Jung, Robert; Kukuk, Martin; Liesenfeld, Roman

Working Paper

Time Series of Count Data : Modelling and Estimation

Economics working paper / Christian-Albrechts-Universität Kiel, Department of Economics, No. 2005,08

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

Time Series of Count Data: Modelling and Estimation

by Robert Jung, Martin Kukuk and Roman Liesenfeld

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

Economics Working Paper
No 2005-08

Time Series of Count Data: Modelling and Estimation

Robert Jung
Department of Economics, Eberhard-Karls Universität, Tübingen, Germany (robert.jung@uni-tuebingen.de)

Martin Kukuk
Department of Economics, Bayerische Julius-Maximilians-Universität, Würzburg, Germany (martin.kukuk@mail.uni-wuerzburg.de)

Roman Liesenfeld
Department of Economics, Christian-Albrechts-Universität, Kiel, Germany (liesenfeld@stat-econ.uni-kiel.de)

First version: August 14, 2004
This version: May 13, 2005

This paper is dedicated to our joint teacher Gerd Ronning in honour of his 65th birthday.

Abstract

This paper compares various models for time series of counts which can account for discreetness, overdispersion and serial correlation. Besides observation- and parameter-driven models based upon corresponding conditional Poisson distributions, we also consider a dynamic ordered probit model as a flexible specification to capture the salient features of time series of counts. For all models, we present appropriate efficient estimation procedures. For parameter-driven specifications this requires Monte Carlo procedures like simulated Maximum likelihood or Markov Chain Monte-Carlo. The methods including corresponding diagnostic tests are illustrated with data on daily admissions for asthma to a single hospital.

KEY WORDS: Efficient Importance Sampling; GLARMA; Markov Chain Monte-Carlo; Observation-driven model; Parameter-driven model; Ordered Probit.

1 Introduction

Time series of observed counts arise in a wide variety of contexts including studies of incidences of a certain disease (see Zeger, 1988 and Davis et al. 1999) or of discrete transaction price movements on financial markets (see Liesenfeld et al. 2005 and Rydberg and Shephard, 2003). A successful model for such series should take the following features regularly found in the data into account: (1) a sometimes rather pronounced dependence structure; and (2) extra binomial variation or overdispersion relative to the mean of the series. Moreover, in a regression context an easy to use link function allowing for a straightforward interpretation of the effects of the covariates is also a highly desirable modelling feature. Statistical models suitable for dependent counts are usually classified as being either observation- or parameter-driven in nature. (While in the latter case a latent dynamic process governs the conditional mean function, the dependence structure in the former class is introduced via the incorporation of lagged values of the observed counts directly into the mean function of the model.) A survey of the rather heterogeneous literature in this area can be found in the monographs of Cameron and Trivedi (1998) and Kedem and Fokianos (2002).

An interesting member of the class of observation driven models is undoubtedly the group of integer valued autoregressive moving average (INARMA) models. Its decisive feature is the use of appropriate thinning operations replacing the scalar multiplications in the Gaussian ARMA framework of time series consisting of continuous data. Theoretical models covering a wide range of possible correlation structures combined with equidispersed as well as overdispersed discrete marginal distributions are available in the literature (see McKenzie, 2003 for an up to date survey). However, only a limited range of models has been systematically analyzed in terms of their practical applicability so far (see Jung and Tremayne, 2005, for a survey on some recent work in this area). In particular, regression models that are able to cope with the two aforementioned features of the data seem not to be readily available yet.

Another group of models in the observation-driven class are the generalized linear autoregressive moving average (GLARMA) models proposed, e.g., by Davis et al. (1999,2003) and Shephard (1995). They extend the familiar generalized linear models framework to allow for serial correlation as well as extra binomial variation in the data by specifying the log of the conditional mean process as a linear function of previous counts. In contrast to the INARMA specifications, it is straightforward to include covariates into those models. An additional appeal of the GLARMA models is that their efficient estimation by maximum likelihood (ML) is easy to implement. A similar group of observation-driven models only recently proposed by Heinen (2003) are the autorgressive conditional Poisson (ACP) models.

The benchmark model in the class of parameter-driven specifications has been introduced by

Zeger (1988). This model has successfully been applied to a wide range of predominantly biometric problems. It extends the generalized linear models by incorporating into the conditional mean function a latent autoregressive process which evolves independently of the past observed counts. This process introduces autocorrelation as well as overdispersion into the model. The main problem with this class of stochastic autorgressive mean (SAM) specifications is that their efficient estimation is not straightforward and typically requires methods based on Monte Carlo (MC) integration. Examples of such methods for fitting dynamic SAM models are the Monte-Carlo EM approach proposed by Chan and Ledolter (1995), the Monte Carlo Newton Raphson method of Kuk and Chen (1997), and the ML procedure based on Richard and Zhang's (2004) efficient importance sampling (EIS) applied by Jung and Liesenfeld (2001). However, as noted by Davis et al. (2003), such estimation procedures are not yet rontinely available, especially, in a context involving many covariates, and realistically long and numerous time series.

The purpose of this paper is twofold. Firstly, we propose to illustrate how to handle the estimation problem of the Zeger-type SAM models by using EIS which represents a highly flexible MC integration procedure which is easy to implement even for many covariates and long time series for counts. In particular, we show how use EIS to perform not only a classical ML estimation, but also a Markov Chain Monte-Carlo (MCMC) Bayesian posterior analysis of the SAM models. Secondly, we compare this parameter-driven SAM model with observation-driven alternatives with respect to their ability to account for dynamic and distributional properties of count data. Especially, we consider Heinen's (2003) ACP approch and the GLARMA model of Davis el al. (2003) as well as a highly flexible approach based on an ordered probit specification. In order to assess the dynamic and the distributional properties of the fitted models we report on useful diagnostics.

The models and methods proposed in this paper are illustrated with data of daily admissions for asthma to a single hospital (at Campbelltown) in the Sydney metropolitan area from 1 January 1990 to 31 December 1993, giving a sample of 1461 observations. This data set has been previously analyzed by Davis et al. (1999) using a GLARMA model and in Davis et al. (2000) using a generalized linear model analysis. Panel (a) of Figure 1 shows the time series plot of the daily number of asthma presentations. The observed counts vary from 0 to 14 with no deterministic trend evident. The mean of the series is 1.94 and its variance is 2.71 indicating the possible presence of extra binomial variation in the marginal distribution of the data. The autocorrelation properties of the data are depicted in Panel (c) and (d) of Figure 1. The counts as well as the squared counts exhibit a relatively small but significant positive autocorrelation. Additional features of the data that should be accounted for in an empirical analysis and which have been revealed by the analysis in Davis et al. (1999) include seasonal cycles and day of the week effects due to higher number of admittances on Sundays and Mondays.

The remainder of the paper is organized as follows. Section 2 presents the parameter-driven SAM model for count data and corresponding efficient estimation procedures. In Section 3, we discuss the observation-driven ACP and GLARMA model. Dynamic ordered probit models as a possible alternative to standard count data models are introduced in Section 4. Section 5 concludes.

2 Poisson model with a stochastic autoregressive mean

2.1 Specification

A prominent dynamic count data model is Zeger's (1988) parameter-driven Poisson model with a stochastic autoregressive mean. Let $\{y_t, t : 1 \to T\}$ denote a univariate time series of counts, let $\{x_t\}$ be a sequence of k-dimensional vectors of covariates, and let $\{u_t\}$ denote a latent non-negative stochastic process. Then the conditional distribution of $y_t|(x_t, u_t)$ is assumed to be Poisson with mean $\mu_t = \exp\{x_t'\varphi\}u_t$ denoted by

$$y_t|x_t, u_t \sim \text{Po}(\exp\{x_t'\varphi\}u_t), \qquad t : 1 \to T, \tag{1}$$

where $\varphi = (\varphi_1, ..., \varphi_k)'$ is a vector of regression parameters. The latent process u_t is typically introduced to account for possible overdispersion and serial correlation within the data. A convenient assumption successfully used, e.g., by Chan and Ledolter (1995), Kuk and Cheng (1997) and Jung and Liesenfeld (2001) is that $\lambda_t = \ln(u_t)$ follows a Gaussian first-order autoregressive process, satisfying

$$\lambda_t = \delta\lambda_{t-1} + \nu\epsilon_t, \qquad \epsilon_t \sim iid\text{N}(0, 1). \tag{2}$$

To ensure stationarity of λ_t, it is assumed that $|\delta| < 1$. Note that for $\delta = 0$ and $\nu \to 0$, the latent process vanishes and we obtain a standard Poisson regression model. A complete description of the statistical properties of the SAM model (1) and (2) is provided by Davis et al. (1999).

Unlike the observation driven Poisson model discussed below, the mean function of the parameter-driven SAM model is equipped with a separate dynamic latent error term. This gives rise to a straightforward interpretation of the effects of the covariates on the count process in the SAM model. Furthermore, its stochastic properties are easy to derive. On the other hand, the dynamic latent process leads to a likelihood function which depends upon high-dimensional integrals so that its efficient estimation is not straightforward. However, nowadays, a variety of efficient classical estimation procedures (including the Monte Carlo EM algorithm of Chan and Ledolter, 1995 and the Monte Carlo Newton Raphson method of Kuk and Cheng, 1997) are available to estimate dynamic latent variable models like the SAM model within minutes. Here, we propose to use the EIS procedure either to obtain ML parameter estimates and/or to perform a Bayesian MCMC analysis of the SAM model. EIS is a MC integration technique developed by Richard and Zhang (2004) for the evaluation of high-dimensional integrals. Apart from its adaptability for a classical as well as for a Bayesian analysis of count data models with dynamic latent variables, EIS has the attractive feature to be highly generic, since its basic structure does not depend upon a specific model. (This is illustrated, e.g., in Jung and Liesenfeld (2001) where EIS is applied for a classical analysis of the SAM model using different specifications of the latent process obtained by different orders of serial dependence and different distributional assumptions.)

