Drescher, Daniel

Working Paper

Alternative distributions for observation driven count series models

Economics working paper / Christian-Albrechts-Universität Kiel, Department of Economics, No. 2005,11

Provided in cooperation with:
Christian-Albrechts-Universität Kiel (CAU)

Alternative distributions for observation driven count series models

by Daniel Drescher

C|A|U
Christian-Albrechts-Universität Kiel
Department of Economics

Economics Working Paper

No 2005-11

Alternative distributions for observation driven count series models

Daniel Drescher
Department of Econometrics and Statistics,
Christian-Albrechts-Universität, Kiel, Germany (drescher.d@gmx.net)

20 July 2005

Abstract

Observation-driven models provide a flexible framework for modelling time series of counts. They are able to capture a wide range of dependence structures. Many applications in this field of research are concerned with count series whose conditional distribution given past observations and explanatory variables is assumed to follow a Poisson distribution. This assumption is very convenient since the Poisson distribution is simple and leads to models which are easy to implement. On the other hand this assumption is often too restrictive since it implies equidispersion, the fact that the conditional mean equals the conditional variance. This assumption is often violated in empirical applications. Therefore more flexible distributions which allow for overdispersion or underdispersion should be used. This paper is concerned with the use of alternative distributions in the framework of observation-driven count series models. In this paper different count distributions and their properties are reviewed and used for modelling. The models under consideration are applied to a time series of daily counts of asthma presentations at a Sydney hospital. This data set has already been analyzed by Davis et al. (1999, 2000). The Poisson-GLARMA model proposed by these authors is used as a benchmark. This paper extends the work of Davis et al. (1999) to distributions which are nested in either the generalized negative binomial or the generalized Poisson distribution. Additionally the maximum likelihood estimation for observation-driven models with generalized distributions is presented in this paper.

Keywords: Count series, observation-driven models, GLARMA, dicrete distributions

JEL classification: C13, C22, C25

1 Introduction

A series of observations $\{Y_t\}_{t=0}^{T}$ indexed by the time t is called a count series if $Y_t \in \mathbb{N}_0 \; \forall t$. Count series arise typically in applications where the number of certain events or happenings during a certain time period is the object of interest. Examples of count series are the number of incidences of a certain disease (poliomyelitis in the U.S.: Zeger, 1988 or asthma cases in a hospital in Campbelltown: Davis et al. 1999) and the number of transaction price movements of a certain financial instrument measured in multiples of the minimum tick size (IBM stocks at the New York Stock Exchange: Rydberg and Shephard, 2003; Henkel shares at Frankfurt Stock Exchange: Liesenfeld and Pohlmeier, 2003 and options on shares of the Bayer AG: Czado and Kolbe, 2004).

A wide variety of models for count series have been discussed in literature among others Markov chains (Cox and Miller, 1965), hidden Markov models (MacDonald and Zucchini, 1997), discrete and integer valued ARMA type models (DARMA and INARMA, Jacobs and Lewis, 1978; Alzaid and Al-Osh, 1988, 1990) and models based on a extension of the generalized linear models proposed e.g. by Zeger (1988), Shephard (1995) and Davis et al. (1999). Following the systematic introduced by Cox (1981), these models can be characterized either as observation-driven or parameter-driven. This classification is based on the theory and terminology of state-space models. State space models consist of two constituting equations referred to as observation and state equation. The observation equation constitutes the distribution of the output variable given the state variable. The state equation specifies the behaviour of the state variable. For both parameter- and observation-driven models the observation equation is the same. The differences between these models are based on the specification of the state equation. Parameter-driven models induce the serial dependence by a latent variable which evolve independently of the past observations of the outcome variable. Davis at al. (1999) point out that for parameter-driven models the statistical properties are easy to derive as well as the fact that the regression parameters can be interpreted in a meaningful way. On the other hand estimation for these models requires computational effort since the likelihood function is not easily calculated. In observation-driven models the serial dependence is introduced by specifying the state variable explicitly as a function of past outcomes. According to Davis at al. (1999) prediction of counts and the calculation of the likelihood function is easy for observation-driven models. On the other hand the disadvantages of this model type are the fact that statistical characteristics, e.g. stationarity and ergodicity, are difficult to derive and interpretation of the parameters is not simple.

This paper focus on the use of different discrete distributions for observation-driven models. The models proposed in this paper are illustrated with data of the daily number of asthma cases in a hospital in Campbelltown (Sydney, Australia). These data have been already analyzed by Davis at al. (1999). In their work Davis at al. (1999) propose a Poisson-GLARMA model using seasonal variables as regressors and with lags 1, 3, 7, 10 for the AR component and no moving average component. This model will be used as a benchmark for models using alternative count distributions. In order to make the different approaches comparable the same regressors and the same AR components will be used across all models.
The major objective of this paper is to illustrate the impact of using alternative distributions in observation-driven models for count series on the ability of the models to fit the data. The paper is organized as follows.Section 2 presents the general set up for observation driven models. Section 3 presents different discrete distributions and some of their properties. Estimation and inference issues are discussed in section 4. The results of fitting the models to the number of asthma presentations at a hospital in Campbelltown are presented in Section 5 while section 6 concludes.

2 The Model

In the general setup it is assumed that given the past history (denoted by $\mathcal{F}_{t-1}$), $Y_t \mid \mathcal{F}_{t-1} \sim g(\mu_t, \lambda_t)$, where $g(\cdot)$ is the probability mass function. The first two conditional moments are given by $E\left[Y_t \mid \mathcal{F}_{t-1}\right] = \mu_t$ and $Var\left[Y_t \mid \mathcal{F}_{t-1}\right] = \sigma_t^2$. In order to include external regressors the conditional mean is connected to the process W_t via a link function. In this paper for all models under consideration the same link function $\mu_t = \exp(W_t)$ is used where $W_t = x_t^\top \beta + Z_t$. In this model x_t is a series of regressor variables with dimension $r \times 1$. Using the model proposed by Davis et al. (2003) the term Z_t which is responsible for the correlation structure of Y_t can be expressed as an infinite sum of past errors $Z_t = \sum_{i=1}^{\infty} \pi_i \epsilon_{t-i}$ where:

$$\epsilon_t = \frac{Y_t - \mu_t}{\sigma_t} . \qquad (1)$$

The infinite moving average can be specified in terms of a finite number of parameters. One way to parameterize the moving average weights π_i is to express them as the coefficients of an autoregressive-moving average filter:

$$\pi(B) = \sum_{i=1}^{\infty} \pi_i B^i = \frac{\theta(B)}{\phi(B)} - 1$$

where:

$$\begin{aligned}\theta(B) &= 1 + \theta_1 B^1 + \theta_2 B^2 + \cdots + \theta_q B^q \\ \phi(B) &= 1 - \phi_1 B^1 - \phi_2 B^2 - \cdots - \phi_p B^p.\end{aligned}$$

are polynomials in the backshift operator (B) with having all their zeros outside the unit circle. Therefore Z_t can be expressed by:

$$Z_t = \sum_{i=1}^{p} \phi_i \, (Z_{t-i} + \epsilon_{t-i}) + \sum_{i=1}^{q} \theta_{t-i} \, \epsilon_{t-i}. \tag{2}$$

In this set up different distributions with probability mass functions $g(\cdot)$ can be used for modelling count series. In the following section the count distributions under consideration will be presented.

3 Distributions

As already mentioned in the previews section Y_t, μ_t, σ_t^2, λ_t depend on time. In order to have easy notation the subscript will be ignored in this section.

Poisson Distribution
The standard choice in many applications for count data is the Poisson distribution with the pmf given by:

$$P(Y = y) = \frac{e^{-\mu} \mu^y}{y!} \, .$$

A special characteristic of the Poisson model is equidispersion, the fact that mean and variance are equal: $\mu = \sigma^2$. This fact is often too restrictive. In many applications one can observe overdispersion. This observation is the justification for using more complicated count distributions.

Negative Binomial Distribution
An alternative model which allows for overdispersion is the negative binomial distribution, with the following probability mass function:

$$P(Y = y) = \frac{\Gamma(y + \lambda)}{\Gamma(1 + y)\Gamma(\lambda)} \left(\frac{\lambda}{\lambda + \mu}\right)^{\lambda} \left(\frac{\mu}{\lambda + \mu}\right)^{y} \tag{3}$$

and the variance given by: $\sigma^2 = \mu(1 + \mu/\lambda)$. There exist two different types of negative binomial distributions which differ with respect to the definition of λ. The first model (NB-I) uses $\lambda = \mu/\alpha_0$ while the second model (NB-II) uses $\lambda = 1/\alpha_0$. This affects the variance function. While for the NB-I model the variance is given by: $\sigma^2 = \mu(1 + \alpha_0)$ in the NB-II model the variance is a second degree polynomial of the mean: $\sigma^2 = \mu(1 + \alpha_0\mu)$. Therefore a NB-II regression e.g. proposed by Lawless (1987) is able to model overdispersion in a more flexible way. It is important to know that the negative binomial distribution is not part of the one parameter exponential family used in the standard theory of generalized linear models e.g. McCullagh and Nelder(1989). Nevertheless the negative binomial distribution is part of a more general definition of the exponential family introduced by Jørgenson (1986).

Generalized Negative Binomial Distribution

The two types of negative binomial distributions can be nested in a more general approach proposed by Cameron and Trivedi (1986) or Saha and Dong (1997). The probability mass function of the generalized negative binomial model uses the same two parameter probability mass function as given in expression (3). It gains more flexibility by including another parameter in the definition of λ:

$$\lambda = \frac{\mu^{\alpha_1}}{\alpha_0}.$$

In this distribution the variance is given by $\sigma^2 = \mu + \alpha_0\mu^{2-\alpha_1}$. The NB-I can be obtained by setting $\alpha_1 = 1$ while the NB-II model can be found in the case of $\alpha_1 = 0$. In the case of $\alpha_0 \to 0$ one gets the standard Poisson model.

Geometric Distribution

In addition to the two types of negative binomial distributions there exists another discrete distribution which is nested in the generalized negative binomial distribution. In the case of $\lambda = 1$ expression (3) results in:

$$P(Y = y) = \left(\frac{1}{1+\mu}\right)\left(\frac{\mu}{1+\mu}\right)^y.$$

The variance in this case is given as a second order polynomial of the mean: $\sigma^2 = \mu(1+\mu)$. It is easy to see that the Geometric distribution is also nested in the NB-II distribution in the case $\alpha_0 = 1$.

