Supplementary materials for this article are available at http://pubs.amstat.org/toc/jasa/104/488.

Poisson Autoregression

Konstantinos FOKIANOS, Anders RAHBEK, and Dag TJØSTHEIM

In this article we consider geometric ergodicity and likelihood-based inference for linear and nonlinear Poisson autoregression. In the linear case, the conditional mean is linked linearly to its past values, as well as to the observed values of the Poisson process. This also applies to the conditional variance, making possible interpretation as an integer-valued generalized autoregressive conditional heteroscedasticity process. In a nonlinear conditional Poisson model, the conditional mean is a nonlinear function of its past values and past observations. As a particular example, we consider an exponential autoregressive Poisson model for time series. Under geometric ergodicity, the maximum likelihood estimators are shown to be asymptotically Gaussian in the linear model. In addition, we provide a consistent estimator of their asymptotic covariance matrix. Our approach to verifying geometric ergodicity proceeds via Markov theory and irreducibility. Finding transparent conditions for proving ergodicity turns out to be a delicate problem in the original model formulation. This problem is circumvented by allowing a perturbation of the model. We show that as the perturbations can be chosen to be arbitrarily small, the differences between the perturbed and nonperturbed versions vanish as far as the asymptotic distribution of the parameter estimates is concerned. This article has supplementary material online.

KEY WORDS: ϕ irreducibility; Asymptotic theory; Count data; Generalized linear model; Geometric ergodicity; Integer generalized autoregressive conditional heteroscedasticity; Likelihood; Noncanonical link function; Observation-driven model; Poisson regression.

1. INTRODUCTION

In this article we study ergodicity and likelihood inference for a specific class of generalized autoregressive conditional heteroscedasticity (GARCH)-type Poisson time series models. Models for time series of counts have been considered by several authors (see, e.g., Kedem and Fokianos 2002, chap. 4, for a comprehensive account). The most popular model among these authors is the log-linear model. If it is assumed that Y_t is conditionally Poisson-distributed with mean λ_t, then most existing models are based on regressing $\log \lambda_t$—the canonical link parameter—on past values of the response and/or covariates. As it has been developed by Fokianos and Kedem (2004), these models fall within the broad class of generalized linear time series models, and their analysis is based on partial likelihood inference. Estimation, diagnostics, model assessment, and forecasting are implemented in a straightforward manner, with the computation carried out in various existing statistical computing environments. Empirical evidence shows that both positive and negative association can be taken into account by a suitable parameterization of the model (see, e.g., Zeger and Qaqish 1988).

An element largely missing in these developments has been the possibility of an AR feedback mechanism in $\{\lambda_t\}$. Such a feedback is a key feature in state-space models, such as the GARCH model for volatility. These models generally are expected to be more parsimonious. In the present work we studied AR models of λ_t, both linear and nonlinear. More specifically, we regressed λ_t on past values of the observed process and past values of λ_t itself. This type of process has been considered by Rydberg and Shephard (2000), Streett (2000), and, more recently, Ferland, Latour, and Oraichi (2006). In this article we propose two classes of models. The first class is a simple linear model that postulates that the conditional mean of the Poisson observed time series is a linear function of its past values and lagged values of the observations. This model can be motivated by the arguments of Rydberg and Shephard (2000), who showed that it is a reasonable approximation for inference about the number of trades within a short time interval. In fact, this is a generalized linear model for time series of counts, but with an identity link—that is, a noncanonical link function. The second class of models generalizes the linear model by imposing a nonlinear structure on both past values of λ_t and lagged values of Y_t. A specific example of this class is the Poisson analog of the so-called "exponential AR" model (see Haggan and Ozaki 1981).

A fundamental problem in the analysis of these models is proving the geometric ergodicity of both the observed $\{Y_t\}$ and the latent process $\{\lambda_t\}$. This problem can be bypassed by slightly perturbing the models and proving that their perturbed versions are geometrically ergodic under simple restrictions on the parameter space. The perturbation idea is a form of regularization—an idea similar in spirit to the analysis of ill-posed problems. We develop likelihood inference in detail for the linear model by showing that the difference between the perturbed and unperturbed versions can be made arbitrarily small provided that the perturbation decreases. We show that the corresponding maximum likelihood estimator (MLE) is asymptotically normal. We apply our results to real and simulated data. Nonlinear models for count time series have not been discussed in the literature, and their development adds an additional tool to the analysis of count time series.

The article is organized as follows. Section 2 introduces the relevant models and discusses the link between their perturbed and unperturbed versions. It demonstrates the joint ergodicity of both the observed and unobserved process for both linear and nonlinear perturbed models. Section 3 develops likelihood

Konstantinos Fokianos is Associate Professor, Department of Mathematics & Statistics, University of Cyprus, 1678 Nicosia, Cyprus (E-mail: *fokianos@ucy.ac.cy*). Anders Rahbek is Professor, Department of Economics, University of Copenhagen and CREATES, Copenhagen, Denmark (E-mail: *rahbek@econ.ku.dk*). Dag Tjøstheim is Professor, Department of Mathematics, University of Bergen, N-5008 Bergen, Norway (E-mail: *Dag.Tjostheim@math.uib.no*). The work of Rahbek was supported by Danish Social Sciences Research Council (project 2114-04-001). The authors thank the associate editor and two reviewers for constructive criticism and helpful remarks; the participants at the CFE08 conference and seminar participants at LSE and Bologna University; Neil Shephard, Andrew Patton, and Theis Lange for their comments; and K. Brännäs for providing the transactions data.

Journal of the American Statistical Association
December 2009, Vol. 104, No. 488, Theory and Methods
DOI: 10.1198/jasa.2009.tm08270

inference for the linear model and discusses some aspects of inference for the nonlinear model, with a special focus on the exponential AR model. Section 3 presents several simulations and some real data analysis. The Appendix provides proofs of some of the theoretical results. (For all of the proofs and further empirical results, see this article's online supplemental materials at *http://pubs.amstat.org/toc/jasa/104/488*.)

2. MODEL SPECIFICATION & ERGODICITY RESULTS

Suppose that $\{Y_t\}$ is a time series of counts and that $\mathcal{F}_t^{Y,\lambda}$ represents the σ-field generated by $\{Y_0, \ldots, Y_t, \lambda_0\}$, that is, $\mathcal{F}_t^{Y,\lambda} = \sigma(Y_s, s \le t, \lambda_0)$, where $\{\lambda_t\}$ is a Poisson intensity process introduced later.

2.1 Linear Model

Consider the model given by

$$Y_t|\mathcal{F}_{t-1}^{Y,\lambda} \sim \text{Poisson}(\lambda_t), \qquad \lambda_t = d + a\lambda_{t-1} + bY_{t-1}, \qquad (1)$$

for $t \ge 1$, where the parameters d, a, b are assumed to be positive. In addition, assume that λ_0 and Y_0 are fixed. For the Poisson distribution, the conditional mean is equal to the conditional variance, that is, $\mathrm{E}[Y_t|\mathcal{F}_{t-1}^{Y,\lambda}] = \mathrm{Var}[Y_t|\mathcal{F}_{t-1}^{Y,\lambda}] = \lambda_t$. Therefore it is tempting to call (1) an INGARCH(1, 1) (integer GARCH) model, because its structure parallels that of the customary GARCH model (see Bollerslev 1986). But the proposed modeling is based on the evolution of the mean of the Poisson, not on its variance.

For technical reasons having to do with the proofs of asymptotic normality of parameter estimates, it is advantageous to rephrase model (1). More specifically, it is desirable to express the sequence of independent Poisson drawings—that is, the first equation of (1)—more explicitly in terms of random variables, like the defining equation in a GARCH model giving the relationship between the observations and the conditional variance. To achieve this, for each time point t, introduce a Poisson process $N_t(\cdot)$ of unit intensity. Then the first equation of (1) can be restated in terms of these Poisson processes by assuming that Y_t is equal to the number of events, say $N_t(\lambda_t)$, of $N_t(\cdot)$ in the time interval $[0, \lambda_t]$. Therefore, let $\{N_t(\cdot), t = 1, 2, \ldots\}$ be a sequence of independent Poisson processes of unit intensity and rephrase (1) as

$$Y_t = N_t(\lambda_t), \qquad \lambda_t = d + a\lambda_{t-1} + bY_{t-1}, \qquad (2)$$

for $t \ge 1$ and with Y_0 and λ_0 fixed. We use this notation throughout, and emphasize that we can always recover (1) by (2).

Model (1) [or its rephrasing (2)] is related to the theory of generalized linear models for time series (see Kedem and Fokianos 2002, ch. 1 and 4). In particular, the random component of the model corresponds to the Poisson distribution that belongs to the exponential family of distributions. The link function is taken to be the identity, whereas the systematic component is the time-dependent random vector $(1, \lambda_{t-1}, Y_{t-1})'$. Thus (2) is a noncanonical link model for time series of counts. Note that even though the vector of time-dependent covariates that influences the evolution of (1) is composed by the unobserved process λ_t, the linear model still belongs to the class of observation driven models as defined by Cox (1981). This is because the unobserved process λ_t can be expressed as a function of past values of the observed process Y_t, after repeated substitution. Observation-driven models for time series of counts have been studied by several authors, including Zeger and Qaqish (1988), Li (1994), and, more recently, Brumback et al. (2000), Fahrmeir and Tutz (2001), Davis, Dunsmuir, and Streett (2003), and Jung, Kukuk, and Liesenfeld (2006). But a log-linear model for the mean of the observed process is usually assumed, the structure of which is determined by past values of the response, moving average terms, and other explanatory variables. With the exception of Davis, Dunsmuir, and Streett (2003), who considered a simple but important case of a log-linear model, none of thos authors addressed the problem of ergodicity of the joint process (Y_t, λ_t).

