

Blockwise empirical likelihood for time series of counts

Rongning Wu^{a,*}, Jiguo Cao^b

^a*Department of Statistics and Computer Information Systems, Baruch College,
The City University of New York, New York, New York 10010, U.S.A.*

^b*Department of Statistics and Actuarial Science, Simon Fraser University,
Burnaby, B.C. Canada V5A 1S6*

Abstract

Time series of counts has a wide variety of applications in real life. Analyzing time series of counts requires accommodations for serial dependence, discreteness, and overdispersion of data. In this paper, we extend blockwise empirical likelihood (Kitamura, 1997) to the analysis of time series of counts under a regression setting. In particular, our contribution is the extension of Kitamura's (1997) method to the analysis of nonstationary time series. Serial dependence among observations is treated nonparametrically using blocking technique; and overdispersion in count data is accommodated by the specification of variance-mean relationship. We establish consistency and asymptotic normality of the maximum blockwise empirical likelihood estimator. Simulation studies show that our method has a good finite sample performance. The method is also illustrated by analyzing two real data sets: monthly counts of poliomyelitis cases in the U.S.A. and daily counts of non-accidental deaths in Toronto, Canada.

Keywords: Autocorrelation, Generalized linear model, Latent process, Nonstationarity, Overdispersion, Regression analysis.

AMS 2010 subject classifications: primary 62M10, secondary 62E20, 62G05.

1. Introduction

The analysis of time series of counts has wide applicability in the real world, and thus is of interest to many researchers. Generalized linear models (GLMs) are a powerful device for analyzing time series of counts in a regression setting provided that serial dependence among observed

*Corresponding author.

Email addresses: rongning.wu@baruch.cuny.edu (Rongning Wu), jiguo_cao@sfu.ca (Jiguo Cao)

counts is suitably treated. For such models, serial dependence is usually modeled through the link function. Various such approaches have been proposed in the literature, which are categorized into two types of models (Cox, 1981): observation-driven models and parameter-driven models. In observation-driven models, serial dependence is introduced by the explicit dependence of current observation made on past outcomes. In contrast, parameter-driven models specify an autocorrelated latent process that evolves independently of the observed process. Both model specifications are also able to accommodate overdispersion in count data—the phenomenon that the variance is larger than the mean. We refer the reader to Davis et al. (1999) for a comprehensive discussion and comparison of the two types of models.

In such a modeling framework, the conditional distribution of observed counts given past outcomes (for observation-driven models) or latent process (for parameter-driven models) is usually assumed to come from the exponential family. When modeling count data, the Poisson distribution is often the natural choice; for example, Poisson log-linear models have been studied by Gourieroux et al. (1984), Zeger (1988), Zeger and Qaqish (1988), and Davis et al. (1999, 2000). Alternatively, other distributions such as negative binomial and binomial are also applicable; see e.g. Davis and Wu (2009).

However, imposing distributional assumption on observed counts is somewhat restrictive. In this paper, taking a quasi-likelihood approach (Wedderburn, 1974; McCullagh and Nelder, 1989), we relax this assumption and specify only the form of the variance-mean relationship to allow for overdispersion. Moreover, since modeling serial dependence is not our primary interest, we treat serial dependence nonparametrically using blocking technique (Künsch, 1989) instead of specifying dependence structure. We then use an empirical likelihood method for the analysis of time series of counts.

Empirical likelihood, a nonparametric inference method originally introduced by Owen (1988, 1990, 1991) for independent data, has been extensively studied in the literature. A good review can be found in, for example, Owen (2001). In order to apply empirical likelihood to time series data, serial dependence among observations needs to be taken into account. It is well known that empirical likelihood has sampling properties similar to those of bootstrap. Hence, strategies widely used in bootstrap literature for dependent data, such as block, sieve and local bootstraps, can be

adapted for empirical likelihood. With these devices, empirical likelihood provides an alternative to conducting statistical inferences for time series data. Monti (1997), using a frequency domain analysis, studied empirical likelihood confidence regions for parameters of stationary time series models. Kitamura (1997) studied empirical likelihood for weakly dependent stationary processes using the blocking technique to preserve the dependence structure. Chuang and Chan (2002) developed empirical likelihood method for (unstable) autoregressive models based on Mykland's (1995) dual likelihood approach. Chan and Ling (2006) proposed an empirical likelihood approach for GARCH models with and without unit roots, and obtained asymptotic properties of the corresponding empirical likelihood ratio statistics. Nordman et al. (2007) proposed a modified blockwise empirical likelihood method for confidence interval estimation of the mean of stationary but potentially long-range dependent processes. Bravo (2009) studied the blockwise generalized empirical likelihood method for non-linear dynamic moment conditions models.

Most of the work in the literature deals with stationary and continuous-valued time series when studying empirical likelihood. In the present paper, we extend Kitamura's (1997) blockwise empirical likelihood method to regression analysis of time series of counts, where observations are nonnegative and integer-valued. Although Kitamura (1997) covered discrete time series, stationarity was assumed, whereas the time series we study is nonstationary due to the nonrandom covariate setting. This paper also sheds more light on applications of empirical likelihood to discrete-valued time series. The rest of the paper is outlined as follows. In Section 2 we provide a brief review of empirical likelihood for independent and identically distributed (i.i.d.) data. Then we apply the method to time series of counts under a regression setting, and propose a maximum blockwise empirical likelihood estimator for covariate coefficients. In Sections 3 we establish asymptotic results for the estimator. Simulation studies are presented in Section 4 to evaluate the finite sample performance of the estimation method, and empirical examples are provided in Section 5. Section 6 concludes the paper, and the Appendix includes technical proofs.

2. Empirical likelihood for time series of counts

Empirical likelihood was initially introduced by Owen (1988). For i.i.d. observations y_1, \dots, y_n from a random variable Y with unknown distribution function F , the empirical likelihood ratio is

defined as

$$R(F) = \frac{L(F)}{L(F_n)} = \prod_{i=1}^n np_i,$$

where $L(F) = \prod_{i=1}^n p_i$ is the nonparametric likelihood with $p_i = dF(y_i)$ being the probability of $Y = y_i$, and $F_n(\cdot) = \frac{1}{n} \sum_{i=1}^n 1_{\{y_i \leq \cdot\}}$ is the empirical distribution function, which maximizes $L(F)$ subject to the constraints $p_i \geq 0$ and $\sum_{i=1}^n p_i = 1$.

Suppose a parameter θ of F , which we are interested in estimating, satisfies an ℓ -dimensional equation $E_F[\mathbf{g}(Y, \theta)] = \mathbf{0}$. Then, its sample version $\sum_{i=1}^n p_i \mathbf{g}(y_i, \theta) = \mathbf{0}$ should be imposed as an additional constraint on p_i 's. To obtain the profile empirical likelihood ratio function

$$\mathcal{R}(\theta) = \sup \left\{ \prod_{i=1}^n np_i : p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \mathbf{g}(y_i, \theta) = \mathbf{0} \right\},$$

the Lagrange multiplier (LM) method may be used. Let

$$\mathcal{L}(\theta) = \sum_{i=1}^n \log p_i + \alpha \left(1 - \sum_{i=1}^n p_i \right) - n\beta' \sum_{i=1}^n p_i \mathbf{g}(y_i, \theta),$$

where α and β are Lagrange multipliers. It can be shown that the maximum is attained when $\alpha = n$ and

$$p_i = \frac{1}{n} \frac{1}{1 + \beta' \mathbf{g}(y_i, \theta)}.$$

Here, being a function of θ , $\beta = \beta(\theta)$ is a solution to

$$\sum_{i=1}^n \frac{\mathbf{g}(y_i, \theta)}{1 + \beta' \mathbf{g}(y_i, \theta)} = \mathbf{0}.$$

Qin and Lawless (1994) showed that a continuous differentiable solution exists provided that the matrix $\sum_{i=1}^n \mathbf{g}(y_i, \theta) \mathbf{g}'(y_i, \theta)$ is positive definite. Therefore, we can write the minus log profile empirical likelihood function as

$$\mathcal{L}_E(\theta) = -\log \mathcal{R}(\theta) = \sum_{i=1}^n \log [1 + \beta'(\theta) \mathbf{g}(y_i, \theta)].$$

The minimizer of $\mathcal{L}_E(\boldsymbol{\theta})$ is called the maximum empirical likelihood estimator of $\boldsymbol{\theta}$.