In the following two subsections, we provide a description of EIS and its application to the efficient estimation of the SAM model. For the general theory of EIS, see Richard and Zhang (2004).

2.2 ML-EIS estimation

Estimation of the parameters $\theta = (\varphi', \delta, \nu)'$ in the parameter driven model (1) and (2) by direct numerical maximization of the likelihood function is difficult since the likelihood does not have a closed-form solution and cannot be evaluated by standard numerical procedures. In particular, let $Y = \{y_t\}_{t=1}^T$, $\Lambda = \{\lambda_t\}_{t=1}^T$, $X = \{x_t\}_{t=1}^T$, $Y_s = \{y_\tau\}_{\tau=1}^s$ and $\Lambda_s = \{\lambda_\tau\}_{\tau=1}^s$. Then, the likelihood is the following T-fold integral:

$$L(\theta; Y, X) = \int f(Y, \Lambda | X, \theta) d\Lambda = \int \prod_{t=1}^T f_t(y_t, \lambda_t | \Lambda_{t-1}, Y_{t-1}, X, \theta) d\Lambda, \tag{3}$$

where $f(Y, \Lambda | X, \theta)$ represents the joint conditional density for (Y, Λ) given X, which can be factorized into the sequence of conditional densities $f_t(\cdot)$ for (y_t, λ_t) given $(\Lambda_{t-1}, Y_{t-1}, X)$. For convenience we set the initial value λ_0 equal to $\mathrm{E}(\lambda_t) = 0$. The assumptions relative to the dynamic structure of the latent process and the conditional distribution of counts lead to the following additional factorization:

$$f_t(y_t, \lambda_t | \Lambda_{t-1}, Y_{t-1}, X, \theta) = g_t(y_t | \lambda_t, x_t, \theta) p_t(\lambda_t | \lambda_{t-1}, \theta), \tag{4}$$

where $g_t(\cdot)$ denotes the conditional density of y_t given (λ_t, x_t) and $p_t(\cdot)$ the conditional density of λ_t given λ_{t-1}. These densities are given by

$$g_t(y_t | \lambda_t, x_t, \theta) \propto \exp(-\exp\{x_t'\varphi + \lambda_t\}) \exp\{\lambda_t\}^{y_t} \tag{5}$$

$$p_t(\lambda_t | \lambda_{t-1}, \theta) \propto \exp\{-\frac{1}{2\nu^2}(\lambda_t - \delta\lambda_{t-1})^2\}, \tag{6}$$

where multiplicative factors which do not depend upon λ_t are omitted.

A natural tool to approximate this intractable likelihood is provided by MC integration. In particular, the direct MC evaluation of the likelihood based on the factorizations (3) and (4) is given by

$$\bar{L}_N(\theta; Y, X) = \frac{1}{N} \sum_{i=1}^{N} \left[\prod_{t=1}^{T} g_t(y_t | \tilde{\lambda}_t^{(i)}, x_t, \theta) \right], \tag{7}$$

where $\{[\tilde{\lambda}_1^{(i)}, ..., \tilde{\lambda}_T^{(i)}], i : 1 \to N\}$ are N independent trajectories from the natural sampler given by the sequence of p_t densities. However, it is well-known that such an MC approximation is typically highly inefficient with a very large MC sampling variance. Essentially, this follows from the fact that the λ_t's sampled from the p_t densities do not bear any resemblance to the true values of the latent process under which the counts y_t are obtained. A dramatic illustration of the resulting inefficiency is provided by Danielsson and Richard (1993).

In order to resolve this efficiency problem, EIS replaces the initial sampling densities p_t by a sequence of auxiliary importance samplers $\{m_t(\lambda_t | \lambda_{t-1}, a_t)\}_{t=1}^{T}$ indexed by the auxiliary parameters $a = \{a_t\}_{t=1}^{T}$. Typically, the class of samplers m_t include parametric extensions of the initial samplers p_t. For any choice of the auxiliary parameters, the integral in (3) can be rewritten as

$$L(\theta; Y, X) = \int \prod_{t=1}^{T} \frac{f_t(y_t, \lambda_t | \Lambda_{t-1}, Y_{t-1}, X, \theta)}{m_t(\lambda_t | \lambda_{t-1}, a_t)} \prod_{t=1}^{T} m_t(\lambda_t | \lambda_{t-1}, a_t) d\Lambda, \tag{8}$$

and the corresponding importance sampling MC estimate is given by

$$\tilde{L}_N(\theta; Y, X, a) = \frac{1}{N} \sum_{i=1}^{N} \left[\prod_{t=1}^{T} \frac{f_t(y_t, \breve{\lambda}_t^{(i)} | \breve{\Lambda}_{t-1}^{(i)}, Y_{t-1}, X, \theta)}{m_t(\breve{\lambda}_t^{(i)} | \breve{\lambda}_{t-1}^{(i)}, a_t)} \right], \tag{9}$$

where $\{[\breve{\lambda}_1^{(i)}, ..., \breve{\lambda}_T^{(i)}], i : 1 \to N\}$ are N independent trajectories from the sequence of importance sampling densities m_t. Then, for a given parametric class of samplers m_t, EIS aims at selecting a_t's that minimize the MC sampling variance of the MC approximation (9), which is tantamount to selecting a_t's such that in Equation (9) the denominator $\prod_t m_t$ be as close as possible to being proportional to the numerator $\prod_t f_t$. The sequential implementation of this minimization is based on density kernels $k_t(\lambda_t, \lambda_{t-1}, a_t)$ of the m_t densities, satisfying

$$m_t(\lambda_t | \lambda_{t-1}, a_t) = \frac{k_t(\lambda_t, \lambda_{t-1}, a_t)}{\chi_t(\lambda_{t-1}, a_t)}, \quad \text{where} \quad \chi_t(\lambda_{t-1}, a_t) = \int k_t(\lambda_t, \lambda_{t-1}, a_t) d\lambda_t, \tag{10}$$

and EIS requires solving a back-recursive sequence of least-squares problems of the form

$$\begin{aligned} (\hat{c}_t, \hat{a}_t) \quad = \quad \arg\min_{c_t, a_t} \sum_{i=1}^{N} \Bigg\{ & \ln \Big[f_t\big(y_t, \tilde{\lambda}_t^{(i)} | \tilde{\Lambda}_{t-1}^{(i)}, Y_{t-1}, X, \theta\big) \cdot \chi_{t+1}\big(\tilde{\lambda}_t^{(i)}, \hat{a}_{t+1}\big) \Big] \\ & - c_t - \ln k_t\big(\tilde{\lambda}_t^{(i)}, \tilde{\lambda}_{t-1}^{(i)}, a_t\big) \Bigg\}^2, \end{aligned} \tag{11}$$

for $t : T \to 1$, where $\chi_{T+1}(\cdot) \equiv 1$. (A weighted least squares version of (11) is provided in Richard and Zhang, 2004.) The N independent trajectories $\{[\tilde{\lambda}_1^{(i)}, ..., \tilde{\lambda}_T^{(i)}], i : 1 \to N\}$ are drawn from the sequence of p_t densities, and the c_t's are constants to be estimated jointly with the a_t's. Finally, the MC EIS estimate of the likelihood is obtained by substituting $\hat{a} = \{\hat{a}_t\}_{t=1}^T$ for a in (9) and ML-EIS estimates of θ are obtained by maximizing $\tilde{L}_N(\theta; Y, X, \hat{a})$ with respect to θ using a standard numerical optimizer. (For a detailed description of the implementation of EIS see the Appendix.)

EIS can also be used to compute filtered estimates of the latent λ_t or of functions thereof (see, e.g., Jung and Liesenfeld, 2001). Filtering enables us to perform diagnostic tests, e.g., based on the standardized (Pearson) residuals $z_t = [y_t - \mathrm{E}(y_t|Y_{t-1}, X_t)]/\mathrm{var}(y_t|Y_{t-1}, X_t)^{1/2}$, where the conditional moments under the SAM model are given by

$$\mathrm{E}(y_t|Y_{t-1}, X_t) = \exp\{x_t'\varphi\}\mathrm{E}(\exp\{\lambda_t\}|Y_{t-1}, X_{t-1})$$

$$\mathrm{var}(y_t|Y_{t-1}, X_t) = \exp\{x_t'\varphi\}\left[\mathrm{E}(\exp\{\lambda_t\}|Y_{t-1}, X_{t-1}) + \exp\{x_t'\varphi\}\mathrm{var}(\exp\{\lambda_t\}|Y_{t-1}, X_{t-1})\right] .$$

If the model is correctly specified, z_t has mean zero and unit variance and is serially uncorrelated in the first- and second-order moments.