Second-order properties of model (1) have been studied by Rydberg and Shephard (2000), and Streett (2000) have demonstrated the existence and uniqueness of a stationary distribution. Recently, Ferland, Latour, and Oraichi (2006) considered model (1) in a more general form and showed that the process Y_t is stationary provided that $0 \le a + b < 1$, by using a different technique than that of Streett (2000). In particular, $\mathrm{E}[Y_t] = \mathrm{E}[\lambda_t] \equiv \mu = d/(1 - a - b)$, and its autocovariance function is given by

$$\mathrm{Cov}[Y_t, Y_{t+h}] = \begin{cases} \dfrac{(1-(a+b)^2+b^2)\mu}{1-(a+b)^2}, & h = 0 \\[2ex] \dfrac{b(1-a(a+b))(a+b)^{h-1}\mu}{1-(a+b)^2}, & h \ge 1. \end{cases}$$

In addition, it was shown that all moments of model (1) are finite if and only if $0 \le a + b < 1$. Given that

$$\mathrm{Var}[Y_t] = \mu\left(1 + \frac{b^2}{1-(a+b)^2}\right),$$

we conclude that $\mathrm{Var}[Y_t] \ge \mathrm{E}[Y_t]$ with equality when $b = 0$. Thus including the past values of Y_t in the evolution of λ_t leads to overdispersion, a frequent phenomenon in count time series data.

2.2 Ergodicity of a Perturbed Model

We first note that $\{\lambda_t\}$, as defined by (2), is a Markov chain. Consider the skeleton $\lambda_t = d + a\lambda_{t-1}$ of (2). Here $\lambda^* = d/(1 - a)$ is the solution of $\lambda = d + a\lambda$, that is, a fix point of the mapping $f(\lambda) = d + a\lambda$. Streett (2000) demonstrated that if $0 \le a + b < 1$, then there exists a stationary initial distribution for $\{\lambda_t\}$, by showing that the point λ^* is reachable, where (cf. Meyn and Tweedie 1993, p. 131) a point λ^1 in the state space Λ is reachable if for every neighborhood O of λ^1, $\sum_n \mathrm{P}^n(\lambda, O) > 0$, $\lambda \in \Lambda$. Here $\mathrm{P}^n(\lambda, A) = \mathrm{P}(\lambda_n \in A|\lambda_0 = \lambda)$ is the nth step transition probability of $\{\lambda_t, t \ge 0\}$.

Lemma A.1 in the Appendix establishes that $\{\lambda_t\}$ is open set irreducible on $[\lambda^*, \infty)$, provided that $a < 1$. But in our approach, to show geometric ergodicity, we need ϕ irreducibility, where ϕ is the Lebesgue measure with support $[k, \infty)$, for some $k \ge \lambda^*$. Establishing ϕ irreducibility is equivalent to showing that $\sum_n \mathrm{P}^n(\lambda, A) > 0$ if $\phi(A) > 0$. The problem is that some sets A of positive Lebesgue measure are not open sets (e.g., the set of all irrationals in $[\lambda^*, \infty)$), and Lemma A.1 cannot be used for such nonopen sets. We could add an assumption stating that the chain is not allowed to stay on sets of Lebesgue measure zero, but because it is not clear how restrictive such an assumption is, we chose to avoid that issue by resorting to the perturbed chain (Y_t^m, λ_t^m), defined by

$$Y_t^m = N_t(\lambda_t^m),$$
$$\lambda_t^m = d + a\lambda_{t-1}^m + bY_{t-1}^m + \varepsilon_{t,m}, \tag{3}$$

with λ_0^m, Y_0^m fixed, and

$$\varepsilon_{t,m} = c_m 1(Y_{t-1}^m = 1)U_t, c_m > 0, \qquad c_m \to 0, \text{ as } m \to \infty,$$

where $1(\cdot)$ is the indicator function and $\{U_t\}$ is a sequence of iid uniform random variables on $(0, 1)$ such that the $\{U_t\}$ is independent of $\{N_t(\cdot)\}$. The introduction of $\{U_t\}$ makes it possible to establish irreducibility. Another possibility would be to try an approach to prove ergodicity that does not depend on ϕ irreducibility (see, e.g., Aue, Berkes, and Horváth 2006; Mikosch and Straumann 2006). But the structure of our model is different, and we have not successfully used such an approach. Note that $\{\lambda_t^m\}$ is still a Markov chain.

The perturbation in (3) is a purely auxiliary device to obtain ϕ irreducibility. The U_t's can be thought of as pseudo-observations generated by the uniform law. As we make clear in the subsequent proofs, the perturbation can be introduced in many other ways; for instance, it is sufficient to set $\{U_t\}$ to be an iid sequence of positive random variables with bounded support possessing density on the positive real axis with respect to the Lebesgue measure and finite fourth moment. In addition, the form of the likelihood functions for $\{Y_t\}$ and $\{Y_t^m\}$ in terms of dependence on $\{\lambda_t\}$ will be the same for models (2) and (3). Note that both $\{Y_t\}$ and $\{Y_t^m\}$ can be identified with the observations in the likelihood, but they cannot be identified as stochastic variables because they are generated by different models. Technically speaking, with the introduction of $\{\varepsilon_{t,m}\}$, the process $\{\lambda_t^m\}$ is made into a T chain with a continuous component, where open set irreducibility implies measure-theoretic irreducibility. We have the following results concerning ergodicity of model (3), the proof of which is deferred to the Appendix.

Proposition 2.1. Consider model (3) and suppose that $0 < a + b < 1$. Then the process $\{\lambda_t^m, t \geq 0\}$ is a geometrically ergodic Markov chain with finite moments of order k for an arbitrary k.

The foregoing proposition is useful in obtaining geometric ergodicity of the joint process $(Y_t^m, U_t, \lambda_t^m)$. In fact, it can been shown that the joint process is $V_{(Y,U,\lambda)}$ geometrically ergodic with $V_{Y,U,\lambda}(Y, U, \lambda) = 1 + Y^k + \lambda^k + U^k$ (for a definition, see Meyn and Tweedie 1993, p. 355). In particular, Proposition 2.2 shows the existence of moments of $(Y_t^m, U_t, \lambda_t^m)$, for any k. The inferential results for the $\{Y_t^m\}$ process depend on proving that geometric ergodicity of the $\{\lambda_t^m\}$ series implies geometric ergodicity of the chain $\{(Y_t^m, U_t, \lambda_t^m)\}$. The proof of the following proposition closely follows the arguments of Meitz and Saikkonen (2008) (see also Carrasco and Chen 2002); its proof is given in the Appendix.

Proposition 2.2. Consider model (3) and suppose that the conditions of Proposition 2.1 hold. Then the process $\{(Y_t^m, \lambda_t^m, U_t), t \geq 0\}$ is a $V_{(Y,U,\lambda)}$ geometrically ergodic Markov chain with $V_{Y,U,\lambda}(Y, U, \lambda) = 1 + Y^k + \lambda^k + U^k$.

The following lemma quantifies the difference between (2) and (3) as $m \to \infty$ such that $c_m \to 0$, and shows that essentially the perturbed model can be made arbitrarily close to the unperturbed model. The rephrasing of model (1) as model (2) is very useful in the proof of this lemma.

Lemma 2.1. Suppose that (Y_t, λ_t) and (Y_t^m, λ_t^m) are defined by (2) and (3), respectively. If $0 \leq a + b < 1$, then the following statements hold:

1. $|\mathrm{E}(\lambda_t^m - \lambda_t)| = |\mathrm{E}(Y_t^m - Y_t)| \leq \delta_{1,m}$.
2. $\mathrm{E}(\lambda_t^m - \lambda_t)^2 \leq \delta_{2,m}$.
3. $\mathrm{E}(Y_t^m - Y_t)^2 \leq \delta_{3,m}$.

Here $\delta_{i,m} \to 0$ as $m \to \infty$ for $i = 1, 2, 3$. Furthermore, with m sufficiently large,

$$|\lambda_t^m - \lambda_t| \leq \delta \qquad \text{and} \qquad |Y_t^m - Y_t| \leq \delta, \quad \text{for any } \delta > 0$$

almost surely.

2.3 Nonlinear Models

A simple generalization of the linear model (2) is given by

$$Y_t = N_t(\lambda_t), \qquad \lambda_t = f(\lambda_{t-1}) + b(Y_{t-1}), \tag{4}$$

for $t \geq 1$, where $f(\cdot)$ and $b(\cdot)$ are known functions up to an unknown finite-dimensional parameter vector and $f, b : R^+ \to R^+$. The initial values Y_0 and λ_0 are fixed. It can be seen that (2) is a special case of (4) on defining $f(x) = d + ax$ and $b(x) = bx$, with $d, a, b > 0$, and $x \geq 0$.

There are many examples of nonlinear time series models (see Tong 1990 and Fan and Yao 2003 for comprehensive reviews). Such models, which have not been considered in the literature in the context of generalized linear models for count time series, provide a flexible framework for studying dependent count data. Consider, for example, the so-called "exponential" AR model,

$$Y_t = N_t(\lambda_t),$$
$$\lambda_t = (a + c\exp(-\gamma\lambda_{t-1}^2))\lambda_{t-1} + bY_{t-1}. \tag{5}$$

This model parallels the structure of the traditional exponential AR model (see Haggan and Ozaki 1981). Comparing recursions (4) and (5) clearly shows that $f(x) = (a + c\exp(-\gamma x^2))x$ and $b(x) = bx$, $a, c, b, \gamma > 0$ and $x \geq 0$.

Model (4) is related to a time series following generalized linear models, as described by Kedem and Fokianos (2002). Specifically, for (5), the Poisson assumption guarantees that the random component belongs to the exponential family of distributions, while the link function is equal to the identity, as before. If γ is known, then the systematic component of the model consists of the vector $(\lambda_{t-1}, \lambda_{t-1}\exp(-\gamma\lambda_{t-1}^2), Y_{t-1})'$; otherwise, (5) does not belong to the class of generalized linear models.

Assume that the following conditions hold for $f(\cdot)$ and $b(\cdot)$:

Assumption NL.

1. There exists a unique solution of the equation $\lambda = f(\lambda)$, denoted by $\lambda^\star$.
2. With λ positive real and κ a positive integer, $f(\lambda)$ is increasing in λ for $\lambda > \lambda^\star$ and $b(\kappa)$ is increasing in κ such that $b(\kappa) \geq \beta^\star\kappa$, $\beta^\star > 0$.

