Independence, successive and conditional likelihood for time series of counts

Sørensen, Helle

Published in:
Journal of Statistical Planning and Inference

DOI:
10.1016/j.jspi.2018.09.002

Publication date:
2019

Document version
Peer reviewed version

Independence, successive and conditional likelihood for time series of counts

Helle Sørensen

Department of Mathematical Sciences, University of Copenhagen, Denmark

email: helle@math.ku.dk

January 2018

SUMMARY: Serial correlation and overdispersion must be handled properly in analyses of time series of counts, and parameter-driven models combine an underlying latent process with a conditional log-linear Poisson model (given the latent process) for that purpose. Regression coefficients have direct interpretations, but likelihood inference is not straight-forward. We consider a two-step procedure for estimation: First regression parameters are estimated from the marginal distribution; second parameters concerning the latent process are estimated with composite likelihood methods, based on low-order simultaneous or conditional distributions. Confidence intervals are computed by bootstrap. Properties of estimators are examined and compared to other methods in three simulation studies, and the methods are applied to two data sets from the literature concerning hospital admission related to asthma and traffic deaths.

KEY WORDS: Bootstrap; Composite likelihood; Overdispersion; Generalized linear mixed model; Serial correlation

1. Introduction

Occurrences of diseases or other events are often registered over time, for example as daily, weekly or monthly counts during some period. Apart from bare surveillance it may be of interest to examine the potential association between occurrence and one or more covariates measured in the same period, or to study trends and seasonal patterns. Hence, regression models for counts are needed that take into account the time series structure of the data.

There is a vast literature for Gaussian time series, but Gaussian methods are only appropriate when counts are of reasonable size. For small counts, corresponding to rare events, the Poisson distribution is a natural starting point, but simple log-linear Poisson regression models must be accommodated to incorporate serial correlation and overdispersion. In the following, let $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ denote the time series of covariate vectors and outcomes. Each y_t is thought of as the realization of a random variable Y_t, and we are thus interested in the distribution of $Y = (Y_1, \ldots, Y_n)$ given x.

Two types of models are often distinguished: observation-driven models and parameter-driven models (Cox, 1981; Zeger, 1988). They differ in their accessibility, both regarding interpretation and statistical inference. Briefly, observation-driven models introduce correlation over time by specifying the conditional distribution of Y_t given past values of $Y_1, \ldots, Y_{t-1}$ and covariates. The maximum likelihood principle is directly applicable (Davis et al., 2005; Fokianos et al., 2009; Davis et al., 2003), but interpretation of regression coefficients is more subtle. Parameter-driven models, on the other hand, introduce a latent process and use a conditional log-linear Poisson model for the covariate-response relationship given the latent process. Interpretation of the regression coefficients from the simple Poisson regression is thereby maintained, however it comes at the expense of a more difficult estimation problem.

Since our aim is to understand associations between counts and covariates, we prefer parameter-driven models. More specifically, let $\alpha = (\alpha_1, \ldots, \alpha_n)$ be an unobserved process and assume that, conditional on α, the random variables $Y_1, \ldots, Y_t$ are independent with Y_t Poisson distributed with mean $\exp(x_t'\beta + \alpha_t)$. As an example, y could be the number of cases of a disease, x could be the usage of antibiotics, and α could represent an underlying and unobserved risk process incorporating unobserved environmental variables (Hay and Pettitt, 2001). With brief notation, and if γ denotes parameters that determine the distribution of

α, the likelihood for (β, γ) is given by

$$L(\beta,\gamma) = p_{\beta,\gamma}(y|x) = \int p_{\beta,\gamma}(y,\alpha|x)\,d\alpha = \int p_\beta(y|\alpha,x)p_\gamma(\alpha)\,d\alpha \tag{1}$$

where arguments tell whether p denotes a marginal, simultaneous or conditional density. The integral is n-dimensional and does not have an explicit solution, so maximum likelihood estimation is not readily possible.

Several estimation approaches have been suggested in the literature, starting from Zeger (1988) who used an iterative scheme with quasi-likelihood estimation for β (for fixed γ) and method-of-moment estimation for γ (for fixed β). Later, focus changed to approximations to the likelihood: Chan and Ledolter (1995) used the EM algorithm and computed the expectation in the E-step by Markov chain Monte Carlo (MCMC) simulation, Durbin and Koopman (1997) used MCMC simulation for an approximating linear Gaussian model in a more general state space setting, and Davis and Rodriguez-Yam (2005) and Jung et al. (2006) used importance sampling techniques. The papers just mentioned rely on approximations of the complete likelihood. A simpler alternative is to use composite likelihood methods (Varin et al., 2011). In particular, Davis and Yau (2011) considered pseudo likelihoods corresponding to pairs of counts of at most lag k (consecutive pairwise likelihood, CPL, of order k). Parameter-driven models are hierarchical generalized linear models in the sense of Lee and Nelder (1996), so h-likelihood methods could be applied; see Lee and Nelder (2001b) for applications to longitudinal and spatial data. Bayesian analysis was presented by Hay and Pettitt (2001).

The method in this paper consists of two steps, both relying on composite likelihoods. First, the marginal density is used to estimate the regression coefficients. This corresponds to a working assumption of independence and could also be interpreted as CPL of order zero (if "pairs of lag zero" is interpreted as single observations). It can be carried out with standard software for generalized linear mixed models. Second, parameters determining the distribution of α are estimated via another composite likelihood, either derived from the simultaneous distribution of k successive counts or derived from the conditional distribution of Y_t given the past m observations. The latter corresponds to a working assumption that Y is Markov of order m. Both pseudo likelihoods are computed by MCMC simulations, but only require simulation of $(\alpha_1,\ldots,\alpha_k)$ and $(\alpha_1,\ldots,\alpha_{m+1})$, respectively. If α is an AR(1) process,

as it is often assumed, then the pseudo likelihoods corresponding to $k = 2$ and $m = 1$ can be used. The marginal variance of α can be estimated as part of either first or second step.

Step 1 provides consistent estimators for the regression parameters of interest, so at first sight step 2 with estimation of nuisance parameters seems unnecessary. However, step 1 does not provide valid standard errors or confidence intervals, so they are computed by parametric bootstrap, for which we need a full data generating model. Since β is estimated for each bootstrap sample it is of great importance that our routine in step 1 is fast and safe. An important part of our numerical studies consists of examining whether the bootstrap confidence intervals have the appropriate coverage, even in cases where the dependence structure is estimated with large uncertainty.

In summary, the aim of the paper is threefold: (1) Comparison of successive and conditional likelihood estimators for γ; (2) validation of coverage for bootstrap confidence intervals for regression coefficients; (3) comparison of our estimators, in particular for β, to those from other methods. Estimation of beta is simple and valid, and our numerical experiments with an AR(1) model for α show that (1) estimators obtained from successive and conditional likelihood are almost identical; (2) confidence intervals have acceptable coverage rates—also for misspecified modes—except when the correlation parameter, denoted ϕ, in the latent AR(1) process is close to 1; and (3) successive and conditional likelihood compares well to other methods for estimation of the regression coefficients, but underestimates ϕ when ϕ is large.

The rest of the paper is organized as follows: Estimation procedures are described in detail in Section 2 and tested on simulated data in Section 3. In Section 4 the methods are applied to two datasets known from the literature. Concluding remarks are given in Section 5.