Now, let y_1, \dots, y_n be observations from a time series of counts $\{Y_t\}$, and suppose that, for each $t = 1, 2, \dots$, \mathbf{x}_t is an observed ℓ -dimensional covariate, which is assumed to be nonrandom. In some cases \mathbf{x}_t may depend on the sample size n and form a triangular array \mathbf{x}_{nt} . Our primary objective is the regression analysis of Y_t on \mathbf{x}_t . We consider the following GLM

$$\log \mu_t(\boldsymbol{\theta}) = \mathbf{x}_t' \boldsymbol{\theta} \quad \text{and} \quad \text{Var}(Y_t) = V(\mu_t(\boldsymbol{\theta})), \quad (1)$$

where $\boldsymbol{\theta}$ is the parameter vector of interest, $\mu_t(\boldsymbol{\theta}) = E(Y_t)$, and $V(\cdot)$ is some function whose form is supposed to be known. We will consider two forms of $V(\cdot)$ that are most widely postulated in the literature, namely $V(z) = \phi z$ and $V(z) = z + \sigma^2 z^2$, where ϕ and σ^2 are unknown scale parameters. The mean and variance of Y_t are not free of the fixed covariate \mathbf{x}_t , and hence the time series $\{Y_t\}$ is in general nonstationary unless there exists only a constant mean term in the regression.

To ensure the validity of statistical inferences with respect to $\boldsymbol{\theta}$ using the empirical likelihood method, we need to appropriately treat serial dependence among the observed counts. This is done by utilizing Künsch's (1989) blocking technique, which was proposed for the purpose of extending the bootstrap method of estimating standard errors to general stationary observations. The technique was also used by Kitamura (1997) for studying empirical likelihood for weakly dependent stationary processes. However, the results obtained by Kitamura (1997) cannot be straightforwardly carried over to the current nonstationary setting.

To estimate $\boldsymbol{\theta}$, first note that (1) suggests the following score equations:

$$\sum_{t=1}^n \frac{y_t - \mu_t}{V(\mu_t)} \frac{\partial \mu_t}{\partial \boldsymbol{\theta}} = \mathbf{0}.$$

For $t = 1, \dots, n$, we define

$$\mathbf{g}(y_t, \boldsymbol{\theta}) = \frac{y_t - \mu_t}{V(\mu_t)} \frac{\partial \mu_t}{\partial \boldsymbol{\theta}}.$$

Let m and d be integers that depend on the sample size n , and we assume that, as $n \rightarrow \infty$, $m = o(n^{1/2-1/\tau})$ where $\tau > 2$ and $d/m \rightarrow c_0$ for some constant $c_0 \in (0, 1]$. Let B_i denote

the block of m consecutive observations $(y_{(i-1)d+1}, \dots, y_{(i-1)d+m})$ for each $i = 1, \dots, q$. Here $q = [(n - m)/d] + 1$, with $[\]$ standing for the integer part, is the total number of blocks. By definition, m is the block width and d is the separation between successive block starting points. For $i = 1, \dots, q$, now let

$$\mathbf{T}_i(\boldsymbol{\theta}) = \phi_m(B_i, \boldsymbol{\theta}) = \frac{1}{m} \sum_{t=1}^m \mathbf{g}(y_{(i-1)d+t}, \boldsymbol{\theta}).$$

Then, $E[\mathbf{g}(Y_t, \boldsymbol{\theta}_0)] = \mathbf{0}$ implies $E[\mathbf{T}_i(\boldsymbol{\theta}_0)] = \mathbf{0}$ for each $i = 1, \dots, q$. For a given sample size n , the dependency of $\{\mathbf{T}_i(\boldsymbol{\theta})\}$ is weaker than that of $\{Y_t\}$. In particular, when $\{Y_t\}$ is strong mixing, $\{\mathbf{T}_i(\boldsymbol{\theta})\}$ is also strong mixing with a smaller mixing coefficient and is asymptotic independent (Politis and Romano, 1992). That is, the blocks can be treated as if they were independent when n is large enough. Thus, we incorporate serial dependence by replacing $\mathbf{g}(y_t, \boldsymbol{\theta})$ with $\mathbf{T}_i(\boldsymbol{\theta})$, and formulate the profile blockwise empirical likelihood ratio function

$$\mathcal{R}_B(\boldsymbol{\theta}) = \sup \left\{ \prod_{i=1}^q p_i : p_i \geq 0, \sum_{i=1}^q p_i = 1, \sum_{i=1}^q p_i \mathbf{T}_i(\boldsymbol{\theta}) = \mathbf{0} \right\}.$$

Analogously, let

$$\mathcal{L}_{BE}(\boldsymbol{\theta}) = \sum_{i=1}^q \log[1 + \boldsymbol{\beta}'(\boldsymbol{\theta}) \mathbf{T}_i(\boldsymbol{\theta})]$$

be the minus log profile blockwise empirical likelihood function, where $\boldsymbol{\beta}$ is a solution to

$$\sum_{i=1}^q \frac{\mathbf{T}_i(\boldsymbol{\theta})}{1 + \boldsymbol{\beta}' \mathbf{T}_i(\boldsymbol{\theta})} = \mathbf{0}.$$

The maximum blockwise empirical likelihood estimator (MBELE) of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_n$, is defined as the minimizer of $\mathcal{L}_{BE}(\boldsymbol{\theta})$, or equivalently the maximizer of $\mathcal{R}_B(\boldsymbol{\theta})$.

3. Main results

Let θ_0 be the true parameter value. We assume that $\{Y_t\}$ is a strongly mixing (α -mixing) process in the sense that

$$\alpha(h) := \sup_n \sup_k \sup_{A \in \mathcal{F}_{n,-\infty}^k, B \in \mathcal{F}_{n,k+h}^\infty} |\mathbf{P}(AB) - \mathbf{P}(A)\mathbf{P}(B)|$$

goes to zero as $h \rightarrow \infty$, where $\mathcal{F}_{n,-\infty}^k$ and $\mathcal{F}_{n,k+h}^\infty$ are σ -fields generated by $\{Y_t, t \leq k\}$ and $\{Y_t, t \geq k+h\}$, respectively, given sample size n and an integer k . In addition, we assume that
 (A) the mixing coefficient $\alpha(h)$ satisfies $\sum_{h=1}^\infty \alpha(h)^{\frac{\lambda}{\lambda+2}} < \infty$ for some constant $\lambda > 0$, and
 (B) the process $\{Y_t\}$ has finite fourth moment.

We also introduce conditions on the covariates \mathbf{x}_t to facilitate asymptotic analysis. The conditions depend on the form of variance function. For the variance function $\text{Var}(Y_t) = \phi\mu_t$, we have

$$\mathbf{g}(y_t, \theta) = \mathbf{x}_t(y_t - \mu_t)/\phi = \mathbf{x}_t[y_t - \exp(\mathbf{x}_t'\theta)]/\phi,$$

and

$$\mathbf{T}_i(\theta) = \frac{1}{\phi m} \sum_{t=1}^m \mathbf{x}_{(i-1)d+t} [y_{(i-1)d+t} - \exp(\mathbf{x}_{(i-1)d+t}'\theta)].$$

We assume that

(C) as $n \rightarrow \infty$,

$$\begin{aligned} -\frac{1}{\phi n} \sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' \exp(\mathbf{x}_t' \theta_0) &\rightarrow \mathbf{\Lambda}, \\ \frac{1}{\phi^2 n} \sum_{i=1}^n \sum_{j=1}^n \mathbf{x}_i \mathbf{x}_j' \text{Cov}(Y_i, Y_j) &\rightarrow \mathbf{\Omega}, \end{aligned}$$

where $\mathbf{\Omega}$ is positive definite. Moving on to the variance function $\text{Var}(Y_t) = \mu_t + \sigma^2 \mu_t^2$, we have for $t = 1, \dots, n$,

$$\mathbf{g}(y_t, \theta) = \frac{\mathbf{x}_t[y_t - \exp(\mathbf{x}_t'\theta)]}{1 + \sigma^2 \exp(\mathbf{x}_t'\theta)},$$

and for $i = 1, \dots, q$,

$$\mathbf{T}_i(\boldsymbol{\theta}) = \frac{1}{m} \sum_{t=1}^m \frac{\mathbf{x}_{(i-1)d+t}}{1 + \sigma^2 \exp(\mathbf{x}'_{(i-1)d+t} \boldsymbol{\theta})} [y_{(i-1)d+t} - \exp(\mathbf{x}'_{(i-1)d+t} \boldsymbol{\theta})].$$

The corresponding conditions on the covariates \mathbf{x}_t are

(C') as $n \rightarrow \infty$,

$$\begin{aligned} -\frac{1}{n} \sum_{t=1}^n \frac{\mathbf{x}_t \mathbf{x}_t' \exp(\mathbf{x}_t' \boldsymbol{\theta}_0)}{1 + \sigma^2 \exp(\mathbf{x}_t' \boldsymbol{\theta}_0)} &\rightarrow \tilde{\boldsymbol{\Lambda}}, \\ \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \frac{\mathbf{x}_i \mathbf{x}_j'}{[1 + \sigma^2 \exp(\mathbf{x}_i' \boldsymbol{\theta}_0)][1 + \sigma^2 \exp(\mathbf{x}_j' \boldsymbol{\theta}_0)]} \text{Cov}(Y_i, Y_j) &\rightarrow \tilde{\boldsymbol{\Omega}}, \end{aligned}$$

where $\tilde{\boldsymbol{\Omega}}$ is positive definite.