3. For some $\alpha_2 > 0$, $f(\lambda_2) - f(\lambda_1) \leq \alpha_2(\lambda_2 - \lambda_1)$, for all $\lambda_1, \lambda_2 \geq \lambda^\star$ with $\lambda_2 \geq \lambda_1$.
4. For some $\beta_2 > 0$ such that $\alpha_2 + \beta_2 < 1$, $b(\kappa_2) - b(\kappa_1) \leq \beta_2(\kappa_2 - \kappa_1)$, $\kappa_2 \geq \kappa_1$.

To prove ϕ irreducibility, we introduce an ϵ-perturbed model,

$$\begin{aligned} Y_t^m &= N_t(\lambda_t^m), \\ \lambda_t^m &= f(\lambda_{t-1}^m) + b(Y_{t-1}^m) + \varepsilon_{t,m}, \qquad t \geq 1, \end{aligned} \tag{6}$$

where $\varepsilon_{t,m}$ is defined as in (3). Then the following results hold (for proofs see this article's online supplemental materials at *http://pubs.amstat.org/toc/jasa/104/488.*)

Proposition 2.3. Consider model (6) and assume that Assumption NL holds. Then the process $\{\lambda_t^m, t \geq 0\}$ is a geometrically ergodic Markov chain with finite moments of order k, for an arbitrary k.

Proposition 2.4. Consider model (6) and assume that the conditions of Proposition 2.3 hold. Then the process $\{(Y_t^m, U_t, \lambda_t^m), t \geq 0\}$ is a $V_{(Y,U,\lambda)}$ geometrically ergodic Markov chain with $V_{Y,U,\lambda}(Y, U, \lambda) = 1 + Y^k + U^k + \lambda^k$.

We conclude this section with the following corollary for the perturbed exponential AR model,

$$\begin{aligned} Y_t^m &= N_t(\lambda_t^m), \\ \lambda_t^m &= \left(a + c\exp(-\gamma(\lambda_{t-1}^m)^2)\right)\lambda_{t-1}^m + bY_{t-1}^m + \varepsilon_{t,m}, \\ &\quad t \geq 1. \end{aligned} \tag{7}$$

Corollary 2.1. Consider the exponential AR model (7). Assume that $0 < a + b < 1$. Then the process $\{(Y_t^m, U_t, \lambda_t^m), t \geq 0\}$ is a $V_{(Y,U,\lambda)}$ geometrically ergodic Markov chain with $V_{Y,U,\lambda}(Y, U, \lambda) = 1 + Y^k + U^k + \lambda^k$.

3. LIKELIHOOD INFERENCE

Denote by $\boldsymbol{\theta}$ the three-dimensional vector of unknown parameters [i.e., $\boldsymbol{\theta} = (d, a, b)'$], and write the true value of the parameter as $\boldsymbol{\theta}_0 = (d_0, a_0, b_0)'$. Then the conditional likelihood function for $\boldsymbol{\theta}$ based on (2), given the starting value λ_0 in terms of the observations $Y_1, \ldots, Y_n$, is given by

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{n} \frac{\exp(-\lambda_t(\boldsymbol{\theta}))\lambda_t^{Y_t}(\boldsymbol{\theta})}{Y_t!}.$$

Here we have used the Poisson assumption, $\lambda_t(\boldsymbol{\theta}) = d + a\lambda_{t-1}(\boldsymbol{\theta}) + bY_{t-1}$ by (2) and $\lambda_t = \lambda_t(\boldsymbol{\theta}_0)$. Thus the log-likelihood function is given, up to a constant, by

$$l(\boldsymbol{\theta}) = \sum_{t=1}^{n} l_t(\boldsymbol{\theta}) = \sum_{t=1}^{n} (Y_t \log \lambda_t(\boldsymbol{\theta}) - \lambda_t(\boldsymbol{\theta})), \tag{8}$$

and the score function is defined by

$$\begin{aligned} \mathbf{S}_n(\boldsymbol{\theta}) &= \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ &= \sum_{t=1}^{n} \left(\frac{Y_t}{\lambda_t(\boldsymbol{\theta})} - 1\right) \frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \end{aligned} \tag{9}$$

where $\partial\lambda_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is a three-dimensional vector with components given by

$$\begin{aligned} &\frac{\partial \lambda_t}{\partial d} = 1 + a\frac{\partial \lambda_{t-1}}{\partial d}, \qquad \frac{\partial \lambda_t}{\partial a} = \lambda_{t-1} + a\frac{\partial \lambda_{t-1}}{\partial a}, \\ &\frac{\partial \lambda_t}{\partial b} = Y_{t-1} + a\frac{\partial \lambda_{t-1}}{\partial b}. \end{aligned} \tag{10}$$

The solution of the equation $\mathbf{S}_n(\boldsymbol{\theta}) = 0$, if it exists, yields the conditional MLE of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}$. Furthermore, the Hessian matrix for model (2) is obtained by further differentiation of the score equations (9),

$$\begin{aligned} \mathbf{H}_n(\boldsymbol{\theta}) &= -\sum_{t=1}^{n} \frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}'} \\ &= \sum_{t=1}^{n} \frac{Y_t}{\lambda_t^2(\boldsymbol{\theta})} \left(\frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)' \\ &\quad - \sum_{t=1}^{n} \left(\frac{Y_t}{\lambda_t(\boldsymbol{\theta})} - 1\right) \frac{\partial^2 \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}'}. \end{aligned} \tag{11}$$

The assumptions of Proposition 2.2 guarantee geometric ergodicity of the perturbed model (Y_t^m, λ_t^m). In addition, Lemma 2.1 shows that λ_t^m approaches λ_t, for large m. Thus it is rather natural to use the ergodic properties of the perturbed process (Y_t^m, λ_t^m) to study the asymptotic properties of the MLEs for process (2). Toward this end, we define the counterparts of expressions (8)–(11) for model (3).

The likelihood function, say L^m, including the pseudo-observations $U_1, U_2, \ldots, U_n$, is given by

$$L^m(\boldsymbol{\theta}) = \prod_{t=1}^{n} \frac{\exp(-\lambda_t^m(\boldsymbol{\theta}))(\lambda_t^m(\boldsymbol{\theta}))^{Y_t^m}}{Y_t^m!} \prod_{t=1}^{n} f_u(U_t),$$

by the Poisson assumption and the asserted independence of U_t from $(Y_{t-1}^m, \lambda_{t-1}^m)$. Here $f_u(\cdot)$ denotes the uniform density and $\lambda_t^m(\boldsymbol{\theta}) = d + a\lambda_{t-1}^m(\boldsymbol{\theta}) + bY_{t-1}^m + \varepsilon_{t,m}$ as given by (3). We note that $L(\boldsymbol{\theta})$ and $L^m(\boldsymbol{\theta})$ have identical forms, except that (Y_t, λ_t) is replaced by (Y_t^m, λ_t^m). Therefore, $\mathbf{S}_n^m(\boldsymbol{\theta})$ and $\mathbf{H}_n^m(\boldsymbol{\theta})$ have the same form as (9) and (11), with recursions defined by (10), but with (Y_t, λ_t) replaced by (Y_t^m, λ_t^m). The solution of the equation $\mathbf{S}_n^m(\boldsymbol{\theta}) = 0$ is denoted by $\hat{\boldsymbol{\theta}}^m$. (See this article's online supplemental materials at *http://pubs.amstat.org/toc/jasa/104/488* for a detailed exposition.)

3.1 Asymptotic Theory

To study the asymptotic properties of the MLE $\hat{\boldsymbol{\theta}}$ for the linear model (2), we derive and use the asymptotic properties of the MLE $\hat{\boldsymbol{\theta}}^m$ for the perturbed linear model (3). The main tool for linking $\hat{\boldsymbol{\theta}}$ to $\hat{\boldsymbol{\theta}}^m$ is prop. 6.3.9 of Brockwell and Davis (1991). Accordingly, we first show that $\hat{\boldsymbol{\theta}}^m$ is asymptotically normal, where to prove consistency and asymptotic normality, we take advantage of the fact that the log-likelihood function is three times differentiable, applying lemma 1 of Jensen and Rahbek (2004). We then show that the score function, the information matrix, and the third derivatives of the perturbed likelihood function tend to the corresponding quantities of the unperturbed likelihood function (8), which allows us to use Brockwell and Davis (1991, prop. 6.3.9). To formulate the end result, we introduce the lower and upper values of each component of $\boldsymbol{\theta}$,

$\delta_L < d_0 < \delta_U$, $\alpha_L < a_0 < \alpha_U < 1$, and $\beta_L < b_0 < \beta_U$, and in terms of these define

$$O(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta} | 0 < \delta_L \le d \le \delta_U, \\ 0 < \alpha_L \le a \le \alpha_U < 1 \text{ and} \\ 0 < \beta_L \le b \le \beta_U\}. \tag{12}$$

Then the following theorem regarding the properties of the MLE $\hat{\boldsymbol{\theta}}$ holds.

