Statistical Methodology 31 (2016) 73–90

Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

Autoregressive conditional negative binomial model applied to over-dispersed time series of counts

Cathy W.S. Chen [a,*], Mike K.P. So [b], Jessica C. Li [a], Songsak Sriboonchitta [c]

[a] *Department of Statistics, Feng Chia University, Taiwan*

[b] *Department of Information Systems, Business Statistics and Operations Management, Hong Kong University of Science and Technology, Hong Kong*

[c] *School of Economics, Chiang Mai University, Thailand*

ARTICLE INFO

Article history:
Received 11 September 2015
Received in revised form
31 January 2016
Accepted 1 February 2016
Available online 9 February 2016

Keywords:
Poisson autoregressive model
Negative binomial
Integer-valued time series
MCMC
Over-dispersion

ABSTRACT

Integer-valued time series analysis offers various applications in biomedical, financial, and environmental research. However, existing works usually assume no or constant over-dispersion. In this paper, we propose a new model for time series of counts, the autoregressive conditional negative binomial model that has a time-varying conditional autoregressive mean function and heteroskedasticity. The location and scale parameters of the negative binomial distribution are flexible in the proposed set-up, inducing dynamic over-dispersion. We adopt Bayesian methods with a Markov chain Monte Carlo sampling scheme to estimate model parameters and utilize deviance information criterion for model comparison. We conduct simulations to investigate the estimation performance of this sampling scheme for the proposed negative binomial model. To demonstrate the proposed approach in modelling time-varying over-dispersion, we consider two types of criminal incidents recorded by New South Wales (NSW) Police Force in Australia. We also fit the autoregressive conditional Poisson model to these two datasets. Our results demonstrate that the proposed negative binomial model is preferable to the Poisson model.

* Corresponding author.
E-mail address: chenws@mail.fcu.edu.tw (C.W.S. Chen).

http://dx.doi.org/10.1016/j.stamet.2016.02.001
1572-3127/

1. Introduction

There has been considerable and growing interest in integer-valued time series modelling. The literature has developed many methods for time series of counts, because of their wide range of applications, including those in epidemiology, finance, disease modelling, and environmental science. In the area of accident prevention, Johansson [19] studies time series counts to assess the effect of lowered speed limits on the number of road casualties. Zeger [29] and Brännäs and Johansson [2] examine the time series of monthly cases of polio in the US. In the area of crimes, Zhu et al. [33] use data on computer-aided dispatch calls, restricting their attention to a drug monthly data series. Xu et al. [28] study time series of weekly dengue cases in Singapore. Time sequence of counts, which exhibits temporal dependence, also appears in many other fields, such as statistical quality control [27] and insurance [32]. In finance, other than the applications in [3,4], counts arise in market microstructure when we look at tick-by-tick data, and the price process for a stock can be viewed as a sum of discrete price changes. The daily number of these price changes constitutes a time series of counts. Freeland and McCabe [15] model a monthly count dataset of claimants to determine wage loss benefits.

The majority of existing count models assume that the observations follow a Poisson distribution conditioned on an intensity process that drives the dynamics of the model (see [8,12–14,9,10,26]). Using a Poisson distribution implicitly implies equi-dispersion, i.e. the mean and variance are equal. However, a typical feature in a time series of counts is autocorrelation and over-dispersion, where the variance is larger than the mean. Thus, it has become increasingly popular to handle serial correlation and over-dispersion in data. Both of these features are addressed simultaneously by using an autoregressive conditional Poisson (ACP) model proposed in [18]. This model also allows unconditional over-dispersion and heteroskedasticity.

In order to highlight over-dispersion, asymmetry, structural change, and a large proportion of zeros for time series of counts, Chen and Lee [5] propose a class of generalized Poisson autoregressive models that captures asymmetric and nonlinear responses through a switching mechanism. Wang et al. [26] study the use of a Poisson distribution conditioned on an accompanying intensity process, which is equipped with a two-regime structure according to the magnitude of the lagged observations. Jung, Kukuk, and Liesenfeld [20] consider Zeger's [29] parameter-driven Poisson model with a stochastic autoregressive mean (SAM)—a simple alternative to the SAM model is the observation-driven ACP model introduced by Heinen [18].

To describe the time series count data, many papers also use the integer-valued autoregressive conditional heteroscedastic INARCH(p) and INGARCH(p, q) models, which are similar to ACP(p, q) model. Weiß [27] proposes the subfamily of integer-valued generalized autoregressive conditional heteroscedasticity INGARCH(p, 0) models where the autocorrelation structure is shown to be closely related to that of the standard INAR(p) models. Zhu and Wang [33] present a mixture of the INARCH model, which consists of a mixture of stationary or non-stationary INARCH components. The advantages of the mixture model over the single-component model include the ability to handle multimodality and non-stationary components. Fokianos et al. [13] recommend to make an inference for linear and non-linear Poisson autoregression. In the linear case, the conditional mean is linked linearly to its past values and the observed values of the Poisson process. This also applies to the conditional variance, which implies an INGARCH process.

To deal with under-dispersion, Zhu [31] offers the INGARCH model based on the generalized Poisson (GP) distribution to account for both over-dispersion and under-dispersion. To handle the conditional equi-dispersion, over-dispersion, and under-dispersion, Xu et al. [28] set up a dispersed integer-valued autoregressive conditional heteroscedastic (DINARCH) model. Xu et al. [28] extend the conditional variance of the INARCH model, by adding a new indicator to make the conditional variance more flexible, as well as by using negative binomial distribution, the double Poisson distribution of Efron [11], and the generalized Poisson distribution of Consul and Jain [7]. Zhu [30] presents a negative binomial INGARCH (NBINGARCH) to deal with both over-dispersion and potential extreme observations at the same time. Mukhopadhyay and Banerjee [24] investigate sequential negative binomial problems and illustrate the use of count data models in statistical ecology.

It may be too restrictive to have only one parameter in a Poisson distribution for both mean and variance. To accommodate time-varying over-dispersion, we propose a new model in this paper for