability of the forecast error for the forecasting statistic* (26) *has the following form*:

$$\tilde{r}(\tau) = 1 - \sum_{i_t,\ldots,i_T=0}^{+\infty} h_{i_T,m}(T, T+\tau) \prod_{s=1}^{n} \prod_{t=1}^{T} \frac{\exp\{-\lambda_{s,t}\}\lambda_{s,t}^{l_{s,i_t}}}{l_{s,i_t}!},$$

where m is determined by the equation $l_m = \tilde{X}_{T+\tau}$.

Proof. Calculate the probability of the forecast error using the total probability formula:

$$\begin{aligned} \tilde{r}(\tau) &= P\{\tilde{X}_{T+\tau} \neq X_{T+\tau}\} = 1 - P\{\tilde{X}_{T+\tau} = X_{T+\tau}\} \\ &= 1 - \sum_{i_t,\ldots,i_T}^{+\infty} P\{\tilde{X}_{T+\tau} = X_{T+\tau} | X_T = l_{i_T}, \ldots, X_1 = l_{i_1}\} \\ &\quad \times P\{X_T = l_{i_T}, \ldots, X_1 = l_{i_1}\}. \end{aligned} \tag{27}$$

According to (26):

$$P\{\tilde{X}_{T+\tau} = X_{T+\tau} | X_T = l_{i_T}, \ldots, X_1 = l_{i_1}\}$$
$$= \sum_{k=0}^{+\infty} h_{i_T,k}(T, T+\tau) I\left\{\arg\max_{l_j \in L} \tilde{h}_{i_T,j}(T, T+\tau) = l_k\right\} = h_{i_T,m}(T, T+\tau), \quad (28)$$

where $l_m = \tilde{X}_{T+\tau}$. We have from Theorem 2 that $P\{X_T = l_{i_T}, \ldots, X_1 = l_{i_1}\} = \prod_{s=1}^{n} \prod_{t=1}^{T} \frac{\exp\{-\lambda_{s,t}\}\lambda_{s,t}^{l_{s,i_t}}}{l_{s,i_t}!}$, then substituting (28) into (27) we obtain:

$$\tilde{r}(\tau) = 1 - \sum_{i_t=0,\, t=1,\ldots,T}^{+\infty} h_{i_T,m}(T, T+\tau) \prod_{s=1}^{n} \prod_{t=1}^{T} \frac{\exp\{-\lambda_{s,t}\}\lambda_{s,t}^{l_{s,i_t}}}{l_{s,i_t}!}.$$

□

6. Results of Computer Experiments

6.1. *Experiments on Simulated Data*

We consider the model (2)–(5) with the following values of parameters: $n = 3$, $K = 1$, $S = \{1, 2, 3\}$, $U(1) = \emptyset$, $U(2) = \{1\}$, $U(3) = \{2\}$, $\varphi_1(t) = 1$, $\beta_s \equiv 0$, $\theta_1 = (-0.1, 1.4)'$, $\theta_2 = (-0.2, -0.1, 1.2)'$, $\theta_3 = (-0.15, -0.2, 1.5)'$, $T = 20$, $\tau \in \{1, 2, \ldots, 5\}$. Statistical estimators of the parameters obtained by $T = 20$ observations are

$$\hat{\theta}_1 = (-0.04, 1.4)', \qquad \hat{\theta}_2 = (-0.4, -0.04, 1.16)', \qquad \hat{\theta}_3 = (-0.01, -0.22, 1.24)'.$$

Figure 4 plots dependence of the mean square error of the parameter estimators

$$\hat{\upsilon} = \hat{E}\{\|\hat{\theta} - \theta\|^2\} = \frac{1}{M} \sum_{k=1}^{M} \|\hat{\theta}^{(k)} - \theta\|^2$$

on the observation time T (T varies from 20 to 240) that was estimated by $M = 5000$ Monte Carlo replications, where $\hat{\theta}^{(k)}$ is the estimate for the kth realization. Figure 4 illustrates the property of consistency of the MLE $\hat{\theta}$.

To illustrate performance of the forecasting statistic (26), we use the relative forecast error:

$$\kappa = \frac{1}{\tau n} \sum_{t=1}^{\tau} \sum_{s=1}^{n} \kappa_{s,t}, \qquad \kappa_{s,t} = \left| \frac{\hat{x}_{s,t} - x_{s,t}}{\overline{x}_s} \right|, \quad (29)$$

where $\overline{x}_s = \frac{1}{T} \sum_{t=1}^{T} x_{s,t}$, $\{x_{s,t} : t = 1, \ldots, T;\ s \in S\}$ is the observed data, which is used to construct the forecasting statistic, $\{x_{s,t} : t = T+1, \ldots, T+\tau;\ s \in S\}$ are values which are needed to be forecasted, $\{\hat{x}_{s,t} : t = T+1, \ldots, T+\tau;\ s \in S\}$ are estimates calculated