Theorem 3.1. Consider model (2) and assume that at the true value $\boldsymbol{\theta}_0$, $0 < a_0 + b_0 < 1$. Then there exists a fixed open neighborhood $O = O(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ [see (12)]—such that with probability tending to 1, as $n \to \infty$, the log-likelihood function (8) has a unique maximum point, $\hat{\boldsymbol{\theta}}$. Furthermore, $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically normal,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} \mathcal{N}(0, \mathbf{G}^{-1}),$$

where the matrix $\mathbf{G}$ is as defined in Lemma 3.1. A consistent estimator of $\mathbf{G}$ is given by $\mathbf{G}_n(\hat{\boldsymbol{\theta}})$, where

$$\mathbf{G}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \mathrm{Var}\left[\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big| \mathcal{F}_{t-1}\right] \\ = \sum_{t=1}^{n} \frac{1}{\lambda_t(\boldsymbol{\theta})} \left(\frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) \left(\frac{\partial \lambda_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)'.$$

To prove this theorem, we need a series of results. Lemma 3.1 shows that the limiting conditional information matrix of model (3) tends to another matrix that plays the role of the conditional information matrix for model (2). In Lemmas 3.2–3.4, the conditions (A.1), (A.2), and (A.3) of lemma 1 of Jensen and Rahbek (2004) are verified for the perturbed model (3). In addition, these lemmas show that the score function, the Hessian matrix and the third derivative of the log-likelihood function of the perturbed model tend to their counterparts of the unperturbed model. The proof of Lemma 3.2, which illustrates the methodology, is provided in the Appendix; proofs of the other lemmas are detailed in this article's supplemental materials, at *http://pubs.amstat.org/toc/jasa/104/488.*

Lemma 3.1. Define the matrixes

$$\mathbf{G}^m(\boldsymbol{\theta}) = \mathrm{E}\left(\frac{1}{\lambda_t^m}\left(\frac{\partial \lambda_t^m}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \lambda_t^m}{\partial \boldsymbol{\theta}}\right)'\right) \quad \text{and}$$

$$\mathbf{G}(\boldsymbol{\theta}) = \mathrm{E}\left(\frac{1}{\lambda_t}\left(\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}}\right)'\right).$$

Under the assumptions of Theorem 3.1, the foregoing matrixes evaluated at the true value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ satisfy $\mathbf{G}^m \to \mathbf{G}$, as $m \to \infty$. In addition, $\mathbf{G}^m$ and $\mathbf{G}$ are positive definite.

Lemma 3.2. Under the assumptions of Theorem 3.1, the score functions defined by (9) and its counterpart for the perturbed model (3) and evaluated at the true value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ satisfy the following:

1. $\frac{1}{\sqrt{n}}\mathbf{S}_n^m \xrightarrow{D} \mathbf{S}^m := \mathcal{N}(0, \mathbf{G}^m)$, as $n \to \infty$ for each $m = 1, 2, \ldots$,
2. $\mathbf{S}^m \xrightarrow{D} \mathcal{N}(0, \mathbf{G})$ as $m \to \infty$,
3. $\lim_{m\to\infty} \limsup_{n\to\infty} \mathrm{P}(\|\mathbf{S}_n^m - \mathbf{S}_n\| > \varepsilon\sqrt{n}) = 0$, for any $\varepsilon > 0$.

Lemma 3.3. Under the assumptions of Theorem 3.1, the Hessian matrixes defined by (11) and its counterpart for the perturbed model (3) and evaluated at the the true value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ satisfy the following:

1. $\frac{1}{n}\mathbf{H}_n^m \xrightarrow{P} \mathbf{G}^m$ as $n \to \infty$ for each $m = 1, 2, \ldots$,
2. $\mathbf{G}^m \to \mathbf{G}$, as $m \to \infty$,
3. $\lim_{m\to\infty} \limsup_{n\to\infty} \mathrm{P}(\|\mathbf{H}_n^m - \mathbf{H}_n\| > \varepsilon n) = 0$, for any $\varepsilon > 0$.

Lemma 3.4. With the neighborhood $O(\boldsymbol{\theta}_0)$ defined in (12), under the assumptions of Theorem 3.1, it holds that

$$\max_{i,j,k=1,2,3} \sup_{\boldsymbol{\theta}\in O(\boldsymbol{\theta}_0)} \left| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^3 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} \right| \le M_n := \frac{1}{n} \sum_{t=1}^{n} m_t,$$

where $\boldsymbol{\theta}_i$ for $i = 1, 2, 3$ refers to $\boldsymbol{\theta} = d$, $\boldsymbol{\theta} = a$, and $\boldsymbol{\theta} = b$, respectively. In addition,

$$m_t = C(Y_t\mu_{3t} + \mu_{3t} + Y_t\mu_{2t}\mu_{1t} + Y_t\mu_{1t}^3),$$

$$\mu_{it} = \beta_U \sum_{j=1}^{t-i} k_{j,i}\alpha_U^{j-1} Y_{t-i-j},$$

$$k_{j,1} = j,\ k_{j,2} = j(j+1) \text{ and } k_{j,3} = j(j+1)(j+2).$$

Correspondingly define M_n^m, m_t^m, and μ_{it}^m in terms of Y_t^m. Then

1. $M_n^m \xrightarrow{P} M^m$, as $n \to \infty$ for each $m = 1, 2, \ldots$,
2. $M^m \to M$, as $m \to \infty$, where M is a finite constant,
3. $\lim_{m\to\infty} \limsup_{n\to\infty} \mathrm{P}(|M_n^m - M_n| > \varepsilon n) = 0$, for any $\varepsilon > 0$.

Remark 3.1. Studying the asymptotic properties of the MLE for the nonlinear model (6) proceeds along the same lines as before, but the corresponding analysis is more cumbersome. Consider, for instance, the exponential AR model (5) and put $\boldsymbol{\theta} = (a, c, b, \gamma)'$. Then the recursions for calculating the score are given by

$$\frac{\partial \lambda_t}{\partial a} = \left(1 - 2\gamma c\lambda_{t-1}\exp(-\gamma\lambda_{t-1}^2)\frac{\partial \lambda_{t-1}}{\partial a}\right)\lambda_{t-1} \\ + (a + c\exp(-\gamma\lambda_{t-1}^2))\frac{\partial \lambda_{t-1}}{\partial a},$$

$$\frac{\partial \lambda_t}{\partial c} = \left(1 - 2\gamma c\lambda_{t-1}\frac{\partial \lambda_{t-1}}{\partial c}\right)\exp(-\gamma\lambda_{t-1}^2)\lambda_{t-1} \\ + (a + c\exp(-\gamma\lambda_{t-1}^2))\frac{\partial \lambda_{t-1}}{\partial c},$$

$$\frac{\partial \lambda_t}{\partial b} = a\frac{\partial \lambda_{t-1}}{\partial b} \\ + (1 - 2\gamma\lambda_{t-1}^2)c\exp(-\gamma\lambda_{t-1}^2)\frac{\partial \lambda_{t-1}}{\partial b} + Y_{t-1},$$

$$\frac{\partial \lambda_t}{\partial \gamma} = -c\exp(-\gamma\lambda_{t-1}^2)\lambda_{t-1}^2\left(\lambda_{t-1} + 2\gamma\frac{\partial \lambda_{t-1}}{\partial \gamma}\right) \\ + (a + c\exp(-\gamma\lambda_{t-1}^2))\frac{\partial \lambda_{t-1}}{\partial \gamma}.$$

We remark that when $c = 0$, the parameter γ is not identifiable. Therefore testing the hypothesis that $c = 0$ by means of the likelihood ratio test, for instance, is more involved but can be done using the methodology outlined by Davies (1987). Although we do not study the asymptotic properties of the MLE for the exponential AR model detail, we believe that analogous results can be proved for models (4) and (6) under Assumption NL. In particular, when $\alpha_2 + \beta_2 < 1$ (see NL4), then Lemma 2.1 can be restated in terms of the nonlinear processes.

4. SIMULATION AND DATA ANALYSIS

Here we report a few simulation results to illustrate the asymptotic normality of the MLE for both the linear and nonlinear models. The MLEs were calculated by optimizing the log-likelihood function (8) using a quasi-Newton method (details available from the authors on request). The simple linear model (2) was generated with $(d_0, a_0, b_0) = (0.3, 0.4, 0.5)$ for different sample sizes. For this choice of parameter values, $a_0 + b_0 = 0.9$, which is close to the sum $\hat{a} + \hat{b}$ of the estimates obtained from the real data example given in Section 4.3. Results for different sets of parameter values were similar, provided that the condition $a_0 + b_0 < 1$ was satisfied, and thus are omitted here.

4.1 Simulations for the Linear Model

Ferland, Latour, and Oraichi (2006) showed that (2) has moments up to second-order identical to those of an ARMA(1, 1) process that satisfies the difference equation

$$(Y_t - \mu) - (a+b)(Y_{t-1} - \mu) = e_t - ae_{t-1},$$

where e_t is a white noise process with $\mathrm{E}[e_t] = \mathrm{Var}[e_t] = d/(1 - a - b)$. This observation points to the potential use of the least squares theory for estimating the parameter vector (d, a, b). We compared the least squares estimators with the MLEs. To initiate the algorithm for optimization of the log-likelihood function (8) using the recursions (10), we obtained the starting values for (d, a, b) by the ARIMA(1, 1) fit to the data.

Table 1 compares the results for the MLE and least squares estimator. The table reports the estimates of the parameters obtained by averaging out the results from all simulations. It also reports the ratio of the mean squared error (MSE) of the conditional least squares estimator (CLSE) to the MSE of the MLE (fifth column). The MSE was calculated using the simulation output. In all cases the MSE of the MLE is lower than that of the CLSE. In addition, we used a likelihood ratio test to examine whether the ratio equals 1; in other words, we considered the simulated estimators from both methods as a sample from the bivariate normal with known mean vector and unknown covariance matrix. Testing whether the diagonal elements are equal (i.e., the ratio of the MSE of the CLSE to the MSE of MLE equals 1) shows that the likelihood ratio test yields very small p-values in all cases, indicating that the CLSE has a greater MSE than the MLE. The last three columns of Table 1 give some summary statistics of the sampling distribution of the standardized MLE. When n is large, the asserted asymptotic normality is supported. Note that for $n = 500$, $\hat{d}$ does not approach normality satisfactorily, but the approximation improves for larger sample sizes.

4.2 Simulations for the Exponential Autoregressive Model

Here we report a small simulation study for the exponential AR model (5). To estimate the parameter vector (a, c, b, γ) of (5), we proceeded as follows. We first fitted the linear model (2) to the data to obtain starting values for both a and b, then set the initial value of c equal to some constant. We generated a grid of values for γ and for each of these values fit model (5) with known γ. Finally, to maximize the log-likelihood function over all (a, c, b, γ), we used as a starting value the γ value that yields the maximum log-likelihood from the previous step together with the corresponding coefficients. We performed maximization numerically as described earlier. Table 2 reports our results; note that the sample size should be reasonably large for an adequate approximation of both c and γ (see this article's supplemental materials at *http://pubs.amstat.org/toc/jasa/104/488*, for more details).

