HERCMA 2001
E. A. Lipitakis (Editor)
2001, L.E.A.
Printed in Hellas

Maximum Likelihood Estimation For Integer Valued Time Series Models

Dimitris Karlis and Evdokia Xekalaki, *

Abstract **Integer valued time series are useful models for describing dependence structures over time for count data for which classical time series models are inappropriate. Such models have been used for a variety of applications. Difficulties arise in attempting to estimate the parameters via maximum likelihood due to the complicated expressions of the quantities involved. The purpose of this paper is to describe EM type algorithms to estimate the parameters of such models. Their derivation is based on properties of the processes that generated the data. The results are illustrated on a real set of environmental data.**

Keywords: **INAR model; SINAR model; discrete time series models; EM algorithm; environmental data;**

I. Introduction

Time series models with normal data have been considered for a long time and have appeared in standard textbooks in time series analysis. In several cases, the data are clearly non-normal, in the sense that they may be series of counts, proportions, binary outcomes, strictly positive data etc. and thus standard techniques, based on the normality assumption, are not applicable. A variety of models have been proposed for treating such types of time series data. For a comprehensive review on these models the interested reader is referred to Grunwald et al. [7].

Let X_t denote the value of a variable at time t. Let $E(X_t \mid X_{t-1}=x)$ denote the one step ahead conditional mean. The general form of an autoregressive model is

$$E(X_t \mid X_{t-1}=x) = ax+\lambda$$

for real a, λ suitably chosen in order to ensure the stationarity of the series.

In this paper, attention is focused on autore-gressive time series models concerning count data. Usually, such series are referred to as ***integer-valued time series***. The above representation is not suitable for discrete random variables and modified versions of it have been considered in the literature. These are known *as integer-valued autoregressive time series* and are briefly discussed in section *II*. Section *III* concentrates on members of this broad family, based on the Poisson distribution and its alternatives. In section *IV* an EM type algorithm based on the latent structure of autoregressive models is provided to facilitate the estimation task. An application to real data is given in section *V*. The results are briefly discussed in section *VI*.

II. The INAR(1) process

The integer valued autoregressive process of order 1 (INAR(1)) is a sequence of random variables $\{X_t : t=0,\pm1,\pm2,...\}$ which can take only integer values. The INAR(1) process is based on the assumption that the value of the process at time t, denoted as X_t, consists of two parts. The first part is comprised by the survivors of the elements of the process at the preceding point in time *t-1*, denoted by X_{t-1}, each with a probability of survival equal to α . The second part consists of the elements which entered the system in the interval $(t-1,t]$, usually termed as *innovations.* More formally, we can write that

$$X_t = a\circ X_{t-1}+R_t \qquad (1)$$

with the operator '$\circ$' defined by

$$a\circ X = \sum_{i=1}^{X} Y_i = Y\,,$$

where Y_i are independently and identically distributed Bernoulli random variables with $P(Y_i=1)=a=1-P(Y_i=0)$. This operator is known as the *binomial thinning operator.* In fact, the random variable Y defined as above has a binomial distribution with index parameter X and probability of success α.

Definition (1) is due to McKenzie [11] and mimics the well-known autoregressive model for normal data modified to suit the discrete nature of the data. By specifying the distributional form of the innovation term R_t, a large number of different models can arise.

Several authors have proposed generalizations of the model. These can be produced by considering other distributional forms for the summands Y_i in the definition of the thinning operator (e.g. Al-Osh and Aly, [1]) or by considering generalizations of the thinning operator (e.g. Al-Zaid and Al-Osh, [3]).

Let μ and σ^2 denote the mean and the variance of the innovation term, respectively. Then, it can be shown that

$$E(X_t) = aE(X_{t-1})+\mu$$

and

$$Var(X_t) = a^2 Var(X_{t-1}) + a(1-a)E(X_{t-1})+\sigma^2.$$

The conditions for second order stationarity to hold are

* Department of Statistics, Athens University of Economics, 76 Patision Str., 10434 Athens

$$E(X_0)=\frac{\mu}{1-a}, \quad \text{and} \quad Var(X_0)=\frac{\mu+\sigma^2}{1-a^2}$$

The conditional mean and variance of X_t are both linear with respect to X_{t-1}. Grunwald et al. [7] provided conditions for the stationarity of a general non-Gaussian autoregressive model.