2. Statistical model, estimation and inference

Most of the notation was already introduced in the introduction: $y = (y_1, \ldots, y_n)$ is the observed time series of counts, and a realization of $Y = (Y_1, \ldots, Y_n)$, and $x = (x_1, \ldots, x_n)$ is the time series of covariates. Each x_t is a vector of length p, most often including 1 corresponding to an intercept. The covariates can be of any type: Observed variables, dummies

(for weekday effects, say), time itself to allow for a trend (possibly scaled), or sine/cosine functions to allow for seasonal patterns.

2.1 *Conditional Poisson model*

Primary interest is in understanding the effect of covariates on the distribution of the outcome, and we model the conditional distribution of Y given x with a hierarchical log-linear Poisson model, assuming that an underlying, unobserved process $\alpha = (\alpha_1, \ldots, \alpha_n)$ drives the dynamics of Y in interplay with the covariates. The latent process introduces over-dispersion as well as serial correlation into the distribution of Y.

The model consists of two parts. One part concerns the distribution of outcomes given the latent process and covariates: Conditionally on the complete α series, the random variables $Y_1, \ldots, Y_n$ are independent and Y_t is Poisson distributed with mean

$$\mathrm{E}(Y_t|x, \alpha) = \exp(x_t'\beta + \alpha_t). \tag{2}$$

In particular, the conditional distribution of Y_t given the complete α series only depends on the current value, α_t, and the marginal distribution of Y_t thus only depends on the distribution of α via its marginal distribution.

The second part concerns the distribution of the latent process: The α series is assumed to be a second order stationary Gaussian process with zero mean (i.e., mean is incorporated in the intercept of the regression model). We write $\tau^2 = \mathrm{Var}(\alpha_t)$ for the marginal variance, ϕ for the parameters that determine the correlation function $\rho(h) = \mathrm{Corr}(\alpha_t, \alpha_{t+h})$, and collect the parameters in $\gamma = (\tau, \phi)$. In our numerical studies we model α as a Gaussian auto-regressive process of order 1, AR(1). Then ϕ is one-dimensional, and we let $\phi = \rho(1)$ such that $\rho(h) = \phi^h$. We denote this model the Poisson auto-regressive model of order one, PAR(1) for short.

The marginal expectation of Y_t is $\mathrm{E}Y_t = \exp(x_t'\beta + \tau^2/2)$, so the interpretation of regression coefficients is the same as in the standard GLM, except for the intercept. Notice that some papers use a parameterization where $\mathrm{E}\alpha_t = -\tau^2/2$; then the intercept parameter in the conditional model can be interpreted as log-expectation of Y_t when all other covariates are zero.

The unknown parameters in the model are β and γ. The regression coefficients are of primary interest whereas entries in γ are regarded as nuisance parameters. Nevertheless we

need estimates for γ in order to generate bootstrap simulations from the estimated model. We suggest to estimate the parameters in two steps: First, the marginal distribution of Y_t is used to estimate β; next, composite likelihood metods are used to estimate γ.

2.2 *Estimation of regression parameters*

For fixed x_t the marginal density of Y_t is given by

$$p_{\beta,\tau}(y_t|x_t) \quad = \int p_\beta(y_t|x_t,\alpha_t)\, p_\tau(\alpha_t)\, d\alpha_t \tag{3}$$

where $p_\beta(y_t|x_t,\alpha_t)$ is the Poisson density with mean $\exp(x_t'\beta+\alpha_t)$, evaluated in y_t, and $p_\tau(\alpha_t)$ is the density for the Gaussian distribution with mean zero and variance τ^2, evaluated in α_t. Notice, that we use $p_\cdot(\cdot|\cdot)$ generically for densities, let the arguments tell whether it is a marginal, conditional or simultaneous density, and give parameter(s) as subscripts.

If the outcomes were independent over time, then the log-likelihood would be

$$\log L_1(\beta,\tau) = \sum_{t=1}^{n} \log p_{\beta,\tau}(y_t|x_t),$$

and we will refer to $\log L_1$ as the independence log-likelihood. The estimator $(\hat\beta,\hat\tau_1)$ is obtained by maximization of $\log L_1$. The outcomes are *not* independent, yet the pseudo score function corresponding to $\log L_1$ is an unbiased estimating function, and $\hat\beta$ is consistent and asymptotically normal if there is sufficient internal replication, for example as for low-order auto-regressive models (Varin et al., 2011). Notice that $\log L_1$ is not identical to the basic GLM log-likelihood that completely ignores α. Davis et al. (2000) showed that even the GLM estimator is consistent, and the estimators are in general close, but since $\log L_1$ uses the correct marginal distribution, we prefer $\hat\beta_1$. The distribution of α only enters into $\log L_1$ via its marginal distribution, so serial correlation is not incorporated. Hence $\log L_1$ only depends on τ and cannot be used to estimate the correlation parameters ϕ.

There is no explicit expression for the integral (3) and thus not for $\log L_1$, but the integrals can be approximated by Laplace or Gauss-Hermite approximations. Such approximations have been implemented for generalized linear mixed models (GLMMs) in software programs like R (`lme4` package, Bates et al. (2015)), SAS (proc glimmix), and Stata (mepoisson), which we can thus rely on: Consider for a moment a Poisson GLMM with log-link, fixed effects of x and random effect of an index variable, e.g., numbered from 1 to n. The log-likelihood for this model is exactly $\log L_1$, so the model fit yields $(\hat\beta,\hat\tau_1)$. The GLMM fit

also supplies standard errors, confidence intervals and hypothesis tests, but—and this is important—they cannot be trusted since they are based on the independence assumption.

In the estimation steps below we keep β fixed at $\hat{\beta}$ (and possibly also τ fixed at $\hat{\tau}_1$), and let $\mu_t = x_t'\hat{\beta}$ be the estimated linear predictor at time t.

2.3 *Estimation of parameters for the latent process via successive likelihood*

Our first suggestion for estimation of γ takes a composite likelihood approach based on successive outcomes. Consider μ_t fixed ($t = 1, \ldots, n$), let $k \geqslant 2$, and consider the density of tuples $(Y_{t+1}, \ldots, Y_{t+k})$ of length k,

$$\begin{aligned} &p_\gamma(y_{t+1}, \ldots, y_{t+k} | \mu_{t+1}, \ldots, \mu_{t+k}) \\ &\qquad = \int \prod_{j=1}^{k} p(y_{t+j} | \mu_{t+j}, \alpha_{t+j})\, p_\gamma(\alpha_{t+1}, \ldots, \alpha_{t+k})\, d(\alpha_{t+1}, \ldots, \alpha_{t+k}), \end{aligned} \tag{4}$$

where dependence on the linear predictors μ_ts is emphasized in the notation. There are $n - k + 1$ tuples $(y_{t+1}, \ldots, y_{t+k})$ of length k. The successive log-likelihood of order k is defined as function that adds the log-densities over all those tuples:

$$\log L_k^{\text{succ}}(\gamma) = \sum_{t=0}^{n-k} \log p_\gamma(y_{t+1}, \ldots, y_{t+k} | \mu_{t+1}, \ldots, \mu_{t+k}).$$

This is not the true log-likelihood for any model since each observation is included in k tuples (except the first and last $k - 1$ observations). Nevertheless, the corresponding pseudo score function would be unbiased if the μ_ts were true rather than estimated.