4.3 Data Example

To illustrate our methodology, we applied models (2) and (5) to the number of transactions per minute for the stock Ericsson B on July 2, 2002. This is a part of a larger data set that includes all of the transactions of this specific stock for the period July 2–22, 2002. There are 460 available observations conveying approximately 8 hours of transactions. Note that the first-minute and last-minute transactions were not taken into account. Figure 1(a, b) shows both the data and the respective autocorrelation function. Even though the data are counts, the plot of the autocorrelation function reveals high dependence between transactions. Note that the mean number of transactions for these particular data was 9.909, with a sample variance of 32.836. This is a case of overdispersion, as discussed in Section 2.1. To model these data, we set $\lambda_0 = 0$ and $\partial\lambda_0/\partial\boldsymbol{\theta} = 0$ for initialization of the recursions, and considered the linear model (2). Maximization of the log-likelihood function (8) yielded the

Table 1. Results of simulation for model (2) when $(d_0, a_0, b_0) = (0.3, 0.4, 0.5)$

Parameters	Sample size	MLE	CLSE	Efficiency	Skewness	Kurtosis	p-value
d_0	500	0.3271	0.3318	1.3957	0.8644	4.5069	0.0012
a_0		0.3923	0.3922	1.4299	−0.0202	3.1440	0.8296
b_0		0.4971	0.4932	1.4610	−0.0268	2.8811	0.9711
d_0	1000	0.3148	0.3180	1.5651	0.4410	3.3538	0.2820
a_0		0.3954	0.3951	1.4111	0.0726	3.1183	0.8832
b_0		0.4985	0.4965	1.4204	−0.0714	2.9575	0.8375

NOTE: The third and fourth columns report the means of the MLE and CLSE. The fifth column reports the ratio of the MSE of CLSE to the MSE of MLE. The other three columns report sample skewness, sample kurtosis, and p-values of a Kolmogorov–Smirnov test statistics (for testing against the standard normal distribution) for the standardized MLE obtained by the simulation. Results are based on 1000 simulations.

Table 2. Mean of estimators and their sampling standard error (in parentheses) for the exponential AR model (5) when $(a_0, c_0, b_0, \gamma_0) = (0.25, 1, 0.65, \gamma_0)$, where $\gamma_0 \in \{0.5, 1, 1.5\}$

MLEs				
$\hat{a}$	$\hat{c}$	$\hat{b}$	$\hat{\gamma}$	True γ
0.2504	1.1645	0.6430	1.8655	1.50
(0.0537)	(0.5348)	(0.0487)	(1.2573)	
0.2500	1.1190	0.6426	1.1542	1
(0.0521)	(0.4431)	(0.0485)	(0.6919)	
0.2467	1.1025	0.6472	0.5419	0.50
(0.0514)	(0.3813)	(0.0463)	(0.2341)	

NOTE: The sample size is $n = 500$, and the results are based on 500 simulations.

following results:

$$\hat{\lambda}_t = \underset{(0.1628)}{0.5808} + \underset{(0.0264)}{0.7445}\hat{\lambda}_{t-1} + \underset{(0.0167)}{0.1986}Y_{t-1}$$

where the standard errors underneath the estimated parameter were computed using the robust sandwich matrix $\mathbf{H}_n(\hat{\boldsymbol{\theta}}) \times \mathbf{G}_n^{-1}(\hat{\boldsymbol{\theta}})\mathbf{H}_n(\hat{\boldsymbol{\theta}})$, where $\mathbf{G}_n(\boldsymbol{\theta})$ is as defined in Theorem 3.1 and $\mathbf{H}_n(\boldsymbol{\theta})$ is given by (11). It is rather interesting to see that $\hat{a} + \hat{b}$ is close to unity; in fact, an approximate 95% confidence interval for the parameter $a + b$ is given by $(0.9097, 0.9765)$. This is very similar to the unit-root case in AR time series in econometric analyses of most financial data, and is similar to the results reported in the IGARCH literature, in which high persistence in the conditional variance is common. Analogous results were reported by Rydberg and Shephard (2000), who also examined transaction data.

To examine the adequacy of the fit, we considered the so-called Pearson residuals, defined by $e_t = (Y_t - \lambda_t)/\sqrt{\lambda_t}$. Under the correct model, the sequence e_t is a white noise sequence with constant variance (see Kedem and Fokianos 2002, sec. 1.6.3). Substitute λ_t by $\lambda_t(\hat{\boldsymbol{\theta}})$ to obtain e_t. Figure 1(c) shows that the prediction given by $\hat{Y}_t = \lambda_t(\hat{\boldsymbol{\theta}})$ approximates the observed process reasonably well. Figure 1(d) illustrates the whiteness of the Pearson residuals by depicting the cumulative periodogram plot (see Brockwell and Davis 1991, sec. 10.2).

Applying the Poisson exponential AR model (5) yields

$$\hat{\lambda}_t = (\underset{(0.0232)}{0.8303} + \underset{(3.0732)}{7.030}\exp(-\underset{(0.0592)}{0.1675}\hat{\lambda}_{t-1}^2))\hat{\lambda}_{t-1} + \underset{(0.0218)}{0.1551}Y_{t-1},$$

where the standard errors refer to those of $\hat{a}$, $\hat{c}$, $\hat{d}$, and $\hat{\gamma}$, respectively. The starting value for estimating γ was determined

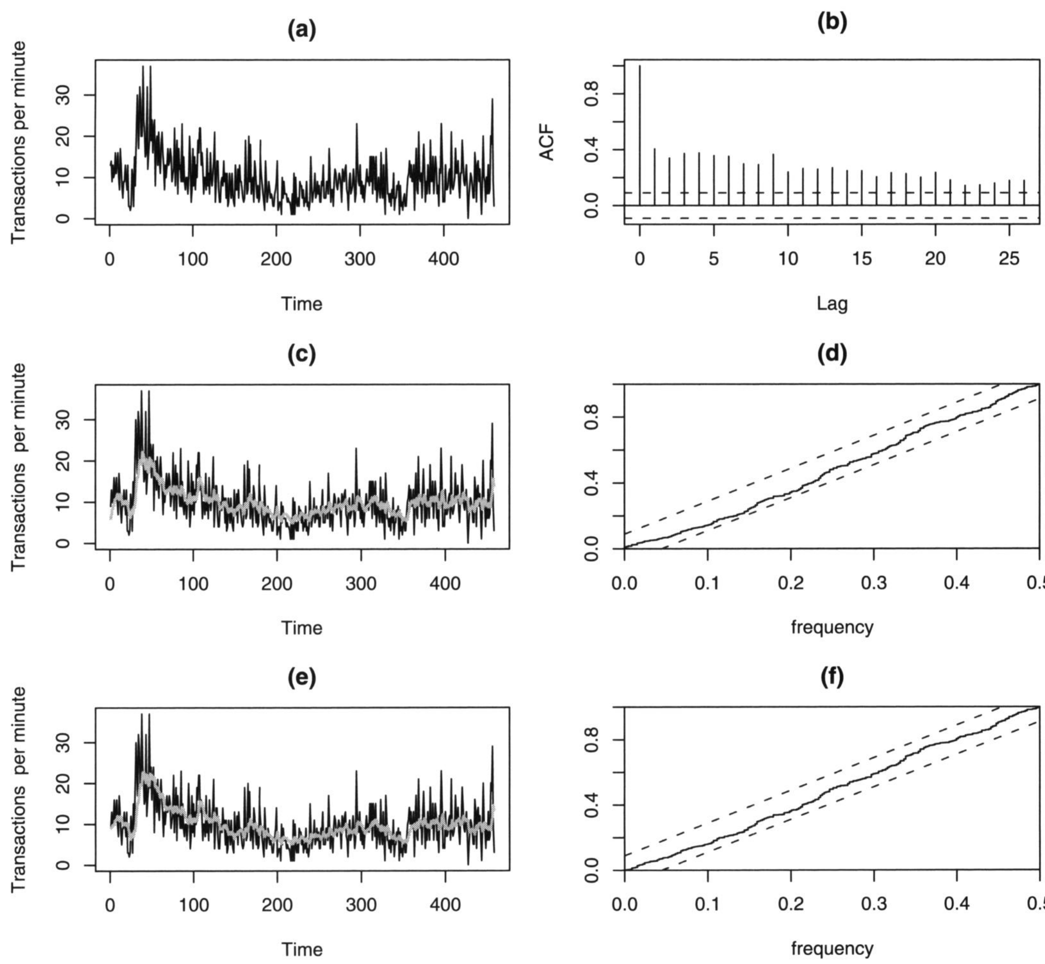

Figure 1. (a) Number of transactions per minute for the Ericsson B stock on July 2, 2002. (b) Autocorrelation function of the transaction data. (c) Observed and predicted (gray) number of transactions per minute using (2). (d) Cumulative periodogram plot of the Pearson residuals from the linear fit. (e) Observed and predicted (gray) numbers of transactions per minute using (5). (f) Cumulative periodogram plot of the raw residuals from the exponential model fit.

by the suggested profiling procedure. More specifically, a grid of values was generated, and the corresponding log-likelihood value was evaluated at these points. For these data, values of the parameter γ are generated from 0.001 to 10 in increments of 0.20; that is, a total of 50 values were obtained. The maximum value of the log-likelihood function was obtained at a value of approximately 0.205 (data not shown). Using this value as a starting point for γ, we fit model (5) to the data. Figure 1(e, f) shows the predicted response (again note that $\hat{Y}_t = \hat{\lambda}_t$) and the cumulative periodogram plot of the Pearson residuals. To compare the models, we calculated the MSE of the Pearson residuals defined by $\sum_{t=1}^{N} e_t^2/(N-p)$, where p is the number of estimated parameters (see Kedem and Fokianos 2002, sec. 1.8). For the linear model, the MSE of the Pearson residuals is approximately 2.367, whereas for the nonlinear model, it is 2.392. This indicates that both of the models yielded similar conclusions. This finding is further supported by the sample correlation coefficient of 0.991 for the Pearson residuals from both models. As a final remark, we note that the numerical results, particularly those for $\hat{d}$, are sensitive to the choice of the initial value λ_0, similar to classic GARCH model estimation.