The INAR(1) process as defined by equation (1) can be viewed as a Galton-Watson branching process with immigration and relates to the notion of discrete self-decomposability defined by Steutel and van Harn [16]. Any member of the class of discrete self-decomposable distributions can be considered as a marginal distribution of an INAR(1) process. Such examples are the models of Al-Osh and Al-Zaid [2] and McKenzie [12] with Poisson and Negative Binomial marginals, respectively.

In practice, there are two different approaches for constructing INAR(1) models. The first assumes a particular form of marginal distribution and subsequently identifies the required form of the distribution of the innovations in order for stationarity to hold. The second approach starts by considering a specific form for the innovation distribution. Note that, in both cases, it is equivalent, but perhaps more cumbersome, to consider the Markov chain transition matrix between X_{t-1} and X_t.

In the sequel, attention is restricted to INAR(1) models obtained via appropriately chosen forms for the innovation distribution.

III. Some INAR(1) Models

A. The Poisson INAR(1) model.

The model was introduced independently by McKenzie [11] and Al-Osh and Al-Zaid [2]. Assuming that the innovation distribution is a Poisson distribution with parameter λ, it follows that the marginal distribution for X_t is a Poisson distribution with parameter $\frac{\lambda}{1-a}$. For this model, the stationary mean and the stationary variance coincide.

From the general definition of the INAR(1) process, one can see that the overdispersion of the series is measured by

$$ID(X_t)=\frac{Var(X_t)}{E(X_t)}=\left(1+\frac{\phi}{1+a}\right), \tag{2}$$

where $1+\phi=\frac{\sigma^2}{\mu}$. Equation (2) relates the overdispersion of the innovation distribution to that of the observed series. Thus, overdispersed (or underdispersed) innovation distributions lead to overdispersed (or underdispersed) time series. However, this overdispersion (or underdispersion) has been reduced due to the 'survival' part of the process.

B. The Mixed Poisson INAR(1) processes

In order to describe a time series with overdispersion one has to consider an overdispersed innovation distribution. The family of mixed Poisson probability distributions is a natural candidate. This family contains probability functions of the form

$$P(x)=\int \frac{e^{-\lambda}\lambda^x}{x!}dG(\lambda),$$

where $G(\lambda)$ is the mixing distribution, which can be a discrete, a continuous or a finite step distribution, i.e. a distribution with positive probability for a finite number of points. Some well known distributions belong to this class, like the negative binomial distribution.

Consider the case of a finite-step mixing distribution. A k-finite Poisson mixture probability function is given by

$$P(x)=\sum_{j=1}^{k} p_j \frac{e^{-\lambda_j}\lambda_j^x}{x!}.$$

where $\sum_{i=1}^{k} p_i=1$, $p_i>0$, for $i=1,\dots,k$ and $\lambda_1<\lambda_2<\dots<\lambda_k$ in order to ensure the identifiability of the mixture.

The INAR(1) model with a 2-finite Poisson mixture innovation has been considered by Franke and Seligmann [6], under the name *Switching INAR model*. For k=1 the model reduces to the simple Poisson INAR(1) model. A natural interpretation of a k-finite mixture is that the entire population consists of k subpopulations with different parameter values. Hence, the innovations can be considered as coming from k distinct subpopulations. A 2-finite mixture model can be regarded as being indicative of the lowest possible degree of inhomogeneity. For further details about finite mixtures, the interested reader is referred to Titterington *et al.* [17] and McLachlan and Peel [14].