The successive log-likelihood must be computed numerically and for many values of γ. Each integral in $\log L_k^{\text{succ}}$ can be written as an expected value with respect to the distribution of k successive α_ts,

$$\log L_k^{\text{succ}}(\gamma) = \sum_{t=0}^{n-k} \log \mathrm{E}_\gamma \prod_{j=1}^{k} p(y_{t+j} | \mu_{t+j}, \alpha_{t+j}),$$

where the expectation is with respect to the distribution of $(\alpha_{t+1}, \ldots, \alpha_{t+k})$ which depends on γ as emphasized by the subscript. We suggest to calculate the integrals/expectations by simple MCMC sampling and use the same samples of α tuples for every term in $\log L_k^{\text{succ}}$. That is, we simulate M independent samples of $(\alpha_1, \ldots, \alpha_k)$ from the distribution of α when the true parameter is γ. If the MCMC samples are denoted $(\alpha_1^m, \ldots, \alpha_k^m)$, $m = 1, \ldots, M$, we

then compute an approximation to $\log L_k^{\text{succ}}(\gamma)$ as

$$\log L_k^{\text{succ}}(\gamma) \approx \sum_{t=0}^{n-k} \log \frac{1}{M} \sum_{m=1}^{M} \prod_{j=1}^{k} p(y_{t+j}|\mu_{t+j}, \alpha_j^m).$$

Finally, this function is maximized with respect to γ providing estimates $\hat{\gamma}_k^{\text{succ}} = (\hat{\tau}_k^{\text{succ}}, \hat{\phi}_k^{\text{succ}})$. We stress that the same simulated values $(\alpha_1^m, \ldots, \alpha_k^m)$ are used for all t, and in order for the numerical approximation to $\log L_k^{\text{succ}}$ to be continuous we also use the same samples for every value of γ. Since the regression parameters are held fixed, the optimization problem is low-dimensional.

Recall that step 1 with estimation of β also provides an estimate of τ denoted $\hat{\tau}_1$. Keeping $\tau = \hat{\tau}_1$ fixed in $\log L_k^{\text{succ}}$ reduces the dimension of the maximization problem by one and makes it even simpler. The corresponding estimator of ϕ is denoted $\tilde{\phi}_k^{\text{succ}}$, and it will be compared to $\hat{\phi}_k^{\text{succ}}$ in our simulation studies. If $\hat{\tau}_1 = 0$, then $\tilde{\phi}_k^{\text{succ}}$ is left undefined.

Choosing the value of k is a matter of identifiability, efficiency, (numerical) precision, and computation time. First, k should be large enough for the model to be identified. The log-likelihood of order k is determined by the distribution of k-tuples of α, so k should be large enough that all elements in γ can be determined from the k-tuple distribution. For example, if α is an AR(q) process then $k \geqslant q + 1$ is needed as the complete distribution is determined by the distribution of $(\alpha_1, \ldots, \alpha_{q+1})$. Similar constraints apply to other models for α. In principle, one would expect the estimators for larger k to be more efficient compared to smaller k, since more features of the distribution are taken into account. On the other hand, correlation in α is blurred by the extra Poisson variation, so the gained efficiency is presumably limited, as was also confirmed by simulation studies in Davis and Yau (2011). Furthermore, the MCMC computation of the k-dimensional integral/expectation, which must be carried out for all t and for many values of γ (for maximization) is obviously easier to do fast and precisely when k is small. Altogether, our suggestion is to use the smallest possible k such that the model is identified.

When α is modeled as an AR(1) process we thus use $k = 2$. We then simulate only pairs (α_1^m, α_2^m), and the pairwise log-likelihood is computed as

$$\log L_2^{\text{succ}}(\gamma) \approx \sum_{t=0}^{n-1} \log \frac{1}{M} \sum_{m=1}^{M} p(y_{t+1}|\mu_{t+1}, \alpha_1^m)\, p(y_{t+2}|\mu_{t+2}, \alpha_2^m)$$

Notice that $\log L_2^{\text{succ}}$ is identical to the consecutive pairwise likelihood of order 2 from Davis and Yau (2011), except that β is not fixed in their approach.

2.4 *Estimation of parameters for the latent process via conditional likelihood*

As an alternative to successive likelihood we also consider a conditional likelihood approach (Azzalini, 1983; Sørensen, 2003). Again we consider μ_ts fixed. First, notice that the true likelihood function can be factorized as

$$L(\gamma) = \prod_{t=1}^{n} p_\gamma(y_t|\mu_1, \ldots, \mu_t, y_1, \ldots, y_{t-1})$$

where term t is the conditional density of Y_t given the complete past (assuming that β and thus μ_ts are known). The conditional likelihood of order m is the approximation where we only condition on the recent m observations, i.e.,

$$L_m^{\text{cond}}(\gamma) = \prod_{t=m+1}^{n} p_\gamma(y_t|\mu_t, y_{t-m}, \ldots, y_{t-1}) = \prod_{t=m+1}^{n} \frac{p_\gamma(y_{t-m}, \ldots, y_t|\mu_{t-m}, \ldots, \mu_t)}{p_\gamma(y_{t-m}, \ldots, y_{t-1}|\mu_{t-m}, \ldots, \mu_{t-1})}$$

where, for simplicity, we have skipped terms corresponding to the first m observations. This would be the true likelihood if Y was a Markov process of order m; which it is not, but the corresponding pseudo score function would be unbiased if the μ_ts were true rather than estimated (Sørensen, 2003). The conditional log-likelihood can be written in terms of successive log-likelihoods for k equal to $m+1$ and m,

$$\log L_m^{\text{cond}}(\gamma) = \log L_{m+1}^{\text{succ}}(\gamma) - \log L_m^{\text{succ}}(\gamma)$$

where, for $m = 1$, we define $\log L_1^{\text{succ}} = \log L_1$ (the independence likelihood). Therefore we need the same type of computations as for $\log L^{\text{succ}}$, and we use the same MCMC simulations for αs in both terms.

The estimates are denoted $\hat{\gamma}_m^{\text{cond}} = (\hat{\tau}_m^{\text{cond}}, \hat{\phi}_m^{\text{cond}})$ if τ is re-estimated using conditional likelihood and $\tilde{\phi}_m^{\text{cond}}$ if $\tau = \hat{\tau}_1$ is kept fixed. The arguments for choosing m are similar to those for choosing k, and we suggest to choose m as small as possible; see also simulations in Azzalini (1983). For an AR(1) specification of the latent process we thus use $m = 1$. As mentioned above

$$\log L_1^{\text{cond}}(\gamma) = \log L_1^{\text{cond}}(\tau, \phi) = \log L_2^{\text{succ}}(\tau, \phi) - \log L_1(\tau),$$

so if we keep $\tau = \hat{\tau}_1$ fixed, then successive likelihood of order 2 and conditional likelihood of order 1 leads to same estimate of ϕ, that is, $\tilde{\phi}_2^{\text{succ}} = \tilde{\phi}_1^{\text{cond}}$.

2.5 *Complete estimation procedure, incl. bootstrap confidence intervals*

Recall from Section 2.2 that the regression coefficients are estimated by maximization of the independence likelihood. This is easy and provides consistent estimators, but the Hessian matrix cannot be used to compute standard errors since $\log L_1$ does not comply with the correlation structure of the model. Instead we rely on the parametric bootstrap via simulations from the model with parameters equal to their estimates.

Several alternatives were suggested for $\gamma = (\tau, \phi)$ in Sections 2.2–2.4: τ could either be estimated from the independence, successive or conditional likelihood, and ϕ could either be estimated from the successive or conditional likelihood (with different values of k or m), and with or without τ fixed at $\hat{\tau}_1$. We compare the different choices for simulated data in Section 3.

After having decided which estimator to use for γ, the complete estimation procedure goes as follows.

(1) Maximize $\log L_1$ wrt. (β, τ) and obtain an estimate $\hat{\beta}$ of the regression coefficients.