APPENDIX

Recall that $\lambda^* = d/(1-a)$ is a fix point of the skeleton $\lambda_t = d + a\lambda_{t-1}$ of (2) and (3). We start by proving that $\{\lambda_t\}$, defined by (2) or (3), is open set irreducible on $[\lambda^*, \infty)$. The proof of the following lemma does not require the ϵ perturbation of (3) but is valid in the presence of such a perturbation.

Lemma A.1. Let $\{\lambda_t\}$ be a Markov chain defined by (2) or (3). If $0 < a < 1$, then every point in $[\lambda^*, \infty)$ is reachable.

Proof. To simplify notation, we state the proof for model (2). Consider a point $c \in [\lambda^*, \infty)$. We may assume that $\lambda_1 > c$, because otherwise we can start from $\lambda_j = d + a\lambda_{j-1} + bY_{j-1} > c$. Consider a path such that $Y_1 = \cdots = Y_i = 0$. Then λ_i approaches λ^* as i increases and $a < 1$, which proves that c is reachable if $c = \lambda^*$. This was already proved previously by Streett (2000). The proof is more difficult if $c > \lambda^*$. The intuitive idea is to consider realizations in which $Y_1 = N$ and subsequent $\{Y_i, i > 1\}$ are 0. Then if (c_1, c_2) is an open interval containing c, λ_i will approach this interval from above in successively smaller steps. If N is chosen appropriately, then one of these steps will be contained in (c_1, c_2). Equivalently, we can prove that one of these steps will be arbitrarily close to c. Let $Y_1 = N$ and $Y_2 = 0, Y_3 = 0, \ldots$. Then, for $j \geq 1$

$$\lambda_{1+j} = \lambda_{1+j}(N) = a^{j-1}(a\lambda_1 + bN) + \frac{1-a^j}{1-a}d.$$

Let $N := N_j$ be the least integer such that

$$\lambda_{1+j}(N) > c, \qquad \lambda_{1+j}(N-1) \leq c.$$

Note that we can always choose λ_1 such that $N - 1 > 1$, so that the reasoning is exactly the same for the ϵ-perturbed chain given in (3). We have $\lambda_{1+j}(N) - \lambda_{1+j}(N-1) = a^{j-1}b$. For any $\delta > 0$, because $a < 1$, we can choose a j such that $a^{j-1}b < \delta$, that is, $c \in [\lambda_{1+j}(N-1), \lambda_{1+j}(N)]$, where the width of the interval is less than δ. Because δ is arbitrary, c can be approximated arbitrarily closely, and c is reachable if $c \in [\lambda^*, \infty)$.

To prove Proposition 2.1, we need to show that $\{\lambda_t^m, t \geq 0\}$ is aperiodic and ϕ irreducible. We also show the existence of a small set C and a test function $V(\cdot)$ that satisfies

$$\mathrm{E}[V(\lambda_{t+1}^m)|\lambda_t^m = \lambda] \leq (1-k_1)V(\lambda) + k_2 1(\lambda \in C) \tag{A.1}$$

for some constants k_1, k_2 such that $0 < k_1 < 1$, $0 < k_2 < \infty$. This implies that the chain $\{\lambda_t^m, t \geq 0\}$ is geometrically ergodic and with a proper choice of V, the kth moment of λ_t^m exists for an arbitrary k (see Meyn and Tweedie 1993). The inequality (A.1) and uniform open set reachability from a compact set $[\lambda^*, K]$, $K > \lambda^*$, can be established for both (2) and (3), but the ϵ perturbation is used to establish the ϕ irreducibility and uniform ϕ reachability required to establish the existence of a small set.

Proof of Proposition 2.1

From Lemma A.1, we have that $\{\lambda_t^m, t \geq 0\}$ as defined by (3) is open set irreducible on $[\lambda^*, \infty)$. Let A be a set in $[k, \infty)$ in the support of ϕ for some $k > \lambda^*$, where ϕ is the Lebesgue measure, such that $\phi(A) > 0$. Let c' be a point in A. Then, using the technique of proof of Lemma A.1, for some j, λ_{j+1}^m will be arbitrarily close to $(c' - d - b)/a$, where $(c' - d - b)/a > \lambda^*$ by choosing k large enough. In particular, j can be chosen so that $|d + a\lambda_{j+1}^m + b - c'| < \epsilon/2$. Therefore, if $D = A \cap B$ with $B = (c' - \delta/2, c' + \delta/2)$ for some small δ and $f_u(\cdot)$ the density of U_t, then the probability of being in A in the next step is

$$P(A) \geq P(D) = \int_D f_u(u)\,du \geq \inf_D f_u(u)\phi(D) > 0,$$

which implies ϕ irreducibility. Note that it follows from this proof that the ϵ perturbation need not be introduced for $Y = 1$, but in fact may be inserted for any Y.

Existence of a small set can be proved by extending and modifying the technique of Lemma A.1. Let C be a compact set, $C = [\lambda^*, K]$ for a finite $K > \lambda^*$. Because $a < 1$ and K is finite, there exists an integer $n = n(\eta)$ such that for given $\eta > 0$, and with a path where $Y_1^m = \cdots = Y_{n-1}^m = 0$, $\lambda_1^m = \lambda$, $|\lambda_n^m - \lambda^*| = a^n|\lambda - \lambda^*| < \eta$ for all $\lambda \in C$. Then with $Y_n^m = N$, $Y_{n+1}^m = 0$, $Y_{n+2}^m = 0, \ldots,$

$$\begin{aligned}\lambda_{n+j}^m &= a^j\lambda_n^m + a^{j-1}bN + \frac{1-a^j}{1-a}d \\ &= a^j(\lambda_n^m - \lambda^*) + a^{j-1}bN + \lambda^*.\end{aligned}$$

Similar to the proof of Lemma A.1, consider an open interval (c_1, c_2) with $c_1 > \lambda^*$, and let N be the least integer such that

$$\mu_{n+j}(N) \equiv a^{j-1}bN + \lambda^* > c_2,$$

where with no loss of generality we may assume that $N > 2$. By choosing j large enough and following the previous arguments, for any $\delta > 0$, there exists an j such that $a^{j-1}b < \delta$. Therefore,

$$c_2 - \delta < \mu_{n+j}(N-1) < c_2.$$

But $\lambda_{n+j}^m(N-1) = \mu_{n+j}(N-1) + a^j(\lambda_n^m - \lambda^*)$, where, by choosing η small enough,

$$c_2 - \delta < \mu_{n+j}(N-1) < \lambda_{n+j}^m(N-1) < c_2,$$

so that for these choices of n, j, N, $\lambda_{n+j}^m(N-1) \in (c_1, c_2)$ for all $\lambda \in C$ and

$$\begin{aligned}\inf_{\lambda \in C} \mathrm{P}^{n+m}(\lambda, (c_1, c_2)) \geq \mathrm{P}(&Y_1^m = 0, \ldots, Y_{n-1}^m = 0, Y_n^m = N, \\ &Y_{n+1}^m = 0, \ldots, Y_{n+j-1}^m = 0) > 0.\end{aligned}$$

This means that the interval (c_1, c_2) is uniformly reachable from all $\lambda \in [\lambda^*, K]$, and arguing as in the foregoing proof of ϕ irreducibility, it follows that an n can be found such that

$$\inf_{\lambda \in C} \mathrm{P}^n(\lambda, A) > 0$$

for a set A of positive Lebesgue measure. This implies that the set C is a small set.

We now show that $\{\lambda_t^m, t \geq 0\}$ is aperiodic. Consider the small set $C = [\lambda^*, K]$. Note that $\phi(C) > 0$, and let $\lambda_{t-1}^m = \lambda \in C$. Then

$\lambda_t^m = d + a\lambda + bY_{t-1}^m + \varepsilon_{t,m}$. If $Y_{t-1}^m = 0$, then $\lambda_t^m = d + a\lambda = \lambda^*(1-a) + a\lambda = \lambda^* + a(\lambda - \lambda^*) \geq \lambda^*$, because $a > 0$. In contrast, $\lambda_t^m - \lambda = \lambda^*(1-a) - \lambda(1-a) = -(1-a)(\lambda - \lambda^*) \leq 0$ for $a < 1$. We can conclude that if $0 < a < 1$, then $\lambda \in C$ and $Y_{t-1}^m = 0$ imply that $\lambda_t^m \in C$, which in turn implies $\mathrm{P}(\lambda, C) \geq \mathrm{P}(Y_{t-1}^m = 0 | \lambda_{t-1}^m = \lambda) = \mathrm{P}(N_t(\lambda) = 0) > 0$. Similarly, $\mathrm{P}^2(\lambda, C) \geq \mathrm{P}(Y_t^m = Y_{t-1}^m = 0 | \lambda_{t-1}^m = \lambda) > 0$. It follows that $\{\lambda_t^m, t \geq 0\}$ is aperiodic by Chan (1990, prop. A1.1).