In order to check the adequacy of the distributional assumptions of the SAM model, we use a generalization of the approach followed, e.g., by Kim et al. (1998) which is based on predicted probabilities and which exploits Rosenblatt's (1952) transformation of an absolutely continuous conditional distribution into a uniform distribution. In particular, a generalization to the discrete case can be based on residuals which are obtained as a sequence of simulated random draws $\tilde{u}_t$ from uniform distributions on the intervals $[c_t^{(l)}, c_t^{(u)}]$, where $c_t^{(u)}$ and $c_t^{(l)}$ are the predicted probabilities that the random variable y_t be less than the actually observed count y_t^{o} and less than $y_t^{\mathrm{o}} - 1$, respectively, i.e.

$$\tilde{u}_t \sim \mathcal{U}(c_t^{(l)}, c_t^{(u)}), \qquad t : 1 \to T, \tag{12}$$

with

$$c_t^{(u)} = P(y_t \leq y_t^{\mathrm{o}}|Y_{t-1}, X_t) \quad \text{and} \quad c_t^{(l)} = P(y_t \leq y_t^{\mathrm{o}} - 1|Y_{t-1}, X_t). \tag{13}$$

If the model is correctly specified, $\tilde{u}_t$ is a serially independent random variable following a uniform distribution on the interval $[0, 1]$. The sequence of conditional probabilities in Equation (13) which can be represented as

$$P(y_t \leq \bar{y}|Y_{t-1}, X_t) = \mathrm{E}\left(\sum_{j=0}^{\bar{y}} \frac{\exp(-\exp\{x_t'\varphi + \lambda_t\})\exp\{\lambda_t\}^j}{j!}\middle| Y_{t-1}, X_t\right), \tag{14}$$

can be produced by EIS integration (see Liesenfeld and Richard, 2003). Using the inverse of a standard normal distribution function denoted by F_N^{-1}, the variable $\tilde{u}_t$ can be mapped into a

N(0, 1)-distribution:

$$z_t^* = F_N^{-1}(\tilde{u}_t). \tag{15}$$

Under the hypothesis that the model is correctly specified, the normalized residuals z_t^* are serially independent variables following a standard normal distribution.

2.3 A Bayesian analysis based on EIS

So far, we have discussed the application of EIS for classical inference of the SAM model. We now show how to use EIS to carry out a Bayesian MCMC posterior analysis of the parameters via Gibbs sampling. Under a Bayesian treatment, the vector of parameters θ is augmented with the vector of the latent process Λ. Then, the Gibbs sampling approach of estimating the SAM model involves drawing from the conditional posterior distribution $f(\theta|Y, X, \Lambda)$ for θ given (Y, X, Λ) and from the conditional distribution $f(\Lambda|Y, X, \theta)$ for Λ given (Y, X, θ). The parameter vector θ is estimated by reporting appropriate statistics for the simulations of $(\theta, \Lambda)|(Y, X)$ from the joint posterior $f(\theta, \Lambda|Y, X)$.

The main difficulty with such an MCMC approach is that of simulating from the conditional posterior $f(\Lambda|Y, X, \theta)$, which is an unknown high-dimensional distribution. This suggests to sample the T-dimensional Λ using a Gibbs sampler based on T univariate conditional posteriors $\lambda_t|\Lambda_{\backslash t}, Y, X, \theta$, where $\Lambda_{\backslash t}$ denotes Λ without the t-th element. However, a disadvantage of this approach is that high correlation between the elements in Λ leads to a very slow convergence of the MCMC algorithm, a particularly severe problem in time-series applications (see, e.g., Shephard and Pitt, 1997).

Here, we propose to use a combination of the EIS sampler with Tierney's (1994) Acceptance-Rejection Metropolis-Hastings (AR-MH) to simulate $\Lambda|Y, X, \theta$ as one block, which eliminates the slow convergence due to high correlation in the Λ-elements. (For a detailed description of the AR-MH procedure, see Chib and Greenberg, 1995.) The basis of such a procedure is the fact that the EIS-sampler provides the best approximation (within a preassigned parametric class of distributions) to the target density $f(\Lambda|Y, X, \theta)$ which has the form

$$f(\Lambda|Y, X, \theta) \propto f(Y, \Lambda|X, \theta) = \prod_{t=1}^{T} f_t(y_t, \lambda_t|\Lambda_{t-1}, Y_{t-1}, X, \theta). \tag{16}$$

Hence, one can expect that the EIS-sampling density provides an efficient proposal distribution for the target density $f(\Lambda|Y, X, \theta)$ within an acceptance-rejection algorithm. The corresponding functional approximation is of the form

$$f(Y, \Lambda|X, \theta) \simeq M(\Lambda) := \prod_{t=1}^{T} m_t(\lambda_t|\lambda_{t-1}, \hat{a}_t)e^{\hat{c}_t}, \tag{17}$$

where $\hat{a}_t$ and $\hat{c}_t$ are the estimated coefficients from the EIS regression (11), and are implicit functions of θ.

In the acceptance-rejection part of the MH-AR algorithm, the EIS sampling densities m_t are used to generate candidate trajectories $\tilde{Z}$ for $\Lambda|(Y, X, \theta)$ until acceptance with probability $\min\{f(Y, \tilde{Z}|X, \theta)/M(\tilde{Z}), 1\}$. Because $M(\Lambda)$ does not bound $f(Y, \Lambda|X, \theta)$, it follows that the target density is not adequately sampled here. However, this can be corrected with an additional Metropolis–Hastings step applied to the Λ–trajectories that come from the acceptance-rejection step. This means that, given the previously sampled trajectory $\tilde{\Lambda}^{(k)}$, the candidate trajectory from the acceptance-rejection step $\tilde{\Lambda}$ is accepted as the next trajectory $\tilde{\Lambda}^{(k+1)}$ with probability

$$\min\left\{\frac{f(Y, \tilde{\Lambda}|X, \theta)\min\left\{f(Y, \tilde{\Lambda}^{(k)}|X, \theta), M(\tilde{\Lambda}^{(k)})\right\}}{f(Y, \tilde{\Lambda}^{(k)}|X, \theta)\min\left\{f(Y, \tilde{\Lambda}|X, \theta), M(\tilde{\Lambda})\right\}}, 1\right\};$$

otherwise $\tilde{\Lambda}^{(k+1)}$ is set equal to $\tilde{\Lambda}^{(k)}$. After a sufficiently long 'burn-in', the draws $\{\tilde{\Lambda}^{(k)}\}$ represent a dependent sample from $f(\Lambda|Y, X, \theta)$. In the application below, the AR-MH step for Λ is repeated 10 times before the parameters are updated in the Gibbs sequence.

An alternative block-sampling procedure which could also be used for sampling Λ is the 'multi-move' sampler of Shephard and Pitt (1997). Using a Taylor expansion, this sampler is based on *local* approximations of the target density. In contrast, EIS provides corresponding *global* approximations, which is important insofar as global approximations typically lead to more efficient samplers than local ones.

To pursue a Bayesian analysis of the parameters θ, we need to specify corresponding prior densities. For δ, we assume a Beta distribution conformably with the stationarity condition $\delta \in (-1, 1)$. In particular, we employ for $(\delta + 1)/2$ a Beta prior with parameters $\delta^{(1)} > 1/2$ and $\delta^{(2)} > 1/2$. In our application we set $\delta^{(1)} = 20$ and $\delta^{(2)} = 1.5$, implying a prior mean of 0.86 and a prior standard deviation of 0.11. The resulting conditional posterior is non-conjugate. To sample from this posterior, we use an independent MH sampler based on a Gaussian proposal density (for details, see Kim et al., 1998). Furthermore, for ν^2 we assume an inverted chi-squared prior with $p_0 s_0/\chi^2_{(p_0)}$. Then the conditional posterior is also an inverted chi-squared distribution with $\nu^2|\Lambda, Y, X, \delta, \varphi \sim [\sum_{t=1}^{T}(\lambda_t - \delta\lambda_{t-1})^2 + p_0 s_0]/\chi^2_{(T+p_0)}$. In the application we set $p_0 = 10$ and $s_0 = 0.01$. Finally, we assume for φ a multivariate Normal prior with zero means and a covariance matrix given by $100 \cdot I$, where I is the identity matrix, reflecting a large prior uncertainty. To sample from the resulting non-conjugate conditional posterior, we follow Chib and Winkelmann (2001) and use a MH algorithm based on a Gaussian proposal density which is found by approximating the target density around its modal value (for details, see Winkelmann, 2003, p. 219).

2.4 Application

ML-EIS estimation results for the parameter driven SAM model (1) and (2) based on a simulation sample size $N = 50$ are given in Table 1. As in Davis et al. (1999) we include in x_t dummy variables for Mondays and Sundays, and Fourier series terms consisting of $\cos(2\pi kt/365)$ and $\sin(2\pi kt/365)$, for $k = 1, 2, 3, 4$, which capture seasonal cycles. The ML estimation requires approximately 31 BFGS iterations and take of the order of 5 minutes on a Pentium IV personal computer for a code written in GAUSS. The parameter estimates are numerically very accurate as indicated by the small numerical MC standard errors which are computed from 20 ML-EIS estimations conducted under different random numbers.

The estimates of δ and ν are given by 0.900 and 0.096 and are statistically significant at the 1 percent level. This indicates the presence of a highly persistent latent process generating overdispersion and positive serial correlation within the count process. This is confirmed by a likelihood ratio statistic testing the hypothesis $\delta = \nu = 0$, resulting in a value of 47.8. Furthermore, our estimates of the parameters measuring the Sunday and Monday effect and the impact of the seasonal patterns are very close to those obtained by Davis et al. (1999). The fitted values from the SAM model are shown in Figure 2 along with the actual counts.