Generalized Poisson Distribution

The pmf of the generalized Poisson distribution proposed by Consul and Jain (1973) is given by:

$$P(Y = y) = \begin{cases} \frac{\kappa[\kappa+\lambda y]^{y-1}e^{-[\kappa+\lambda y]}}{y!} \\ 0 \quad \text{for } y > m, \text{ when } \lambda < 1 \end{cases} \tag{4}$$

with the restrictions: $\kappa > 0$, $\max\left\{-1, -\frac{\kappa}{m}\right\} \leq \lambda \leq 1$ where m is the largest positive integer for which $\kappa + m\lambda > 0$ when $\lambda < 0$. The first two moments are given by $\mu = \kappa/(1-\lambda)$ and $\sigma^2 = \kappa/(1-\lambda)^3 = \mu/(1-\lambda)^2$. A different parametrization proposed by Consul and Famoye (1992) expresses the pmf of the generalized Poisson distribution as a function of its mean:

$$P(Y=y) = \begin{cases} \frac{\mu[\mu+(\rho-1)y]^{y-1}\rho^{-y}e^{-\frac{1}{\rho}[\mu+(\rho-1)y]}}{y!} \\ 0 \quad \text{for } y > m, \text{ when } \rho < 1 \end{cases} \tag{5}$$

with the adjusted restrictions: $\mu > 0$, $\rho \geq \max\left\{\frac{1}{2}, 1-\frac{\mu}{m}\right\}$ where m is the largest positive integer for which $\mu + m(\rho-1) > 0$ when $\rho < 0$. The variance can be expressed as a function of the parameter ρ and the mean: $\sigma^2 = \mu\rho^2$. The fact that the variance is a linear function of the mean seems to be too restrictive. In order to allow for more flexible relations between the mean and the variance a restricted version of the generalized Poisson was proposed by Consul (1989).

Restricted Generalized Poisson Distribution

The restricted generalized Poisson distribution results by substituting $\lambda = \alpha_0\kappa$ for λ in expression (4):

$$P(Y=y) = \begin{cases} \frac{\kappa^y[1+\alpha_0 y]^{y-1}e^{-\kappa[1+\alpha_0 y]}}{y!} \\ 0 \quad \text{for } y > m, \text{ when } \alpha_0 < 0\ . \end{cases} \tag{6}$$

The parameters have the following restrictions: $\kappa > 0$, $\max\left\{-\frac{1}{\kappa}, -\frac{1}{m}\right\} \leq \alpha_0 \leq \frac{1}{\kappa}$, where m is the largest positive integer for which $\kappa(1+m\alpha_0) > 0$ when $\alpha_0 < 0$. The mean and the variance of the generalized Poisson distribution in restricted form are given by $\mu = \kappa/(1-\alpha_0\kappa)$ and $\sigma^2 = \kappa/(1-\alpha_0\kappa)^3 = \mu(1+\alpha_0\mu)^2$ respectively. The variance is a third degree polynomial of the mean. This allows for more flexibility when modelling overdispersion. The restricted generalized Poisson distribution can also be expressed as a function of its mean:

$$\begin{aligned} P(Y=y) &= \kappa^y \frac{(1+\alpha_0 y)^{y-1}e^{-\kappa(1+\alpha_0 y)}}{y!} \\ \kappa &= \frac{\mu}{1+\alpha_0\mu}\ . \end{aligned}$$

The restricted and the unrestricted forms of the generalized Poisson regression model can be nested in a two-parameter hybrid generalized Poisson model proposed by Santos Silva (1997).

Hybrid Generalized Poisson Distribution
The hybrid generalized Poisson (HGP-I) distribution is given by:

$$\begin{aligned} P(Y=y) &= \kappa^y \frac{(1+\lambda y)^{y-1} e^{-\kappa(1+\lambda y)}}{y!} \\ \kappa = \frac{\mu}{1+\lambda\mu} \qquad \lambda &= \alpha_0 \mu^{\alpha_1} \, . \end{aligned}$$

The variance is given by: $\sigma^2 = \mu(1+\lambda\mu)^2$. In this model both parameters α_0 and α_1 need to be estimated. The restricted generalized Poisson model (abbreviated by HGP-II) can be obtained by setting $\alpha_1 = 0$. In the case of $\alpha_1 = -1$ one gets the HGP-III model which is identical with the generalized Poisson model in expression (5). The parameter ρ in the generalized Poisson models is then equivalent to $(1+\alpha_0)$. The standard Poisson model can be obtained by setting $\alpha_0 = 0$.

Modified Borel Distribution
Another distribution of interest is a modified Borel distribution, which is nested in the HGP-I model. In the case of $\lambda = 1$ one finds the following pmf:

$$\begin{aligned} P(Y=y) &= \kappa^y \frac{(y+1)^{y-1} e^{-\kappa(1+y)}}{y!} \\ \kappa &= \frac{\mu}{1+\mu} \, . \end{aligned}$$

This probability mass function is different to the one presented by Borel (1942) and Tanner (1953). The Borel distribution in its original form does not include events $\{Y=0\}$. By shifting the original distribution one finds a modified Borel distribution which is defined on $\mathbb{N}_0$. A random variable following this modified Borel distribution has the variance: $\sigma^2 = \mu(1+\mu)^2$. It is easy to see that the modified Borel distribution is also a special case of the HGP-II model for $\alpha_0 = 1$.

Table 1 given on the next page summarizes of the different distributions and their properties.

	Generalized Poisson					Generalized Negative Binomial				
λ	$\alpha_0\mu^{\alpha_1}$					$\frac{1}{\alpha_0}\mu^{\alpha_1}$				
σ^2	$\mu(1+\lambda\mu)^2$					$\mu(1+\frac{\mu}{\lambda})$				
α_0	free	free	free	1	0	$\to 0$	1	free	free	free
α_1	free	0	-1	0	not relevant		0	1	0	free
Model	HGP-I	HGP-II	HGP-III	mod. Borel	Poisson		Geometric	NB-I	NB-II	GNB
σ^2	$\mu(1+\alpha_0\mu^{1+\alpha_1})^2$	$\mu(1+\alpha_0\mu)^2$	$\mu(1+\alpha_0)^2$	$\mu(1+\mu)^2$	μ		$\mu(1+\mu)$	$\mu(1+\alpha_0)$	$\mu(1+\alpha_0\mu)$	$\mu(1+\alpha_0\mu^{1-\alpha_1})$

Table 1: Overview: Alternative count models.

∞

4 Estimation and Inference

4.1 Maximum Likelihood Estimation

In order to estimate the model parameters the use of a modified maximum likelihood approach is appropriate. The modification of the standard likelihood approach is necessary since the model specifies only a conditional distribution $g(\mu_t, \lambda_t \mid \mathcal{F}_{t-1})$. The partial likelihood function is given by:

$$\mathcal{L}(\delta) = \prod_{t=1}^{T} g(y_t, \mu_t(\delta), \lambda_t(\delta) \mid \mathcal{F}_{t-1}) \ .$$

The corresponding log-likelihood function is given by:

$$\ell(\delta) = \sum_{t=1}^{T} f(y_t, \mu_t(\delta), \lambda_t(\delta) \mid \mathcal{F}_{t-1})$$

where: $f(\cdot) = \log g(\cdot)$. The maximum likelihood estimates can be obtained by solving the score equation:

$$S(\delta) = \sum_{t=1}^{T} \frac{\partial \ell_t(\delta)}{\partial \delta} = 0 \ .$$

In order to calculate the score function one has to take care about the fact that $\ell_t(\delta)$ depends on δ via μ_t and λ_t which itself depends on μ_t, α_0 and α_1. The derivation of the log-likelihood function is discussed in detail in appendix 1. Due to the fact that the score function for observation-driven models is highly nonlinear in δ it cannot be solved analytically. The maximization can be done by using numerical algorithms which are implemented in most of the statistical software packages. The results given in section (5) were obtained by using the BFGS-Algorithm (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) which is implemented in R (http://www.r-project.org).

4.2 Inference

After fitting a model to observed data one likes to answer the question whether or not certain parameters are statistically significant and whether or not one model performs in total better than another one. In order to answer these questions one need knowledge ore at least reasonable assumptions about the distribution of the estimators. In the framework of generalized linear models which uses distributions which belong to the exponential family one can proof that the maximum likelihood estimator $\hat{\delta}$ of the model parameters is consistent and asymptotically normal distributed (e.g. Fahrmeir and Kaufmann, 1985 or McCullagh and Nelder, 1989):

$$\sqrt{n}(\hat{\delta} - \delta_0) \ \xrightarrow{d} \ N(0, \Omega)$$

where:

$$\hat{\Omega} = -\left(\frac{1}{n} \frac{\partial^2 \ell(\delta)}{\partial \delta \partial \delta^\top} \mid_{\delta = \hat{\delta}} \right)^{-1} .$$

The problem in applying these results to the models proposed in this paper is twofold. First some of the distributions used for modelling in this paper are not part of the exponential family. By using these distributions one leaves the framework in which the asymptotical properties were proofed. Second even if a distribution which belongs to the exponential family is used for modelling the validity of the normality assumption for the estimator is still questionable since the model does not deal with independent identically distributed data. Furthermore a central limit theorem for the maximum likelihood estimators is currently not available for the general observation-driven model. Davis et al. (2003) provide proofs for a simple Poisson-GLARMA model which justify the use of the properties above. Based on the results of Davis et al. (2003) it seems to be reasonable to assume the validity of the properties given above for observation-driven models which use distributions which are part of the exponential family. The use of these results for the distributions under consideration in this paper can be justified by the following arguments. The Poisson, the Geometric and the modified Borel distribution belong to the exponential family (e.g. Winkelmann, 2000 chap.2). For given α_0 the NB-II belongs to the exponential family too (e.g. Winkelmann, 2000). The same can be shown for the restricted generalized Poisson distribution (see appendix 2). The NB-II and the HGP-II distribution is nested in the generalized negative binomial and in the HGP-I respectively. The Poisson distribution is nested in all distribution under consideration. These relations show that all distributions under consideration in this paper are connected to exponential family. Due to the absence of a central limit theorem this connection to the exponential family together with the results given by Davis et al. (2003) should be enough to justify the use of the asymptotic results given above for the maximum likelihood estimators of β, ϕ and θ. This assumption is not made for the estimators of α_0 and α_1 therefore in table 3 no statements concerning significance are made for the estimates of these parameters.

5 Example

5.1 Data

In order to illustrate the impact of using alternative distributions the models presented in this paper are applied to a time series of daily asthma incidences in a single hospital in Campbelltown in the metropolitan area of Sydney. The observations were made in the time period from 1 January 1990 to 31 December 1993 on a daily basis, in total 1461 observations. The series is plotted in figure 1. It shows periodical behaviour. Its autocorrelation function given in the same figure shows significant autocorrelations. The data set has been analyzed by Davis et al. (1999) who propose a Poisson-GLARMA model with autoregressive components for lag 1, 3, 7 and 10. Furthermore Davis el al. (1999) include trigonometric functions as external regressors to describe the periodic behaviour in addition to regressors for Sunday and Monday effects. Due to the fact that the Poisson model is easy to implement it will be used as a benchmark in this analysis. In order to make the different approaches comparable the same regressors and the same AR components were used across all models.