Finally, we prove the existence of a test function $V(\cdot)$ such that (A.1) holds. Consider $V(x) = 1 + x^k$. Then

$$\begin{aligned}
\mathrm{E}[V(\lambda_t^m)|\lambda_{t-1}^m = \lambda] &= \mathrm{E}\big[(1 + (\lambda_t^m)^k)|\lambda_{t-1}^m = \lambda\big] \\
&= 1 + \mathrm{E}[(d + a\lambda + bY_{t-1}^m + \varepsilon_{t,m})^k|\lambda_{t-1}^m = \lambda] \\
&= 1 + \sum_{i=0}^{k} \binom{k}{i} (a\lambda)^i (b\lambda)^{k-i} + \sum_{j=0}^{k-1} c_j \lambda^j \\
&= 1 + (a+b)^k \lambda^k + \sum_{j=0}^{k-1} c_j \lambda^j
\end{aligned}$$

for some constants c_j depending on a, b, d, and ϵ. Consider the small set $C = [\lambda^*, K]$ and write

$$1 + (a+b)^k \lambda^k = \left[1 - \frac{\lambda^k[1 - (a+b)^k]}{1 + \lambda^k}\right] \times (1 + \lambda^k)[1(\lambda \in C) + 1(\lambda \in C^c)].$$

For $\lambda \in C^c$, we obtain

$$\sup_{\lambda \in C^c} \left[1 - \frac{\lambda^k[1 - (a+b)^k]}{1 + \lambda^k}\right] \to 1 - [1 - (a+b)^k] = (a+b)^k$$

as K increases. Similarly, by making K large enough, $\sup_{\lambda \in C^c} (\sum_{j=1}^{k-1} c_j \lambda^j / (1 + \lambda^k)) < \delta$, where $0 < \delta < (a+b)^k$. But on C, $1 + (a+b)^k \lambda^k + \sum_{j=1}^{k-1} c_j \lambda^j$ is bounded; therefore, it follows that there exist constants k_1 and k_2 such that $(0 < k_1 < 1, 0 < k_2 < \infty)$

$$\mathrm{E}(V(\lambda_t^m)|\lambda_{t-1}^m = \lambda) \leq (1 - k_1)V(\lambda) + k_2 1(\lambda \in C),$$

which this implies that the chain $\{\lambda_t^m, t \geq 0\}$ is geometrically ergodic and that the kth moment of λ_t^m exists for an arbitrary k.

Proof of Proposition 2.2

We use the method of Meitz and Saikonnen (2008) to show that geometric ergodicity of the $\{\lambda_t^m\}$ process implies geometric ergodicity of the chain $\{(Y_t^m, U_t, \lambda_t^m)\}$, defined by (3). Denote the σ-algebra generated by the past of U_{t+1} and $N_t(\cdot)$ process by $\mathcal{F}_{t-1}$, that is, $\mathcal{F}_t = \sigma(U_{k+1}, N_k, k \leq t)$.

Note that conditional distribution of Y_t^m given $\mathcal{F}_{t-1}$ depends only on λ_t^m. In addition, the conditional distribution function of Y_t^m given $\lambda_t^m = \lambda$ does not depend on t. Furthermore, given the initial state $(Y_0^m, \lambda_0^m, U_1)$, we have $\lambda_1^m = d + a\lambda_0^m + bY_0^m + \varepsilon_{1,m}$, and because the conditional distribution of Y_1^m given $\{Y_0^m, U_1, \lambda_0^m, \lambda_1^m\}$ is Poisson(λ_1^m), conditionally on $\{Y_0^m, U_1, \lambda_0^m, \lambda_1^m\}$, we have $\lambda_2^m = d + a\lambda_1^m + bN_1(\lambda_1^m) + \varepsilon_{2,m}$, where $N_1(\lambda_1^m)$ is Poisson(λ_1^m). Thus $\{\lambda_t^m\}$, considered a component of the trivariate chain $\{(Y_t^m, U_t, \lambda_t^m\}$ from $t \geq 2$, has the same structure as the one-dimensional chain $\{\lambda_t^m\}$, where, conditional on $\mathcal{F}_{t-1}$,

$$\lambda_t^m = d + a\lambda_{t-1}^m + bN_{t-1}(\lambda_{t-1}^m) + \varepsilon_{t,m},$$

with $N_{t-1}(\lambda_{t-1}^m)$ being Poisson(λ_{t-1}^m). It follows that assumption 1 of Meitz and Saikonnen (2008) is fulfilled. Therefore, proposition 1 of Meitz and Saikonnen (2008) shows that $\{Y_t^m, U_t, \lambda_t^m\}$ inherits the geometric ergodicity of $\{\lambda_t^m\}$. Moreover, because we can take $\mathrm{E}_\mu[1 + (d + \lambda_0^m + Y_0^m + \varepsilon_{1,m})^k] < \infty$, where μ is the distribution of the initial value $(Y_0^m, U_1, \lambda_0^m)$, and because for some constant $C(k)$,

$$\begin{aligned}
&\mathrm{E}[1 + (\lambda_{t-1}^m)^k + (Y_{t-1}^m)^k + \varepsilon_{t,m}^k | \lambda_{t-1}^m = \lambda] \\
&\quad = C(k) + \lambda^k + \mathrm{E}[Y_{t-1}^k | \lambda_{t-1} = \lambda] \\
&\quad = C(k) + \lambda^k + \lambda^k + \sum_{i=1}^{k-1} c_i \lambda^i \leq C'(k)(1 + \lambda^k),
\end{aligned}$$

it follows from theorem 2 of Meitz and Saikonnen (2008) (see also prop. 2) that $\{(Y_t^m, U_t, \lambda_t^m)\}$ is $V_{(Y,U,\lambda)}$ geometrically ergodic with $V_{(Y,U,\lambda)}(Y, U, \lambda) = 1 + Y^k + U^k + \lambda^k$.

Proof of Lemma 2.1

From the defining equations (1) and (2), it follows that

$$\lambda_t^m - \lambda_t = a(\lambda_{t-1}^m - \lambda_{t-1}) + b(Y_{t-1}^m - Y_{t-1}) + \varepsilon_{t,m}. \tag{A.2}$$

By taking conditional expectation and using the properties of the Poisson process, we obtain that $\mathrm{E}(\lambda_t^m - \lambda_t) = \sum_{i=0}^{t-1} (a+b)^i \mathrm{E}(\varepsilon_{t-i,m})$. Because $(a+b) < 1$ and $|\mathrm{E}(\varepsilon_{t,m})| \leq c_m$, with $c_m \to 0$ as $m \to \infty$,

$$|\mathrm{E}(\lambda_t^m - \lambda_t)| \leq \frac{c_m}{1 - (a+b)} := \delta_{1,m},$$

which proves the first assertion.

Next consider the second statement. By using (A.2) again,

$$\begin{aligned}
\mathrm{E}(\lambda_t^m - \lambda_t)^2 &= a^2 \mathrm{E}(\lambda_{t-1}^m - \lambda_{t-1})^2 + b^2 \mathrm{E}(Y_{t-1}^m - Y_{t-1})^2 \\
&\quad + 2ab \mathrm{E}(Y_{t-1}^m - Y_{t-1})(\lambda_{t-1}^m - \lambda_{t-1}) \\
&\quad + \mathrm{E}(\varepsilon_{t,m}^2) + 2a\mathrm{E}[(\lambda_{t-1}^m - \lambda_{t-1})\varepsilon_{t,m}] \\
&\quad + 2b\mathrm{E}[(Y_{t-1}^m - Y_{t-1})\varepsilon_{t,m}].
\end{aligned}$$

But for $\lambda_t^m \geq \lambda_t$,

$$\begin{aligned}
&\mathrm{E}((Y_t^m - Y_t)(\lambda_t^m - \lambda_t)) \\
&\quad = \mathrm{E}\big[\mathrm{E}((Y_t^m - Y_t)(\lambda_t^m - \lambda_t))|\mathcal{F}_{t-1}\big] \\
&\quad = \mathrm{E}\big[(\lambda_t^m - \lambda_t)\big(\mathrm{E}(N_t[\lambda_t, \lambda_t^m])\big)\big] = \mathrm{E}(\lambda_t^m - \lambda_t)^2,
\end{aligned}$$

where $\mathcal{F}_{t-1}$ is the σ-algebra generated by $\{(U_{k+1}, N_k, k \leq t-1\}$ and $N_t[\lambda_t, \lambda_t^m]$ is equal to the number of events between λ_t and λ_t^m for the unit intensity Poisson process N_t. (If $\lambda_t^m < \lambda_t$, we work along the same lines.) Again using the properties of the Poisson process, we find that

$$\begin{aligned}
\mathrm{E}(Y_t^m - Y_t)^2 &\leq \mathrm{E}(\lambda_t^m - \lambda_t)^2 + 2|\mathrm{E}(\lambda_t^m - \lambda_t)| \\
&\leq \mathrm{E}(\lambda_t^m - \lambda_t)^2 + 2\delta_{1,m}.
\end{aligned} \tag{A.3}$$

Finally, with K a positive constant,

$$\mathrm{E}(\varepsilon_{t,m}^2) + 2a\mathrm{E}((\lambda_{t-1}^m - \lambda_{t-1})\varepsilon_{t,m}) + 2b\mathrm{E}((Y_{t-1}^m - Y_{t-1})\varepsilon_{t,m}) \leq Kc_m^2.$$

Therefore, by simple recursion,

$$\mathrm{E}(\lambda_t^m - \lambda_t)^2 \leq (a+b)^2 \mathrm{E}(\lambda_{t-1}^m - \lambda_{t-1})^2 + Kc_m^2 + 2b^2\delta_{1,m} \leq \delta_{2,m},$$

where $\delta_{2,m} \to 0$ as $m \to \infty$. This establishes the second assertion and thus the last statement of the lemma. As for the third statement, this follows by (A.3) and the second assertion of the lemma.

