Generalized Autoregressive Moving Average Models

Michael A. BENJAMIN, Robert A. RIGBY, and D. Mikis STASINOPOULOS

A class of generalized autoregressive moving average (GARMA) models is developed that extends the univariate Gaussian ARMA time series model to a flexible observation-driven model for non-Gaussian time series data. The dependent variable is assumed to have a conditional exponential family distribution given the past history of the process. The model estimation is carried out using an iteratively reweighted least squares algorithm. Properties of the model, including stationarity and marginal moments, are either derived explicitly or investigated using Monte Carlo simulation. The relationship of the GARMA model to other models is shown, including the autoregressive models of Zeger and Qaqish, the moving average models of Li, and the reparameterized generalized autoregressive conditional heteroscedastic GARCH model (providing the formula for its fourth marginal moment not previously derived). The model is demonstrated by the application of the GARMA model with a negative binomial conditional distribution to a well-known time series dataset of poliomyelitis counts.

KEY WORDS: Exponential family; Generalized autoregressive conditional heteroscedastic model; Generalized linear model; Negative binomial; Non-Gaussian; Poisson; Time series.

1. INTRODUCTION

This article considers the problem of extending Gaussian autoregressive moving average (ARMA) time series models to a non-Gaussian framework. Two approaches to this problem were originally described by Cox (1981) as observation-driven and parameter-driven models.

Several methods have been used for the parameter-driven (or state-space) modeling approach. Zeger (1988), for example, introduced time dependency through a latent process within a generalized linear model (GLM) (McCullagh and Nelder 1989) context and made adjustments to the marginal variance-covariance matrix to reflect the correlations that exist within the response observations. West, Harrison, and Migon (1985) approached the problem from a Bayesian perspective, using Kalman filtering to model response observations from an exponential family distribution. More recently, state-space models have been used with Markov chain Monte Carlo (MCMC) methods (Durbin and Koopman 1997; Shephard and Pitt 1997) to compute posterior distributions for the parameters of state-space models. But these state-space models can require complicated estimation techniques and occasionally crude approximations. A drawback of MCMC techniques is that convergence issues and inferential theory are as-yet not fully developed. A recent article by Durbin and Koopman (2000) attempted to overcome some of these problems. These authors used a Gaussian (importance) density to fit the non-Gaussian data, thus making the model amenable to the Kalman filter techniques. Their Kalman filter technique is essentially the same as that outlined by Fahrmeir and Wagenfeil (1997); however, Durbin and Koopman corrected for errors induced in their model due to the importance density approximation by using importance sampling techniques.

Considerable work has also been done on observation-driven models (Li 1994; Zeger and Qaqish 1988); this article also focuses on these models. Observation-driven models allow the likelihood of the data to be expressed explicitly for any fixed set of parameter values. They also facilitate model comparison and diagnostics. In the model developed herein, the conditional distribution of the dependent variable at time t, given the previous information set, is modeled by an exponential family distribution. Early work of this nature extends as far back as Heyde and Feigin (1975), who defined what they called conditional exponential family (CEF) distributions. They used simple models that had one autoregressive term. This first-order model was further developed in the work of Heyde (1978), Feigin (1981), and, more recently, by Grunwald, Hyndman, Tedesco, and Tweedie (2000).

Zeger and Qaqish (1988) developed autoregressive exponential family models, in particular autoregressive conditional Poisson and gamma models that extend Feigin's work by using additional autoregressive terms and including past and present covariates. Li (1994) introduced a moving average form of the Zeger and Qaqish (1988) model. Davis, Dunsmuir, and Wang (1999) and Brumback et al. (2000) considered alternative forms of conditional models.

This article extends the work of Zeger and Qaqish (1988) and Li (1994), giving rise to the generalized autoregressive moving average (GARMA) model. This model can accommodate nonstationary behavior, which may arise due to the influence of exogenous variables and also, through the use of "mixed models," allows a more parsimonious parameterization than either the pure autoregressive or pure moving average processes used in earlier observation-driven approaches. Conditions for stationarity and resulting stationary moments are also investigated.