5.2 Computational Issues

Since the maximization of the log-likelihood function is done by an iterative numerical algorithm starting values of δ need to be specified. Similar to Davis et al. (2003) the estimation of the Poisson-GLARMA model was initialized with the standard generalized linear model Poisson regression estimates without the autoregressive moving average terms together with zero initial values for ϵ_t $(t \leq 0)$. The estimation of the time series models using the modified Borel and the Geometric distribution was initialized in the same way. Since the Geometric distribution is nested in the negative binomial distribution type II (NB-II) the estimation of this models was initialized by the estimates of the time series Geometric model and starting value $\alpha_0 = 1$. The estimates of the Geometric time series model were used as initial values for the generalized negative binomial model together with the starting values $\alpha_0 = 1$ and $\alpha_1 = 0$. The estimation of the NB-I model was initialized with standard non-time series negative binomial regression estimates. The Poisson-GLARMA estimates of the mean were used to calculate the initial value of α_0:

$$\hat{\alpha}_0 = \frac{1}{T - r - p - q} \sum_{t=1}^{T} \frac{(y_t - \hat{\mu}_t)^2}{\hat{\mu}_t} - 1 \, .$$

The estimates of the time series modified Borel model together with $\alpha_0 = 1$ or $\alpha_0 = 1,\ \alpha_1 = 0$ were used as initial values for the estimation of the HGP-II model and HGP-I model respectively. The estimation of the HGP-III model was initialized with Poisson-GLARMA estimates. The series of Poisson-GLARMA estimates of the mean was used to calculate the initial value of α_0:

$$\hat{\alpha}_0 = \sqrt{\frac{1}{T-r-p-q}\sum_{t=1}^{T}\frac{(y_t-\hat{\mu}_t)^2}{\hat{\mu}_t}} - 1\ .$$

5.3 Results

The parameter estimates of the models are given in table 3. The estimates of δ do not differ very much accross the models concerning the value of the estimates and their significance. Nevertheless the models using the modified Borel and the Geometric distribution are rather different to all other models under consideration. Another interesting fact is the result that for both the HGP-I and the GNB model the value of $\hat{\alpha}_0$ is close to zero. This gives rise to the presumption that these models do not differ significantly from the Poisson model. The results of a likelihood ratio test presented in table 2 can help answering this question. Based on the results presented in this table one can conclude that starting from a Poisson model all alternatives (HGP-III, HGP-II, HGP-I, NB-I, NB-II and GNB) increase the model fit significantly. Similar results can be found for the likelihood ratio test when starting from the modified Borel or the Geometric distribution. The values of the test statistic are very large for these models, confirming the impression that the model using both the modified Borel and the Geometric distribution are rather different to the other alternatives. Another interesting point is the fact that starting from a HGP-II or NB-II model the use of the HGP-I model or GNB respectively does not increase the model fit significantly. This can be explained by the fact that both the HGP-II and the NB-II distribution already have a flexible variance function which covers a wide range of overdispersion. Therefore the additional parameter leading to the generalized distributions is not necessary.

Additional to the results of the likelihood ratio test table 4 provides the values of different model fit statistics. In this tabel for each model the following quantities are presented: the number of paramters, the value of the log-likelihood function evaluated at the estimates, the likelihood ratio test for a model using the intercept only an the time series model, three types of pseudo R^2 (McFadden, 1973; Maddala, 1983 and Veall and Zimmermann, 1992), the root mean squared error and two information criteria (Akaike, 1973 and Schwarz 1978). From the results given in this table one can see that based on the value of the log-likelihood function evaluated at the estimates and based on the Akaike information criterion (AIC) the HGP-I model fits the data best. In the Schwarz information criterion (BIC) including additional parameters is penalized stronger than in the AIC. Therefore based on this criterion the Poisson model should be chosen. Based on the pseudo R^2 the Poisson model fits the data best while a model selection based on the RMSE will not lead to a unique decision. The values of the fit statistics also confirm the presumption that the modified Borel and the Geometric distribution lead to rather different results. Both models show worse results for all the characteristics mentioned above. Figure 2 and 3 illustrate these findings. These graphics show the estimated mean and variance for each model. The rather different properties of the models using the modified Borel and the Geometric distribution become obvious in these plots. In the plots of these two models the variance exceeds the mean very much.

Another point of interest is the behaviour of the residuals. If the model is specified correctly the estimated residuals will have an expected value of zero and unit variance. Furthermore its first two moments should be uncorrelated. Table 5 summarizes the characteristics of the model residuals. With exception of the models using the modified Borel and the Geometric distributions and the Box-Pierce test on autocorrelation of the squares residuals for lag 10 non of the model residuals show significant deviation from the assumed behaviour. Similar the the previews findings the models using the modified Borel and the Geometric distribution show different residual characteristics. This can be seen as an indication that these models are not sufficient for the data set under consideration. Figure 4 and 5 show the autocorrelation function of the model residuals. None of theses plots shows significant autocorrelations.

Finally one can see that among all models under consideration in this paper pairs of models with equivalent properties can be found. Concerning their characteristics the following pairs of almost equal models can be found: The modified Borel and the Geometric model, the NB-I and the HGP-III, the NB-II and the HGP-II, the HGP-I and the GNB model. The pair wise similarity can be seen by the presence of some type of symmetry around the column containing the results of the Poisson model in tables 3 to 5.

6 Conclusion

This paper analyzes and illustrates the use of alternative discrete distributions in the framework of observation-driven models for count series.
Observation-driven models are a flexible approach for modelling a wide range of serial dependence in count data. Although the Poisson distribution is widely used in modelling count data it has some disadvantages since it assumes equidisperion. This paper shows that the ability of a model to fit the data can be improved significantly by using distributions which are able to cover overdispersion. Among eight alternative distributions the GNB, the NB-II, the HGP-II and the HGP-I can be considered as serious competitors to the approach using the Poisson distribution. On the other hand the Geometric, the NB-I, the modified Borel and the HGP-III distribution does not improve the model since they use less flexible variance functions. Further research can be done in analyzing the impact of using alternative count distributions on out-of-sample forecasting properties of observation-driven models.

Acknowledgements
I gratefully acknowledge the support of William Dunsmuir (Professor and Head of the Department of Statistics at the University of New South Wales) for providing both the data of daily asthma admissions at the hospital in Campbelltown and the S-Plus program for calculating the derivatives of Z_t recursively.

Appendix 1: Estimation

As already mentioned the conditional probablity mass function depends on μ_t and λ_t. Furthermore λ_t depends on α_0, α_1 and μ_t as well. The parameters of interest are collected in the vector $\delta = (\beta^\top, \phi^\top, \theta^\top, \alpha_0, \alpha_1)^\top$ which has dimensions $(r+p+q+2) \times 1$. In order to derive a general expression for the maximum likelihood estimation two additional vectors need to be introduced: $u = (0, 0, \cdots, 0, 1, 0)^\top$, $v = (0, 0, \cdots, 0, 0, 1)^\top$. Both vectors have the same dimensions as δ. The first derivative of the log-likelihood function is given by:

$$
\begin{aligned}
\frac{\partial \ell_t}{\partial \delta} &= \frac{\partial \ell_t}{\partial W_t}\frac{\partial W_t}{\partial \delta} + \frac{\partial \ell_t}{\partial \alpha_0} u + \frac{\partial \ell_t}{\partial \alpha_1} v \\
&= \left[\frac{\partial \ell_t}{\partial \mu_t}\frac{\partial \mu_t}{\partial W_t} + \frac{\partial \ell_t}{\partial \lambda_t}\frac{\partial \lambda_t}{\partial \mu_t}\frac{\partial \mu_t}{\partial W_t}\right]\frac{\partial W_t}{\partial \delta} + \frac{\partial \ell_t}{\partial \lambda_t}\frac{\partial \lambda_t}{\partial \alpha_0} u + \frac{\partial \ell_t}{\partial \lambda_t}\frac{\partial \lambda_t}{\partial \alpha_1} v \ .
\end{aligned}
$$

The second derivative is given by:

$$
\begin{aligned}
\frac{\partial^2 \ell_t}{\partial \delta \partial \delta^\top} &= \frac{\partial^2 \ell_t}{\partial W_t^2}\frac{\partial W_t}{\partial \delta}\left[\frac{\partial W_t}{\partial \delta}\right]^\top + u\frac{\partial^2 \ell_t}{\partial \alpha_0 \partial W_t}\left[\frac{\partial W_t}{\partial \delta}\right]^\top + v\frac{\partial^2 \ell_t}{\partial \alpha_1 \partial W_t}\left[\frac{\partial W_t}{\partial \delta}\right]^\top \\
&+ \frac{\partial \ell_t}{\partial W_t}\frac{\partial^2 W_t}{\partial \delta \partial \delta^\top} + \frac{\partial^2 \ell_t}{\partial \alpha_0 \partial W_t}\frac{\partial W_t}{\partial \delta} u^\top + \frac{\partial^2 \ell_t}{\partial \alpha_0 \partial \alpha_1} v u^\top + \frac{\partial^2 \ell_t}{\partial \alpha_0^2} u u^\top \\
&+ \frac{\partial^2 \ell_t}{\partial \alpha_1 \partial W_t}\frac{\partial W_t}{\partial \delta} v^\top + \frac{\partial^2 \ell_t}{\partial \alpha_0 \partial \alpha_1} u v^\top + \frac{\partial^2 \ell_t}{\partial \alpha_1^2} v v^\top \ .
\end{aligned}
$$

This general representation of the second derivative contains expressions which need to be specified further:

$$
\begin{aligned}
\frac{\partial^2 \ell_t}{\partial W_t^2} &= \frac{\partial^2 \ell_t}{\partial \mu_t^2}\left(\frac{\partial \mu_t}{\partial W_t}\right)^2 + \frac{\partial \ell_t}{\partial \mu_t}\frac{\partial^2 \mu_t}{\partial W_t^2} + \frac{\partial^2 \ell_t}{\partial \lambda_t^2}\left(\frac{\partial \lambda_t}{\partial \mu_t}\frac{\partial \mu_t}{\partial W_t}\right)^2 \\
&+ \frac{\partial \ell_t}{\partial \lambda_t}\left[\frac{\partial^2 \lambda_t}{\partial \mu_t^2}\left(\frac{\partial \mu_t}{\partial W_t}\right)^2 + \frac{\partial \lambda_t}{\partial \mu_t}\frac{\partial^2 \mu_t}{\partial W_t^2}\right] \\
\frac{\partial^2 \ell_t}{\partial \alpha_k \partial W_t} &= \frac{\partial^2 \ell_t}{\partial \lambda_t^2}\frac{\partial \lambda_t}{\partial \alpha_k}\frac{\partial \lambda_t}{\partial \mu_t}\frac{\partial \mu_t}{\partial W_t} + \frac{\partial \ell_t}{\partial \lambda_t}\frac{\partial^2 \lambda_t}{\partial \alpha_k \partial \mu_t}\frac{\partial \mu_t}{\partial W_t} \\
\frac{\partial^2 \ell_t}{\partial \alpha_k \partial \alpha_l} &= \frac{\partial^2 \ell_t}{\partial \lambda_t^2}\frac{\partial \lambda_t}{\partial \alpha_k}\frac{\partial \lambda_t}{\partial \alpha_l} + \frac{\partial \ell_t}{\partial \lambda_t}\frac{\partial^2 \lambda_t}{\partial \alpha_k \partial \alpha_l} \ .
\end{aligned}
$$

For deriving the expressions above it is important to know that based on the link function we have:

$$\frac{\partial \mu_t}{\partial \alpha_k} = \frac{\partial^2 \mu_t}{\partial \alpha_k \partial W_t} = 0 \; .$$