In the case of a single Poisson innovation distribution, the conditional probability function $P(X_t \mid X_{t-1})$ is known to be the convolution of a Poisson and a Binomial distribution (see, e.g. Shumway and Gurland, [15]). In the more general case the conditional distribution is a k-finite mixture of the convolution of a Poisson distribution and a Binomial distribution. More details can be found in Karlis and Xekalaki [9].

IV. Estimation via EM algorithms

Estimation for the Poisson INAR(1) model has been proposed by Al-Osh and Al-Zaid [2] based on least squares estimation via the conditional mean function. Alternatively, the conditional maximum likelihood approach maximizes the likelihood conditional on the initial value x_0. Suppose that we have observed the series $\mathbf{x}_T=(X_0,X_1,X_2,\dots,X_T)$. The conditional likelihood, in the general form, is given by

$$L(\theta;\mathbf{x}_T)=\prod_{t=1}^{T} P^t(X_t=x_t \mid X_{t-1}=x_{t-1},\theta)$$

where $P^t(X_t=x_t \mid X_{t-1}=x_{t-1},\theta)$ denotes the conditional probability of X_t given X_{t-1} and θ is a vector of parameters. The superscript t indicates that the form of the conditional probability may depend on t. Direct maximization of the likelihood is not easy, especially when the dimensionality of θ increases as, for example, in the case of a k-finite Poisson mixture innovation distribution. In the sequel an EM type algorithm is provided for carrying out the estimation procedure.

A. The General EM Algorithm

The impact of the EM algorithm in maximum likelihood (ML) estimation has been tremendous (see for example [5] and [13]). Several problems, which were considered as intractable, can be solved using the easily programmable EM algorithm. Moreover, the algorithm itself admits an interesting statistical interpretation and thus it is not merely a numerical technique.

The basic idea of the EM algorithm lies in the 'missing data' principle. The algorithm is applicable in cases where the data are or can be considered incomplete. The latter implies that a latent structure exists behind the observed data for which the convolution derivation of the INAR model is quite appropriate.

Roughly speaking, one can describe the basis of the algorithm as follows. If the data were complete, then ML estimation would be an easy task, resulting even in closed form solutions. So, the algorithm consists of two steps. At the first step, the expectation step, the missing data are estimated via their conditional expectations given the observed data and the current values of the parameters (E-step). At the second step, the maximization step, the likelihood of the complete data is maximized using the expectations of the previous step (M-step).

The convolutional form of the conditional distribution of a general INAR process makes the algorithm appealing. Let X,Y,Z be three random variables such that $Z=X+Y$. In our case only Z is observable. The algorithm proceeds by estimating the missing data (i.e. X and Y) by their conditional expectations given Z at the E-step and subsequently maximizing the complete data likelihood at the M-step using the estimates for the missing data from the E-step.

B. The Poisson INAR model

In the case of the Poisson INAR model, the conditional probability distribution is the convolution of a binomial distribution with a Poisson distribution and hence

$$P(x_t \mid x_{t-1}, a, \lambda) = P(X_t = x_t \mid X_{t-1} = x_{t-1}, a, \lambda)$$
$$= \sum_{k=0}^{x_t} \frac{\exp(-\lambda)\lambda^k}{k!} \binom{x_{t-1}}{x_t - k} a^{x_t - k}(1-a)^{x_{t-1} - x_t + k}.$$

The calculation of the values of the probability function can be facilitated by a recursive scheme given in Shumway and Gurland [15].

The algorithm has to be constructed so as to estimate, at the E-step, the conditional expectations of $Y_t = a \circ X_{t-1}$ and R_t given the data and the current values of the estimates and to maximize, at the M-step, the complete likelihood. The latter is equivalent to maximizing the likelihood of a binomial distribution and the likelihood of a Poisson distribution, both of which are rather straightforward tasks. Hence the algorithm can be described as follows.