These conditions hold for a wide range of covariates; for example, a trend function $\mathbf{x}_t = \mathbf{f}(t/n)$, where $\mathbf{f}(\cdot)$ is a vector-valued continuous function on $[0, 1]$. Using an integral approximation to sums, we obtain that $\boldsymbol{\Lambda} = -\int_0^1 \mathbf{f}(x) \mathbf{f}'(x) \exp[\mathbf{f}'(x) \boldsymbol{\theta}_0] / \phi \, dx$. For further examples and discussion, see Davis et al. (2000) and Davis and Wu (2009).

The asymptotic results of $\hat{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\beta}}_n := \boldsymbol{\beta}(\hat{\boldsymbol{\theta}}_n)$ are stated in the following theorems.

Theorem 1. *If the conditions (A), (B), and (C) are satisfied, then*

$$\begin{pmatrix} \sqrt{nm}^{-1} \hat{\boldsymbol{\beta}}_n \\ \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \end{pmatrix} \xrightarrow{d} N \left(\mathbf{0}, \begin{pmatrix} \mathbf{V}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_\theta \end{pmatrix} \right)$$

where $\mathbf{V}_\theta = (\boldsymbol{\Lambda}' \boldsymbol{\Omega}^{-1} \boldsymbol{\Lambda})^{-1}$ and $\mathbf{V}_\beta = c_0^2 \boldsymbol{\Omega}^{-1} (\mathbf{I} - \boldsymbol{\Lambda} \mathbf{V}_\theta \boldsymbol{\Lambda}' \boldsymbol{\Omega}^{-1})$.

Proof. See Appendix. □

REMARK 1. One virtue of specifying the variance function as $V(z) = \phi z$ is that estimation of the scale parameter ϕ is not required for estimating covariate coefficient vector $\boldsymbol{\theta}$ and even for computing its asymptotic variance matrix. However, a ϕ estimate is required in order to estimate the asymptotic variance of $\hat{\boldsymbol{\beta}}_n$. Since $\text{Var}(Y_t) = \phi \mu_t$ under this specification, we may estimate ϕ by $\hat{\phi} = \sum_{t=1}^n (Y_t - \hat{\mu}_t)^2 / \sum_{t=1}^n \hat{\mu}_t$.

REMARK 2. It may be of interest to also conduct statistical inferences on the scale parameter ϕ . In this regard, we need an additional estimating function for ϕ . In the independent data case, for example, the constraint $\sum_{t=1}^n \{(y_t - \mu_t)^2 / [\phi V(\mu_t)] - 1/\phi\} = 0$ was used in Kolaczyk (1994, Section 4.2). Alternatively, we may construct estimating functions based on Nelder and Pregibon's (1987) extended quasi-likelihood. See also Godambe and Thompson (1989) and Chen and Cui (2003).

Theorem 2. *If the conditions (A), (B), and (C') are satisfied, then*

$$\begin{pmatrix} \sqrt{nm}^{-1} \hat{\beta}_n \\ \sqrt{n}(\hat{\theta}_n - \theta_0) \end{pmatrix} \xrightarrow{d} N \left(\theta, \begin{pmatrix} \tilde{V}_\beta & \mathbf{0} \\ \mathbf{0} & \tilde{V}_\theta \end{pmatrix} \right),$$

where $\tilde{V}_\theta = (\tilde{\Lambda}' \tilde{\Omega}^{-1} \tilde{\Lambda})^{-1}$ and $\tilde{V}_\beta = c_0^2 \tilde{\Omega}^{-1} (\mathbf{I} - \tilde{\Lambda} \tilde{V}_\theta \tilde{\Lambda}' \tilde{\Omega}^{-1})$.

The proof is similar to that of Theorem 1 and hence is omitted.

As in the independent data case, for testing $H_0 : \theta = \theta_0$, we can use the blockwise empirical likelihood ratio statistic $W_{BE}(\theta_0)$ defined by

$$W_{BE}(\theta_0) = 2\mathcal{L}_{BE}(\theta_0) - 2\mathcal{L}_{BE}(\hat{\theta}_n). \quad (2)$$

Theorem 3. *Under the assumptions of Theorem 1, $W_{BE}(\theta_0) \rightarrow \chi_\ell^2$ in distribution when H_0 is true.*

Proof. See Appendix. □

Theorem 4. *Under the assumptions of Theorem 2, $W_{BE}(\theta_0) \rightarrow \chi_\ell^2$ in distribution when H_0 is true.*

The proof is similar to that of Theorem 3 and hence is omitted.

4. Simulation studies

Simulation studies were conducted to evaluate the finite sample performance of the proposed maximum blockwise empirical likelihood estimator. We generated data from two models respectively, which specify $\{Y_t\}$ as a sequence of independent random variables conditional on a latent

process, with Poisson and geometric densities for the conditional distribution. More precisely, the two models are:

Model 1. $Y_t \mid \alpha_t \sim Po(\lambda_t)$ and $\log(\lambda_t) = \mathbf{x}_t' \boldsymbol{\theta} + \alpha_t$.

Model 2. $Y_t \mid \alpha_t \sim Geo(p_t)$ and $\log[(1 - p_t)/p_t] = \mathbf{x}_t' \boldsymbol{\theta} + \alpha_t$.

The latent process $\{\alpha_t\}$ was specified by an AR(1) process: $\alpha_t - \mu_\alpha = \rho(\alpha_{t-1} - \mu_\alpha) + \zeta_t$, where $\{\zeta_t\} \sim \text{i.i.d. } N(0, \varsigma^2)$. In order to satisfy the condition $\log \mu_t(\boldsymbol{\theta}) = \mathbf{x}_t' \boldsymbol{\theta}$, we set $\mu_\alpha = -\sigma_\alpha^2/2$, where $\sigma_\alpha^2 = \varsigma^2/(1 - \rho^2)$ is the variance of α_t . Otherwise, the intercept of the linear predictor is not identifiable. The value of the noise's standard deviation ς was set to 0.1, and three values of ρ , 0.1, 0.5, and 0.8, were considered. The covariate vector \mathbf{x}_t was defined to include a standardized trend and two harmonic function components, namely $\mathbf{x}_t = (1, t/n, \cos(2\pi t/6), \sin(2\pi t/12))'$, while the true coefficient vector was taken to be $\boldsymbol{\theta}_0 = (1, 1, 1, 0.5)'$. For each model, we considered both forms of variance function: $V(z) = \phi z$ and $V(z) = z + \sigma^2 z^2$. In addition, we took several combinations of m and d values to study their effect on estimation.

For each case, we simulated 100 replications and computed the empirical mean, bias, standard deviation (SD), and root mean squared error (RMSE) of the estimates. The estimation technique performed well in all cases overall, and the estimates followed the asymptotic normal law. For succinctness, we only report here partial results, which are primarily based on samples of size 500, to illustrate the effect of a single factor while holding the others identical. Tables 1 to 4 summarize the simulation results for Model 1 with variance function $V(z) = \phi z$. Other results, which displayed a similar pattern, are included in the supplementary file and are also available from the authors upon request.