The results of the Bayesian posterior analysis based on the MCMC-EIS sampling scheme are summarized in Table 2. These are obtained from 10,000 Gibbs iterations on the parameters where the first 1000 are discarded. The table shows the posterior means and standard deviations together with the corresponding MC standard errors. The MC standard errors are computed using a spectral estimator, as proposed by Shephard and Pitt (1997). In particular, for M draws of the parameters $\{\theta^{(k)}, k : 1 \to M\}$ the MC standard errors are the square root of the diagonal elements of

$$J_M = \frac{1}{M}\Big[\Gamma_0 + \frac{2M}{M-1}\sum_{\ell=1}^{L_M} K\Big(\frac{\ell}{L_M}\Big)\Gamma_\ell\Big], \quad \text{where} \quad \Gamma_\ell = \frac{1}{M}\sum_{k=\ell+1}^{M}(\theta^{(k)} - \bar{\theta})(\theta^{(k-\ell)} - \bar{\theta})', \qquad (18)$$

L_M is the bandwidth, and $K(\cdot)$ represents the Parzen kernel.

The small values of these MC standard errors shown in Table 2 indicate that the MCMC-EIS procedure works reasonably efficient for the SAM model. This is confirmed by the autocorrelation functions of the Gibbs draws of the parameters (not presented here). They indicate that there is no significant autocorrelation at lags larger than 150 for the critical parameters δ and ν, and no correlation at lags larger than 20 for the remaining parameters. Finally, note that the MCMC-EIS estimates of all parameters are very close to ML-EIS estimates. Since we assumed, especially for φ, very uninformative priors, the quasi-identical estimation results indicate that the likelihood is very informative about the week effects and the seasonal cycles.

Table 3 presents the results of the diagnostic tests based on the standardized Pearson residuals

z_t and the normalized residuals z_t^* from the ML-EIS estimation. The Ljung-Box statistic $\text{LB}_{30}(\cdot)$ for the residuals and the squared residuals including 30 lags, indicates that the SAM models successfully accounts for the observed dynamics in the first and second-order moments of the counts. Furthermore, the Jarque-Bera statistic $\text{JB}(\cdot)$ for z_t^* has a marginal significance level of 8.2 percent indicating that normality cannot be rejected at the 5-percent level. The time series and quantile-quantile plot of z_t^* is displayed in Figure 2. The results suggest that even if the model performs quite well, it seems to have slight problems to approximate the distribution of the counts near the origin as well as in the right tail.

3 Autoregressive conditional Poisson model

3.1 Specification and Estimation

Although MC-techniques like EIS and MCMC are routinely used nowadays, the efficient estimation of the parameter-driven SAM model and the implementation of diagnostic tests requires some computational effort. A simple alternative to the SAM model is Heinen's (2003) observation-driven ACP model. Like all observation-driven models, the ACP specification is designed to allow the likelihood to be evaluated easily.

Let Y_{t-1} denote the information available on the series of counts y_t up to and including time $t-1$. In the simplest model without any covariates, the counts are assumed to be Poisson

$$y_t | Y_{t-1} \sim \text{Po}(\mu_t) \ , \tag{19}$$

with an autoregressive conditional mean or intensity μ_t

$$\text{E}(y_t | Y_{t-1}) \equiv \mu_t = \omega + \sum_{j=1}^{p} \alpha_j y_{t-j} + \sum_{j=1}^{q} \beta_j \, \mu_{t-j} \tag{20}$$

and positive α_j's, β_j's and ω, ensuring the non-negativity of μ_t. The ACP model (19) and (20) is similar in spirit to the autoregressive conditional duration (ACD) model of Engle and Russell (1998) or the generalized autoregressive conditional heteroskedasticity (GARCH) model of Bollerslev (1986). In particular, in all these specifications, the autoregressive structure is introduced by an observable recursion on lagged endogenous variables. Heinen (2003) shows, that as long as the sum of the autoregressive coefficients is less than one, the ACP specification (19) and (20) is stationary and the expression for the unconditional mean of the counts is identical to the mean of a Gaussian ARMA process.

In the following we focus on the most commonly used ACP(1,1) model with conditional mean equation equal to

$$\mu_t = \omega + \alpha \, y_{t-1} + \beta \, \mu_{t-1} \ . \tag{21}$$

It can be shown (see Heinen, 2003) that unconditionally the variance of the ACP(1,1) model is always greater than its mean as long as $\alpha \neq 0$. The ACP(1,1) is therefore able to cope with extra binomial variation in an observed series of counts. The autocorrelation properties of the ACP(1,1) model are summarized by its autocorrelation function which is given by

$$corr(y_t, y_{t-s}) = (\alpha+\beta)^{s-1} \frac{\alpha[1-\beta(\alpha+\beta)]}{1-(\alpha+\beta)^2+\alpha^2}, \qquad s = 1, 2, 3, \ldots . \tag{22}$$

Further stochastic properties of the ACP model as well as some generalizations are provided in Heinen (2003). Note, that under the non-negativity and stationarity conditions for the ACP model only positive serial correlation is possible. Hence, in contrast to the SAM specification, a stationary ACP model does not satisfy the desideratum for dynamic count data models to allow for positive as well as negative serial correlation. However, in most applications of count data models, including that to the asthma data considered here, the case of negative serial correlation is irrelevant. Especially, the sample autocorrelation function of the asthma data depicted in Figure 1 clearly shows, that in our application this case can be be ignored. Moreover, the restriction to positive autocorrelations is a property that the ACP has in common with (G)ARCH-type models which are very successfully applied to accommodate the time varying volatility found in many economic time series.

An alternative observation-driven model for stationary count processes which would allow for positive and negative serial correlation is, e.g., that proposed in Davis et al. (1999, 2003), where the log of the Poisson intensity is assumed to be a linear function of lagged standardized counts. In particular, they use instead of the specification (20) a GLARMA specification based on the following recursion for the log intensity:

$$\ln \mu_t \equiv w_t = \omega + \sum_{j=1}^{p} \alpha_j (w_{t-j} + \xi_{t-j}) + \sum_{j=1}^{q} \beta_j \, \xi_{t-j} \tag{23}$$

where $\xi_t = (y_t - \mu_t)/\mu_t^{\rho}$, $\rho \in (0, 1]$. However, note that the stochastic properties for such a log-linear alternative are typically more difficult to analyze than those for the ACP model.

Using an exponential link, the ACP model can easily be extended to include covariates

$$\mathrm{E}(y_t|Y_{t-1}, x_t) \equiv \mu_t^* = \mu_t \cdot \exp(x_t'\varphi) \ , \tag{24}$$

where $\{x_t, t : 1 \to T\}$ is a sequence of k-dimensional vectors of covariates (without a constant regressor) and φ an appropriate parameter vector.

Estimation of the parameter $\theta = (\varphi', \alpha, \beta, \omega)'$ of the ACP(1,1) model including covariates is carried out by maximizing the log-likelihood function using numerical techniques routinely available in standard software packages. The contribution of the t-th observation to the log-likelihood is given by

$$l_t(\theta) = y_t \ln(\mu_t \cdot \exp(x_t'\varphi)) - \mu_t \cdot \exp(x_t'\varphi) - \ln(y_t!) \ , \tag{25}$$

where the initial values in the recursion (21) μ_0 and y_0 are set equal to the sample mean of the counts. Heinen (2003) shows that the resulting ML estimators are consistent irrespectively of the correctness of the distributional assumption employed in (19). Note that like the SAM model, the ACP specification could also be estimated by a Bayesian posterior analysis using, e.g., the Gibbs sampling scheme proposed by Bauwens and Lubrano (1998) for GARCH-type models.

For diagnostic checking of the assumed dynamic structure in the mean and variance one can use the standardized (Pearson) residuals

$$z_t = \frac{y_t - \mu_t^*}{\sqrt{\mu_t^*}}, \tag{26}$$

where the square root of μ_t^* represents the conditional standard deviation of y_t under the ACP model. If the model is correctly specified, these residuals should have mean zero and variance one and no significant serial correlation in the first and second-order moments. As for the SAM model, one can compute the normalized residuals $z_t^* = F_N^{-1}(\tilde{u}_t)$, where $\tilde{u}_t$ is given by Equations (12) and (13), which should be $iid\mathrm{N}(0, 1)$ under the correct specification.

3.2 Application

ML estimation of the observation driven ACP(1,1) model including the same covariates as the SAM model was carried out in GAUSS with the results displayed in Table 4. The parameters α and β are estimated to be 0.058 and 0.811 respectively. Both parameters are significantly different from zero at the 1-percent level and positive, indicating the presence of a positive correlation structure in the data. This result, which is consistent with the persistence found under the SAM model, is confirmed by a likelihood ratio statistic testing the joint hypothesis $\alpha = \beta = 0$. The observed value of the test statistic is found to be 42.2 and significant at the 1-percent level. Our estimates of the Sunday and Monday dummies as well as the trigonometric regressors are in very close accordance to those obtained under the SAM model and to those obtained by Davis et al. (1999, Table 4) based on a GLARMA specification with a recursion of the form (23). Finally, we note that the fitted log-likelihood is slightly better under the SAM than under the ACP model, while the log-likelihood value of -2444.9 obtained for the preferred GLARMA specification of Davis et al. (1999) with lags 1,3,7,10 for the AR components and no MA component is slightly lower than that for the ACP model. (For further comparisons of the fit of the ACP and SAM model one could use the formal tools for comparing non-nested models proposed by Kim et al., 1998. In particular, they applied Bayes factors and (simulated) likelihood-ratio tests for non-nested models to compare GARCH and stochastic volatility models.)