Similar to ℓ_t the residuals ϵ_t depend on δ via μ_t and λ_t which itself depend on α_0 and α_1. These facts must be considered when calculating the derivatives:

$$\begin{aligned}
\frac{\partial \epsilon_t}{\partial \delta} &= \frac{\partial \epsilon_t}{\partial W_t}\frac{\partial W_t}{\partial \delta} + \frac{\partial \epsilon_t}{\partial \alpha_0} u + \frac{\partial \epsilon_t}{\partial \alpha_1} v \\
&= \left[\frac{\partial \epsilon_t}{\partial \mu_t}\frac{\partial \mu_t}{\partial W_t} + \frac{\partial \epsilon_t}{\partial \lambda_t}\frac{\partial \lambda_t}{\partial \mu_t}\frac{\partial \mu_t}{\partial W_t}\right]\frac{\partial W_t}{\partial \delta} + \frac{\partial \epsilon_t}{\partial \lambda_t}\frac{\partial \lambda_t}{\partial \alpha_0} u + \frac{\partial \epsilon_t}{\partial \lambda_t}\frac{\partial \lambda_t}{\partial \alpha_1} v \\
\frac{\partial^2 \epsilon_t}{\partial \delta \partial \delta^\top} &= \frac{\partial^2 \epsilon_t}{\partial W_t^2}\frac{\partial W_t}{\partial \delta}\left[\frac{\partial W_t}{\partial \delta}\right]^\top + u\frac{\partial^2 \epsilon_t}{\partial \alpha_0 \partial W_t}\left[\frac{\partial W_t}{\partial \delta}\right]^\top + v\frac{\partial^2 \epsilon_t}{\partial \alpha_1 \partial W_t}\left[\frac{\partial W_t}{\partial \delta}\right]^\top \\
&+ \frac{\partial \epsilon_t}{\partial W_t}\frac{\partial^2 W_t}{\partial \delta \partial \delta^\top} + \frac{\partial^2 \epsilon_t}{\partial \alpha_0 \partial W_t}\frac{\partial W_t}{\partial \delta} u^\top + \frac{\partial^2 \epsilon_t}{\partial \alpha_0 \partial \alpha_1} v u^\top + \frac{\partial^2 \epsilon_t}{\partial \alpha_0^2} u u^\top \\
&+ \frac{\partial^2 \epsilon_t}{\partial \alpha_1 \partial W_t}\frac{\partial W_t}{\partial \delta} v^\top + \frac{\partial^2 \epsilon_t}{\partial \alpha_0 \partial \alpha_1} u v^\top + \frac{\partial^2 \epsilon_t}{\partial \alpha_1^2} v v^\top \; .
\end{aligned}$$

In the expression of the second derivative of ϵ_t some components need to be specified more precisely:

$$\begin{aligned}
\frac{\partial^2 \epsilon_t}{\partial W_t^2} &= \frac{\partial^2 \epsilon_t}{\partial \mu_t^2}\left(\frac{\partial \mu_t}{\partial W_t}\right)^2 + \frac{\partial \epsilon_t}{\partial \mu_t}\frac{\partial^2 \mu_t}{\partial W_t^2} + \frac{\partial^2 \epsilon_t}{\partial \lambda_t^2}\left(\frac{\partial \lambda_t}{\partial \mu_t}\frac{\partial \mu_t}{\partial W_t}\right)^2 \\
&+ \frac{\partial \epsilon_t}{\partial \lambda_t}\left[\frac{\partial^2 \lambda_t}{\partial \mu_t^2}\left(\frac{\partial \mu_t}{\partial W_t}\right)^2 + \frac{\partial \lambda_t}{\partial \mu_t}\frac{\partial^2 \mu_t}{\partial W_t^2}\right] \\
\frac{\partial^2 \epsilon_t}{\partial \alpha_k \partial W_t} &= \frac{\partial^2 \epsilon_t}{\partial \lambda_t^2}\frac{\partial \lambda_t}{\partial \alpha_k}\frac{\partial \lambda_t}{\partial \mu_t}\frac{\partial \mu_t}{\partial W_t} + \frac{\partial \epsilon_t}{\partial \lambda_t}\frac{\partial^2 \lambda_t}{\partial \alpha_k \partial \mu_t}\frac{\partial \mu_t}{\partial W_t} \\
\frac{\partial^2 \epsilon_t}{\partial \alpha_k \partial \alpha_l} &= \frac{\partial^2 \epsilon_t}{\partial \lambda_t^2}\frac{\partial \lambda_t}{\partial \alpha_k}\frac{\partial \lambda_t}{\partial \alpha_l} + \frac{\partial \epsilon_t}{\partial \lambda_t}\frac{\partial^2 \lambda_t}{\partial \alpha_k \partial \alpha_l} \; .
\end{aligned}$$

These results depend on the derivatives of ϵ_t with respect to μ_t and λ_t.

Based on the definition of ϵ_t its derivatives are given by:

$$\begin{aligned}
\frac{\partial \epsilon_t}{\partial \mu_t} &= -\frac{1}{\sigma_t} - \frac{1}{2}\frac{\epsilon_t}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \mu_t} \\
\frac{\partial \epsilon_t}{\partial \lambda_t} &= -\frac{1}{2}\frac{\epsilon_t}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \lambda_t} \\
\frac{\partial^2 \epsilon_t}{\partial \mu_t^2} &= \frac{1}{\sigma_t^3}\frac{\partial \sigma_t^2}{\partial \mu_t} + \frac{3}{4}\frac{\epsilon_t}{\sigma_t^4}\left(\frac{\partial \sigma_t^2}{\partial \mu_t}\right)^2 - \frac{1}{2}\frac{\epsilon_t}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \mu_t^2} \\
\frac{\partial^2 \epsilon_t}{\partial \lambda_t^2} &= \frac{3}{4}\frac{\epsilon_t}{\sigma_t^4}\left(\frac{\partial \sigma_t^2}{\partial \lambda_t}\right)^2 - \frac{1}{2}\frac{\epsilon_t}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \lambda_t^2} \; .
\end{aligned}$$

For deriving these results the following facts are useful:

$$\begin{aligned}
\frac{\partial \mu_t}{\partial W_t} &= \frac{\partial^2 \mu_t}{\partial W_t^2} = \mu_t \\
\frac{\partial}{\partial \mu_t}\,\sigma_t &= \frac{\partial}{\partial \mu_t}\,\sqrt{\sigma_t^2} = \frac{1}{2}\frac{1}{\sigma_t}\frac{\partial \sigma_t^2}{\partial \mu_t} \; .
\end{aligned}$$

The derivatives of ℓ_t and ϵ_t depend on the derivatives of W_t which are given by:

$$\begin{aligned}
\frac{\partial W_t}{\partial \delta} &= x_t^\top + \frac{\partial Z_t}{\partial \delta} \\
\frac{\partial^2 W_t}{\partial \delta \partial \delta^\top} &= \frac{\partial^2 Z_t}{\partial \delta \partial \delta^\top} \; .
\end{aligned}$$

Based on its definition given in expression (2) the derivatives of Z_t need to be calculated recursively (see Davis et al, 2004):

$$\begin{aligned}
\frac{\partial Z_t}{\partial \delta} &= \sum_{i=1}^{p}\left[\frac{\partial \phi_i}{\partial \delta}\;(Z_{t-i} + \epsilon_{t-i}) + \phi_i\;\left(\frac{\partial Z_{t-i}}{\partial \delta} + \frac{\partial \epsilon_{t-i}}{\partial \delta}\right)\right] \\
&+ \sum_{i=1}^{q}\left[\frac{\partial \theta_i}{\partial \delta}\;\epsilon_{t-i} + \theta_i\;\frac{\partial \epsilon_{t-i}}{\partial \delta}\right] \; .
\end{aligned}$$

In particular (as shown by Davis et al., 2004):

$$
\begin{aligned}
\frac{\partial Z_t}{\partial \beta_k} &= \sum_{i=1}^{p} \phi_i \left(\frac{\partial Z_{t-i}}{\partial \beta_k} + \frac{\partial \epsilon_{t-i}}{\partial \beta_k} \right) + \sum_{i=1}^{q} \theta_i \frac{\partial \epsilon_{t-i}}{\partial \beta_k} \\
\frac{\partial Z_t}{\partial \phi_k} &= Z_{t-k} + \epsilon_{t-k} + \sum_{i=1}^{p} \phi_i \left(\frac{\partial Z_{t-i}}{\partial \phi_k} + \frac{\partial \epsilon_{t-i}}{\partial \phi_k} \right) + \sum_{i=1}^{q} \theta_i \frac{\partial \epsilon_{t-i}}{\partial \phi_k} \\
\frac{\partial Z_t}{\partial \theta_k} &= \sum_{i=1}^{p} \phi_i \left(\frac{\partial Z_{t-i}}{\partial \theta_k} + \frac{\partial \epsilon_{t-i}}{\partial \theta_k} \right) + \epsilon_{t-k} + \sum_{i=1}^{q} \theta_i \frac{\partial \epsilon_{t-i}}{\partial \theta_k} \\
\frac{\partial Z_t}{\partial \alpha_k} &= \sum_{i=1}^{p} \phi_i \left(\frac{\partial Z_{t-i}}{\partial \alpha_k} + \frac{\partial \epsilon_{t-i}}{\partial \alpha_k} \right) + \sum_{i=1}^{q} \theta_i \frac{\partial \epsilon_{t-i}}{\partial \alpha_k} \ .
\end{aligned}
$$

The second derivatives of Z_t need to be calculated recurslively too:

$$
\begin{aligned}
\frac{\partial^2 Z_t}{\partial \delta \partial \delta^\top} &= \sum_{i=1}^{p} \left[\frac{\partial \phi_i}{\partial \delta} \left(\frac{\partial Z_{t-i}}{\partial \delta^\top} + \frac{\partial \epsilon_{t-i}}{\partial \delta^\top} \right) + \left(\frac{\partial Z_{t-i}}{\partial \delta} + \frac{\partial \epsilon_{t-i}}{\partial \delta} \right) \frac{\partial \phi_i}{\partial \delta^\top} \right] \\
&+ \sum_{i=1}^{p} \phi_i \left(\frac{\partial^2 Z_{t-i}}{\partial \delta \partial \delta^\top} + \frac{\partial^2 \epsilon_{t-i}}{\partial \delta \partial \delta^\top} \right) + \sum_{i=1}^{q} \left[\frac{\partial \theta_i}{\partial \delta} \frac{\partial \epsilon_{t-i}}{\partial \delta^\top} + \frac{\partial \epsilon_{t-i}}{\partial \delta} \frac{\partial \theta_i}{\partial \delta^\top} \right] \\
&+ \sum_{i=1}^{q} \theta_i \frac{\partial^2 \epsilon_{t-i}}{\partial \delta \partial \delta^\top}
\end{aligned}
$$

in particular for certain k and l:

$$
\begin{aligned}
\frac{\partial^2 Z_t}{\partial \beta_k \partial \beta_l} &= \sum_{i=1}^{p} \phi_i \left(\frac{\partial^2 Z_{t-i}}{\partial \beta_k \partial \beta_l} + \frac{\partial^2 \epsilon_{t-i}}{\partial \beta_k \partial \beta_l} \right) + \sum_{i=1}^{q} \theta_i \frac{\partial^2 \epsilon_{t-i}}{\partial \beta_k \partial \beta_l} \\
\frac{\partial^2 Z_t}{\partial \beta_k \partial \phi_l} &= \frac{\partial Z_{t-l}}{\partial \beta_k} + \frac{\partial \epsilon_{t-l}}{\partial \beta_k} + \sum_{i=1}^{p} \phi_i \left(\frac{\partial^2 Z_{t-i}}{\partial \beta_k \partial \phi_l} + \frac{\partial^2 \epsilon_{t-i}}{\partial \beta_k \partial \phi_l} \right) + \sum_{i=1}^{q} \theta_i \frac{\partial^2 \epsilon_{t-i}}{\partial \beta_k \partial \phi_l} \\
\frac{\partial^2 Z_t}{\partial \beta_k \partial \theta_l} &= \sum_{i=1}^{p} \phi_i \left(\frac{\partial^2 Z_{t-i}}{\partial \beta_k \partial \theta_l} + \frac{\partial^2 \epsilon_{t-i}}{\partial \beta_k \partial \theta_l} \right) + \frac{\partial \epsilon_{t-l}}{\partial \beta_k} + \sum_{i=1}^{q} \theta_i \frac{\partial^2 \epsilon_{t-i}}{\partial \beta_k \partial \theta_l} \\
\frac{\partial^2 Z_t}{\partial \beta_k \partial \alpha_l} &= \sum_{i=1}^{p} \phi_i \left(\frac{\partial^2 Z_{t-i}}{\partial \beta_k \partial \alpha_l} + \frac{\partial^2 \epsilon_{t-i}}{\partial \beta_k \partial \alpha_l} \right) + \sum_{i=1}^{q} \theta_i \frac{\partial^2 \epsilon_{t-i}}{\partial \beta_k \partial \alpha_l}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 Z_t}{\partial\phi_k\partial\phi_l} &= \frac{\partial Z_{t-k}}{\partial\phi_l} + \frac{\partial\epsilon_{t-k}}{\partial\phi_l} + \frac{\partial Z_{t-l}}{\partial\phi_k} + \frac{\partial\epsilon_{t-l}}{\partial\phi_k} + \sum_{i=1}^{p}\phi_i\left(\frac{\partial^2 Z_{t-i}}{\partial\phi_k\partial\phi_l} + \frac{\partial^2\epsilon_{t-i}}{\partial\phi_k\partial\phi_l}\right) \\
&+ \sum_{i=1}^{q}\theta_i\frac{\partial^2\epsilon_{t-i}}{\partial\phi_k\partial\phi_l} \\
\frac{\partial^2 Z_t}{\partial\phi_k\partial\theta_l} &= \frac{\partial Z_{t-k}}{\partial\theta_l} + \frac{\partial\epsilon_{t-k}}{\partial\theta_l} + \sum_{i=1}^{p}\phi_i\left(\frac{\partial^2 Z_{t-i}}{\partial\phi_k\partial\theta_l} + \frac{\partial^2\epsilon_{t-i}}{\partial\phi_k\partial\theta_l}\right) \\
&+ \frac{\partial\epsilon_{t-l}}{\partial\phi_k} + \sum_{i=1}^{q}\theta_i\frac{\partial^2\epsilon_{t-i}}{\partial\phi_k\partial\theta_l} \\
\frac{\partial^2 Z_t}{\partial\phi_k\partial\alpha_l} &= \frac{\partial Z_{t-k}}{\partial\alpha_l} + \frac{\partial\epsilon_{t-k}}{\partial\alpha_l} + \sum_{i=1}^{p}\phi_i\left(\frac{\partial^2 Z_{t-i}}{\partial\phi_k\partial\alpha_l} + \frac{\partial^2\epsilon_{t-i}}{\partial\phi_k\partial\alpha_l}\right) + \sum_{i=1}^{q}\theta_i\frac{\partial^2\epsilon_{t-i}}{\partial\phi_k\partial\alpha_l} \\
\frac{\partial^2 Z_t}{\partial\theta_k\partial\theta_l} &= \sum_{i=1}^{p}\phi_i\left(\frac{\partial^2 Z_{t-i}}{\partial\theta_k\partial\theta_l} + \frac{\partial^2\epsilon_{t-i}}{\partial\theta_k\partial\theta_l}\right) + \frac{\partial\epsilon_{t-k}}{\partial\theta_l} + \frac{\partial\epsilon_{t-l}}{\partial\theta_k} + \sum_{i=1}^{q}\theta_i\frac{\partial^2\epsilon_{t-i}}{\partial\theta_k\partial\theta_l} \\
\frac{\partial^2 Z_t}{\partial\theta_k\partial\alpha_l} &= \sum_{i=1}^{p}\phi_i\left(\frac{\partial^2 Z_{t-i}}{\partial\theta_k\partial\alpha_l} + \frac{\partial^2\epsilon_{t-i}}{\partial\theta_k\partial\alpha_l}\right) + \frac{\partial\epsilon_{t-k}}{\partial\alpha_l} + \sum_{i=1}^{q}\theta_i\frac{\partial^2\epsilon_{t-i}}{\partial\theta_k\partial\alpha_l} \\
\frac{\partial^2 Z_t}{\partial\alpha_k\partial\alpha_l} &= \sum_{i=1}^{p}\phi_i\left(\frac{\partial^2 Z_{t-i}}{\partial\alpha_k\partial\alpha_l} + \frac{\partial^2\epsilon_{t-i}}{\partial\alpha_k\partial\alpha_l}\right) + \sum_{i=1}^{q}\theta_i\frac{\partial^2\epsilon_{t-i}}{\partial\alpha_k\partial\alpha_l}\ .
\end{aligned}
$$

The results given above are general in the sense that expressions which depend on the form of the distribution function used for modelling are not specified in detail. It is obvious that the derivatives of ℓ_t with respect to μ_t and λ_t depend on the functional form of the distribution function. The derivatives of λ_t itself depend on the specific distribution function too. Finally the derivatives of ϵ_t depend on the derivatives of the variance function (see expression 7) which is specified by the choice of the distribution. Since all distributions under consideration in this paper are nested either in the generalized negative binomial distribution or in the hybrid generalized Poisson distribution of type I, the remaining derivatives are given for these two general cases only. The results for all other distributions under consideration can be found by substituting the corresponding values of α_0 and α_1 in the following expressions and simplifying the results by using basic mathematical transformations.

Generalized Negative Binomial Distribution:

$$\begin{aligned}
\ell_t(\delta) &= \psi^{(0)}(y_t+\lambda_t) - \psi^{(0)}(1+y_t) - \psi^{(0)}(\lambda_t) + y_t \log(\mu_t) + \lambda_t \log(\lambda_t) \\
&- (y_t+\lambda_t)\log(\lambda_t+\mu_t)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ell_t}{\partial \mu_t} &= \frac{\lambda_t(y_t-\mu_t)}{\mu_t(\lambda_t+\mu_t)} \\
\frac{\partial \ell_t}{\partial \lambda_t} &= \psi^{(1)}(y_t+\lambda_t) - \psi^{(1)}(\lambda_t) + \log\left(\frac{\lambda_t}{\lambda_t+\mu_t}\right) - \frac{y_t-\mu_t}{\lambda_t+\mu_t} \\
\frac{\partial^2 \ell_t}{\partial \mu_t^2} &= -\frac{\lambda_t y_t}{\mu_t^2(\lambda_t+\mu_t)} - \frac{\lambda_t(y_t-\mu_t)}{\mu_t(\lambda_t+\mu_t)^2} \\
\frac{\partial^2 \ell_t}{\partial \lambda_t^2} &= \psi^{(2)}(y_t+\lambda_t) - \psi^{(2)}(\lambda_t) + \frac{\lambda_t y_t + \mu_t^2}{\lambda_t(\lambda_t+\mu_t)^2} \\
\frac{\partial \lambda_t}{\partial \mu_t} &= \alpha_1 \frac{\lambda_t}{\mu_t}\ , \qquad \frac{\partial \lambda_t}{\partial \alpha_0} = -\frac{\lambda_t}{\alpha_0}\ , \quad \frac{\partial \lambda_t}{\partial \alpha_1} = \lambda_t W_t \\
\frac{\partial^2 \lambda_t}{\partial \mu_t^2} &= \alpha_1(\alpha_1-1)\frac{\lambda_t}{\mu_t^2}\ , \quad \frac{\partial^2 \lambda_t}{\partial \alpha_0^2} = \frac{2\lambda_t}{\alpha_0^2}\ , \quad \frac{\partial^2 \lambda_t}{\partial \alpha_1^2} = \lambda_t W_t^2\ , \quad \frac{\partial^2 \lambda_t}{\partial \alpha_0 \partial \alpha_1} = -\frac{\lambda_t}{\alpha_0} W_t \\
\frac{\partial \sigma_t^2}{\partial \lambda_t} &= -\frac{\mu_t^2}{\lambda_t^2}\ , \qquad \frac{\partial \sigma_t^2}{\partial \mu_t} = 1 + 2\frac{\mu_t}{\lambda_t} \\
\frac{\partial^2 \sigma_t^2}{\partial \lambda_t^2} &= \frac{2\mu_t^2}{\lambda_t^3}\ , \qquad \frac{\partial^2 \sigma_t^2}{\partial \mu_t^2} = \frac{2}{\lambda_t}
\end{aligned}$$

where:

$$\psi^{(j)}(x) = \frac{\partial^j}{\partial x^j} \log \Gamma(x)\ .$$

Hybrid Generalized Poisson Distribution Type I:

$$
\begin{aligned}
\ell_t(\delta) &= y_t \log\left[\frac{\mu_t(1+\lambda_t y_t)}{1+\lambda_t\mu_t}\right] - \frac{\mu_t(1+\lambda_t y_t)}{1+\lambda_t\mu_t} - \log\left[1+\lambda_t y_t\right] - \log\left[y_t!\right] \\
\frac{\partial \ell_t}{\partial \mu_t} &= \frac{y_t-\mu_t}{\mu_t(1+\lambda_t\mu_t)^2} \\
\frac{\partial \ell_t}{\partial \lambda_t} &= \frac{(y_t-\mu_t)^2}{(1+\lambda_t y_t)(1+\lambda_t\mu_t)^2} - \frac{y_t}{1+\lambda_t y_t} \\
\frac{\partial \ell_t^2}{\partial \mu_t^2} &= -\frac{y_t}{\mu_t^2(1+\lambda_t\mu_t)^2} - \frac{2\lambda_t}{\mu_t(1+\lambda_t\mu_t)^3} \\
\frac{\partial^2 \ell_t}{\partial \lambda_t^2} &= -\frac{y_t(y_t-\mu_t)^2}{(1+\lambda_t y_t)^2(1+\lambda_t\mu_t)^2} - \frac{2\mu_t(y_t-\mu_t)^2}{(1+\lambda_t y_t)(1+\lambda_t\mu_t)} + \frac{y_t^2}{(1+\lambda_t y_t)^2} \\
\frac{\partial \lambda_t}{\partial \mu_t} &= \alpha_1\frac{\lambda_t}{\mu_t}\ , \qquad \frac{\partial \lambda_t}{\partial \alpha_0} = \mu_t^{\alpha_1}\ , \quad \frac{\partial \lambda_t}{\partial \alpha_1} = \lambda_t W_t \\
\frac{\partial^2 \lambda_t}{\partial \mu_t^2} &= \alpha_1(\alpha_1-1)\frac{\lambda_t}{\mu_t^2}\ , \ \frac{\partial^2 \lambda_t}{\partial \alpha_0^2} = 0\ , \quad \frac{\partial^2 \lambda_t}{\partial \alpha_1^2} = \lambda_t W_t^2\ , \ \frac{\partial^2 \lambda_t}{\partial \alpha_0 \partial \alpha_1} = \mu_t^{\alpha_1} W_t \\
\frac{\partial \sigma_t^2}{\partial \lambda_t} &= 2\mu_t^2(1+\lambda_t\mu_t)\ , \ \frac{\partial \sigma_t^2}{\partial \mu_t} = (1+\lambda_t\mu_t)(1+3\lambda_t\mu_t) \\
\frac{\partial^2 \sigma_t^2}{\partial \lambda_t^2} &= 2\mu_t^3\ , \qquad \frac{\partial^2 \sigma_t^2}{\partial \mu_t^2} = 2\lambda_t(2+3\lambda_t\mu_t)\ .
\end{aligned}
$$

Appendix 2: Exponential Family

Following the notation used by Dobson 1991, a density or probability mass function $g(y, \vartheta)$ belongs to the exponential family with the natural parameter ϑ if it has the form:

$$g(y, \vartheta) = \exp\left\{a(y)b(\vartheta) + c(\vartheta) + d(y)\right\} \tag{7}$$

for some known functions $a(\cdot)$, $b(\cdot)$, $d(\cdot)$ and $d(\cdot)$.