Table 1 shows the effect of various selections of the block width m . As we can see, although all selected m values yield a satisfactory result, the intermediate values 8 and 5 perform better. This is not always the case, however. The optimal selection of m , which depends on the nature of a study, is as difficult a problem as bandwidth selection in nonparametric smoothing. As a practical guideline, the block width can be selected using K -fold cross validation. To be specific, the data are partitioned into K folds. In each of the K experiments, we use $K - 1$ folds for training and the remaining one for testing. Let $\boldsymbol{\theta}_{(-j)}$ be the parameter estimates from the $K - 1$ folds of data after removing the j^{th} fold. And let $\mathbf{T}_\ell(\boldsymbol{\theta}_{(-j)}) = 1/m \sum_{t=1}^m \mathbf{g}(y_{(\ell-1)d+t}, \boldsymbol{\theta}_{(-j)})$, where $\ell \in \mathcal{A}^j$, the

set of indices such that all $y_{(\ell-1)d+t}$, $t = 1, \dots, m$, belong to the j^{th} fold of data. Then, the block width m is selected by minimizing

$$CV(m) = \frac{1}{K} \sum_{j=1}^K \sum_{\ell \in \mathcal{A}^j} \log \left[1 + \boldsymbol{\beta}'_{(-j)} \mathbf{T}_\ell(\boldsymbol{\theta}_{(-j)}) \right].$$

In our simulation studies $K = 10$ was used, which is a common practice in the statistics literature. For the simulated data, on which Table 1 is based, the selection of m based on ten-fold cross validation agrees with that obtained using the RMSE criterion.

Table 2 shows the effect of different selections of d ; since the block width is 8, the first panel of $d = 8$ implies that there is no overlap among blocks. It is well known that estimation is more efficient when the blocks overlap, with which our results are in agreement: the smaller d , the better estimation. Table 3 shows the effect of sample size. It is quite clear that the estimates perform better as sample size becomes larger. Table 4 shows the effect of the selection of ρ . It is not surprising that there is not much difference in estimation between the cases of $\rho = 0.1$ and $\rho = 0.5$. In both cases, the dependence among the observed data is not so strong that the block width of 5 is sufficient to accommodate the dependence. However, for the case of $\rho = 0.8$ there exists some bias in the coefficient estimates of the intercept and linear trend. This is because, as the dependence among the data becomes stronger, larger block width becomes necessary in order to account for the autocorrelation and ensure valid estimation.

The asymptotic standard deviation of $\hat{\boldsymbol{\theta}}$ can be obtained using Theorem 1. To be specific, firstly we evaluate the first two moments of the observed process Y_t as follows: $\mu_t = \exp(\mathbf{x}'_t \boldsymbol{\theta}_0)$, $Var(Y_t) = \mu_t + \mu_t^2 Var(e^{\alpha_t})$, and $Cov(Y_i, Y_j) = \mu_i \mu_j Cov(e^{\alpha_i}, e^{\alpha_j})$ for $i \neq j$. But, the autocovariance functions of $\{e^{\alpha_t}\}$ are explicitly given in terms of those of $\{\alpha_t\}$, namely $Cov(e^{\alpha_i}, e^{\alpha_j}) = \exp(Cov(\alpha_i, \alpha_j)) - 1$ for all i and j , where $Cov(\alpha_i, \alpha_j) = \rho^{|i-j|} \sigma_\alpha^2$ for the specified AR(1) latent process. Therefore, the first two moments of Y_t are readily computed. Then, the asymptotic

Table 1: The effect of m ; Model 1, $V(z) = \phi z$, $n = 500$, $\rho = 0.5$, and $d = 2$.

$$m = n^{5/12} \approx 13$$

	True	Mean	Bias*10 ²	SD	RMSE
θ_1	1.000	1.000	-0.009	0.050	0.050
θ_2	1.000	0.999	-0.143	0.076	0.076
θ_3	1.000	1.003	0.336	0.039	0.039
θ_4	0.500	0.499	-0.095	0.037	0.037

$$m = n^{3/8} \approx 10$$

	True	Mean	Bias*10 ²	SD	RMSE
θ_1	1.000	1.005	0.489	0.054	0.054
θ_2	1.000	0.996	-0.435	0.078	0.078
θ_3	1.000	0.996	-0.387	0.036	0.036
θ_4	0.500	0.500	-0.027	0.042	0.042

$$m = n^{1/3} \approx 8$$

	True	Mean	Bias*10 ²	SD	RMSE
θ_1	1.000	0.999	-0.122	0.048	0.048
θ_2	1.000	1.002	0.157	0.072	0.072
θ_3	1.000	1.002	0.224	0.030	0.030
θ_4	0.500	0.500	0.041	0.037	0.037

$$m = n^{1/4} \approx 5$$

	True	Mean	Bias*10 ²	SD	RMSE
θ_1	1.000	1.004	0.350	0.044	0.045
θ_2	1.000	0.997	-0.346	0.070	0.070
θ_3	1.000	0.995	-0.498	0.035	0.036
θ_4	0.500	0.500	-0.017	0.037	0.037

$$m = n^{1/6} \approx 3$$

	True	Mean	Bias*10 ²	SD	RMSE
θ_1	1.000	0.991	-0.865	0.050	0.050
θ_2	1.000	1.010	0.963	0.077	0.078
θ_3	1.000	1.007	0.740	0.034	0.034
θ_4	0.500	0.505	0.480	0.041	0.041

variance of $\hat{\boldsymbol{\theta}}$ is derived using the formula

$$\left\{ \left[\sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' \exp(\mathbf{x}_t' \boldsymbol{\theta}_0) \right]' \left[\sum_{i=1}^n \sum_{j=1}^n \mathbf{x}_i \mathbf{x}_j' \text{Cov}(Y_i, Y_j) \right]^{-1} \left[\sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' \exp(\mathbf{x}_t' \boldsymbol{\theta}_0) \right] \right\}^{-1}.$$

The asymptotic standard deviation of $\hat{\boldsymbol{\theta}}$ is $(0.047, 0.067, 0.031, 0.037)'$ for the cases in Tables 1 and 2, which does not depend on the selection of m or d . In Table 1 the empirical standard deviations when $m = 8$ or 5 agree best with the asymptotic ones, which again supports the use of the RMSE and ten-fold cross validation as good criteria for the selection of block width. On the other hand, in Table 2 the empirical and asymptotic standard deviations are in the best agreement when $d = 2$, which further illustrates the higher efficiency of overlapped blocks in the estimation. The asymptotic standard deviations for the cases in Tables 3 and 4 are reported correspondingly in the tables.

5. Applications

We applied blockwise empirical likelihood estimation to two empirical data sets: the polio data, which consist of monthly counts of poliomyelitis cases in the U.S.A. from the year 1970 to 1983 as reported by the Centers for Disease Control, and the daily counts of non-accidental deaths from the year 1987 to 1996 in Toronto, Canada.

5.1. American monthly counts of poliomyelitis cases

The polio data have been analyzed by Zeger (1988), Chan and Ledolter (1995), Kuk and Cheng (1997), Davis et al. (1999, 2000), and Davis and Wu (2009), among other researchers. Of central interest in the data is whether or not the rate of polio infection has been decreasing since 1970. In the literature, the covariates are usually specified by

$$\mathbf{x}_t = (1, s/1000, \cos(2\pi s/12), \sin(2\pi s/12), \cos(2\pi s/6), \cos(2\pi s/6))',$$

where $s = t - 73$ is used to locate the intercept term on January 1976. Zeger (1988) fitted to the data a parameter-driven model with AR(1) latent process and two log-linear models with variance

Table 2: The effect of d ; Model 1, $V(z) = \phi z$, $n = 500$, $\rho = 0.5$, and $m = 8$.

$d = 8$					
	True	Mean	Bias*10 ²	SD	RMSE
θ_1	1.000	1.004	0.432	0.051	0.051
θ_2	1.000	0.989	-1.074	0.081	0.081
θ_3	1.000	1.005	0.461	0.032	0.032
θ_4	0.500	0.504	0.371	0.040	0.041

$d = 4$					
	True	Mean	Bias*10 ²	SD	RMSE
θ_1	1.000	1.000	-0.006	0.051	0.051
θ_2	1.000	1.005	0.474	0.079	0.079
θ_3	1.000	0.999	-0.094	0.032	0.032
θ_4	0.500	0.497	-0.317	0.036	0.036

$d = 2$					
	True	Mean	Bias*10 ²	SD	RMSE
θ_1	1.000	0.999	-0.122	0.048	0.048
θ_2	1.000	1.002	0.157	0.072	0.072
θ_3	1.000	1.002	0.224	0.030	0.030
θ_4	0.500	0.500	0.041	0.037	0.037

Table 3: The effect of sample size n ; Model 1, $V(z) = \phi z$, $\rho = 0.5$, $m = 5$, and $d = 2$.