The quantile-quantile plot of the normalized residuals z_t^* displayed in Figure 3 indicates that the ACP model seems to have the same problems as the SAM specification with the approximation of the distribution of the counts near the origin as well as in the right tail. Diagnostic checks for the ACP residuals are presented in Table 5. The Jarque-Bera test rejects normality at the 5-percent level, while the Ljung-Box test for the normalized and standardized Pearson residuals shows that the ACP model fits the dynamics of data in the first and second-order moments. Finally, it should be mentioned that the diagnostic checks of the residuals from the GLARMA specification of Davis et al. (1999) (not presented here) show nearly the same results as for the ACP and SAM model.

Taken all together, the empirical results for the asthma counts suggest that the ACP as well as the SAM specification fit the dynamic behavior in the first and second-order moments very well and the distributional properties reasonably well. However, both models seem to have slight problems in the tails. In order to further improve the approximation of the distributional characteristics one could substitute the Poisson distribution by a more flexible count data distribution. In the following section, we consider a flexible alternative based on an ordered probit approach.

4 Dynamic ordered probit models for count data

4.1 Specification and Estimation

Analyzing count data, the Poisson distribution, which treats the data as generated by an underlying point process, is usually the first choice. As an alternative Cameron and Trivedi (1998, section 3.6) suggest inter alia the use of ordered probit models. Instead of modelling the discrete data by an underlying point process, they can be interpreted as being the result of a continuous latent process. In our application the latent variable could be interpreted as intensity of asthma inducing influences that on crossing a threshold leads to an increase of one in the number of observed asthma incidences. This implies that the unobservable intensity level y_t^* in combination with thresholds determine the observable discrete count categories y_t according to

$$y_t = \begin{cases} 0, & \text{if} \quad -\infty < y_t^* \leq \gamma_1 \\ 1, & \text{if} \quad \gamma_1 < y_t^* \leq \gamma_2 \\ & \vdots \\ K, & \text{if} \quad \gamma_K < y_t^* < \infty \end{cases}, \tag{27}$$

where the thresholds γ_j are unknown parameters to be estimated. In the application below we use $K + 1 = 8$ categories, where the last category $y_t = K$ summarizes K and more asthma incidences.

Considering the modelling of y_t without any covariates, the ordered probit approach represents the *saturated* model. The count categories are fully described by the $K + 1$ occurrence probabilities $p_j = P(y_t = j) = \Phi(\gamma_{j+1}) - \Phi(\gamma_j)$, $j : 0 \to K$, where Φ denotes the cdf of the standard normal distribution. The ML-estimates of the occurrence probabilities p_j are the corresponding relative frequencies $\hat{p}_j$, which can be used to obtain the ML-estimates of the thresholds according to $\hat{\gamma}_j = \Phi^{-1}(\sum_{h=0}^{j-1} \hat{p}_h)$. (Any other distribution function could be used replacing Φ in the corresponding expressions leading to different estimated thresholds only.) Hence, the estimates for the threshold parameters are those values of the γ_js that equate predicted and actual probabilities. Accordingly, compared to conventional count data models, the ordered probit approach provides much more flexibility with respect to the adaption of the distributional properties of the data – a flexibility which is akin to that of non-parametric approaches and which allows to capture possible over- and underdispersion in the count data. (As noted by Cameron and Trivedi, 1998, an additional advantage of the ordered probit approach is that it is applicable to count data with negative count like in the application of Hausman et al., 1992 who modeled discrete changes of stock prices.) On the other hand note that the ordered probit model for counts completely ignores the feature of count data of being cardinal which might lead to an efficiency loss relative to an analysis based on standard count data models taking this property into account.

Using a (linear) regression function for the latent process y_t^*, the saturated ordered probit can easily be extended to include covariates x_t and to accommodate for positive as well as for negative serial correlation. In particular, we consider the following specification:

$$y_t^* = \mu_t + x_t'\varphi + e_t, \qquad e_t \sim iid\mathrm{N}(0,1), \tag{28}$$

with

$$\mu_t = \alpha y_{t-1} + \beta \mu_{t-1}, \tag{29}$$

where φ is a vector of regression parameters without an intercept. The intercept term in the specification for μ_t is also set equal to zero, which is imposed for identifiability reasons. (As discussed in Kukuk, 1994, the ordered probit probabilities are invariant with respect to monotone transformations of (28) and of the thresholds γ_j.) According to this autoregressive conditional ordered probit (ACOP) model the dynamics of the count data variables, which cannot be attributed to the covariates, are captured by the observed recursion (29). Hence, the structure of the ACOP is similar to that of the ACP model discussed in section 3. Furthermore, note that in contrast to the ACP and SAM model, the dynamic and distributional features of the ACOP specification are not directly linked to each other. This means that the parameters of the ACOP model α and β, which govern the dynamic behavior, do not have a direct impact on the properties of the unconditional distribution of the count categories. In contrast, the corresponding parameters in the SAM (δ) and the ACP model (α, β) directly enter the Poisson parameter, which determines the distributional behavior of the counts and which generates a close link between the dynamics and the distribution. Hence, one can expect that the ACOP model accommodates more easily to the dynamic and distributional behavior of the data.

Estimation of the observation-driven ACOP model (27) – (29) can be performed by maximizing the log-likelihood function using standard numerical techniques. Defining a set of dummy variables $y_{tj} = 1$ if $y_t = j$ for $j = 0, 1, 2 \ldots, K$ (the last category summarizes K and more incidences), the contribution of the t-th observation is given by

$$l_t(\theta) = \sum_{j=0}^{K} y_{tj} \cdot \ln \left[\Phi(\gamma_{j+1} - \mu_t - x_t'\varphi) - \Phi(\gamma_j - \mu_t - x_t'\varphi) \right],$$

with $\gamma_0 = -\infty$ and $\gamma_{K+1} = \infty$ and $\theta = (\varphi', \alpha, \beta, \gamma_1, ..., \gamma_K)'$. The initial values in the recursion (29) μ_0 and y_0 are set equal to zero and the sample mean, respectively.

For diagnostic tests of the ACOP model one can use the standardized Pearson residuals $z_t = [y_t - \mathrm{E}(y_t|Y_{t-1}, x_t)]/\mathrm{var}(y_t|Y_{t-1}, x_t)^{1/2}$, where the conditional mean under the ACOP model is

$$\mathrm{E}(y_t|Y_{t-1}, x_t) = \sum_{j=0}^{K} j \cdot \left[\Phi(\gamma_{j+1} - \mu_t - x_t'\varphi) - \Phi(\gamma_j - \mu_t - x_t'\varphi) \right],$$

and the conditional variance is obtained analogously. If the model is correctly specified, z_t has mean zero and unit variance and is serially uncorrelated in the first- and second-order moments.

An alternative which can be expected to be a more powerful diagnostic test of the ACOP, can be constructed using the residuals of the following state vector:

$$s_t = (s_{t1}, ..., s_{tK})' = \begin{cases} (0, 0, \ldots, 0)', & \text{if} \quad y_t = 0 \\ (1, 0, \ldots, 0)', & \text{if} \quad y_t = 1 \\ \quad \vdots & \\ (0, 0, \ldots, 1)', & \text{if} \quad y_t = K \end{cases},$$

with conditional probabilities

$$\pi_{tj} = P(s_{tj} = 1|Y_{t-1}, x_t) = \Phi(\gamma_{j+1} - \mu_t - x_t'\varphi) - \Phi(\gamma_j - \mu_t - x_t'\varphi).$$

The conditional expectation of the state vector s_t is given by $\mathrm{E}(s_t|Y_{t-1}, x_t) = \pi_t = (\pi_{t1}, ..., \pi_{tK})'$ and the conditional variance covariance matrix by $\mathrm{var}(s_t|Y_{t-1}, x_t) = \mathrm{diag}(\pi_t) - \pi_t\pi_t'$. Then, the corresponding standardized residuals are obtained as

$$v_t = \mathrm{var}(s_t|Y_{t-1}, x_t)^{-1/2}[s_t - \mathrm{E}(s_t|Y_{t-1}, x_t)], \tag{30}$$

where $\mathrm{var}(s_t|Y_{t-1}, x_t)^{-1/2}$ denotes the inverse of the Cholesky factor of the conditional variance covariance matrix. Under a correct specification, these residuals should be serially uncorrelated in the first and second-order moments with the following unconditional moments: $\mathrm{E}(v_t) = 0$ and $\mathrm{var}(v_t) = I$. The joint hypothesis that there is no serial correlation in v_t can be tested by the multivariate version of the Portmanteau statistic proposed by Hosking (1980)

$$Q(L) = T \sum_{\ell=1}^{L} \mathrm{tr}\big[\Gamma_v(\ell)'\Gamma_v(0)^{-1}\Gamma_v(\ell)\Gamma_v(0)^{-1}\big], \tag{31}$$

where $\Gamma_v(\ell) = \sum_{t=\ell+1}^{T} v_t v'_{t-\ell}/(T-\ell-1)$. Under the null hypothesis, $Q(L)$ is asymptotically χ^2-distributed with degrees of freedom equal to the difference between $K^2 \cdot L$ and the number of parameters to be estimated.