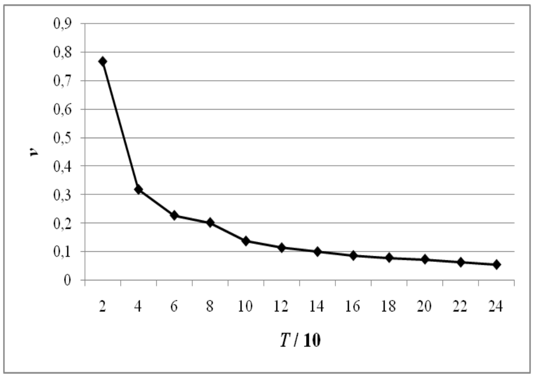

Fig. 4. Dependence of the mean square risk on T.

Table 1
Values of the relative forecast error for the simulated data.

$T = 20$, $\tau = 1$	$T = 20$, $\tau = 3$	$T = 20$, $\tau = 5$
0.09	0.22	0.35

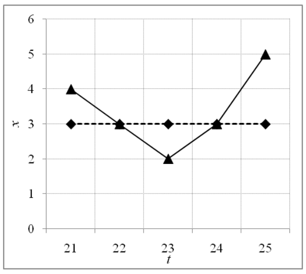

Fig. 5. Forecasting for $s = 1$.

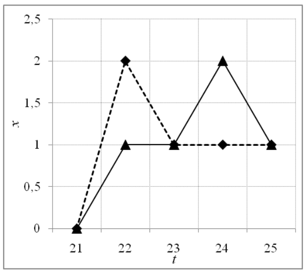

Fig. 6. Forecasting for $s = 2$.

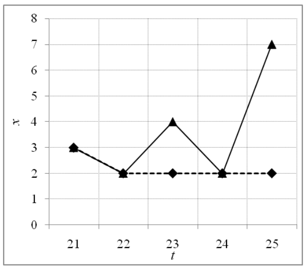

Fig. 7. Forecasting for $s = 3$.

Simulated data Optimal forecast

by the forecasting statistic (26). The values of the relative forecast error (29) are presented in the Table 1.

Figures 5–7 show simulated data for 3 sites and computed "plug-in" forecasts in $\tau = 5$ steps ahead at future time points: $t \in \{21, 22, 23, 24, 25\}$.

6.2. *Experiments on Real Data*

Experiments were carried out on real data that describes the incidence rate of children leukemia in 3 sites ($n = 3$) of Republic of Belarus for 25 years ($T = 25$). We consider the model (2)–(5) with the following values of parameters: $K = 1$, $S = \{1, 2, 3\}$, $U(1) = \emptyset$, $U(2) = \{1\}$, $U(3) = \{2\}$, $\varphi_1(t) = 1$, $\beta_s \equiv 0$. Statistical estimators of the parameters ob-

Table 2
Values of the relative forecast error for the real data.

$T = 20, \tau = 5$	$T = 22, \tau = 3$	$T = 24, \tau = 1$
0.33	0.32	0.35

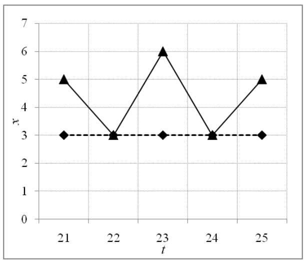

Fig. 8. Forecasting for $s = 1$.

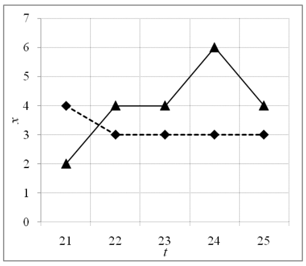

Fig. 9. Forecasting for $s = 2$.

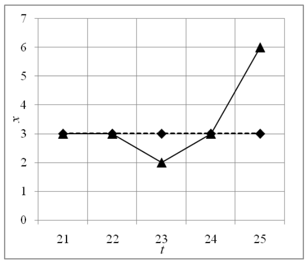

Fig. 10. Forecasting for $s = 3$.

Real data Optimal forecast

tained by $T = 25$ observations are

$$\hat{\theta}_1 = (-0.06, 1.5)', \quad \hat{\theta}_2 = (-0.01, -0.02, 1, 51)', \quad \hat{\theta}_3 = (0.07, 0.05, 0.82)'.$$

To study the performance of the constructed forecasts for the considered model, experiments were conducted on real data for different sizes of learning samples ($T = 20, 22, 24$). The values of the relative forecast error (29) for different values of τ are presented in the Table 2.

Figures 8–10 show real data and computed "plug-in" forecasts in $\tau = 5$ steps ahead at future time points: $t \in \{21, 22, 23, 24, 25\}$.