$n = 200$						
	True	Mean	Bias* 10^2	SD	RMSE	ASD
θ_1	1.000	1.001	0.128	0.072	0.072	0.077
θ_2	1.000	0.986	-1.362	0.107	0.108	0.114
θ_3	1.000	1.003	0.332	0.051	0.051	0.051
θ_4	0.500	0.506	0.566	0.066	0.066	0.060

$n = 500$						
	True	Mean	Bias* 10^2	SD	RMSE	ASD
θ_1	1.000	1.004	0.350	0.044	0.045	0.047
θ_2	1.000	0.997	-0.346	0.070	0.070	0.067
θ_3	1.000	0.995	-0.498	0.035	0.036	0.031
θ_4	0.500	0.500	-0.017	0.037	0.037	0.037

$n = 1000$						
	True	Mean	Bias* 10^2	SD	RMSE	ASD
θ_1	1.000	1.000	-0.001	0.035	0.035	0.033
θ_2	1.000	1.002	0.158	0.055	0.055	0.048
θ_3	1.000	1.000	-0.000	0.023	0.023	0.022
θ_4	0.500	0.502	0.231	0.029	0.029	0.026

Table 4: The effect of ρ ; Model 1, $V(z) = \phi z$, $n = 500$, $m = 5$, and $d = 2$.

$\rho = 0.1$						
	True	Mean	Bias* 10^2	SD	RMSE	ASD
θ_1	1.000	1.004	0.384	0.046	0.046	0.045
θ_2	1.000	0.994	-0.641	0.072	0.072	0.065
θ_3	1.000	1.003	0.347	0.034	0.034	0.031
θ_4	0.500	0.501	0.122	0.037	0.037	0.037

$\rho = 0.5$						
	True	Mean	Bias* 10^2	SD	RMSE	ASD
θ_1	1.000	1.004	0.350	0.044	0.045	0.047
θ_2	1.000	0.997	-0.346	0.070	0.070	0.067
θ_3	1.000	0.995	-0.498	0.035	0.036	0.031
θ_4	0.500	0.500	-0.017	0.037	0.037	0.037

$\rho = 0.8$						
	True	Mean	Bias* 10^2	SD	RMSE	ASD
θ_1	1.000	1.015	1.540	0.058	0.060	0.064
θ_2	1.000	0.980	-2.040	0.098	0.100	0.103
θ_3	1.000	0.997	-0.310	0.032	0.033	0.032
θ_4	0.500	0.496	-0.351	0.040	0.040	0.039

Table 5: Estimates of coefficients in the polio data.

		θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
MBELE 1	est.	0.20	-5.79	-0.13	-0.47	0.12	-0.38
	s.e.	0.226	4.584	0.164	0.173	0.115	0.118
MBELE 2	est.	0.20	-4.88	-0.13	-0.42	0.10	-0.37
	s.e.	0.237	4.701	0.188	0.195	0.151	0.153
Zeger 1	est.	0.15	-4.80	-0.15	-0.53	0.18	-0.43
	s.e.	0.11	1.94	0.13	0.15	0.14	0.14
Zeger 2	est.	0.15	-4.28	-0.14	-0.49	0.18	-0.42
	s.e.	0.10	2.06	0.14	0.15	0.14	0.14

functions $V(z) = \phi z$ and $V(z) = z + \sigma^2 z^2$. Note that the log-linear models ignore the dependence structure among the data. Davis et al. (2000) and more recently Davis and Wu (2009) fitted parameter-driven Poisson and negative binomial models to the counts, respectively, and studied the standard generalized linear model estimation of parameters. Other models have also been proposed by researchers, e.g. the observation-driven model of Davis et al. (1999).

With the same covariates, we applied the blockwise empirical likelihood method to coefficient estimation with both mean-variance specifications. For block width selection, we used ten-fold cross validation introduced in Section 4, resulting in $m = 12$ and $m = 14$ respectively for the succeeding study. Since fully overlapped blocks utilize more efficiently the information on the dependence structure of data, we set $d = 1$. The corresponding coefficient estimates are reported in Table 5. For comparison, the results from Zeger’s (1988) log-linear models are also included in the table.

The standard errors of $\hat{\theta}$ in the 3rd and 5th rows are obtained using Theorems 1 and 2 respectively, assuming that the counts follow a parameter-driven Poisson model with AR(1) errors as in Zeger (1988) and Davis et al. (2000). Under this parametric structure, the autocovariance function of the time series can be easily estimated, and then standard errors are readily computed using the asymptotic variance formulas given in the theorems. Specifically, the observed counts Y_1, \dots, Y_n were modeled by $Y_t | \epsilon_t \sim Po(\epsilon_t \exp(\mathbf{x}_t' \theta))$, where $\{\epsilon_t\}$ is a strictly stationary positive AR(1) process. We estimated $Cov(Y_i, Y_j)$ using the values of $\hat{\sigma}_\epsilon^2 = 0.77$ and $\hat{\rho}_\epsilon(k) = 0.77^k$ for

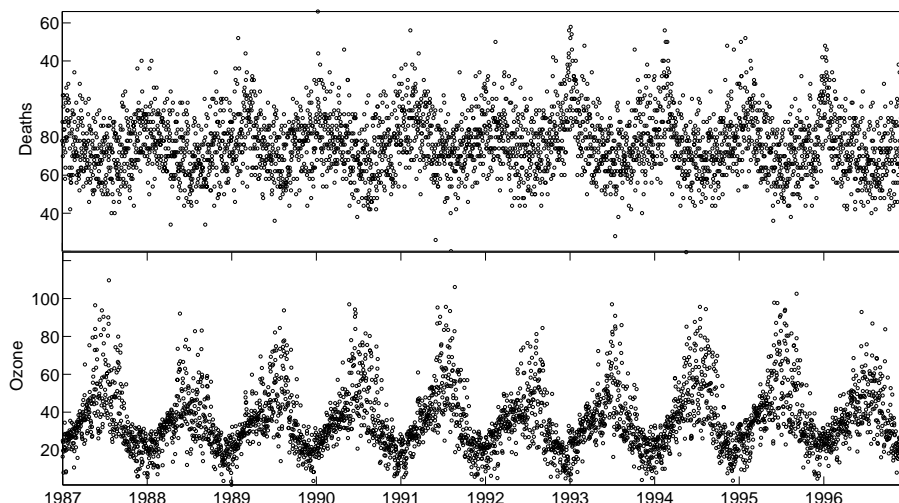


Figure 1: Daily counts of non-accidental deaths from 1987 to 1996 in Toronto. The bottom panel shows the associate daily one-hour-maximum ozone.

$k = 1, \dots, n$ as reported in Zeger (1988), and then plugged the figures into the formula for the estimated standard error.

The results in the table show that the maximum blockwise empirical likelihood estimates of coefficients are comparable to the estimates based on log-linear models with serial dependence ignored. However, the negative trend was not significant using the standard errors based on the blockwise empirical likelihood method, whereas false significance would be obtained if using the standard errors based on log-linear models. That is, the standard error was underestimated without accounting for the serial dependence among the data. Our conclusion is in agreement with that of Davis et al. (2000) and Davis and Wu (2009).

5.2. *Toronto daily death counts*

Inhaling ground-level ozone can result in a number of health effects in the general population. Some of these effects include induction of respiratory symptoms, decrements in lung function and inflammation of airways. Figure 1 displays the daily counts of non-accidental deaths from 1987 to 1996 in Toronto, as well as the daily one-hour-maximum level of ozone. Our objective is to determine whether the amount of daily ozone has any effect on mortality, after taking account of a

Table 6: Estimates of coefficients in the Toronto daily deaths data.

	θ_1	θ_2	θ_3	θ_4
est.	3.85	$6.45 * 10^{-4}$	$9.97 * 10^{-2}$	$3.44 * 10^{-2}$
s.e.	$6.63 * 10^{-3}$	$1.81 * 10^{-4}$	$4.90 * 10^{-3}$	$3.16 * 10^{-3}$

seasonal trend, by fitting the following model

$$\log\{E(\text{Mortality}_t)\} = \theta_1 + \theta_2 \times \text{Ozone}_t + \theta_3 \times \cos(2\pi t/365) + \theta_4 \times \sin(2\pi t/365)$$

where Mortality_t denotes the daily counts of non-accidental deaths in Toronto, which was assumed to follow a Poisson distribution with possible overdispersion. The covariate Ozone_t represents the daily one-hour-maximum level of ozone, and the last two Fourier terms represent the seasonal trend.