4.2 Application

ML-estimation of the dynamic ACOP model including the same covariates as the SAM and ACP model was carried out in GAUSS. The results are given in Table 6. The estimated boundary partitions $\{\gamma_j\}$ are almost equally spaced. Furthermore, the estimates of the parameters α and β in the recursion (29) are given by 0.055 and 0.761 and are statistically significant at any conventional significance level. The likelihood ratio statistic for the hypotheses $\alpha = \beta = 0$ results in a value of 33.78. This indicates a strong positive serial correlation in the latent process y_t^* which is not captured by the included covariates and is consistent with the results for the SAM and ACP model. The estimated impacts of the weekday dummies and the seasonal components on y_t^* are also in close accordance to those obtained under the SAM and ACP specification. However, note that under the ACOP model the asymptotic standard errors for the parameter estimates associated with the covariates are uniformly larger than the ML standard errors for the SAM model (see, Table 1) and those for the ACP model (see, Table 4). These larger standard errors can be interpreted as the result of the efficiency loss due to the fact that the ordinal ACOP model ignores the cardinal meaning of the count data. On the other hand, a comparison of the values for the Schwarz information criterion, which are given by 4957.2 (ACOP), 4981.8 (ACP), and 4979.7 (SAM), indicates that the ACOP provides, as expected, a much better fit to the data than the pure count data models.

Similar to the SAM and ACP specifications, the diagnostic tests based on the standardized Pearson residuals z_t do not indicate any deficiency of the ACOP model: The sample mean and variance of the z_t's are -0.0015 and 0.9683, and the Ljung-Box Statistics for z_t and z_t^2 including 30 lags are 19.35 and 34.79 with p-values of 0.932 and 0.250. Hence, the ACOP model explains the serial correlation of the counts in the first and second-order moments. The sample mean $\hat{\mu}_v$ and the covariance matrix $\hat{\Sigma}_v$ of the standardized state residuals v_t as defined in Equation (30) are given by

$$\hat{\mu}_v = (0.009, 0.008, -0.001, -0.010, -0.030, -0.034, -0.028)'$$

and

$$\hat{\Sigma}_v = \begin{pmatrix} 1.00 & -0.00 & -0.00 & -0.00 & 0.01 & 0.02 & 0.01 \\ & 1.00 & 0.00 & 0.00 & 0.02 & 0.03 & 0.02 \\ & & 1.00 & 0.00 & 0.03 & 0.03 & 0.03 \\ & & & 1.02 & 0.03 & 0.03 & 0.03 \\ & & & & 0.86 & 0.03 & 0.03 \\ & & & & & 0.76 & 0.02 \\ & & & & & & 0.68 \end{pmatrix}.$$

The fact that $\hat{\mu}_v$ is very close to zero suggests that the stochastic behavior of the state vector s_t in the first-order moment is well explained by the ACOP model. This implies that the empirical frequencies of the count categories are closely approximated by the theoretical probabilities under the ACOP model which shows that it fits the distributional properties very well. The deviations of $\hat{\Sigma}_v$ from the unity matrix, especially for the three largest count categories, reveals that, the ACOP model has slight problems to account for the joint variation of the binary state variables in s_t. However, note that this joint variation is related to aspects of the stochastic behavior which are typically not particularly relevant for the modelling of count data.

The multivariate Portmanteau statistic $Q(L)$ for the vector of standardized state residuals including 10 lags is 605.3 with a p-value smaller than 0.0001 and the corresponding statistic for the observed state vector s_t is given by 1326.0. This indicates that the ACOP explains some of the observed serial dependence in the data but not all. A further inspection of the Ljung-Box statistics for the individual elements of the vector of the standardized state residuals v_{tj} (not presented here) revealed that there is nearly no significant serial correlation in the v_{tj} and v_{tj}^2. Hence, the serial dependence in v_t detected by the multivariate Portmanteau statistic seems to be related to a highly non-linear serial dependence in the count-categories, which is not captured by the ACOP model.

5 Conclusions

This paper compares stochastic models for time series of counts and presents appropriate procedures to estimate these models and to perform diagnostic tests. In particular, we consider Zeger's (1988) parameter-driven Poisson model with a stochastic autoregressive mean (SAM) and the observation-driven conditional autoregressive poisson (ACP) model introduced by Heinen (2003). While the ACP model, like all observation-driven models, is designed to allow the likelihood to be evaluated easily, the likelihood evaluation for the SAM model requires high-dimensional integration. To address this integration problem, we propose to use the efficient importance sampling (EIS) procedure of Richard and Zhang (2004). EIS can be used, to estimate the SAM by ML as well as to carry out a Bayesian MCMC posterior analysis of the parameters via Gibbs sampling. As a flexible alternative to the ACP and SAM model, which are based on a conditional Poisson distribution, we consider a dynamic ordered probit model as a specification to capture the salient features of time series of counts. In particular, we propose an autoregressive conditional ordered probit (ACOP) with an observable autoregressive conditional mean of the underlying latent process.

The models, the corresponding estimation procedures and the diagnostics are illustrated with data on daily admissions for asthma to a single hospital in the Sydney metropolitan area from 1 January 1990 to 31 December 1993, with a sample of 1461 observations. All considered models include explanatory variables for a Sunday effect, a Monday effect and a seasonal pattern. The empirical results reveal that the estimated impact of the explanatory variables under the SAM model is very close to those under the ACP model. Furthermore, both specifications can account for the serial correlation in the first and second-order moments of the count data and provide, except for slight problems in tails, a reasonable approximation of the distributional properties. Finally, the ACOP closely approximates the empirical distribution and the dynamics in the mean and variance of the count data and confirms the results about the impact of the weekday dummies and seasonal components on the admission for asthma obtained under the ACP and SAM specification.

Appendix

Implementation of the EIS-algorithm

This appendix details the EIS implementation for the observation-driven SAM model (1) and (2). Using a parametric extension of the initial samplers p_t given by Equation (6), the corresponding density kernel of the Gaussian importance samplers m_t can be written as

$$k_t(\lambda_t, \lambda_{t-1}, a_t) \propto p_t(\lambda_t|\lambda_{t-1}, \theta) \exp\left\{-\frac{1}{2}\left[\alpha_t \lambda_t^2 - 2\beta_t \lambda_t\right]\right\}, \tag{A.1}$$

with $a_t = (\alpha_t, \beta_t)'$. The associated integrating constant is

$$\chi_t(\lambda_{t-1}, a_t) \propto \exp\left\{-\frac{1}{2}\left[\frac{\delta\lambda_{t-1}}{\nu^2} - \frac{\kappa_t^2}{\sigma_t^2}\right]\right\}, \tag{A.2}$$

where κ_t and σ_t^2 are the mean and variance of the importance sampler m_t, which are given by $\sigma_t^2 = \nu^2/(1 + \nu^2\alpha_t)$ and $\kappa_t = \sigma_t^2(\beta_t + \delta\lambda_{t-1}/\nu^2)$. Note, that under this parametrization of m_t the initial sampler p_t cancels out in the EIS-regressions (11). In particular, the EIS-regression for period t is a linear least-squares problem with a dependent variable given by

$$-\exp\{x_t'\beta + \tilde{\lambda}_t^{(i)}\} + y_t\tilde{\lambda}_t^{(i)} + \ln \chi_{t+1}(\tilde{\lambda}_t^{(i)}, \hat{a}_{t+1}),$$

and the regressors: intercept, $\tilde{\lambda}_t^{(i)}$, $[\tilde{\lambda}_t^{(i)}]^2$.

Based on these functional forms, the EIS-MC evaluation of the likelihood requires the following simple steps:

Step (1) Generate N independent trajectories from the p_t densities.

Step (2) Use these trajectories to run for each $t : T \to 1$ the EIS regression.

Step (3) Use the estimated regression coefficients to obtain the means κ_t and variances σ_t^2 of the Gaussian EIS samplers m_t.

Step (4) Generate N independent trajectories from the m_t densities which are used to evaluate the likelihood according to (9).

In order to achieve maximally efficient EIS-samplers only a small number of iterations on the EIS-algorithm is required, where the initial sampling densities p_t are replaced by the previous stage importance sampler.

Acknowledgement

The authors would like to thank William Dunsmuir for supplying us with the data on asthma hospitalisation.

References

Bauwens, L., Lubrano, M. (1998). Bayesian inference on GARCH models using the Gibbs sampler. *Econometrics Journal* **1** C23-46.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31** 307-327.

Cameron, A.C., Trivedi, P.K. (1998). *Regression Analysis of Count Data.* Cambridge University Press, Cambridge.

Chan, K.S., Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association* **90** 242-251.

Chib, S., Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician* **49** 327-335.

Chib, S., Winkelmann, R. (2001). Markov-Chain Monte Carlo analysis of correlated count data. *Journal of Business and Economic Statistics* **19** 428-435.

Danielsson, J., Richard, J.F. (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics* **8** 153-173.

Davis, R.A.; Dunsmuir, W.T.M. and Wang, Y. (1999). Modelling time series of counts. In: Ghosh, S. (Ed.) *Asymptotics, nonparametrics, and time series: A tribute to Madan Lal Puri*, 63-113. Marcel Dekker, New York.

Davis, R.A.; Dunsmuir, W.T.M. and Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika* **87** 491-505.

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

Engle, R. F., Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* **66** 1127-1162.

Hausman, J.A., Lo, A.W. and MacKinlay, A.C. (1992). An ordered probit analysis of transaction stock prices. *Journal of Financial Economics* **31** 319-379.