7. Conclusion

The Poisson conditional autoregressive model of spatio-temporal data is developed. It is proved that under the model the observed process is the nonhomogeneous vector Markov chain with countable space of states. Probabilistic properties of this model are studied: the formulae for calculation of the one-step transition probability matrix, the current probability distribution in a separate site, expectation and variance for each component of Markov chain are given. An algorithm for computing the maximum likelihood estimators is developed; asymptotic properties of estimators are studied. The forecasting statistic that minimizes the probability of the forecasting error is built. The forecast error is calculated; "plug-in" forecasting statistic is constructed in the case of unknown parameters. The computer experiments are carried out on simulated and real medical data.

Finally, let us mention, that the authors intend to use the developed in this paper model (2)–(4) for robust clustering as in Kharin and Zhuk (1998), robust forecasting as in Kharin (2005), Kharin and Voloshko (2011), and for robust sequential hypotheses testing as in Kharin (2013a).

Acknowledgements. Research is partially supported by the project ISTC-1910 with the International Science and Technology Center.

References

Borovkov, A.A. (1998). *Mathematical Statistics*. Gordon & Breach, New York, 592 pp.

Fokianos, K., Tjøstheim, D. (2012). Nonlinear Poisson autoregression. *Annals of the Institute of Statistical Mathematics*, 64(6), 1205–1225.

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

Gelfand, A.E., Diggle, P.J., Fuentes, M., Guttorp, P. (2010). *Handbook of Spatial Statistics*. Taylor & Francis, Boca Raton. 608 pp.

Grob-KlubMann, A., Hautsch, N. (2013). Predicting bid-ask spreads using long-memory autoregressive conditional Poisson models. *Journal of Forecasting*, 32(8), 724–742.

Kang, J., Lee, S. (2014). Parameter change test for Poisson autoregressive models. *Scandinavian Journal of Statistics: Theory and Applications*, 41(4), 1136–1152.

Kemeny, J.G., Snell, J.L. (1976). *Finite Markov Chains*. Springer, New York, 224 pp.

Kharin A. Y. (2005). Robust Bayesian prediction under distortions of prior and conditional distributions. *Journal of Mathematical Sciences*, 126 (1), 992–997.

Kharin Y. (2011). Robustness of the mean square risk in forecasting of regression time series. *Communications in Statistics – Theory and Methods*, 40(16), 2893-2906.

Kharin A. Y. (2013a). Robustness of sequential testing of hypotheses on parameters of M-valued random sequences. *Journal of Mathematical Sciences*, 189(6), 924–931.

Kharin Y. (2013b). *Robustness in Statistical Forecasting*. Springer, New York, 350 pp.

Kharin Y., Voloshko V.A. (2011). Robust estimation of AR coefficients under simultaneously influencing outliers and missing values. *Journal of Statistical Planning and Inference*, 141(9), 3276–3288.

Kharin Y., Zhuk E. (1998). Filtering of multivariate samples containing outliers for clustering. *Pattern Recognition Letters*, 19, 1077–1085.

Mariella, L., Tarantino, M. (2010). Temporal conditional auto-regressive model: a new autoregressive matrix. *Austrian Journal of Statistics*, 3, 223–244.

Pragarauskaite J., Dzemyda G. (2013). Markov models in analysis of frequent patterns in financial data. *Informatica*, 24(8), 87–102.

Quintian H., Calvo-Rolle J.L., Corchado E. (2014). A hybrid regression system based on local models for solar energy prediction. *Informatica*, 25(2), 265–282.

Rodrigues, E.R. (2013). *Application of Discrete-Time Markov Chains and Poisson Processes to Air Pollution Modeling and Studies*. Springer, New York, 107 pp.

Yu. Kharin, Director of the Research Institute for Applied Problems of Mathematics and Informatics, head of the Department of Mathematical Modeling and Data Analysis at Belarusian State University, Minsk, Belarus. He has DrSc in Mathematics. He is a Correspondent member of the National Academy of Sciences of Belarus. His research fields are applied mathematics and statistics, computer science and information technology. The subject of his research interests is related to the topical directions of applied mathematics and computer science: development of mathematical models, methods, algorithms and software for robust pattern recognition and statistical data analysis to create computer systems and information technologies.

M. Zhurak, junior research associate in the Research Institute for Applied Problems of Mathematics and Informatics, received MS degree in applied mathematics and computer science from the Faculty of Applied Mathematics and Computer Science of the Belarusian State University, Minsk, Belarus. Her research interests include statistical analysis of models based on spatio-temporal data.

Statistinė laiko-erdvės duomenų, grindžiamų Puasono sąlyginiu autoregresiniu modeliu, analizė

Yuriy KHARIN, Maryna ZHURAK

Pasiūlytas Puasono sąlyginis autoregresinis laiko-erdvės duomenų modelis. Pateikta šio modelio Markovo savybė ir tikimybinės jo savybės. Sukonstruotas algoritmas šio modelio didžiausio tikėtinumo vertinimui. Nustatyta optimali prognozavimo statistika minimizuojanti prognozės klaidos tikimybę. Pateikti eksperimentų su modeliuotais ir realiais duomenimis rezultatai.