(2) Compute the linear predictors $\mu_t = x_t'\hat{\beta}_t$, consider them fixed, and use the selected estimation procedure for γ to obtain an estimate $\hat{\gamma}$.

(3) Make R simulated trajectories of (α, Y) from the model with parameters $(\beta, \gamma) = (\hat{\beta}, \hat{\gamma})$, and with covariates x as for the original data. Repeat step (1) for each simulated dataset, and denote the corresponding estimate by $\hat{\beta}^r$.

(4) Compute a 95% confidence interval for each β_j based on the simulations. As a standard, since $\hat{\beta}$ is consistent and asymptotically normal for low-order auto-regressive models, we use $\hat{\beta}_j \pm 1.96\,\mathrm{sd}(\hat{\beta}_j^r)$ where $\mathrm{sd}(\hat{\beta}_j^r)$ is the standard deviation across the R bootstrap estimates. This has the advantage that R can be chosen relatively small compared to confidence intervals relying on quantile computations (Davison and Hinkley, 1997).

Since step (1) is carried out for each bootstrap sample it is a great advantage that it does not involve time-consuming (and approximate) MCMC simulations. Step (2) is only carried out for the observed data, except if one is also interested in confidence intervals for the entries in γ, in which case step (2) should be invoked in step (3) for each bootstrap sample. Estimation of the auto-correlation parameter ϕ turns out not to precise in all situations, and

it is not obvious how much this affects the validity of the bootstrap confidence intervals. We therefore examine the actual coverage in simulation studies in the next section.

We used R (R Core Team, 2016) for all computations, more specifically the `glmer` function (Bates et al., 2015) in step (1) and the `optim` function with method `L-BFGS-B` in step (2). The online supplement shows code used for the analyses. The CPU times reported in Sections 3.1 and 4.1 are obtained on a MacBook Pro with a 2.3 GHz Intel Core i7 processor and 8 GB memory. No attempts were made to optimize runtime; in particular bootstrap computations could easily be run in parallel.

3. Simulation studies

3.1 *Comparison of successive and conditional likelihood estimators*

The main purpose of our first simulation study is to compare our different strategies for estimation. The data generating model is the PAR(1) model with true parameters $\tau = \phi = 0.5$ for the latent process and a conditional Poisson model of the form

$$\mathrm{E}(Y_t|x, \alpha) = \exp(\beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \alpha_t)$$

where $x_{t,1} = (t-1)/n$ corresponds to a trend, and $x_{t,2}$ is binary with $P(x_{t,2} = 1) = 1 - P(x_{t,2} = 0) = 0.25$, independently over time, corresponding to some event that happens by chance. The length of the time series is $n = 200$, and the true values of the regression coefficients are

$$\beta_0 = 0.5, \quad \beta_1 = 0.6931, \quad \beta_2 = 0.2231.$$

For the chosen parameter values the marginal mean at the beginning of the time series is $\mathrm{E}Y_1 = \exp(\beta_0 + \tau^2/2) = 1.87$, it doubles over the time range ($e^{0.6931} = 2$), and the event increases expectation by 25% ($e^{0.2231} = 1.25$).

We simulated 1000 time series from the data generating model. For each dataset we computed the estimators $\hat{\beta}_j$ ($j = 0, 1, 2$) and $\hat{\tau}_1$ from the independence likelihood, $(\hat{\tau}_2^{\mathrm{succ}}, \hat{\phi}_2^{\mathrm{succ}})$ from the successive likelihood of order 2, $(\hat{\tau}_1^{\mathrm{cond}}, \hat{\phi}_1^{\mathrm{cond}})$ from the conditional likelihood of order 1, and $\tilde{\phi}_2^{\mathrm{succ}} = \tilde{\phi}_1^{\mathrm{cond}}$ by keeping $\hat{\tau} = \hat{\tau}_1$ fixed. We used $M = 1000$ for computation of the log-likelihoods. We also computed 95% symmetry-based confidence intervals for each regression parameter and each estimation strategy. We used $R = 100$ bootstrap samples, which was

Table 1

Results from 1000 simulated time series from a PAR(1) model. Estimates are computed by maximization of $\log L_1$ *(independence),* $\log L_2^{succ}$ *(successive),* $\log L_1^{cond}$ *(conditional), and* $\log L_2^{succ}$ *or* $\log L_1^{cond}$ *with* τ *fixed at* $\hat{\tau}_1$ *(succ./cond.,* $\tau = \hat{\tau}_1$*).*

True		Independence	Successive	Conditional	Succ./cond., $\tau = \hat{\tau}_1$
β_0=0.5	Mean	0.507	—	—	—
	SD	0.162	—	—	—
	Coverage	—	0.944	0.943	0.945
$\beta_1 = 0.6931$	Mean	0.677	—	—	—
	SD	0.259	—	—	—
	Coverage	—	0.939	0.935	0.939
$\beta_2 = 0.2231$	Mean	0.220	—	—	—
	SD	0.129	—	—	—
	Coverage	—	0.934	0.933	0.931
$\tau = 0.5$	Mean	0.481	0.480	0.481	—
	SD	0.064	0.069	0.070	—
$\phi = 0.5$	Mean	—	0.478	0.478	0.475
	SD	—	0.200	0.201	0.200
Average CPU time			15.1 s	16.7 s	13.3 s

deemed large enough since we only need estimated standard deviations of the bootstrap distributions.

Results regarding mean, standard deviations and coverage are listed in Table 1. All parameters are estimated with no or small bias in this set-up. Results are very similar for the different estimators for τ and ϕ, not only in distribution as indicated by means and standard deviations, but also for each realization: Pairwise Pearson correlations are above 0.95 for the three estimators of τ and above 0.99 for the three estimators of ϕ. Hence, confidence intervals based on different estimators and the corresponding coverage rates are also almost identical between methods. All coverage rates are above 0.93, but a bit below the nominal level.

Histograms for $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\tau}_2^{\text{succ}}$ and $\hat{\phi}_2^{\text{succ}}$ are shown in Figure 1. Vertical red and blue lines show true values and means over the 1000 simulated datasets, respectively. Most importantly, notice that the distribution of each $\hat{\beta}_j$ is symmetric (as expected); hence the symmetry-based confidence intervals are appropriate.

Average CPU time for analysis (estimation and confidence intervals) is listed in the

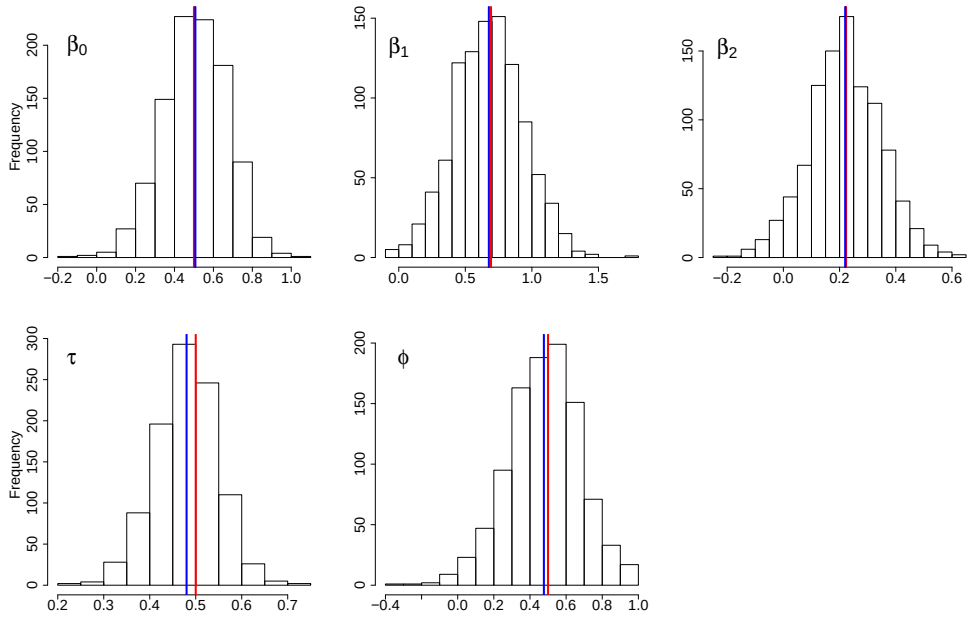

Figure 1. Histograms of estimates from 1000 simulated data sets from the PAR(1) model. We used $\hat{\tau}_2^{\text{succ}}$ and $\hat{\phi}_2^{\text{succ}}$ as estimates of parameters in the latent process. Vertical red lines indicate true values, and vertical blue lines indicate means over the simulated datasets.