For the probability mass function of the restricted generalized Poisson dsitribution given in expression (6) one finds:

$$\begin{aligned} g(y) &= \frac{\kappa^y \left[1+\alpha_0 y\right]^{y-1} e^{-\kappa[1+\alpha_0 y]}}{y!} \\ &= \exp\left\{ y \log(\kappa) + \log\left(\frac{\left[1+\alpha_0 y\right]^{y-1}}{y!}\right) - \kappa\left[1+\alpha_0 y\right] \right\} \\ &= \exp\left\{ y \left[\, \log(\kappa) - \kappa\, \alpha_0 \,\right] - \kappa + \log\left(\frac{\left[1+\alpha_0 y\right]^{y-1}}{y!}\right) \right\} \end{aligned}$$

Comparing this result with expression (7) it can be seen that:

$$\begin{aligned} \vartheta &= \kappa \\ b(\vartheta) &= \log(\kappa) - \kappa\, \alpha_0 \\ c(\vartheta) &= -\kappa \\ d(y) &= \log\left(\frac{\left[1+\alpha_0 y\right]^{y-1}}{y!}\right) \quad . \end{aligned}$$

Therefore for given α_0 the restricted generalized Poisson distribution belongs to the exponential family.

Appendix 3: Tables and Plots

	Model under H_1					
Model under H_0	HPG-III		HPG-II		HPG-I	
Poisson	3.526	*	6.930	*	9.534	*
mod. Borel			1381.0	*	1383.6	*
HGP-III					6.008	*
HGP-II					2.604	
	NB-I		NB-II		GNB	
Poisson	3.474	*	6.760	*	9.500	*
Geometric			508.09	*	510.83	*
NB-I					6.026	*
NB-II					2.740	

Table 2: LRT for alternative count models ($*$ significance to the 5 % level).

	HGP-I	HGP-II	HGP-III	mod. Borel	Poisson	Geometric	NB-I	NB-II	GNB
Intercept	0.532 *	0.533 *	0.534 *	0.531 *	0.532 *	0.534 *	0.534 *	0.533 *	0.532 *
Sunday effect	0.235 *	0.237 *	0.236 *	0.223	0.240 *	0.228 *	0.236 *	0.237 *	0.235 *
Monday effect	0.245 *	0.245 *	0.241 *	0.272	0.244 *	0.256 *	0.241 *	0.245 *	0.246 *
$\cos(2t\pi/365)$	-0.165 *	-0.162 *	-0.162 *	-0.151	-0.163 *	-0.157 *	-0.162 *	-0.162 *	-0.165 *
$\sin(2t\pi/365)$	0.358 *	0.361 *	0.360 *	0.357 *	0.362 *	0.358 *	0.360 *	0.361 *	0.357 *
$\cos(4t\pi/365)$	-0.065	-0.066	-0.065	-0.057	-0.067	-0.061	-0.065	-0.066	-0.065
$\sin(4t\pi/365)$	0.014	0.020	0.023	0.012	0.021	0.016	0.023	0.020	0.014
$\cos(6t\pi/365)$	-0.079 *	-0.080 *	-0.078 *	-0.072	-0.080 *	-0.076	-0.078 *	-0.080 *	-0.079 *
$\sin(6t\pi/365)$	0.007	0.008	0.009	-0.008	0.009	0.000	0.009	0.008	0.007
$\cos(8t\pi/365)$	-0.144 *	-0.149 *	-0.150 *	-0.120	-0.152 *	-0.134 *	-0.150 *	-0.149 *	-0.144 *
$\sin(8t\pi/365)$	-0.058	-0.057	-0.056	-0.049	-0.057	-0.054	-0.056	-0.057	-0.058
ϕ_1	0.041 *	0.048 *	0.049 *	0.080	0.047 *	0.066	0.049 *	0.048 *	0.041 *
ϕ_3	0.045 *	0.050 *	0.050 *	0.100	0.049 *	0.071	0.050 *	0.050 *	0.047 *
ϕ_7	0.058 *	0.061 *	0.061 *	0.142	0.059 *	0.096	0.061 *	0.061 *	0.058 *
ϕ_{10}	0.042 *	0.043 *	0.042 *	0.080	0.041 *	0.063	0.042 *	0.043 *	0.042 *
α_0	0.005	0.023	0.035				0.070	0.046	0.008
α_1	1.785								-1.964

Table 3: Parameter estimates for different count models for asthma data ($*$ significance to the 5% level).

	HGP-I	HGP-II	HGP-III	mod. Borel	Poisson	Geometric	NB-I	NB-II	GNB
Paramters	17	16	16	15	15	15	16	16	17
$\ell(\hat{\delta})$	-2440.125	-2441.427	-2443.129	-3131.935	-2444.892	-2695.557	-2443.155	-2441.512	-2440.142
LRT*	290.956	288.352	284.948	38.442	357.410	115.478	286.580	289.866	292.606
R^2_{VZ}	0.213	0.211	0.209	0.032	0.251	0.093	0.210	0.212	0.214
R^2_{Mad}	0.181	0.179	0.177	0.026	0.217	0.076	0.178	0.180	0.181
R^2_{McF}	0.056	0.056	0.055	0.006	0.068	0.021	0.055	0.056	0.057
RMSE	1.473	1.471	1.471	1.481	1.471	1.475	1.471	1.471	1.473
AIC	4914.250	4914.854	4918.258	6293.87	4919.784	5421.114	4918.310	4915.024	4914.284
BS	5004.127	4999.444	5002.848	6373.173	4999.087	5500.417	5002.900	4999.614	5004.161

Table 4: Model fit statistics (* value of the likelihood ratio test between a model containing only the intercept as regressor and the time series model).

	HGP-I	HGP-II	HGP-III	mod. Borel	Poisson	Geometric	NB-I	NB-II	GNB
Mean	0.0009	0.0002	-0.0003	0.0002	0.0061	-0.0002	-0.0003	0.0002	0.0009
Variance	0.9927	0.9821	1.0031	0.1373	1.0746	0.3752	1.0041	0.9839	0.9937
Lag	Box-Pierce test statsistic on autocorrelation of the residuals								
10	8.066	8.548	8.962	5.664	8.985	6.388	8.961	8.564	8.105
20	10.738	11.220	11.697	7.465	11.717	8.589	11.695	11.238	10.736
30	14.670	15.207	15.668	12.391	15.681	12.947	15.666	15.225	14.667
50	31.407	31.448	31.540	34.269	31.609	31.424	31.538	31.453	31.408
100	78.514	78.812	78.883	84.971	78.960	79.734	78.883	78.818	78.501
Lag	Box-Pierce test statistic on autocorrelation of the squared residuals								
10	20.236 *	21.125 *	21.824 *	45.720 *	21.645 *	25.602 *	21.811 *	21.148 *	20.295 *
20	23.588	24.643	25.416	52.182 *	25.250	28.426	25.404	24.670	23.647
30	30.813	31.590	32.092	61.139 *	31.919	36.210	32.079	31.605	30.867
50	59.670	58.420	57.409	103.506 *	57.361	70.260 *	57.400	58.372	59.762
100	111.104	112.085	113.134	148.724 *	113.193	113.730	113.133	112.128	111.127

Table 5: Residual characteristics (∗ significance to the 5 % level).

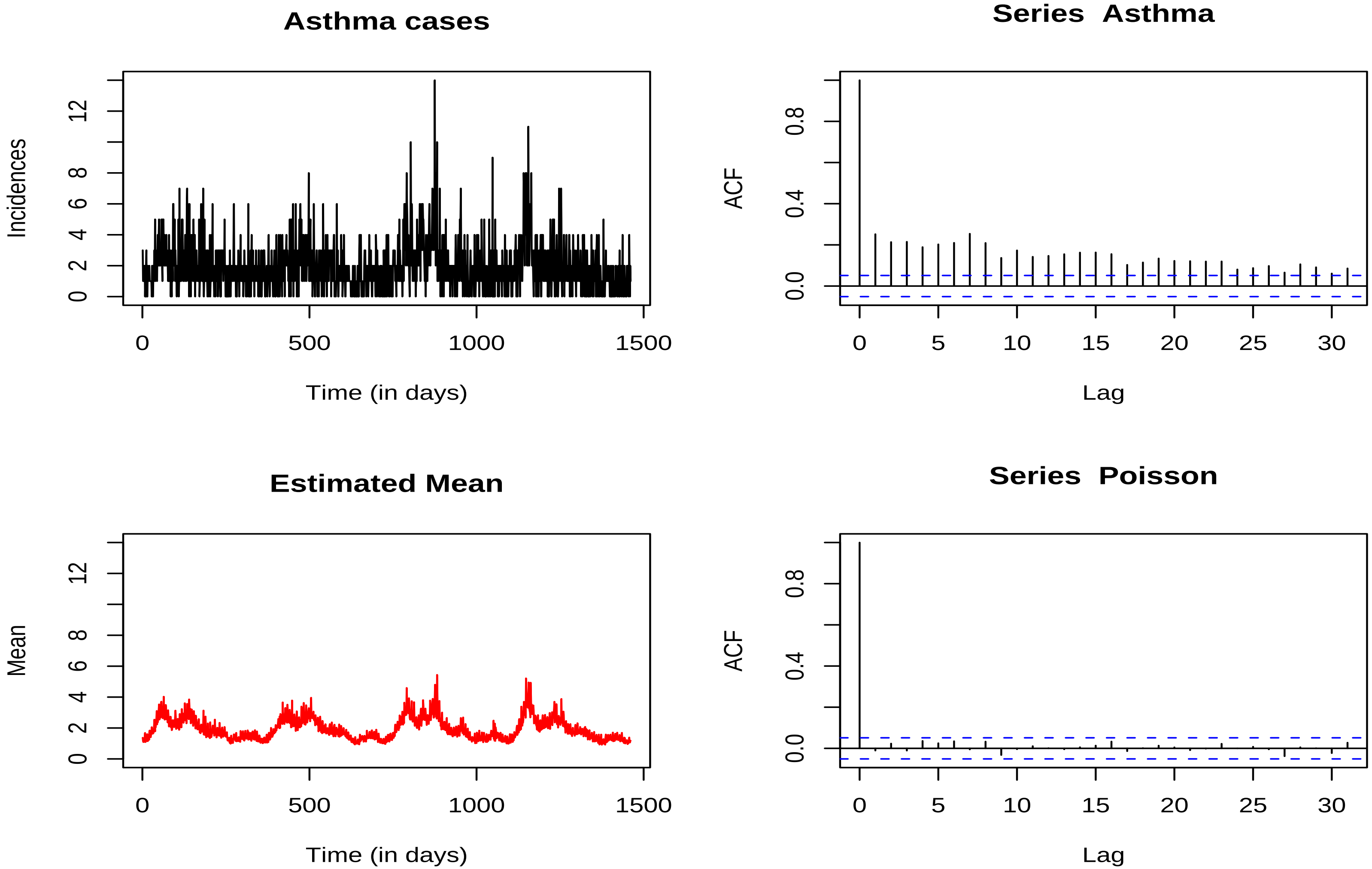

Figure 1: Astma series (top left) its autocorrelation function (top right) the series $\hat{\mu}_t$ for an Poisson-GLARMA model (bottom left) and the autocorrelation function of the estimated Pearson residuals (bottom rigth).