The blockwise empirical likelihood method was applied to estimate the coefficients in the above model. The ten-fold cross validation chose the optimal block width $m = 27$. The estimated coefficients are displayed in Table 6. The fitted linear coefficient for the ozone effect is $\theta_2 = 6.45 \times 10^{-4}$, representing about a 0.06% increase in population mortality associated with a one part-per-billion increase in ozone. The parametric bootstrap (Efron and Tibshirani, 1993) was used to obtain the standard errors and 95% confidence intervals for the parameter estimates. The 95% confidence interval for the ozone effect is $[2.60, 9.43] \times 10^{-4}$. This indicates that the ozone may have a significant effect on population mortality, although, as pointed out by a referee, this effect may be caused by some hidden variables (the temperature for instance), since the amount of ozone is highly correlated with the temperature.

6. Concluding remarks

In this paper, taking a quasi-likelihood approach, we have applied empirical likelihood to the analysis of time series of counts under a regression setting. Serial dependence among observations was treated nonparametrically using blocking technique, instead of modeling the dependence structure; and overdispersion in count data was accommodated by the specification of variance-

mean relationship. We have derived consistency and asymptotic normality of the proposed maximum blockwise empirical likelihood estimator of covariate coefficients. To illustrate the method, we simulated data from two parameter-driven models with Poisson and geometric densities, respectively. Two specifications of variance-mean relationship were considered that are most widely postulated in the literature. In all cases, simulation results showed very good finite sample performance of the estimator. We also applied the blockwise empirical likelihood method to two empirical data sets.

In the literature, applications of empirical likelihood to time series are usually confined to the cases of stationary and continuous-valued time series. The present paper shows that empirical likelihood is also promising for analyzing nonstationary and/or discrete-valued time series data.

Appendix

We first provide two lemmas used to establish Theorem 1.

Lemma 1. $\beta(\theta_0)$ converges to zero in probability.

Proof. The function $\beta(\theta)$ solves the equations

$$\frac{1}{q} \sum_{i=1}^q \frac{\mathbf{T}_i(\theta)}{1 + \beta' \mathbf{T}_i(\theta)} = \mathbf{0}. \quad (3)$$

Similar to Owen (1990), we write $\beta(\theta_0) = u\xi$, where $u \geq 0$ and $\|\xi\| = 1$ with $\|\cdot\|$ denoting the Euclidean norm. Then, we have

$$\begin{aligned} 0 &= \frac{1}{q} \xi' \sum_{i=1}^q \frac{\mathbf{T}_i(\theta_0)}{1 + u \xi' \mathbf{T}_i(\theta_0)} \\ &= \frac{1}{q} \xi' \sum_{i=1}^q \mathbf{T}_i(\theta_0) - \frac{u}{q} \xi' \sum_{i=1}^q \frac{\mathbf{T}_i(\theta_0) \mathbf{T}_i'(\theta_0)}{1 + u \xi' \mathbf{T}_i(\theta_0)} \xi \\ &\leq \xi' \left[\frac{1}{q} \sum_{i=1}^q \mathbf{T}_i(\theta_0) \right] - \frac{u}{q} \xi' \sum_{i=1}^q \frac{\mathbf{T}_i(\theta_0) \mathbf{T}_i'(\theta_0)}{1 + u \max_{1 \leq i \leq q} \|\mathbf{T}_i(\theta_0)\|} \xi. \end{aligned}$$

Therefore,

$$\begin{aligned}
& u \left\{ \boldsymbol{\xi}' \left[\frac{m}{q} \sum_{i=1}^q \mathbf{T}_i(\boldsymbol{\theta}_0) \mathbf{T}_i'(\boldsymbol{\theta}_0) \right] \boldsymbol{\xi} - m \boldsymbol{\xi}' \left[\frac{1}{q} \sum_{i=1}^q \mathbf{T}_i(\boldsymbol{\theta}_0) \right] \max_{1 \leq i \leq q} \|\mathbf{T}_i(\boldsymbol{\theta}_0)\| \right\} \\
& \leq m \boldsymbol{\xi}' \left[\frac{1}{q} \sum_{i=1}^q \mathbf{T}_i(\boldsymbol{\theta}_0) \right].
\end{aligned} \tag{4}$$

The term on the right-hand side of (4) has an order of $O_P(mn^{-1/2})$. To see this, note that a mixing process is near-epoch dependent in L_p norm on itself for all $p \geq 1$. Applying a central limit theorem for near-epoch dependent processes (Davidson, 1992), we have

$$\sqrt{mq} \left[\frac{1}{q} \sum_{i=1}^q \mathbf{T}_i(\boldsymbol{\theta}_0) \right] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}/c_0).$$

Since $mq/n \rightarrow 1/c_0$,

$$\sqrt{n} \left[\frac{1}{q} \sum_{i=1}^q \mathbf{T}_i(\boldsymbol{\theta}_0) \right] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}).$$

That is, $\sum_{i=1}^q \mathbf{T}_i(\boldsymbol{\theta}_0)/q = O_P(n^{-1/2})$. In addition, since $m = o(n^{1/2-1/\tau})$, it is easily shown using Lemma 3.2 of Künsch (1989) that

$$\max_{1 \leq i \leq q} \|\mathbf{T}_i(\boldsymbol{\theta}_0)\| = o(n^{1/2}m^{-1})$$

almost surely. Therefore, the second term in the braces on the left-hand side of (4) has an order of $o_P(1)$. Also, it can be shown by Theorem 3 of de Jong (1995) that

$$\frac{m}{q} \sum_{i=1}^q \mathbf{T}_i(\boldsymbol{\theta}_0) \mathbf{T}_i'(\boldsymbol{\theta}_0) \xrightarrow{P} \boldsymbol{\Omega}/c_0,$$

from which it follows that the first term in the braces on the left-hand side has an order of $O_P(1)$.

Putting all the pieces together, we obtain

$$u \leq \frac{O_P(mn^{-1/2})}{O_P(1) - o_P(1)} = O_P(mn^{-1/2}).$$

That is, $\beta(\theta_0) \rightarrow \mathbf{0}$ in probability. \square

Lemma 2. (i) $\hat{\theta}_n$ is consistent, and (ii) $\hat{\beta}_n$ converges to zero in probability.

Proof. (i) By definition of $\mathbf{g}(y_t, \theta)$ and assumptions in Theorem 1, it is easy to check that Assumptions (ii)–(iv) of Kitamura’s (1997) Theorem 1 are satisfied. In addition, we can put $\theta \in [-M, M]^\ell$ for sufficiently large $M > 0$ such that Assumption (i) holds, and then let $M \rightarrow \infty$. Then, the result follows from the same lines as in Kitamura (1997).

(ii) The result follows immediately from Lemma 1, Lemma 2(i), and the continuous mapping theorem. \square

Proof of Theorem 1. To establish the asymptotic normality of $\hat{\theta}_n$ and $\hat{\beta}_n$, we adapt the arguments of Qin and Lawless (1994) for i.i.d. data, and define

$$\begin{aligned}\mathbf{Q}_{1n}(\theta, \beta) &= \frac{1}{q} \sum_{i=1}^q \frac{\mathbf{T}_i(\theta)}{1 + \beta' \mathbf{T}_i(\theta)}, \\ \mathbf{Q}_{2n}(\theta, \beta) &= \frac{1}{q} \sum_{i=1}^q \frac{1}{1 + \beta' \mathbf{T}_i(\theta)} \left(\frac{\partial \mathbf{T}_i(\theta)}{\partial \theta} \right)' \beta.\end{aligned}$$

