STATISTICAL ANALYSIS FOR SEMIBINOMIAL CONDITIONALLY NONLINEAR AUTOREGRESSIVE TIME SERIES

YURIY KHARIN and VALERIY VOLOSHKO
Belarusian State University, Minsk, Belarus
e-mail: kharin@bsu.by

While the classical models like ARIMA for real-valued time series are well-studied, there is still a lack of adequate models for discrete time series arising in economics, finance, bioinformatics and other applications [1]. In particular, the models are needed for over- and underdispersed time series of counts $x_t \in A = \{0, 1, \ldots, N\}, t \in \mathbb{Z}$. For this purpose we propose the following semibinomial conditionally nonlinear autoregression $P$-CNAR($s$) of order $s \in \mathbb{N}$:

$$x_t | x_{t-1}, x_{t-2}, \cdots \sim P(\theta),$$

where $\theta = F(\sum_{i=1}^{m} a_i \psi_i(x_{t-1}, \ldots, x_{t-s}))$, $F : \mathbb{R} \to [0, 1]$ is some c.d.f., $\psi_i : A^s \to \mathbb{R}$ are base functions, $P$ is a family of probability distributions on $A$, parameterized by $\theta \in [0, 1]$, $a = (a_i) \in \mathbb{R}^m$ is a model parameter. We call the family $P$ semibinomial and assume that it satisfies some regularity conditions and the property $E \xi = N \theta$, $\xi \sim P(\theta)$. In particular, we focus on exponential families of the form $P\{\xi = x\} = \exp(h(x) + \eta x - \psi(\eta))$, $\eta = \eta(\theta) \in \mathbb{R}$ ($h(x) = \ln(Nx)$ corresponds to the binomial family).

Following [2], we construct a frequencies-based estimator (FBE) $\hat{a}$ and study its properties w.r.t. the classical MLE. FBE $\hat{a}$ is shown to be consistent, asymptotically normal and effective, and to have some computational advantages compared to the MLE: explicit form, fast recursive re-computability under model extension (adding new base function $\psi_{m+1}$), and less restrictive uniqueness conditions. We also propose an explicit construction for a wide class of underdispersed exponential families $P$, as well as an empirical choice of basis $\{\psi_i\}$.

Theoretical results are illustrated by computer experiments on simulated and real (genetic and financial) data.

References