last line in Table 1. As expected, successive likelihood is faster than conditional likelihood (one term instead of two per observation), and it is faster to keep τ fixed than to re-estimate it. However, the differences are small since the main burden lies in the bootstrap step, i.e., step (3) from Section 2.5, which is the same for all methods.

We studied other parameter values and always got almost identical results for successive and conditional likelihood. In Section 3.2 below we therefore only use one pair of estimators for the latent process. We have chosen $(\hat{\tau}_2^{\text{succ}}, \hat{\phi}_2^{\text{succ}})$ since it is simpler to compute than $(\hat{\tau}_2^{\text{cond}}, \hat{\phi}_2^{\text{cond}})$ and since is provides an estimate of ϕ even when $\hat{\tau}_1 = 0$.

3.2 *Sensitivity against distribution of latent process*

The purpose of the next study is to examine the performance of the estimators for varying degrees of overdispersion and correlation in the latent process as well as under misspecification of the data generating model. In all scenarios we used $M = 1000$ and $R = 100$.

First, we used the same data generating process as in Section 3.1, except that we varied the value of ϕ between -0.5 and 0.9. Results are presented in Figure 2. Top plots and the

two leftmost bottom plots show boxplots for estimators with red dashed lines corresponding to true values. Estimators for regression parameters are unbiased in all scenarios. Standard deviation increases for $\hat{\beta}_0$ and $\hat{\beta}_1$, but not for $\hat{\beta}_2$, when ϕ increases. This is not surprising since there is less information in the time series about level and trend when data are highly correlated, whereas the effect of the event is less prone to this information loss as it happens at random. The estimator $\hat{\tau}_2^{\text{succ}}$ underestimates τ in all cases, and worst when ϕ is large. The median for $\hat{\phi}_2^{\text{cond}}$ is close to the true value of ϕ in all scenarios, but for ϕ large the distribution is strongly skewed to the left — as is natural due to the upper bound of ϕ at 1.

The bottom right plot of Figure 2 shows the coverage rates for the symmetry-based bootstrap confidence intervals (solid lines) and for the confidence intervals based on the independence likelihood (dashes lines). As expected, the confidence intervals based on independence have wrong coverage rates when ϕ is not close to zero (except for β_2). The bootstrap confidence intervals are of more interest. When ϕ is 0.5 or smaller, the coverage is reasonably close to the nominal level of 0.95, albeit always to the lower side. For $\phi = 0.75$ and particularly for $\phi = 0.9$, however, the coverage rates are considerably lower, and the intervals are thus not valid as 95% confidence intervals. The reason is to be found in the distribution of the estimators for τ and ϕ: When too small estimates of τ and ϕ are used in the bootstrap simulations, then variation of $\hat{\beta}_0$ and $\hat{\beta}_1$ is strongly underestimated, cf. the two leftmost top plots.

Second, we kept $\phi = 0.5$ fixed and compared results for $\tau = 0.3, 0.5, 0.7$, see Figure 3. Many results are similar to those for fixed τ: Regression coefficients are estimated without bias and most precisely when τ is small (also $\hat{\beta}_2$ in this case, since a large τ implies larger variation in general). The correlation parameter is estimated with only little bias in all three scenarios, but the variation is large when τ is small (small overall variability implies little information about correlation). The standard deviation τ is underestimated in all three scenarios. Consequently, coverage rates are in most cases too small, yet above 0.93, and in most cases far better than those from the independence fit.

So far, we used the PAR(1) model as data generating process such that the simulation model and the estimation model are in accordance. We now consider two scenarios with misspecification of the latent process distribution. The conditional distribution of Y given

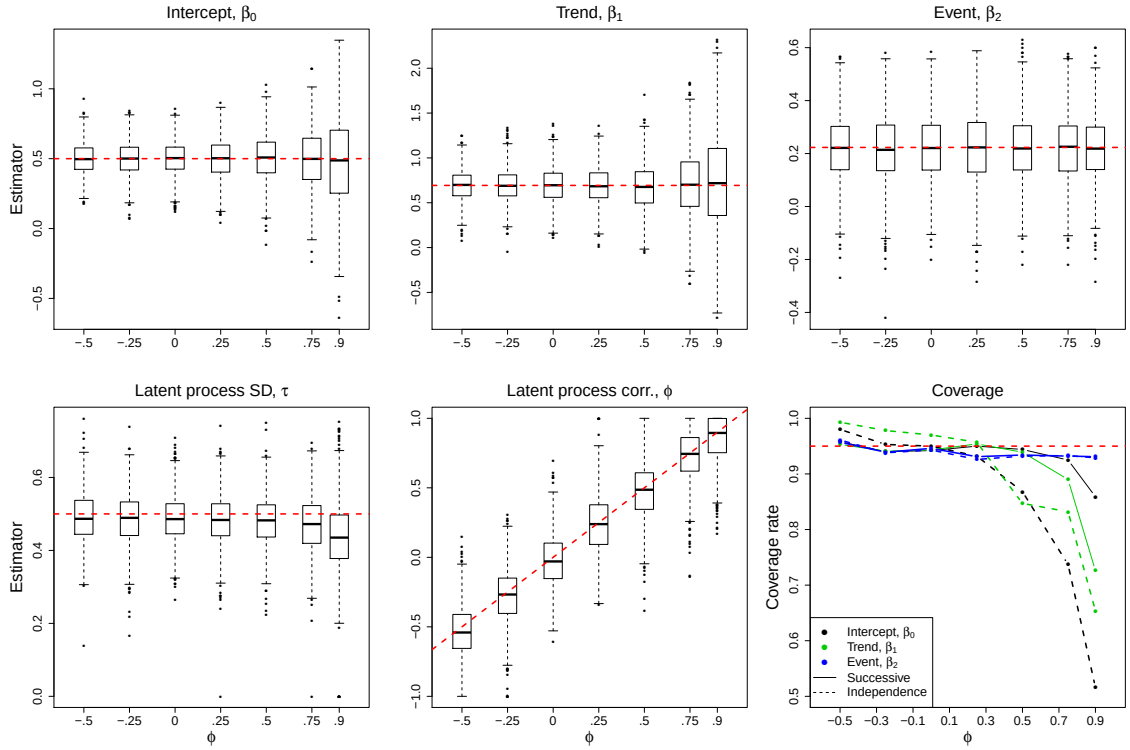

Figure 2. Boxplots for 1000 estimators in the PAR(1) model (first five plots) and coverage of the confidence intervals for the regression coefficients (bottom right) for seven different values of the correlation parameter ϕ. True values of parameters and the nominal level of the confidence intervals (95%) are shown with dashed red lines.