Then, $\mathbf{Q}_{1n}(\hat{\theta}_n, \hat{\beta}_n) = \mathbf{0}$ and $\mathbf{Q}_{2n}(\hat{\theta}_n, \hat{\beta}_n) = \mathbf{0}$. By Lemma 2, we can expand $\mathbf{Q}_{1n}(\hat{\theta}_n, \hat{\beta}_n)$ and $\mathbf{Q}_{2n}(\hat{\theta}_n, \hat{\beta}_n)$ around $(\theta_0, \mathbf{0})$, obtaining

$$\mathbf{0} = \mathbf{Q}_{1n}(\theta_0, \mathbf{0}) + \frac{\partial \mathbf{Q}_{1n}(\theta_0, \mathbf{0})}{\partial \beta'} \hat{\beta}_n + \frac{\partial \mathbf{Q}_{1n}(\theta_0, \mathbf{0})}{\partial \theta} (\hat{\theta}_n - \theta_0) + \mathbf{R}_{1n}, \quad (5)$$

$$\mathbf{0} = \mathbf{Q}_{2n}(\theta_0, \mathbf{0}) + \frac{\partial \mathbf{Q}_{2n}(\theta_0, \mathbf{0})}{\partial \beta'} \hat{\beta}_n + \frac{\partial \mathbf{Q}_{2n}(\theta_0, \mathbf{0})}{\partial \theta} (\hat{\theta}_n - \theta_0) + \mathbf{R}_{2n}, \quad (6)$$

where \mathbf{R}_{1n} and \mathbf{R}_{2n} are remainders. The first derivatives of \mathbf{Q}_{1n} and \mathbf{Q}_{2n} with respect to θ and β' at $(\theta, \mathbf{0})$ are given by

$$\begin{aligned}\frac{\partial \mathbf{Q}_{1n}(\theta, \mathbf{0})}{\partial \theta} &= \frac{1}{q} \sum_{i=1}^q \frac{\partial \mathbf{T}_i(\theta)}{\partial \theta}, & \frac{\partial \mathbf{Q}_{1n}(\theta, \mathbf{0})}{\partial \beta'} &= -\frac{1}{q} \sum_{i=1}^q \mathbf{T}_i(\theta) \mathbf{T}_i'(\theta), \\ \frac{\partial \mathbf{Q}_{2n}(\theta, \mathbf{0})}{\partial \theta} &= \mathbf{0}, & \frac{\partial \mathbf{Q}_{2n}(\theta, \mathbf{0})}{\partial \beta'} &= \frac{1}{q} \sum_{i=1}^q \left(\frac{\partial \mathbf{T}_i(\theta)}{\partial \theta} \right)'. \end{aligned}$$

Therefore, equations (5) and (6) can be rewritten as

$$\begin{pmatrix} m\partial\mathbf{Q}_{1n}/\partial\beta' & \partial\mathbf{Q}_{1n}/\partial\theta \\ \partial\mathbf{Q}_{2n}/\partial\beta' & \mathbf{0} \end{pmatrix}_{(\theta_0, \mathbf{0})} \begin{pmatrix} m^{-1}\widehat{\beta}_n \\ \widehat{\theta}_n - \theta_0 \end{pmatrix} = \begin{pmatrix} -\mathbf{Q}_{1n}(\theta_0, \mathbf{0}) - \mathbf{R}_{1n} \\ -m^{-1}\mathbf{R}_{2n} \end{pmatrix}. \quad (7)$$

It can be shown that $\mathbf{R}_{1n} = o_P(n^{-1/2})$ and $\mathbf{R}_{2n} = o_P(mn^{-1/2})$. Moreover, let

$$\mathbf{S}_n = \begin{pmatrix} m\partial\mathbf{Q}_{1n}/\partial\beta' & \partial\mathbf{Q}_{1n}/\partial\theta \\ \partial\mathbf{Q}_{2n}/\partial\beta' & \mathbf{0} \end{pmatrix}_{(\theta_0, \mathbf{0})}.$$

It is easily seen that

$$\mathbf{S}_n \xrightarrow{P} \begin{pmatrix} -\Omega/c_0 & \Lambda \\ \Lambda' & \mathbf{0} \end{pmatrix}.$$

Hence, it follows from (7) that

$$\begin{aligned} \begin{pmatrix} \sqrt{n}m^{-1}\widehat{\beta}_n \\ \sqrt{n}(\widehat{\theta}_n - \theta_0) \end{pmatrix} &= \mathbf{S}_n^{-1} \begin{pmatrix} -\sqrt{n}\mathbf{Q}_{1n}(\theta_0, \mathbf{0}) + o_P(1) \\ o_P(1) \end{pmatrix} \\ &= \left[\begin{pmatrix} -c_0\Omega^{-1}(\mathbf{I} - \Lambda\mathbf{V}_\theta\Lambda'\Omega^{-1}) & -\Omega^{-1}\Lambda\mathbf{V}_\theta \\ -\mathbf{V}_\theta\Lambda'\Omega^{-1} & \mathbf{V}_\theta/c_0 \end{pmatrix} + o_P(1) \right] \\ &\quad \times \begin{pmatrix} -\sqrt{n}\mathbf{Q}_{1n}(\theta_0, \mathbf{0}) + o_P(1) \\ o_P(1) \end{pmatrix} \\ &\xrightarrow{d} N \left(\mathbf{0}, \begin{pmatrix} c_0^2\Omega^{-1}(\mathbf{I} - \Lambda\mathbf{V}_\theta\Lambda'\Omega^{-1}) & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_\theta \end{pmatrix} \right), \end{aligned}$$

where the weak convergence results from $\sqrt{n}\mathbf{Q}_{1n}(\theta_0, \mathbf{0}) \xrightarrow{d} N(\mathbf{0}, \Omega)$. That is,

$$\begin{pmatrix} \sqrt{n}m^{-1}\widehat{\beta}_n \\ \sqrt{n}(\widehat{\theta}_n - \theta_0) \end{pmatrix} \xrightarrow{d} N \left(\mathbf{0}, \begin{pmatrix} \mathbf{V}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_\theta \end{pmatrix} \right).$$

This completes the proof. □

Proof of Theorem 3. On the one hand,

$$\begin{aligned}
\mathcal{L}_{BE}(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\beta}}_n) &= \sum_{i=1}^q \log [1 + \widehat{\boldsymbol{\beta}}_n' \mathbf{T}_i(\widehat{\boldsymbol{\theta}}_n)] \\
&= \widehat{\boldsymbol{\beta}}_n' \sum_{i=1}^q \mathbf{T}_i(\boldsymbol{\theta}_0) + \widehat{\boldsymbol{\beta}}_n' \sum_{i=1}^q \mathbf{V}_i(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
&\quad - \frac{1}{2} \widehat{\boldsymbol{\beta}}_n' \sum_{i=1}^q \mathbf{T}_i(\boldsymbol{\theta}_0) \mathbf{T}_i'(\boldsymbol{\theta}_0) \widehat{\boldsymbol{\beta}}_n + o_P(1) \\
&= qm(m^{-1} \widehat{\boldsymbol{\beta}}_n') \mathbf{Q}_{1n}(\boldsymbol{\theta}_0, \mathbf{0}) + qm(m^{-1} \widehat{\boldsymbol{\beta}}_n') \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
&\quad - \frac{qm}{2} (m^{-1} \widehat{\boldsymbol{\beta}}_n') (\boldsymbol{\Omega}/c_0) (m^{-1} \widehat{\boldsymbol{\beta}}_n) + o_P(1) \\
&= c_0 qm \mathbf{Q}_{1n}'(\boldsymbol{\theta}_0, \mathbf{0}) \boldsymbol{\Omega}^{-1} (\mathbf{I} - \boldsymbol{\Lambda} \mathbf{V}_{\boldsymbol{\theta}} \boldsymbol{\Lambda}' \boldsymbol{\Omega}^{-1}) \mathbf{Q}_{1n}(\boldsymbol{\theta}_0, \mathbf{0}) \\
&\quad + c_0 qm \mathbf{Q}_{1n}'(\boldsymbol{\theta}_0, \mathbf{0}) \boldsymbol{\Omega}^{-1} (\mathbf{I} - \boldsymbol{\Lambda} \mathbf{V}_{\boldsymbol{\theta}} \boldsymbol{\Lambda}' \boldsymbol{\Omega}^{-1}) \boldsymbol{\Lambda} (\mathbf{V}_{\boldsymbol{\theta}} \boldsymbol{\Lambda}' \boldsymbol{\Omega}^{-1}) \mathbf{Q}_{1n}(\boldsymbol{\theta}_0, \mathbf{0}) \\
&\quad - \frac{c_0 qm}{2} \mathbf{Q}_{1n}'(\boldsymbol{\theta}_0, \mathbf{0}) \boldsymbol{\Omega}^{-1} (\mathbf{I} - \boldsymbol{\Lambda} \mathbf{V}_{\boldsymbol{\theta}} \boldsymbol{\Lambda}' \boldsymbol{\Omega}^{-1}) \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} (\mathbf{I} - \boldsymbol{\Lambda} \mathbf{V}_{\boldsymbol{\theta}} \boldsymbol{\Lambda}' \boldsymbol{\Omega}^{-1}) \\
&\quad \times \mathbf{Q}_{1n}(\boldsymbol{\theta}_0, \mathbf{0}) + o_P(1) \\
&= \frac{c_0 qm}{2} \mathbf{Q}_{1n}'(\boldsymbol{\theta}_0, \mathbf{0}) (\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1} \boldsymbol{\Lambda} \mathbf{V}_{\boldsymbol{\theta}} \boldsymbol{\Lambda}' \boldsymbol{\Omega}^{-1}) \mathbf{Q}_{1n}(\boldsymbol{\theta}_0, \mathbf{0}) + o_P(1).
\end{aligned}$$