E-step: Using the current values of the estimates, say a^{old}, λ^{old}, calculate

$$s_t = E(R_t \mid x_t, x_{t-1}, a^{old}, \lambda^{old})$$
$$= \sum_{z=0}^{x_t} z \frac{\frac{\exp(-\lambda^{old})(\lambda^{old})^z}{z!} \binom{x_{t-1}}{x_t - z} (a^{old})^{x_t - z}(1-a^{old})^{x_{t-1} - x_t + z}}{P(x_t \mid x_{t-1}, a^{old}, \lambda^{old})}$$
$$= \lambda^{old} \frac{P(x_t - 1 \mid x_{t-1}, a^{old}, \lambda^{old})}{P(x_t \mid x_{t-1}, a^{old}, \lambda^{old})},$$

for $t=1, \ldots, T$.

The conditional expectation of Y_t given the data and the current values of the estimates can be determined by simple subtraction, as

$$w_t = E(Y_t \mid x_t, x_{t-1}, a^{old}, \lambda^{old}) = x_t - s_t.$$

M-Step: Update the parameter estimates using

$$\lambda^{new} = \frac{\sum_{t=1}^{T} s_t}{T} \quad \text{and} \quad a^{new} = \frac{\sum_{t=1}^{T} w_t}{\sum_{t=1}^{T} x_{t-1}}.$$

Stop iterating when some convergence criterion is satisfied, otherwise, go back to the E-step.

Clearly, the algorithm can be easily implemented in any computer machine. It has all the advantages and the disadvantages of the EM algorithm, as, for example, it provides estimates in the admissible range, it has monotonic but slow convergence etc.

Note that, in the case of conditional ML estimation, the starting point X_0 is considered known. For the case of unconditional ML estimation, one can use different values for X_0 in order to maximize the conditional likelihood via the EM algorithm considered, and then keep the value for which the global maximum was obtained as an estimate for X_0.

C. The Mixed Poisson INAR Process

For the Mixed Poisson INAR model, the algorithm is quite complicated since the form of the innovation distribution is not simple and there exist no closed form expressions for its ML estimates. However, one can proceed by combining the convolution representation and standard results for EM algorithms for finite mixture distributions. The conditional distribution is now given by

$$P_k(x_t \mid x_{t-1}, \theta) = P_k(X_t = x_t \mid X_{t-1} = x_{t-1}, \theta)$$
$$= \sum_{i=1}^{k} p_i P(X_t = x_t \mid X_{t-1} = x_{t-1}, a, \lambda_i)$$

where $\theta = (a, p_1, \ldots, p_{k-1}, \lambda_1, \ldots, \lambda_k)$ denotes the vector of parameters, $p_i > 0$, for $i=1, \ldots, k$ with $\sum_{i=1}^{k} p_i = 1$, and $\lambda_1 < \lambda_2 < \ldots < \lambda_k$. It is clear that, for $k=1$, the simple Poisson-binomial distribution arises. The EM algorithm is as follows.

E-step: Using the current values of the estimates, say θ^{old}, calculate the weight w_{ij} as the posterior probability that the i-th observation belongs to the j-th subpopulation, i.e.

$$w_{tj} = \frac{p_j^{old} P(x_t \mid x_{t-1}, a^{old}, \lambda_j^{old})}{P_k(x_t \mid x_{t-1}, \theta^{old})},$$

$j=1,\ldots,k$, $t=1,\ldots,T$. In addition, calculate the posterior expectations for the random variable R_t given the observed value x_t and assuming that it belongs to the j-th component, which is given by

$$s_{tj} = E(R_t \mid x_t, x_{t-1}, a^{old}, \lambda_j^{old}) = \lambda_j^{old} \frac{P(x_t - 1 \mid x_{t-1}, a^{old}, \lambda_j^{old})}{P(x_t \mid x_{t-1}, a^{old}, \lambda_j^{old})},$$

for $j=1,\ldots,k$, $t=1,\ldots,T$.