(x, α) is unchanged, see (2). In the first scenario we used t-distributed innovations in the latent AR(1) process instead of Gaussian innovations. More specifically, we generated the latent process as

$$\alpha_t = \phi\alpha_{t-1} + \tilde{\sigma}e_t$$

where e_ts are independent and t-distributed with four degrees of freedom, $\phi = 0.5$ and $\tilde{\sigma} = 0.3062$. Then the innovations have the same variance as in the Gaussian model used so far ($\tau = \phi = 0.5$). In the second scenario we used a moving-average process of order one, MA(1), for the latent process. The parameters were chosen such that $\text{Corr}(\alpha_t, \alpha_{t+1}) = 0.5$ and $\text{Var}(\alpha_t) = 0.25$ as in our standard PAR(1) set-up. Table 2 shows results for the regression coefficients. Estimators are unbiased, and coverage rates are close to the nominal level in both scenarios.

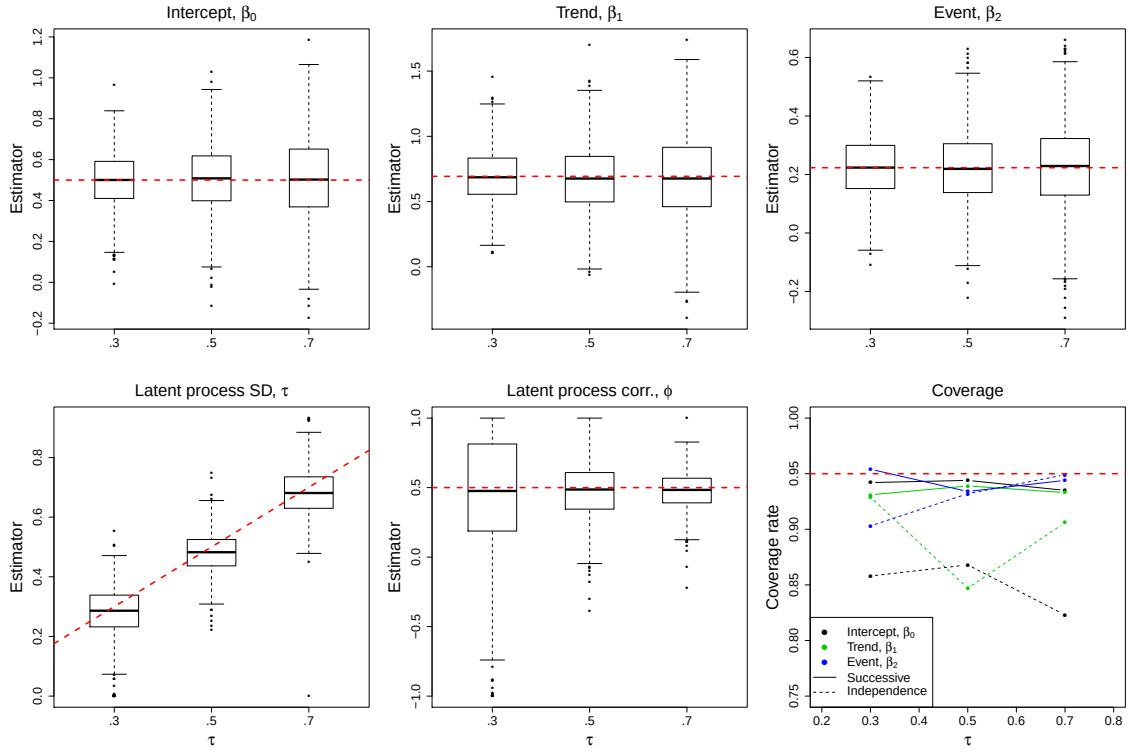

Figure 3. Boxplots for 1000 estimators in the PAR(1) model (first five plots) and coverage of the confidence intervals for the regression coefficients (bottom right) for three different values of the overdispersion parameter τ. True values of parameters and the nominal level of the confidence intervals (95%) are shown with dashed red lines.

Table 2

Results from 1000 simulated time series, based on independence/successive likelihood assuming the PAR(1) model. The latent processes were either generated as auto-regressive processes with t-distributed innovations or as MA(1) processes.

True		t innovations	MA(1)
	Mean	0.490	0.502
β_0=0.5	SD	0.164	0.150
	Coverage	0.946	0.954
	Mean	0.693	0.693
$\beta_1 = 0.6931$	SD	0.267	0.238
	Coverage	0.932	0.964
	Mean	0.212	0.217
$\beta_2 = 0.2231$	SD	0.132	0.132
	Coverage	0.952	0.937

Table 3

Parameter values for simulations in Section 3.3. Parameters in the latent process are linked by $\sigma^2 = (1-\phi^2)\tau^2$.

	β	ϕ	σ	τ
Scenario 1	−0.613	−0.5	1.236	1.427
Scenario 2	0.150	0.5	0.619	0.715
Scenario 3	0.373	0.9	0.111	0.255

3.3 *Comparison to AIS and CPL*

The purpose of our final simulation study is to compare the performance for our estimators and estimators from the literature. The set-up is equivalent to one in Davis and Yau (2011), namely a simple PAR(1) model without covariates (except for a constant), i.e.,

$$\mathrm{E}(Y_t|\alpha_t;\beta) = \exp(\beta + \alpha_t).$$

Recall that we parameterize the AR(1) process by the marginal standard deviation τ and the correlation parameter ϕ. Davis and Yau (2011) used the standard deviation of the independent innovations in the AR(1) process, denoted σ, instead of τ. The two standard deviations are linked by the equation $\sigma^2 = (1-\phi^2)\tau^2$, and we give the results for both parameterizations below when possible.

We considered three sets of parameters, see Table 3. The scenarios are identical to three of the nine scenarios in Table 5 from Davis and Rodriguez-Yam (2005) and Table 3 from Davis and Yau (2011), and the marginal mean of Y_t is 1.5 in all three scenarios. We used $n = 500$ and generated 500 datasets in each scenario (also as in the above-mentioned papers).

Bias and standard deviations are reported in Table 4 for several estimators: those obtained from the independence, successive and conditional likelihoods in this paper as well as those obtained by AIS (Davis and Rodriguez-Yam, 2005) and CPL_1 (Davis and Yau, 2011). Results for AIS and CPL_1 are copied from Davis and Yau (2011). Recall that AIS uses importance sampling approximations to the complete likelihood (1) rather than a lower-order composite likelihood, whereas CPL_1 (consecutive pairwise likelihood of order 1) is equivalent to our successive likelihood of order 2, except that all parameters are estimated using the successive likelihood instead of keeping β fixed at the independence-based estimator.