Heinen, A. (2003). Modelling time series count data: An autoregressive conditional Poisson model. Core discussion paper No. 2003-63.

Hosking, J.R.M. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association* **75** 602-608.

Jung, R.C., Liesenfeld, R. (2001). Estimating time series models for count data using efficient importance sampling. *Allgemeines Statistisches Archiv* **85** 387-407.

Jung, R.C., Tremayne, A.R. (2005). Estimation and testing in time series models with integer support. In: Gregori, D. et al. (eds.) *Correlated Data Modeling, Proceedings Turin 2004*, 51-71. Franco Angeli, Milano.

Kedem, B., Fokianos, K. (2002) *Regression Models for Time Series Analysis.* John Wiley & Sons, Hoboken, NJ.

Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* **65** 361-393.

Kuk, A.Y.C., Chen, Y.W. (1997). The Monte Carlo Newton-Raphson algorithm. *Journal of Statistical and Computational Simulation* **59** 233-250.

Kukuk, M. (1994). Distributional aspects in latent variable models. *Statistical Papers* **35** 231 - 242.

Kukuk, M. (2002). Indirect estimation of (latent) linear models with ordinal regressors. A Monte Carlo study and some empirical illustrations. *Statistical Papers* **43** 379 - 399.

Liesenfeld, R., Nolte, I. and Pohlmeier, W. (2005). Modelling financial transaction price movements: A dynamic integer count data model. Working paper, Universität Konstanz.

Liesenfeld, R., Richard, J.F. (2003). Univariate and multivariate stochastic volatility models: Estimation and diagnostics. *Journal of Empirical Finance* **10** 505-531.

McKenzie, E. (2003). Discrete variate time series. In: Shanbhag, D.N. and Rao, C.R. (Eds.) *Handbook of Statistics*, Volume 21, 573-606. Elsevier, Amsterdam.

Richard, J.F., Zhang, W. (2004). Efficient high-dimensional Monte Carlo importance sampling. Working Paper, University of Pittsburgh.

Rosenblatt, M.(1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics* **23** 470-472.

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

Shephard, N. (1995). Generalized linear autorgressions. Working Paper, Nuffield College, Oxford.

Shephard, N., Pitt, M.K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* **84** 653-667.

Tierney, L. (1994). Markov Chain for exploring posterior distributions. *The Annals of Statistics* **21** 1701-1762.

Winkelmann, R. (2003). *Econometric Analysis of Count Data* (4th edition). Springer, New York.

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

Table 1. ML-EIS Parameter Estimates of the SAM model

Parameters	Estimates	Asy. s.e.	MC s.e.
δ	0.9003	0.0384	0.00055
ν	0.0961	0.0243	0.00043
Intercept	0.5175	0.0324	0.00008
Sunday effect	0.2287	0.0451	0.00002
Monday effect	0.2323	0.0469	0.00002
$\cos(2\pi t/365)$	-0.1629	0.0277	0.00001
$\sin(2\pi t/365)$	0.3583	0.0334	0.00002
$\cos(4\pi t/365)$	-0.0654	0.0267	0.00002
$\sin(4\pi t/365)$	0.0170	0.0229	0.00002
$\cos(6\pi t/365)$	-0.0761	0.0204	0.00002
$\sin(6\pi t/365)$	0.0062	0.0240	0.00001
$\cos(8\pi t/365)$	-0.1419	0.0226	0.00004
$\sin(8\pi t/365)$	-0.0524	0.0267	0.00002
Log–Likelihood	-2442.49		0.0446

The ML-EIS estimates are based on a MC-sample size of $N = 50$ and three EIS iterations. Asymptotic (statistical) standard errors are obtained from a numerical approximation to the Hessian and MC (numerical) standard errors were computed from 20 ML-EIS estimations conducted under different sets of CRNs.

Table 2. MCMC-EIS Posterior Analysis of the SAM model

Parameters	Mean	Std. Deviation	MC s.e.
δ	0.9110	0.0279	0.0024
ν	0.0995	0.0153	0.0018
Intercept	0.5141	0.0402	0.0008
Sunday effect	0.2279	0.0513	0.0004
Monday effect	0.2313	0.0510	0.0004
$\cos(2\pi t/365)$	-0.1631	0.0518	0.0016
$\sin(2\pi t/365)$	0.3602	0.0526	0.0013
$\cos(4\pi t/365)$	-0.0639	0.0506	0.0014
$\sin(4\pi t/365)$	0.0159	0.0491	0.0011
$\cos(6\pi t/365)$	-0.0755	0.0464	0.0012
$\sin(6\pi t/365)$	0.0046	0.0471	0.0009
$\cos(8\pi t/365)$	-0.1398	0.0430	0.0008
$\sin(8\pi t/365)$	-0.0513	0.0434	0.0009

The Posterior means and posterior standard deviations are obtained from 10,000 Gibbs iterations (discarding the first 1000 draws). The MC standard errors are computed using a Parzen window with bandwidth of $L_M = 1000$. The EIS proposal densities for the AR-MH sampler are obtained from EIS-regressions based on a MC sample size $N = 30$ and four EIS iterations.

Table 3. Diagnostics for the SAM model based on the ML-EIS estimates

$LB_{30}(z_t)$	$LB_{30}(z_t^2)$	Skewness (z_t^*)	Kurtosis (z_t^*)	$JB(z_t^*)$	$LB_{30}(z_t^*)$	$LB_{30}({z_t^*}^2)$
20.15	29.40	0.015	3.285	4.991	26.91	27.32
(0.912)	(0.497)			(0.082)	(0.627)	(0.606)

p-values are given in parentheses. The values of the LB_{30} statistic for the counts y_t and the squared counts y_t^2 are 1056.6 and 762.8.

Table 4. ML Parameter Estimates of the ACP model

Parameters	Estimates	Asy. s.e.
α	0.0576	0.0141
β	0.8110	0.0566
ω	0.2132	0.0799
Sunday effect	0.2464	0.0516
Monday effect	0.2364	0.0512
$\cos(2\pi t/365)$	-0.1031	0.0322
$\sin(2\pi t/365)$	0.2451	0.0367
$\cos(4\pi t/365)$	-0.0458	0.0293
$\sin(4\pi t/365)$	0.0211	0.0268
$\cos(6\pi t/365)$	-0.0567	0.0282
$\sin(6\pi t/365)$	0.0119	0.0279
$\cos(8\pi t/365)$	-0.1177	0.0282
$\sin(8\pi t/365)$	-0.0146	0.0290
Log–Likelihood	-2443.16	

Table 5. Diagnostics for the ACP model

$LB_{30}(z_t)$	$LB_{30}(z_t^2)$	Skewness (z_t^*)	Kurtosis (z_t^*)	$JB(z_t^*)$	$LB_{30}(z_t^*)$	$LB_{30}({z_t^*}^2)$
20.12	28.64	0.100	3.276	7.119	18.29	27.78
(0.913)	(0.537)			(0.028)	(0.954)	(0.582)

p-values are given in parentheses. The values of the LB_{30} statistic for the counts y_t and the squared counts y_t^2 are 1056.6 and 762.8.

Table 6. ML estimates of the Ordered Probit (ACOP) model

Parameters	Estimates	Asy. s.e.	Parameters	Estimates	Asy. s.e.
α	0.0550	0.0119	γ_1	-0.4969	0.1009
β	0.7608	0.0597	γ_2	0.4182	0.0996
Sunday effect	0.3010	0.0791	γ_3	1.1473	0.1016
Monday effect	0.3262	0.0803	γ_4	1.7047	0.1056
$\cos(2\pi t/365)$	-0.1399	0.0505	γ_5	2.1763	0.1127
$\sin(2\pi t/365)$	0.3337	0.0512	γ_6	2.5497	0.1224
$\cos(4\pi t/365)$	-0.0635	0.0410	γ_7	2.9997	0.1428
$\sin(4\pi t/365)$	0.0289	0.0957			
$\cos(6\pi t/365)$	-0.0716	0.0383			
$\sin(6\pi t/365)$	0.0104	0.0934			
$\cos(8\pi t/365)$	-0.1440	0.0414			
$\sin(8\pi t/365)$	-0.0253	0.0405			
Log–Likelihood	-2409.38				

Twelve values of y_t greater than 6 are censored to 7.