M-step: Update the parameter estimates using

$$\lambda_j^{new} = \frac{\sum_{t=1}^{T} w_{tj} s_{tj}}{\sum_{t=1}^{T} w_{tj}}, \quad p_j^{new} = \frac{\sum_{t=1}^{T} w_{tj}}{T}, \quad j=1,\ldots,k$$

and

$$a^{new} = \frac{\sum_{t=1}^{T} x_t - \sum_{j=1}^{k}\sum_{t=1}^{T} w_{tj} s_{tj}}{\sum_{t=1}^{T} x_{t-1}}.$$

Stop iterating when some convergence criterion is satisfied, otherwise, go back to the E-step.

More details about the algorithm can be found in Karlis and Xekalaki [9].

Note that since the number of parameters to be estimated is usually large, the algorithm is rather slow. To improve the speed of the algorithm, one may obtain some EM iterations in order to get quite close to the solution and, subsequently, locate the maximum using standard numerical techniques (e.g. Newton-Raphson). It is interesting to note that the first 5 iterations usually lead to estimates very close to the maximum.

An interesting feature of the EM algorithm is that the posterior probabilities calculated at each E-step can be used for clustering the observations, as they are the posterior probabilities that an observation may have came from each of the components.

V. An Application

Each summer, south Europe suffers from fires that destroy forest areas and can be hazardous for residential districts. Last year, huge areas of forests were consumed by fires. A controversial issue concerning fires is whether these are accidentally set, i.e. whether the situation is gathered by pure chance. The INAR(1) model seems to be a useful model for describing the number of fires. According to the model, the number of fires is partially comprised by the number of fires started in the immediately preceding days (assuming sustainment or rekindling). To this number, the number of new fires for several reasons is added. The data that are analyzed in the sequel refer to the number of fires in Greece in the period from July 1, 1998 to August 31, of the same year, thus they consist of T=62 observations. Only fires in forest districts are considered.

If pure chance were the governing factor, the Poisson INAR(1) model could be used for the description of the data. If, however, one considers that factors other than pure chance may have had some contribution to the situation, a mixed Poisson innovation distribution may be considered. Of course, these non-random factors cannot be directly linked with specific causes unless extra information becomes available. They may, however, be regarded as pertaining to the fact that forest areas differ in their susceptibility to fires depending, for instance, on the weather conditions (e.g. heat, windiness) or/and perhaps on some other uncontrollable situations linked with the appeal of their location (e.g. careless picnicking or even arsons by persons hunting for land that can be exploited as a building ground).

To these data (depicted in figure 1), a semi-parametric mixture of Poisson distributions was fitted on the assumption that the number of components is an unknown quantity that ought to be estimated from the data. Of course, even with a continuous mixing distribution, one is able to estimate just a finite version of the mixture (see, e.g. [4], [10]). The estimated number of components, indicates the number of different 'subpopulations' comprising the entire population and can thus be quite helpful for the identification of each observation.

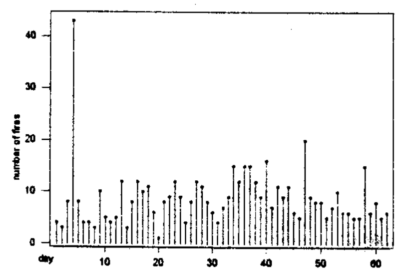

Figure 1. Data concerning the number of fires in Greece for the time period July 1998-August 1998

The mean of the data is 8.75, while the variance equals 33.72, implying a great extent of overdispersion (3.85). This is a strong indication of heterogeneity in the data leading to the conclusion that the assumption of a Poisson innovation distribution does not seem plausible. The autocorrelation coefficient of lag 1 was found to be 0.073. In order to find the semi-parametric ML estimate of the mixing distribution, the EM algorithm of the previous section was used for increasing values of k until the likelihood could not be increased any more. The parameter estimates for several values of k are contained in Table 1.