The results regarding the intercept are very similar for $\hat{\beta}$ computed from the indepen-

Table 4

Results from 500 simulated data series. Results regarding AIS and CPL$_1$ are copied from Davis and Yau (2011). Other estimates are computed by maximization of $\log L_1$ *(independence),* $\log L_2^{succ}$ *(successive),* $\log L_1^{cond}$ *(conditional), and* $\log L_2^{succ}$ *or* $\log L_1^{cond}$ *with* τ *fixed at* $\hat{\tau}_1$ *(succ./cond.,* $\tau = \hat{\tau}_1$*). The bias marked with a star* ($\star$) *is different from the similar number in Davis and Rodriguez-Yam (2005), where is was reported as 0.062.*

	Scenario 1		Scenario 2		Scenario 3	
	Bias	SD	Bias	SD	Bias	SD
β						
AIS	−0.031	0.093	−0.001	0.073	0.003	0.078
CPL$_1$	0.105	0.095	−0.001	0.060	0.146	0.123
Independence	−0.020	0.093	−0.006	0.074	−0.008	0.066
τ						
Independence	−0.006	0.095	−0.003	0.055	−0.004	0.114
Successive	0.066	0.152	−0.0005	0.064	−0.008	0.096
Conditional	0.054	0.151	0.001	0.067	0.003	0.087
ϕ						
AIS	0.022	0.063	0.012	0.091	0.078	0.231
CPL$_1$	−0.005	0.065	−0.003	0.055	0.061	0.032
Successive	−0.003	0.151	0.0001	0.117	−0.248	0.427
Conditional	0.0001	0.092	0.0003	0.117	−0.234	0.412
Succ./cond., $\tau = \hat{\tau}_1$	0.027	0.092	−0.001	0.117	−0.294	0.464
σ						
AIS	0.052*	0.100	0.010	0.061	−0.016	0.062
CPL$_1$	−0.148	0.134	−0.001	0.073	−0.102	0.013
Successive	0.043	0.127	−0.003	0.076	0.013	0.129
Conditional	0.036	0.147	−0.007	0.078	0.014	0.129
Succ./cond., $\tau = \hat{\tau}_1$	−0.001	0.144	−0.009	0.073	0.041	0.137

dence likelihood and the AIS method. Surprisingly, the CPL$_1$ intercept estimator is biased in scenarios 1 and 3.

The estimator $\hat{\tau}_1$ is unbiased in all three scenarios, whereas the estimators based on successive and conditional likelihoods show a small bias in scenario 1. Moreover, $\hat{\tau}_1$ has smaller standard deviation in scenarios 1 and 2. Due to the different parameterization, comparison to AIS and CPL$_1$ is not possible for τ using the results from Davis and Yau (2011). Successive/conditional likelihood and AIS appear to give smaller bias for σ than CPL$_1$, and AIS has the smallest standard deviation.

AIS and especially CPL$_1$ outperform successive and conditional likelihood when it comes to estimation of ϕ; in particular in scenario 3 where ϕ is large, the successive and

conditional likelihood estimators are strongly downwards biased. The difference between successive likelihood and CPL_1 consists of whether β is fixed at $\hat{\beta}_1$ or varies freely during estimation of (τ, ϕ), and even though it was already noted by Davis et al. (2000) that the use of $\hat{\beta}$ leads to substantial bias, the large difference is surprising. However, as noted above, the better estimation of ϕ by CPL_1 comes at the expense of bias for the estimator of β, which is problematic since β is the parameter of interest.

Coverages rates for 95% bootstrap confidence intervals based on successive/conditional estimation range from 94% to 97% in scenarios 1 and 2, but are only 89–90% for scenario 3; hardly surprising considering the severe bias for ϕ and the results from simulations in Section 3.2. Similar evaluations are not available for AIS and CPL_1; however, one could imagine that bootstrap-based confidence intervals for those methods would have better coverage properties because ϕ is not underestimated.

Altogether we conclude that successive/conditional likelihood give competitive estimators for the regression parameter of interest and for the marginal standard deviation in all scenarios, but not for the correlation parameter when there is strong serial correlation.

4. Data applications

4.1 *Sydney asthma data*

Our first application considers daily admissions to Cambelltown Hospital in Sydney from January 1990 to December 1993 with a total of 1461 observations, available in the R package `glarma` associated to Dunsmuir and Scott (2015). The counts range from zero to fourteen, and the mean is 1.94. The time series was previously analyzed by Davis et al. (2000) using an ordinary GLM in combination with empirical moments of Y_t for estimation and correcting for presence of a latent process, and Jung et al. (2006) using their ML-EIS method.

We use the PAR(1) model and include the following covariates: dummies for Monday and Sunday, trend, a humidity variable as constructed in Davis et al. (2000), and pairs $\cos(2\pi kt/365)$ and $\sin(2\pi kt/365)$ for $k = 1, 2, 3, 4$ in order to correct for seasonal variation. The model is not completely identical to previous analyses: Davis et al. (2000) included two more sine/cosine pairs, Jung et al. (2006) did not include the humidity variable, and neither of them included a trend in their final analysis.

Table 5

Results from analyses of the asthma data. The AR(1) process is parameterized differently in Jung (2006) and Davis et al. (2000), so estimates marked with a star ($\star$) are transformations of those listed in the paper.

	Indep., succ.		Jung (2006)		Davis et al. (2000)	
	Estimate	SE	Estimate	SE	Estimate	SE
Sunday	0.227	0.058	0.229	0.045	0.230	0.055
Monday	0.237	0.054	0.232	0.047	0.236	0.055
Trend	0.089	0.093	—	—	—	—
Humidity	0.204	0.070	—	—	0.210	0.066
τ	0.234	—	$0.220^\star$	—	$0.255^\star$	—
ϕ	0.804	—	0.900	—	$0.796^\star$	—

Results concerning the covariates of interest and the latent process are listed in Table 5; estimates of τ and ϕ are based on successive likelihood of order 2 computed with $M = 1000$. Estimates of τ were similar for successive and conditional likelihood, whereas the estimate $\tilde{\phi}_2^{\text{succ}} = \tilde{\phi}_1^{\text{cond}}$ was considerably lower than $\hat{\phi}_2^{\text{succ}}$ and $\hat{\phi}_1^{\text{cond}}$ (0.718 vs. 0.804 and 0.818). In order to compare with results from the previous analyses we list bootstrap standard errors rather than confidence intervals, computed as the standard deviation of estimates over $R = 100$ bootstrap samples (this took 12 minutes). Estimates as well as standard errors are similar across methods; in particular all methods find that the number of daily admissions are significantly larger on Sundays and Mondays compared to other weekdays, and that high humidity increases the number of admissions. The trend, on the other hand, is not significant as was also mentioned by Davis et al. (2000).

4.2 *Van driver road deaths*

The second application uses a dataset consisting of the monthly number of van drivers killed in Great Britain from January 1969 to December 1984 (a total of 192 observations). The dataset is available in the R package `tscount` accompanying Liboschik et al. (2016). Seatbelt legislation was introduced January 31, 1983, and one aim of the analysis is to see if this had an effect on road deaths. The time series is shown in Figure 4 with a red dashed line at the month of the new law.

We consider a PAR(1) model with a dummy for the new law, trend, petrol price, and dummies for each month to allow for seasonal variation. We get the following estimates for

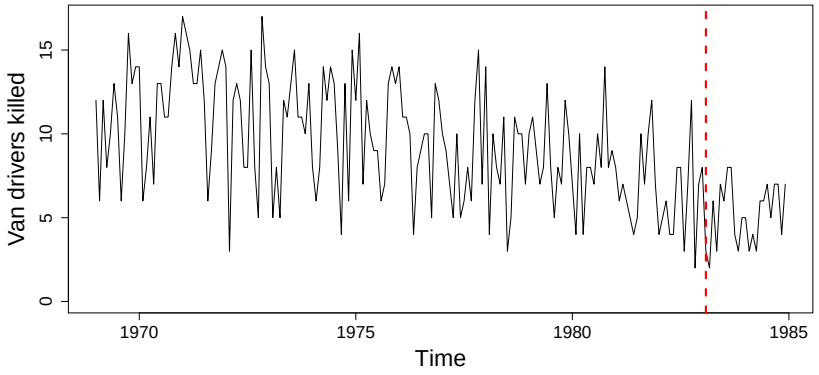

Figure 4. Time series consisting of the number of van drivers killed in traffic in Great Britain each month 1969–1984. The dashed red line indicates introduction of seatbelt legislation.

the latent process, indicating that there is no need for the latent process:

$$\hat{\tau}_1 = 0, \quad \hat{\tau}_2^{\text{succ}} = 0.00023, \quad \hat{\phi}_2^{\text{succ}} = 0.019, \quad \hat{\tau}_2^{\text{cond}} = 0.00039, \quad \hat{\phi}_2^{\text{cond}} = 0.019.$$

In particular, since $\hat{\tau}_1 = 0$, the estimates and standard errors from the independence likelihood coincide with those from an ordinary GLM.