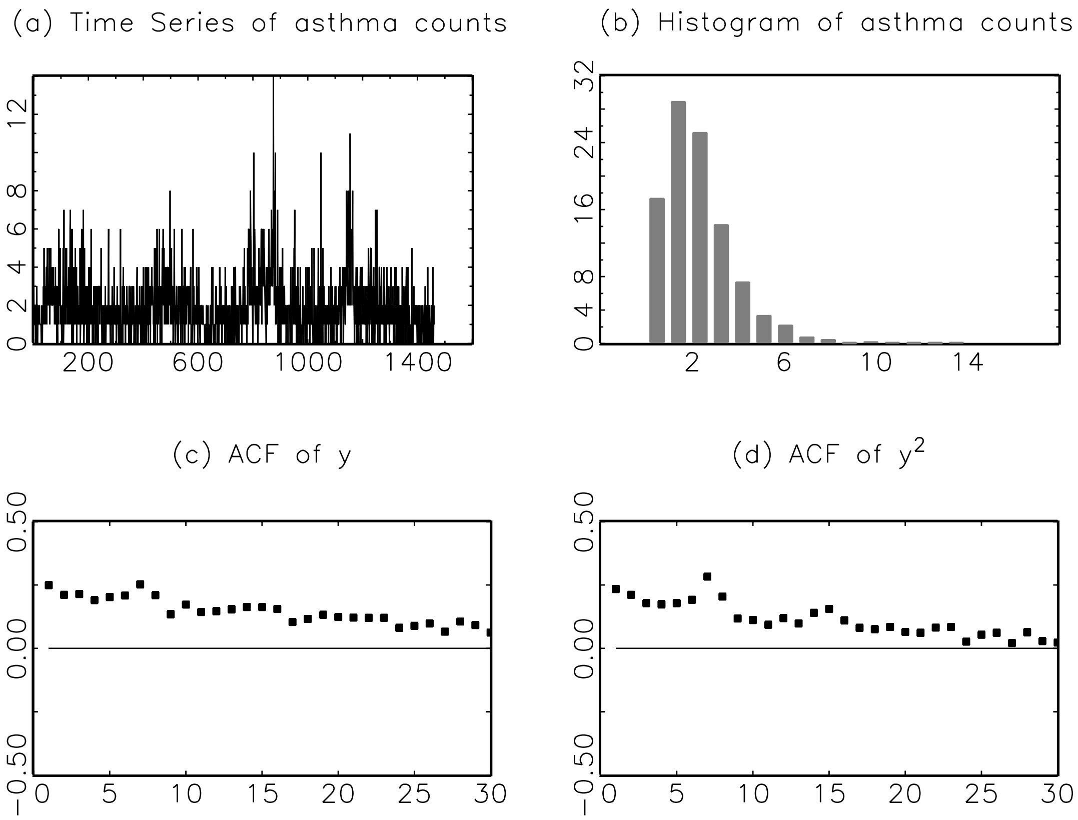


Figure 1. Panel (a): Time series plot of the asthma counts; Panel (b): Histogram of the asthma counts; Panel (c): Sample autocorrelation function of the asthma counts; Panel (d): Sample autocorrelation function of the squared asthma counts.

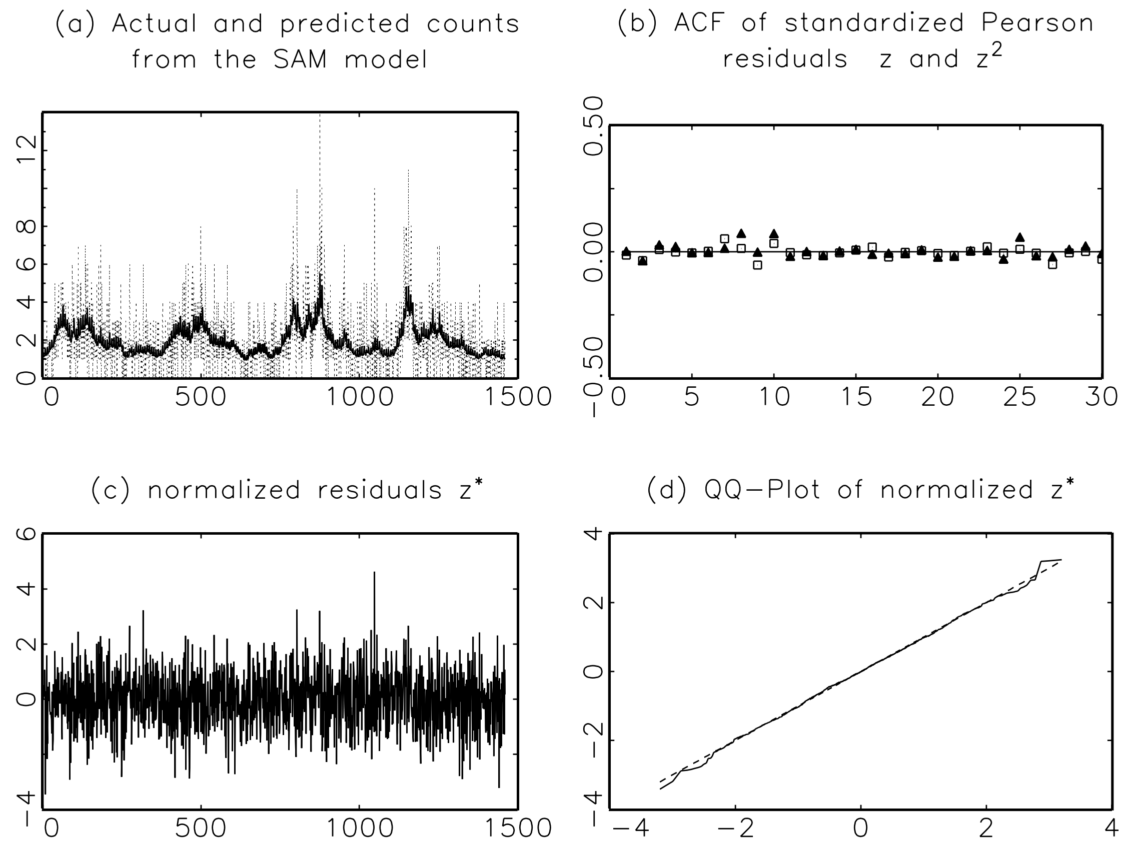
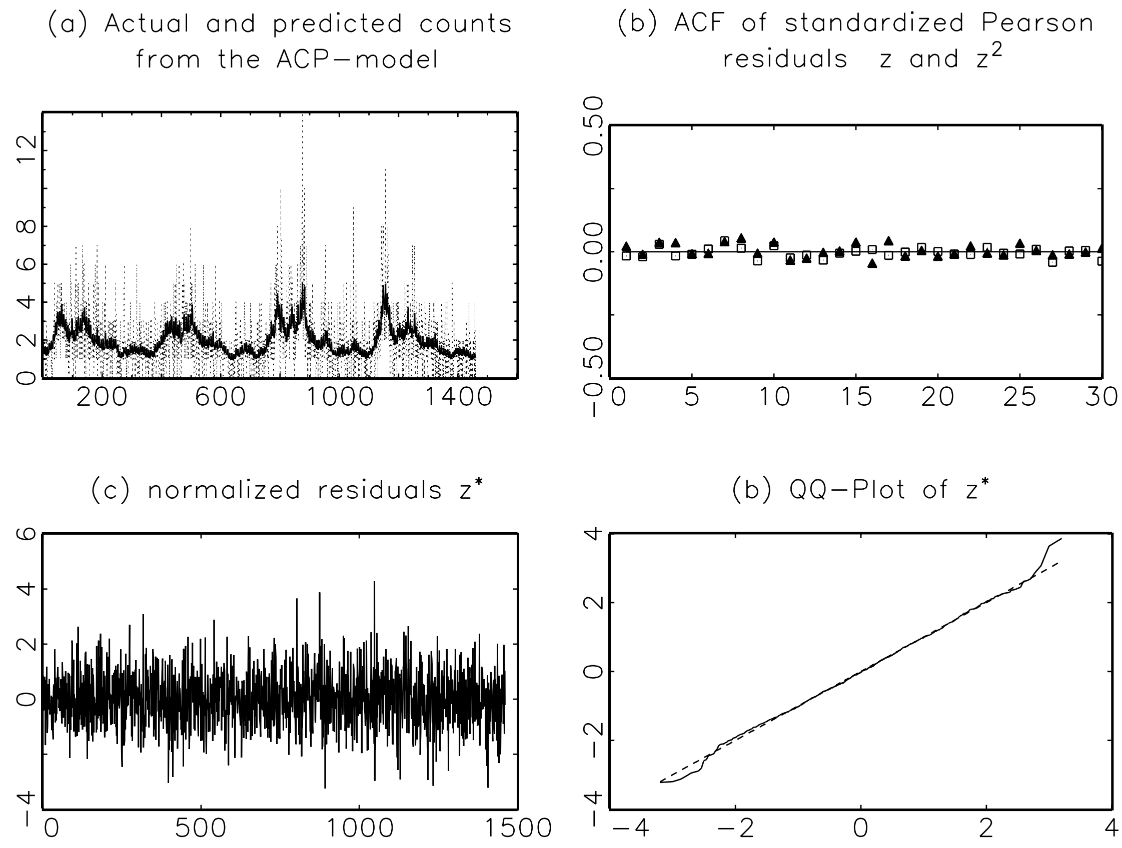


Figure 2. Panel (a): Asthma counts y_t (dotted line) with the conditional mean $\mathrm{E}(y_t|Y_{t-1}, X_t)$ under the SAM model (solid line); Panel (b): Sample autocorrelation function of the standardized Pearson residuals from the SAM z_t (squares) and z_t^2 (triangles); Panel (c): Normalized residuals from the SAM z_t^*; Panel (d): Quantile-quantile plot of the normalized residuals z_t^* from the SAM model (the dashed line plots the quantiles of the standard normal distribution against the quantiles of the standard normal and the solid line plots the sorted values of z_t^* against the quantiles of the standard normal).



Panel (a): Asthma counts y_t (dotted line) with the conditional mean $\mathrm{E}(y_t|Y_{t-1}, X_t)$ under the ACP model (solid line); Panel (b): Sample autocorrelation function of the standardized Pearson residuals from the ACP z_t (squares) and z_t^2 (triangles); Panel (c): Normalized residuals from the ACP z_t^*; Panel (d): Quantile-quantile plot of the normalized residuals z_t^* from the ACP model (the dashed line plots the quantiles of the standard normal distribution against the quantiles of the standard normal and the solid line plots the sorted values of z_t^* against the quantiles of the standard normal).