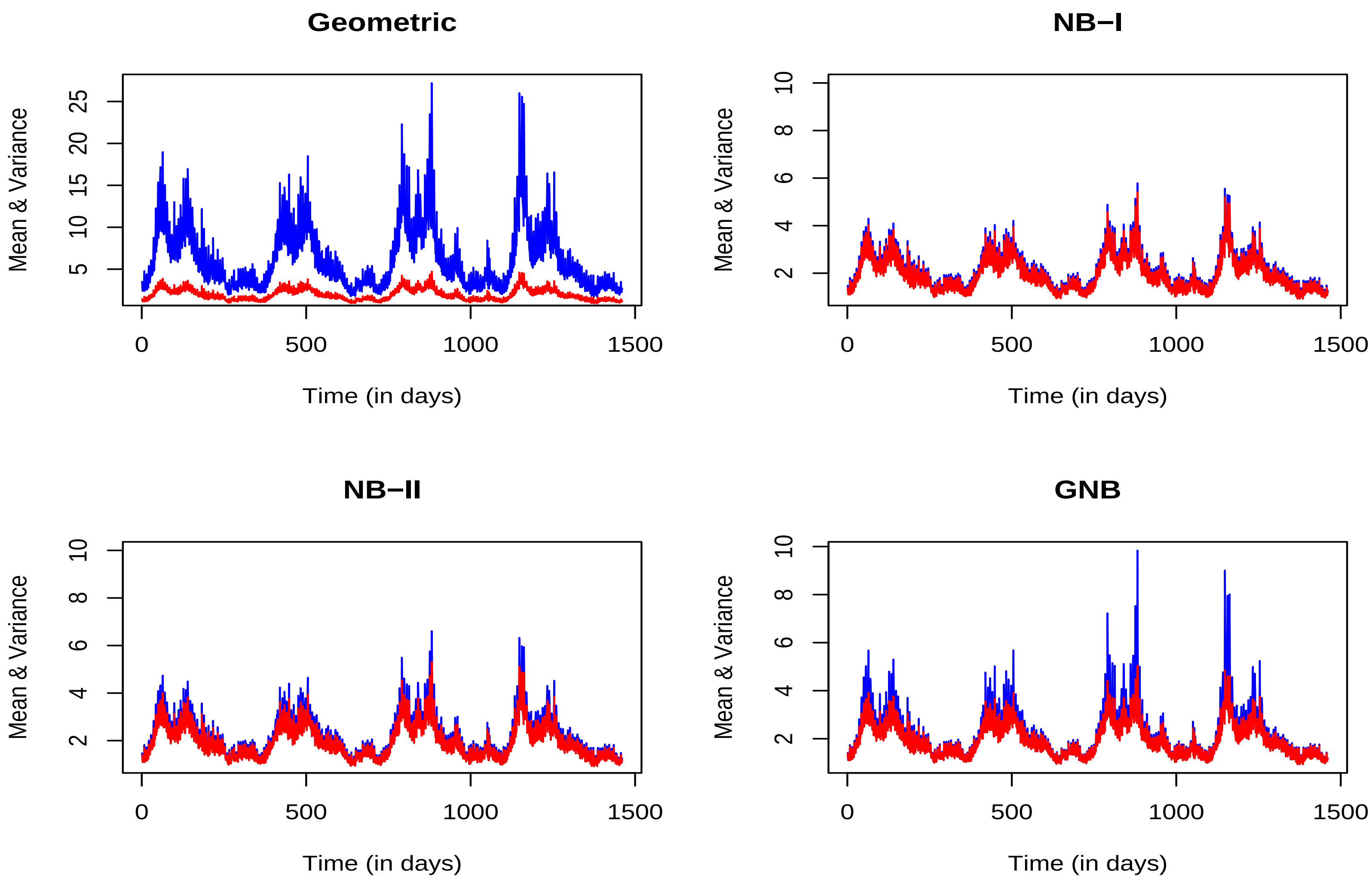

Figure 2: Estimated mean (red) and variance (blue) for count models nested in the generalized negative binomial model.

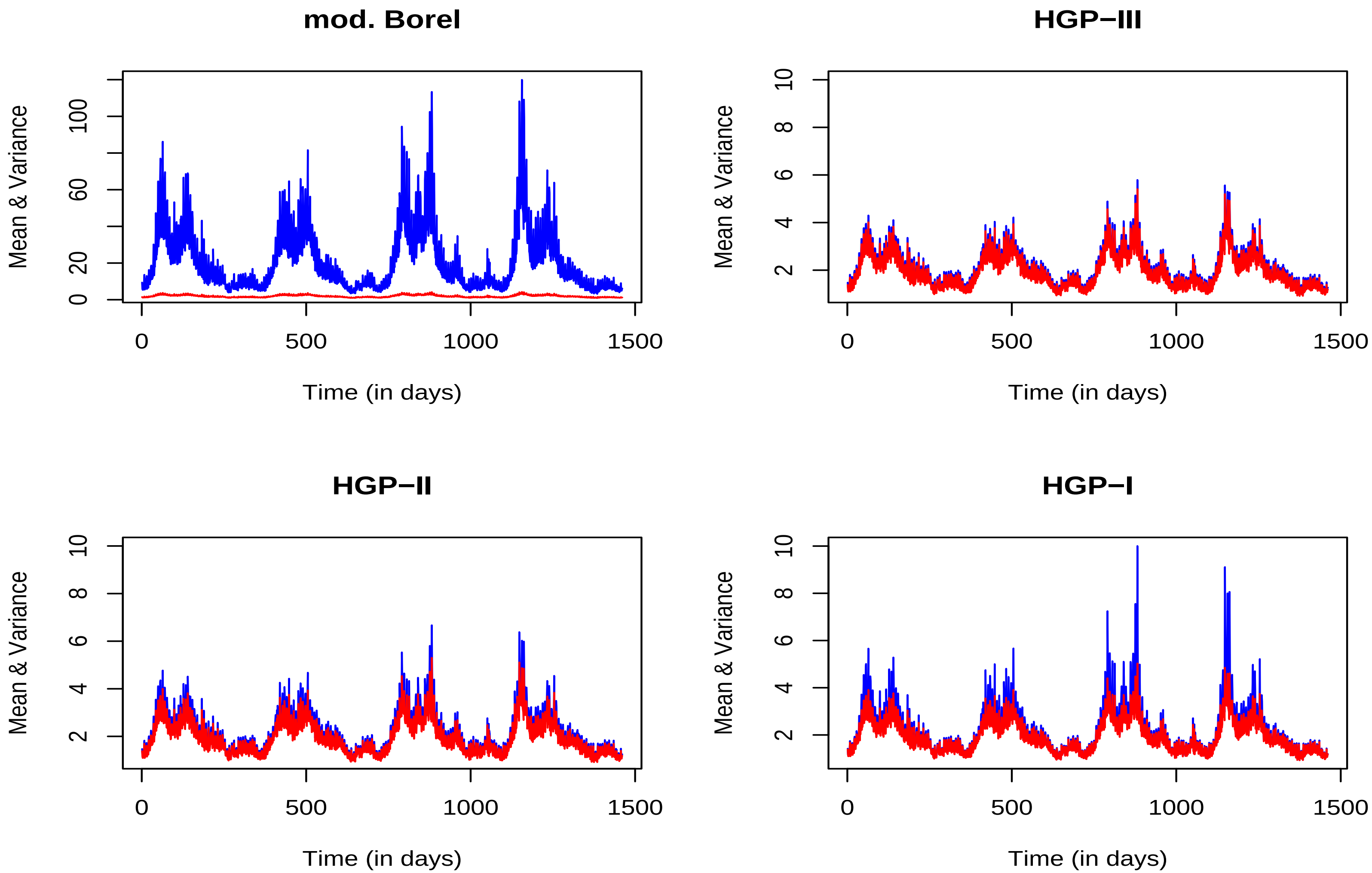

Figure 3: Estimated mean (red) and variance (blue) for count models nested in the hybrid generalized Poisson model type I.