In order to check the validity of the results we also fitted an Gaussian model with log-count as outcome, the same covariates as above, and an AR(1) residual process. Notice from Figure 4 that counts are not very small (and never zero), so a Gaussian model is not completely unreasonable. The correlation parameter in the model is estimated to 0.043, confirming that auto-correlation is weak.

Regarding the effect of seatbelt legislation, the GLM estimate is -0.253 (SE 0.110) and the estimate from the Gaussian model is -0.244 (SE 0.106), corresponding to an reduction in the number of killed van drivers of approximately 22%. This compares well to the estimate -0.276 from Durbin and Koopman (1997) obtained by MCMC approximations to the likelihood in a slightly different model (no standard error was provided), and to the estimate -0.218 (SE 0.129) from Lee and Nelder (2001a) from a model with a time-varying dispersion parameter, but no dependence over time. As opposed to this, the law was estimated to increase the number of deaths, although not significantly, in an observation-driven model in Liboschik et al. (2016).

5. Discussion

We have used composite likelihood techniques for parameter estimation in a class of parameter-driven models for time series of counts. The models are composed of a latent process and a conditional log-linear Poisson time series (given the latent process). Parameters were estimated in two steps: regression coefficients from the marginal density, i.e. under a working assumption of independence, and parameters for the latent process from likelihoods involving low-order simultaneous or conditional distributions. Standard errors and/or confidence intervals for regression coefficients were computed by bootstrap, sampling time series from the estimated model. For our simulations and data applications we modeled the latent process as an AR(1) process, but the methods applies to all second-order stationary Gaussian processes.

Regression parameters are indeed identifiable from the independence likelihood, estimators are asymptotically well-behaved, and standard software for generalized linear mixed models can be used. The independence likelihood uses the correct marginal distribution, and we therefore prefer it to the standard GLM even though the estimates are usually close for non-constant covariates. Our simulation studies showed that the different estimators for the parameters in the latent process are in general strongly correlated. Successive likelihood is simpler to compute, and there is no indication that conditional likelihood is preferable to successive likelihood, so we recommend the successive likelihood approach. We used MCMC simulations for computation of the log-likelihoods, but since only low-dimensional integrals are to be computed this could also be done with Laplace/Gauss-Hermite approximations, thus avoiding extra variation due to simulation.

As expected, regression parameters were estimated without bias, but in situations with strong serial correlation estimators based on successive/conditional likelihoods were severely biased downward, and coverage of bootstrap confidence intervals was too low. The major difference between successive likelihood of order 2 and CPL_1 (consecutive pairwise likelihood of order 1) is whether the regression parameters are held fixed during the estimation of the latent process parameters or not. Apparently, keeping them fixed is beneficial for estimation of serial correlation but introduces bias in the estimators of the regression coefficients. AIS (approximate likelihood by importance sampling) uses the complete likelihood and appears to give the best overall results, but as long as serial correlation is not too strong, successive

likelihood of order 2 and AIS give similar results. This indicates, not surprisingly, that information regarding lags of order larger than one, is useful when correlation is strong, but otherwise not.

Although (marginal) overdispersion is the more common phenomenon for time series of counts, the opposite also occurs. Underdispersion is not accomodated by the PAR models from this paper, but one could substitute the conditional Poisson distribution by a generalization allowing for underdispersion (Shmueli et al., 2005; Consul and Famoye, 1992). The principles for inference, i.e. steps (1)–(4) from Section 2.5, could be carried over to such models; however, step (1) would consist of estimation in generalized Poisson models with random effects for which there appears to be no standard software. Observation-driven models for underdispersed data have been proposed by Zhu (2012a,b).

We did not attempt to prove any theoretical results about asymptotic, but arguments would be in the following directions: As mentioned in Section 2.2, $\hat{\beta}$ is consistent and asymptotically normal because the independence likelihood uses the correct marginal distribution; this follows from the theory for M- and Z-estimators (van der Vaart, 1998, Chapter 5). Similarly, the successive/conditional likelihood for γ—with β fixed at the true parameter β rather then $\hat{\beta}$—use the correct successive/conditional distributions. Plugging in $\hat{\beta}$ requires extra arguments, see van der Vaart (1998, Section 5.4). The standard proof for consistency of the bootstrap confidence interval (i.e., asymptotically correct coverage) would require simultaneous asymptotic normality of $(\hat{\beta}, \hat{\gamma})$, see van der Vaart (1998, Section 23.2).

Acknowledgements

The project was partly funded by the Danish Research Council (grant 0602-02320B). Thanks to René Aakær Jensen who wrote his MSc thesis on conditional likelihood at a early stage of the project and to a reviewer for useful comments.

References

Azzalini, A. (1983). Maximum likelikood estimation of order m for stationary processes. *Biometrika* **70,** 381–387.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67,** 1–48.

Chan, K. S. and Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association* **90,** 242–252.

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

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

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

Davis, R. A., Dunsmuir, W. T. M., and Street, S. B. (2005). Maximum likelihood estimation for an observation driven modeol for Poison counts. *Methodology and Computing in Applied Probability* **7,** 149–159.

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. and Rodriguez-Yam, G. (2005). Estimation for state-space models based on a likelihood approximation. *Statistica Sinica* **15,** 381–406.

Davis, R. A. and Yau, C. Y. (2011). Comments on pairwise likelihood in time series models. *Statistica Sinica* **21,** 255–278.

Davison, A. and Hinkley, D. (1997). *Bootstrap Methods and their Application.* Cambridge University Press, New York.

Dunsmuir, W. T. M. and Scott, D. J. (2015). The glarma package for observation-driven time series regression of counts. *Journal of Statistical Software* **67,** 1–36.

Durbin, J. and Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika* **84,** 669–684.

Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association* **104,** 1430–1439.

Hay, J. and Pettitt, A. (2001). Bayesian analysis of a time series of counts with covariates: an application to the control of an infectious disease. *Biostatistics* **2,** 433–444.

Jung, R. C., Kukuk, M., and Liesenfeld, R. (2006). Time series of count data: modeling,

estimation and diagnostics. *Computational Statistics & Data Analysis* **51,** 2350–2364.

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society, Series B (Methodological)* **58,** 619–678.

Lee, Y. and Nelder, J. A. (2001a). Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika* **88,** 987–1006.

Lee, Y. and Nelder, J. A. (2001b). Modelling and analysing correlated non-normal data. *Statistical Modelling* **1,** 3–16.

Liboschik, T., Fried, R., Fokianos, K., and Probst, P. (2016). *tscount: Analysis of Count Time Series*. R package version 1.3.0.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., and Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54,** 127–142.

Sørensen, H. (2003). Simulated likelihood approximations for stochastic volatility models. *Scandinavian Journal of Statistics* **30,** 257–276.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, New York.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica* **21,** 5–42.

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

Zhu, F. (2012a). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. *Journal of Mathematical Analysis and Applications* **389,** 58–71.

Zhu, F. (2012b). Modeling time series of counts with COM-Poisson INGARCH models. *Mathematical and Computer Modelling* **56,** 191–203.