On the other hand, using the fact that $\sum_{i=1}^q \mathbf{T}_i(\boldsymbol{\theta}_0)/[1 + \boldsymbol{\beta}_0' \mathbf{T}_i(\boldsymbol{\theta}_0)] = \mathbf{0}$, a Taylor expansion about $\boldsymbol{\beta}_0$ at zero yields $m^{-1} \boldsymbol{\beta}_0 = (\boldsymbol{\Omega}/c_0)^{-1} \mathbf{Q}_{1n}(\boldsymbol{\theta}_0, \mathbf{0}) + o_P(n^{-1/2})$. Then,

$$\begin{aligned}
\mathcal{L}_{BE}(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) &= \sum_{i=1}^q \log [1 + \boldsymbol{\beta}_0' \mathbf{T}_i(\boldsymbol{\theta}_0)] \\
&= \boldsymbol{\beta}_0' \sum_{i=1}^q \mathbf{T}_i(\boldsymbol{\theta}_0) - \frac{1}{2} \boldsymbol{\beta}_0' \sum_{i=1}^q \mathbf{T}_i(\boldsymbol{\theta}_0) \mathbf{T}_i'(\boldsymbol{\theta}_0) \boldsymbol{\beta}_0 + o_P(1) \\
&= qm(m^{-1} \boldsymbol{\beta}_0') \mathbf{Q}_{1n}(\boldsymbol{\theta}_0, \mathbf{0}) - \frac{qm}{2} (m^{-1} \boldsymbol{\beta}_0') (\boldsymbol{\Omega}/c_0) (m^{-1} \boldsymbol{\beta}_0) + o_P(1) \\
&= c_0 qm \mathbf{Q}_{1n}'(\boldsymbol{\theta}_0, \mathbf{0}) \boldsymbol{\Omega}^{-1} \mathbf{Q}_{1n}(\boldsymbol{\theta}_0, \mathbf{0}) \\
&\quad - \frac{c_0 qm}{2} \mathbf{Q}_{1n}'(\boldsymbol{\theta}_0, \mathbf{0}) \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \mathbf{Q}_{1n}(\boldsymbol{\theta}_0, \mathbf{0}) + o_P(1) \\
&= \frac{c_0 qm}{2} \mathbf{Q}_{1n}'(\boldsymbol{\theta}_0, \mathbf{0}) \boldsymbol{\Omega}^{-1} \mathbf{Q}_{1n}(\boldsymbol{\theta}_0, \mathbf{0}) + o_P(1).
\end{aligned}$$

So,

$$\begin{aligned}
W_{BE}(\boldsymbol{\theta}_0) &= c_0 q m \mathbf{Q}'_{1n}(\boldsymbol{\theta}_0, \mathbf{0}) \boldsymbol{\Omega}^{-1} \boldsymbol{\Lambda} \mathbf{V}_{\boldsymbol{\theta}} \boldsymbol{\Lambda}' \boldsymbol{\Omega}^{-1} \mathbf{Q}_{1n}(\boldsymbol{\theta}_0, \mathbf{0}) + o_P(1) \\
&= \left[(\boldsymbol{\Omega}/c_0)^{-1/2} \sqrt{mq} \mathbf{Q}_{1n}(\boldsymbol{\theta}_0, \mathbf{0}) \right]' \left[\boldsymbol{\Omega}^{-1/2} \boldsymbol{\Lambda} \mathbf{V}_{\boldsymbol{\theta}} \boldsymbol{\Lambda}' \boldsymbol{\Omega}^{-1/2} \right] \\
&\quad \times \left[(\boldsymbol{\Omega}/c_0)^{-1/2} \sqrt{mq} \mathbf{Q}_{1n}(\boldsymbol{\theta}_0, \mathbf{0}) \right] + o_P(1).
\end{aligned}$$

Note that, $(\boldsymbol{\Omega}/c_0)^{-1/2} \sqrt{mq} \mathbf{Q}_{1n}(\boldsymbol{\theta}_0, \mathbf{0}) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{\ell})$; moreover, $\boldsymbol{\Omega}^{-1/2} \boldsymbol{\Lambda} \mathbf{V}_{\boldsymbol{\theta}} \boldsymbol{\Lambda}' \boldsymbol{\Omega}^{-1/2}$ is symmetric and idempotent with trace equal to ℓ . Therefore, $W_{BE}(\boldsymbol{\theta}_0) \xrightarrow{d} \chi_{\ell}^2$, which is the desired result. \square

Acknowledgments

We thank the two referees for their helpful comments, which led to a significant improvement of the paper. Rongning Wu's work was supported in part by The City University of New York PSC-CUNY Research Award 62561-00 40. Jiguo Cao's work was supported in part by a discovery grant from the Natural Science and Engineering Research Council of Canada (NSERC).

References

- [1] Bravo, F. (2009). Blockwise generalized empirical likelihood inference for non-linear dynamic moment conditions models. *Econometrics Journal* 12, 208–231.
- [2] 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.
- [3] Chan, N. H. and Ling, S. (2006). Empirical likelihood for GARCH models. *Econometric Theory* 22, 403–428.
- [4] Chen, S. X. and Cui, H. (2003). An extended empirical likelihood for generalized linear models. *Statistica Sinica* 13, 69–81.
- [5] Chuang, C. S. and Chan, N. H. (2002). Empirical likelihood for autoregressive models, with applications to unstable time series. *Statistica Sinica* 12, 387–407.
- [6] Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics* 8, 93–115.
- [7] Davidson, J. (1992). A central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes. *Econometric Theory* 8, 313–329.
- [8] Davis, R. A., Dunsmuir, W. T. M., and Wang, Y. (1999). Modeling time series of count data. In *Asymptotics, Nonparametrics and Time Series* (Ghosh, S. ed.), pp 63–114. Marcel Dekker, New York.

- [9] Davis, R. A., Dunsmuir, W. T. M., and Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika* 87, 491–505.
- [10] Davis, R. A. and Wu, R. (2009). A negative binomial model for time series of counts. *Biometrika* 96, 735–749.
- [11] de Jong, R. M. (1995). Laws of large numbers for dependent heterogeneous processes. *Econometric Theory* 11, 347–358.
- [12] Efron, B. and Tibshirani, R. J. (1993). *An Introduction to The Bootstrap*. Chapman and Hall, New York.
- [13] Godambe, V. P. and Thompson, M. E. (1989). An extension of quasi-likelihood estimation. *Journal of Statistical Planning and Inference* 22, 137–152.
- [14] Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood methods: applications to Poisson models. *Econometrica* 52, 701–720.
- [15] Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Annals of Statistics* 25, 2084–2102.
- [16] Kolaczyk, E. D. (1994). Empirical likelihood for generalized linear models. *Statistica Sinica* 4, 199–218.
- [17] Kuk, A. Y. C. and Cheng, Y. W. (1997). The Monte Carlo Newton-Raphson algorithm. *Journal of Statistical Computation and Simulation* 59, 233–250.
- [18] Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1241.
- [19] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. 2nd edition. Chapman and Hall, London.
- [20] Monti, A. C. (1997). Empirical likelihood confidence regions in time series models. *Biometrika* 84, 395–405.
- [21] Mykland, P. A. (1995). Dual likelihood. *Annals of Statistics* 23, 396–421.
- [22] Nelder, J. A. and Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika* 74, 221–232.
- [23] Nordman, D. J., Sibbertsen, P., and Lahiri, S. N. (2007). Empirical likelihood confidence intervals for the mean of a long-range dependent process. *Journal of Time Series Analysis* 28, 576–599.
- [24] Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75, 237–249.
- [25] Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics* 18, 90–120.
- [26] Owen, A. B. (1991). Empirical likelihood for linear models. *Annals of Statistics* 19, 1725–1747.
- [27] Owen, A. B. (2001). *Empirical Likelihood*. Chapman and Hall/CRC.
- [28] Politis, D. N. and Romano, J. P. (1992). A general resampling scheme for triangular arrays of α -mixing random variables with application to the problem of spectral density estimation. *Annals of Statistics* 20, 1985–2007.
- [29] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics* 22, 300–325.
- [30] Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* 61, 439–447.

- [31] Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika* 75, 621–629.
- [32] Zeger, S. L. and Qaqish, B. (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics* 44, 1019–1031.