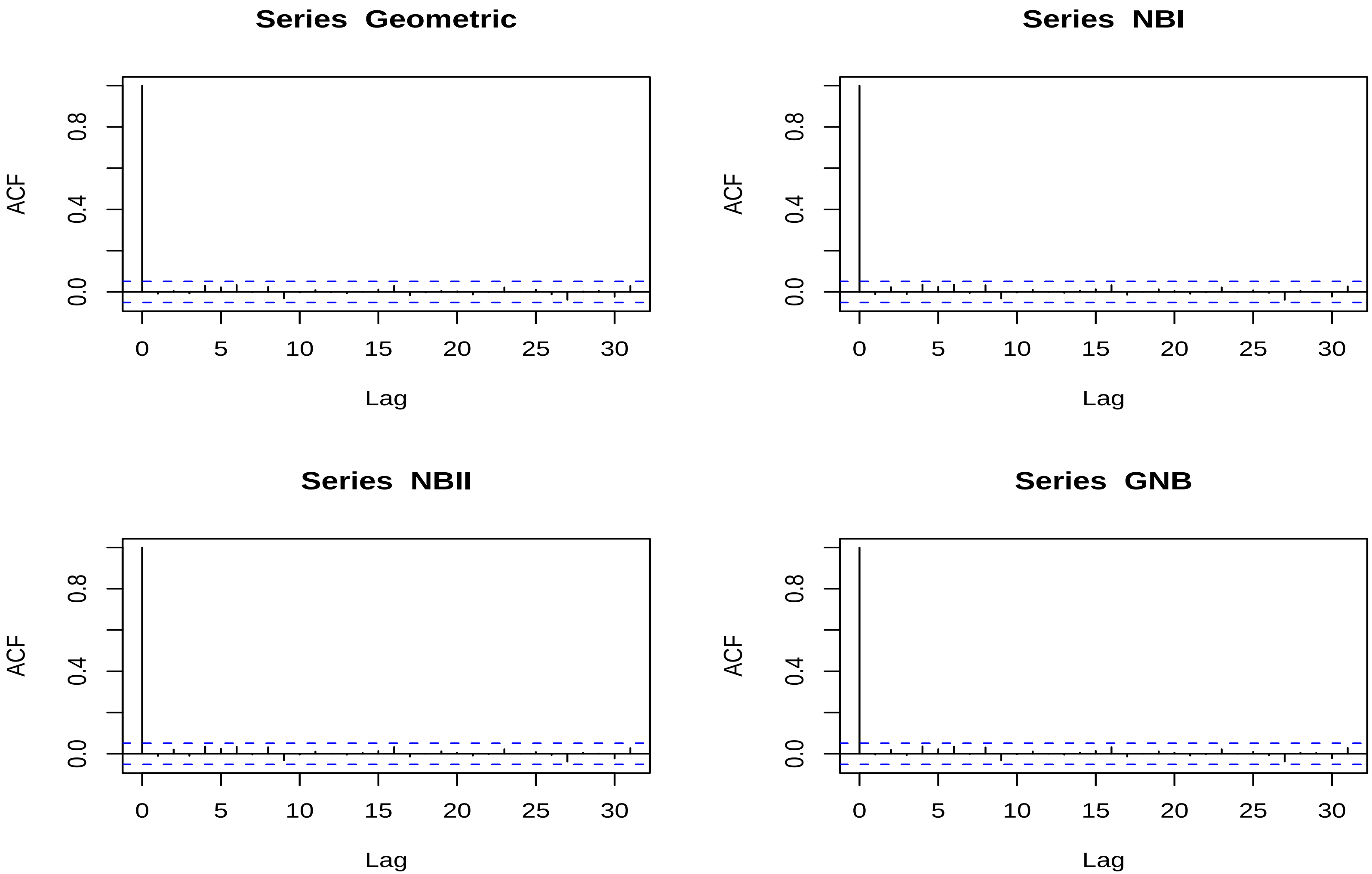

Figure 4: Autocorrelation function for the estimated Pearson residuals of count models nested in the generalized negative binomial model.

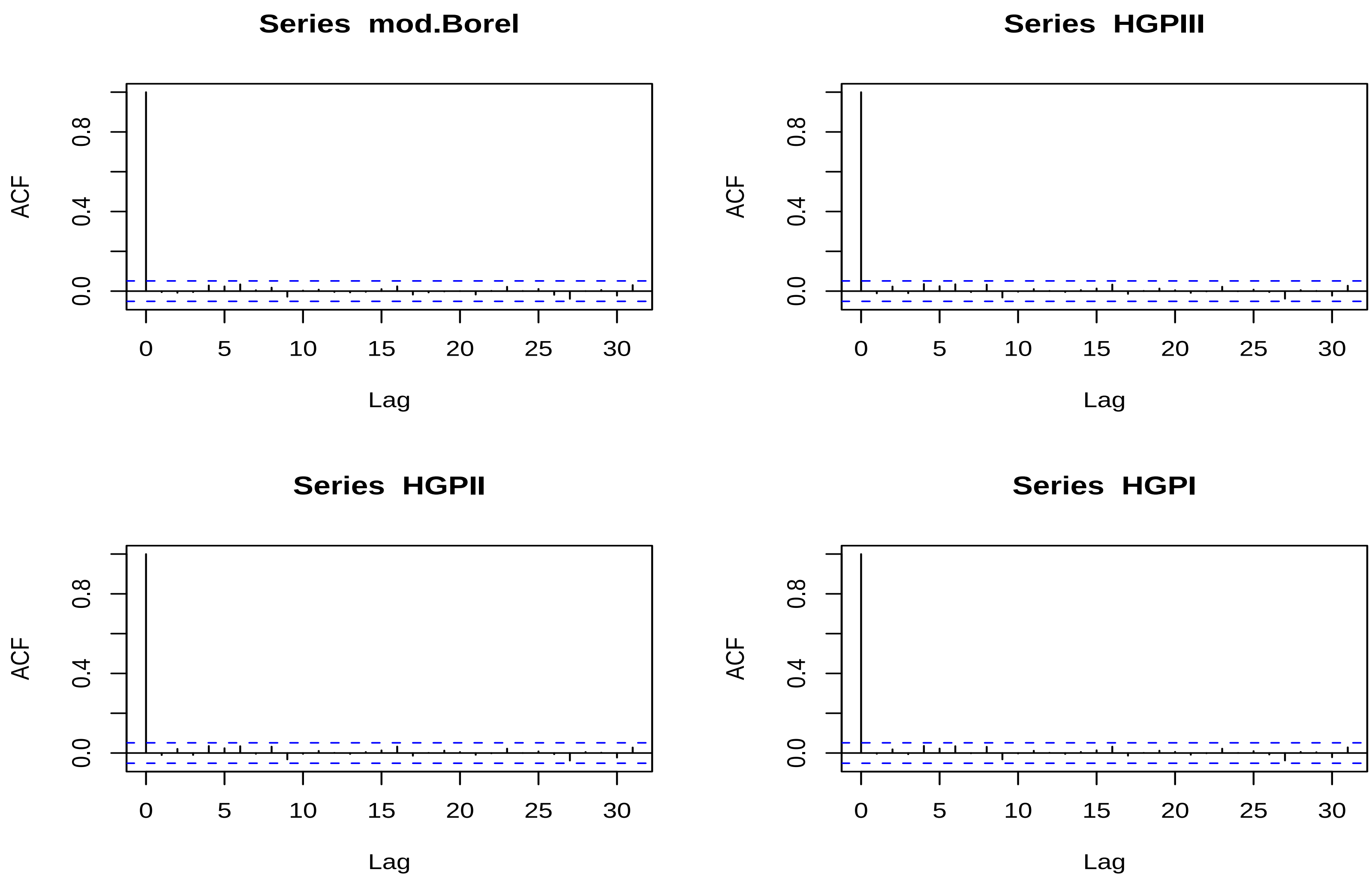

Figure 5: Autocorrelation function for the estimated Pearson residuals of count models nested in the hybrid generalized Poisson model type I.

References

Al-Osh, M. A.; Alzaid, A. A. (1988): Integer-valued moving average (IN-MA) process. *Statistical Papers* **29**, 281-300.

Al-Osh, M. A.; Alzaid, A. A. (1990): An integer-valued p-th order autoregressive structure (INAR(p)) process. *Journal of Applied Probability* **27**, 314-324.

Akaike, H. (1973): Information theory and an extension to the maximum likelihood principle. In: Petroc, B. N.; Kaski, F. (Ed.) *Second International Symposium of Information Theory*, 267-281. Budapest: Akademiai Kiado.

Borel, E. (1942): Sur l'emploi du théorème de bernoulli pour faciliter le calcul d'une infinité de coefficients. Applications au problème de l'attente à un guichet. *Comptes Rendus de l'Académie des Science* **214**, 452-456.

Broyden, C. G. (1970): The convergence of a class of double rank minimization algorithms. *Journal of the Institute of Mathematics and its Applications* **6**, 76-90.

Cameron, A. C.; Trivedi, P. K. (1986): Econometric models based on count data: Comparison and applications of some estimators. *Journal of Applied Econometrics* **1**, 29-53.

Consul, P. J. (1989): *Generalized poisson distributions - properties and applications*. New York: Marcel Decker.

Consul, P. J.; Famoye, F. (1992): Generalized poisson regression model. *Communications in Statistics - Theory and Methods* **21**, 89-109.

Consul, P. J.; Jain, G. C. (1973): A generalization of the poisson distribution. *Technometrics* **15**, 791-799.

Cox, D. R. (1981): Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics* **8**, 93-115.

Czado, C.; Kolbe, A. (2004): Statistical analysis of absolute transaction price changes of options. *Working Paper No.384*, Munich University of Technology. http://www-m4.ma.tum.de/Papers/Czado/cc-pubs.html (20.06.2005)

Davis, R. A.; Dunsmuir, W. T. M.; Streett, S. B. (2003): Observation-driven models for Poisson counts. *Biometrika* **90**, 777-790.

Davis, R. A.; Dunsmuir, W. T. M.; Streett, S. B. (2004): Maximum Likelihood estimation for an observation driven model for poisson counts. Revised and resubmitted to *Methodology, Computing and Applied Probability.*

Davis, R. A.; Dunsmuir, W. T. M.; Wang, Y. (1999): Modelling time series of count data. In: Ghosh, S. (Ed.) *Asymptotics, Nonparametrics and Time Series*, 63-114. New York: Marcel Decker.

Dobson, A.J. (1991): *An Introduction to Generalized Linear Models.* London: Chapman and Hall.

Fahrmeir, L.; Kaufmann, H (1985): Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics* **13**, 342-368.

Famoye, F. (1993): Restricted generalized poisson regression model. *Communications in Statistics - Theory and Methods* **22**, 1335-1354.

Fletcher, R. (1970): A new approach to variable metric algorithms. *Computer Journal* **13**, 317- 322.

Goldfarb, D. (1970): A family of variable metric updates derived by variational means. *Mathematics of Computing* **24**, 23-26.

Jackobs, P. A.; Lewis, W. A. W. (1978): Discrete time series generated by mixtures I: correlation and runs properties. *Journal of the Royal Statistical Society* **B 40**, 94-105.

Jørgenson, B. (1986): Some properties of exponential dispersion models. *Scandinavian Journal of Statistics* **13**, 187-197.

Lawless, J. F. (1987): Negative binomial and mixed poisson regression. *The Canadian Journal of Statistics* **15**, 209-225.

Liesenfeld, R.; Pohlmeier, W. (2003): A dynamic integer count data model for financial transaction prices. *Working Paper*, University of Konstanz. http://econometrics.wiwi.uni-konstanz.de/prof/papers.htm (20.06.2005)

Maddala, G. S. (1983): *Limited-dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.

McCullagh, P.; Nelder, J. A. (1989): *Generalized linear models*. London: Chapmann and Hall.

McFadden, D. (1973): Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka, P. (Ed.) *Frontiers in Econometrics*, 105-142, New York: Academic Press.

Rydberg, T. H.; Shephard, N. (2003): Dynamics of trade-by-trade price movements: Decomposition and models. *Journal of Financial Econometrics* **1**, 2-25.

Santos Silva, J. M. C. (1997): Generalized poisson regression for positive count data. *Communications in Statistics - Simulation and Computation* **26**, 1089-1102.

Saha, A.; Dong, D. (1997): Estimating nested count data models. *Oxford Bulletin in Economics and Statistics* **59**, 423-430.

Schwarz, G. (1978): Estimating the dimension of a model. *Annals of Statistics* **26**, 461-464.

Shanno, D. F. (1970): Conditioning of Quasi-Newton methods for function minimization. *Mathematics of Computing* **24**, 647-656.

Shephard, N. (1995): Generalized linear autoregression. *Working Paper*, University of Oxford. http://www.nuff.ox.ac.uk/users/shephard (20.06.2005).

Tanner, J. C. (1953): A problem of inference between two queues. *Biometrica* **40**, 58-69.

Zeger, S. L. (1988): A regression model for time series for counts. *Biometrika* **75**, 621-629.

Veall, M. R,; Zimmermann, K. F. (1992): Pseudo-R´s in the ordinal probit model. *Journal of Mathematical Sociology* **16**, 332-342.

Winkelmann, R. (2000): *Econometric analysis of count data, 3rd ed.* Berlin: Springer-Verlag.