Number of Components	$\hat{a}$	$\hat{\lambda}_j$	$\hat{p}_j$	Loglikelihood
k=1	0.076	8.098	-	-207.433583
k=2	0.120	7.150 41.835	0.984 0.016	-176.718882
k=3	0.144	5.344 10.493 41.835	0.677 0.307 0.016	-171.273676
k=4	0.162	4.673 8.580 14.631 41.687	0.513 0.430 0.040 0.016	-171.168912

Table 1. The fitted Mixed Poisson INAR(1) models to the data on the number of fires in Greece

Note that, the 4-th observation, whose value is 43 (figure 1), is clearly identified as an outlier by all models with k>1. The mixing proportion is just 1/62 and remains stable with increasing k. For models with more than 4 components, the likelihood stopped increasing. Hence, one may assert that there are 4 components present. One of the components corresponds to the outlier observation as already noted. In figure 2, one can see the component to which each observation belongs according to its largest posterior probability. These probabilities become available after the termination of the EM algorithm. It is interesting to note that observations with the same value may belong to different clusters.

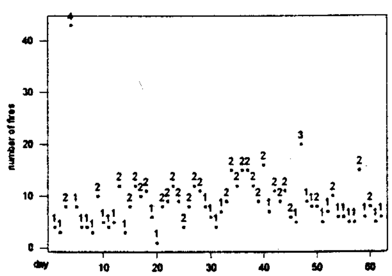

Figure 2. The components in which each observation belongs when the model with 4 components has been fitted

If one wants to test the hypothesis of a k-component model against the alternative of a $(k+1)$- component model, the standard likelihood theory fails. This is well known in mixture theory (see, e.g Lindsay, [10]). A possible solution to the problem can be based on carrying out a bootstrap test as that described in Karlis and Xekalaki [8]. A large number of samples of size T are generated from the finite mixture distribution considered under the null hypothesis using the ML estimates of the model. Then, both models with k and $k+1$ components are fitted via the EM algorithm and the values of the likelihood ratio test (LRT) statistic are compared with the observed value of the LRT statistic calculated from the data. Table 2, summarizes the results for the data considered on the basis of 1000 bootstrap replications. The p-values reported in the third column represent the proportions of times the value of the bootstrapped LRT statistic was greater than the observed value of the statistic. The results support the model with 4 components, despite the very small value of the LRT statistic.

Hypothesis tested	LRT statistic	p-value
H_0: k=1 vs H_1: k=2	61.42	0.000
H_0: k=2 vs H_1: k=3	10.89	0.000
H_0: k=3 vs H_1: k=4	0.21	0.03

Table 2. Likelihood Ratio tests for different models based on bootstrap

Model checking in non-normal autoregressive time series is not an easy task. The goodness of fit of our model was tested following Tsay's [18] bootstrap model checking approach. The sum of absolute residuals between the observed data and the one-step ahead predicted values given by $\sum_{t=1}^{T}|x_t - \hat{x}_t|$ was used as a test statistic. The one-step ahead predicted values were obtained by $\hat{x}_t = \hat{a}x_{t-1} + \hat{\mu}$, where $\hat{\mu} = \sum_{j=1}^{4}\hat{p}_j\hat{\lambda}_j$ and $\hat{a}$ denote the ML estimates of a and μ respectively. The absolute residuals were used instead of the squared residuals in order to reduce the influence of the outlier observation (x_4=43). Then, 1000 series were simulated using the ML parameters of the model and, for each series, the sum of absolute residuals was evaluated, when the hypothesized model was fitted. The observed value of the statistic for our data set (231.46) was not significant (p-value = 0.85). Therefore, the model considered seems to fit the data satisfactorily.

Unconditional ML estimates can also be obtained. All the models fitted above considered the value of X_0 as known (x_0=5). If one wants to derive unconditional ML estimates then the value x_0 of X_0 is unknown and has to be estimated. The 4 component model was fitted for several different values of X_0. In figure 3 one can see that the likelihood is maximized for x_0=0. This value is the estimate of X_0, for this particular data set.