Proof of Lemma 3.2

Equation (9) shows that the score for the perturbed model is given by $\mathbf{S}_n^m(\boldsymbol{\theta}) = \sum_{t=1}^n \partial l_t^m(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, with martingale difference terms defined by

$$\frac{\partial l_t^m}{\partial \boldsymbol{\theta}} = \left(\frac{Y_t^m}{\lambda_t^m} - 1\right)\frac{\partial \lambda_t^m}{\partial \boldsymbol{\theta}} := Z_t^m \frac{\partial \lambda_t^m}{\partial \boldsymbol{\theta}},$$

where $\partial\lambda_t^m/\partial\boldsymbol{\theta}$ is defined analogously to (10). It follows that at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, $\mathrm{E}(\partial l_t^m/\partial\boldsymbol{\theta}|\mathcal{F}_{t-1}) = 0$ and $\mathrm{E}((Z_t^m)^2|\mathcal{F}_{t-1}) = 1/\lambda_t^m$. Furthermore, from (3) and (10),

$$\frac{\partial\lambda_t^m}{\partial a} = \sum_{i=0}^{t-1} a_0^i \lambda_{t-1-i}^m, \qquad \frac{\partial\lambda_t^m}{\partial d} = \frac{1-a_0^t}{1-a_0},$$

and

$$\frac{\partial\lambda_t^m}{\partial b} = \sum_{i=0}^{t-1} a_0^i Y_{t-1-i}^m.$$

Observe that as $a_0, b_0 < 1$, $\mathrm{E}(Y_t^m)^2 < \infty$, and $\mathrm{E}(\lambda_t^m)^2 < \infty$, $(\partial\lambda_t^m/\partial d)^2$, $\mathrm{E}(\partial\lambda_t^m/\partial a)^2$, and $\mathrm{E}(\partial\lambda_t^m/\partial b)^2$ are all finite. Also $1/\lambda t^m \le 1/(d_0 - \eta_1)$ for any small $\eta_1 \ge 0$, where, for m large enough, $\varepsilon_{t,m} = c_m 1(Y_{t-1}^m = 1)U_t \in [-\eta_1, \eta_1]$. From Hölders inequality, we conclude that $\mathrm{E}\|\partial l_t^m/\partial\boldsymbol{\theta}\| < \infty$. Thus $\partial l_t^m/\partial\boldsymbol{\theta}$ is a martingale difference sequence with respect to $\mathcal{F}_t$ and an application of the central limit theorem (Hall and Heyde 1980, cor. 3.1) gives that $n^{-1/2}\mathbf{S}_n^m$ is asymptotically Gaussian with covariance given by the limit,

$$\frac{1}{n}\sum_{t=1}^n \mathrm{E}\left(Z_t^2\left(\frac{\partial\lambda_t^m}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\lambda_t^m}{\partial\boldsymbol{\theta}}\right)' \Bigm| \mathcal{F}_{t-1}\right) \xrightarrow{P} \mathbf{G}^m,$$

by the law of large numbers for geometrically ergodic processes (Jensen and Rahbek 2007). Then the conditional Lindeberg condition holds follows by noting that

$$\frac{1}{n}\sum_{t=1}^n \mathrm{E}\big(\|\partial l_t^m/\partial\boldsymbol{\theta}\|^2 I(\|\partial l_t^m/\partial\boldsymbol{\theta}\| > \sqrt{n}\delta)|\mathcal{F}_{t-1}\big) \le \frac{1}{n^2\delta^2}\sum_{t=1}^n \mathrm{E}(\|\partial l_t^m/\partial\boldsymbol{\theta}\|^4|\mathcal{F}_{t-1}) \to 0,$$

because $\mathrm{E}\|\partial l_t^m/\partial\boldsymbol{\theta}\|^4 < \infty$. This proves the first assertion of the lemma. The second assertion follows by Lemma 3.1.

Now we consider the last conclusion of the lemma. Define Z_t in an analogous way as Z_t^m. Then

$$\begin{aligned}
\frac{1}{\sqrt{n}}(\mathbf{S}_n^m - \mathbf{S}_n) &= \frac{1}{\sqrt{n}}\sum_{t=1}^n\left(\frac{\partial l_t^m}{\partial\boldsymbol{\theta}} - \frac{\partial l_t}{\partial\boldsymbol{\theta}}\right) \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^n\left(Z_t^m\frac{\partial\lambda_t^m}{\partial\boldsymbol{\theta}} - Z_t\frac{\partial\lambda_t}{\partial\boldsymbol{\theta}}\right) \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^n\left[Z_t^m\left(\frac{\partial\lambda_t^m}{\partial\boldsymbol{\theta}} - \frac{\partial\lambda_t}{\partial\boldsymbol{\theta}}\right) + (Z_t^m - Z_t)\frac{\partial\lambda_t}{\partial\boldsymbol{\theta}}\right].
\end{aligned}$$

But for the first summand,

$$\mathrm{P}\left(\left\|\sum_{t=1}^n Z_t^m\left(\frac{\partial\lambda_t^m}{\partial\boldsymbol{\theta}} - \frac{\partial\lambda_t}{\partial\boldsymbol{\theta}}\right)\right\| > \delta\sqrt{n}\right) \le \mathrm{P}\left(\gamma_m\left\|\sum_{t=1}^n Z_t^m\right\| > \delta\sqrt{n}\right) \le \frac{\gamma_{a,m}^2}{\delta^2 n}\sum_{t=1}^n \mathrm{E}\|Z_t^m\|^2 \le C\gamma_{a,m}^2 \to 0,$$

as $m \to \infty$. For the second summand, the same arguments apply because $\mathrm{E}\|\partial\lambda_t/\partial\boldsymbol{\theta}\|^2 < \infty$. In addition, for

$$Z_t^m - Z_t = \frac{\lambda_t(Y_t^m - Y_t) - Y_t(\lambda_t^m - \lambda_t)}{\lambda_t\lambda_t^m},$$

we have $\|Z_t^m - Z_t\| < \gamma_{b,m} \to 0$ as $m \to \infty$. To see this, observe

$$\mathrm{E}\left|\frac{(Y_t^m - Y_t)}{\lambda_t^m}\right| \le \frac{\mathrm{E}|Y_t^m - Y_t|}{(d_0 - \eta_1)} \le C\delta_{1,m}$$

and

$$\mathrm{E}\left|\frac{Y_t(\lambda_t^m - \lambda_t)}{\lambda_t\lambda_t^m}\right| \le \frac{\delta_m \mathrm{E}|Y_t|}{d_0(d_0 - \eta_1)} \le C\delta_m$$

using Lemma 2.1 and the fact that $\mathrm{E}|Y_t| < \infty$.

[Received May 2008. Revised March 2009.]

REFERENCES

Aue, A., Berkes, I., and Horváth, L. (2006), "Strong Approximation for the Sums of Squares of Augmented GARCH Sequences," *Bernoulli*, 12, 583–608.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Data Analysis and Theory* (2nd ed.), New York: Springer.

Brumback, B. A., Ryan, L. M., Schwartz, J. D., Neas, L. M., Stark, P. C., and Burge, H. A. (2000), "Transitional Regression Models With Application to Environmental Time Series," *Journal of the American Statistical Association*, 85, 16–27.

Carrasco, M., and Chen, X. (2002), "Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models," *Econometric Theory*, 18, 1–39.

Chan, K. S. (1990), "Deterministic Stability, Stochastic Stability and Ergodicity," in *Non-Linear Time Series: A Dynamical System Approach*, ed. H. Tong, Oxford: Clarendon Press, Appendix 1.

Cox, D. R. (1981), "Statistical Analysis of Time Series: Some Recent Developments," *Scandinavian Journal of Statistics*, 8, 93–115.

Davies, R. B. (1987), "Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative," *Biometrika*. 74, 33–43.

Davis, R. A., Dunsmuir, W. T. M., and Streett, S. B. (2003), "Observation-Driven Models for Poisson Counts," *Biometrika*, 90, 777–790.

Fahrmeir, L., and Tutz, G. (2001), *Multivariate Statistical Modelling Based on Generalized Linear Models* (2nd ed.), New York: Springer.

Fan, J., and Yao, Q. (2003), *Nonlinear Time Series*, New York: Springer-Verlag.

Ferland, R., Latour, A., and Oraichi, D. (2006), "Integer-Valued GARCH Processes," *Journal of Time Series Analysis*, 27, 923–942.

Fokianos, K., and Kedem, B. (2004), "Partial Likelihood Inference for Time Series Following Generalized Linear Models," *Journal of Time Series Analysis*, 25, 173–197.

Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009), "Poisson Autoregression" (complete version), supplemental materials available at *http://pubs.amstat.org/toc/jasa/104/488*.

Haggan, V., and Ozaki, T. (1981), "Modelling Nonlinear Random Vibrations Using an Amplitude-Dependent Autoregressive Time Series Model," *Biometrika*, 68, 189–196.

Hall, P., and Heyde, C. C. (1980), *Martingale Limit Theory and Its Applications*, New York: Academic Press.

Jensen, S. T., and Rahbek, A. (2004), "Asymptotic Inference for Nonstationary GARCH," *Econometric Theory*, 20, 1203–1226.

——— (2007), "On the Law of Large Numbers for (Geometrically) Ergodic Markov Chains," *Econometric Theory*, 23, 761–766.

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.

Kedem, B., and Fokianos, K. (2002), *Regression Models for Time Series Analysis*, Hoboken, NJ: Wiley.

Li, W. K. (1994), "Time Series Models Based on Generalized Linear Models: Some Further Results," *Biometrics*, 50, 506–511.

Meitz, M., and Saikonnen, P. (2008), "Ergodicity, Mixing and Existence of Moments of a Class of Markov Models With Applications to GARCH and ACD Models," *Econometric Theory*, 24, 1291–1320.

Meyn, S. P., and Tweedie, R. L. (1993), *Markov Chains and Stochastic Stability*, London: Springer.

Mikosch, T., and Straumann, D. (2006), "Stable Limits of Martingale Transforms With Application to the Estimation of GARCH Parameters," *The Annals of Statistics*, 34, 493–522.

Rydberg, T. H., and Shephard, N. (2000), "A Modeling Framework for the Prices and Times of Trades on the New York Stock Exchange," in *Nonlinear and Nonstationary Signal Processing*, eds. W. J. Fitzgerlad, R. L. Smith, A. T. Walden, and P. C. Young, Cambridge: Isaac Newton Institute and Cambridge University Press, pp. 217–246.

Streett, S. (2000), "Some Observation Driven Models for Time Series of Counts," Ph.D. thesis, Colorado State University, Dept. of Statistics.

Tong, H. (1990), *Nonlinear Time Series: A Dynamical System Approach*, New York: Oxford University Press.

Zeger, S. L., and Qaqish, B. (1988), "Markov Regression Models for Time Series: A Quasi-Likelihood Approach," *Biometrics*, 44, 1019–1031.