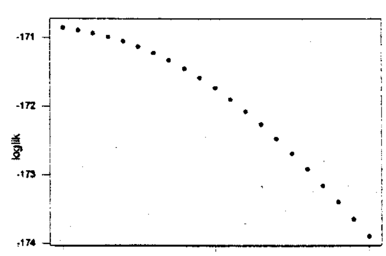

Figure 3. The unconditional loglikelihood function for different values of X_0 for the fire data

VI. Discussion

In this paper, discrete valued autoregressive models were considered. A general mixed Poisson INAR(1) model was defined and an EM type algorithm for conditional and unconditional estimation of the parameters of the models was provided. The model proposed allows for non-parametric estimation of the innovation distribution. The results were illustrated on a real dataset.

VII. References

[1] Al-Osh, M.A. and Aly, E.E.A.A. (1992). First Order Autoregressive Time Series With Negative Binomial and Geometric Marginals. *Communications in Statistics –Theory and Methods*, 21. 2483-2492.

[2] Al-Osh, M.A. and Al-Zaid, A.A. (1987). First Order Integer Valued Autoregressive Process. *Journal of Time Series Analysis*, 8, 261-275.

[3] Al-Zaid, A.A. and Al-Osh, M.A (1993). Some Autoregressive Moving Average Processes with Generalized Poisson Marginal Distributions. *Annals of the Institute Statistics Mathematics*, 45, 223-232.

[4] Bohning, D. (1999). *Computer Assisted Analysis of Mixtures (C.A.M.AN)*. Marcel and Deccer Inc, New York.

[5] Dempster, A.P., Laird, N.M. and Rubin, D. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society*, B 39, 1-38.

[6] Franke, J. and Seligmann, T. (1993). Conditional Maximum Likelihood Estimates for INAR(1) Processes and Their Application to Modeling Epileptic Seizure Counts. In *Developments in Time Series: in honor of Maurice B. Pristley, ed. T. Subba Rao: Chapman and Hall, New York.*

[7] Grunwald, G.K., Hyndman R.J., Tedesco, L. and Tweedie, R.L. (2000). Non-Gaussian Conditional Linear AR(1) Models. *Australian and New Zealand Journal of Statistics*, 42, 479-495.

[8] Karlis. D. and Xekalaki, E. (1999a). On Testing for the Number of Components in Finite Poisson Mixtures. *Annals of the Institute of Statistical Mathematics*, 51, 149-161.

[9] Karlis, D. and Xekalaki, E. (1999b). Maximum Likelihood Estimation for the Poisson-Binomial Distribution via the EM Algorithm with Applications to Discrete Time Series Models. *Technical Report No 62. Dept of Statistics, Athens University of Economics.*

[10] Lindsay, B. (1995). *Mixture Models: Theory, Geometry and Applications*. Regional Conference Series in Probability and Statistics, Vol 5, Institute of Mathematical Statistics and American Statistical Association.

[11] Mc Kenzie, E. (1985). Some Simple Models for Discrete Variable Time Series. *Water Resources Bulletin*, 21, 645-650.

[12] Mc Kenzie, E. (1986). Autoregressive Moving-Average Processes with Negative Binomial and Geometric Marginal Distributions. *Advances in Applied Probability*, 18, 679-695.

[13] McLachlan, G.J. and Krishnan, T. (1997). *The EM Algorithm and Extensions*. Wiley, New York.

[14] McLachlan, G.J. and Peel, D. (2000). *Finite Mixture Model*. Wiley, New York.

[15] Shumway, R. and Gurland, J. (1960). Fitting the Poisson Binomial Distribution. *Biometrics*, 16, 522-534.

[16] Steutel, F. and van Harn, P. (1979). Discrete Analogues of self-decomposability and Stability. *The Annals of Probability*, 7, 893-899.

[17] Titterington, D.M, Smith, A.F.M. and Makov, U.E. (1985). *Statistical Analysis of Finite Mixtures Distributions*. Wiley, New York.

[18] Tsay, R.S. (1992). Model Checking via Parametric Bootstraps in Time Series Analysis. *Applied Statistics*, 41, 1-15.