The GARMA model may be used to model a variety of time-dependent response variables (which also have time-dependent covariates), for example, count data with a conditional Poisson, negative binomial, or binomial distribution or continuous data with a conditional gamma distribution (e.g., the volatility in a GARCH model).

Michael Benjamin is Vice President, Global Portfolio Trading Strategies, Citigroup, London E14 5LB, UK, and Robert Rigby and Mikis Stasinopoulos are Principal Research Fellows at Storm Research Center, London Metropolitan University, London N7 8DB, UK (E-mail: *Michael.Benjamin@citigroup.com; R.Rigby@londonmet.ac.uk; D.Stasinopoulos@londonmet.ac.uk*).

Journal of the American Statistical Association
March 2003, Vol. 98, No. 461, Theory and Methods
DOI 10.1198/016214503388619238

Section 2 defines the GARMA model and highlights its relationship to previous models. Section 3 describes the model-fitting algorithm, as well as model comparison and inference. The fitting algorithm implemented here uses the GLIM statistical package (Francis, Green, and Payne 1993), although other packages with a weighted least squares algorithm, such as S-PLUS, could implement this fitting algorithm. Section 4 considers model properties (in particular, conditions for stationarity), and derives the resulting stationary marginal moments of the dependent variable for the GARMA model with an identity link, using simulation to investigate the nonidentity link situation. Section 5 illustrates the methodology by applying a negative binomial GARMA model to a time series of poliomyelitis counts. The issues of model selection and model diagnostics are discussed. Section 6 gives some concluding remarks, and the Appendix provides proofs.

2. MODEL DEFINITION

2.1 General Model

In the GARMA model, the conditional distribution of each observation y_t, for $t = 1, \ldots, n$ given the previous information set $\underline{\mathbf{H}}_t = \{\underline{\mathbf{x}}_t, \ldots, \underline{\mathbf{x}}_1, y_{t-1}, \ldots, y_1, \mu_{t-1}, \ldots, \mu_1\}$, is assumed to belong to the same exponential family; that is, the conditional density is given by

$$f(y_t|\underline{\mathbf{H}}_t) = \exp\left\{\frac{y_t\vartheta_t - b(\vartheta_t)}{\varphi} + d(y_t, \varphi)\right\}, \qquad (1)$$

where ϑ_t and φ are the canonical and scale parameters, with $b(\cdot)$ and $d(\cdot)$ being specific functions that define the particular exponential family, and $\underline{\mathbf{x}}$ is a vector of r explanatory variables. The terms $\mu_t = b'(\vartheta_t) = E(y_t|\underline{\mathbf{H}}_t)$ and $\text{var}(y_t|\underline{\mathbf{H}}_t) = \varphi v(\mu_t) = \varphi b''(\vartheta_t)$ $(t = 1, \ldots, n)$ represent the conditional mean and variance of y_t given $\underline{\mathbf{H}}_t$. The notation is the same as for GLMs with independent observations (McCullagh and Nelder 1989), but here conditional rather than marginal distributions are modeled.

As with the standard GLM, μ_t is related to the predictor, η_t, by a twice-differentiable one-to-one monotonic function g, which is called the *link function*. Unlike the standard GLM linear predictor $\eta_t = \underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}}$, where $\underline{\boldsymbol{\beta}}' = (\beta_1, \beta_2, \ldots, \beta_r)$, here there is an additional component, τ_t, that allows autoregressive moving average terms to be included additively in the predictor. Past values of the time-dependent covariates, $\underline{\mathbf{x}}$, are also included in the predictor. Thus a general model for μ_t is given by $g(\mu_t) = \eta_t = \underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} + \tau_t$, with

$$\tau_t = \sum_{j=1}^{p} \phi_j \mathcal{A}(y_{t-j}, \underline{\mathbf{x}}_{t-j}, \underline{\boldsymbol{\beta}}) + \sum_{j=1}^{q} \theta_j \mathcal{M}(y_{t-j}, \mu_{t-j}), \qquad (2)$$

where $\mathcal{A}$ and $\mathcal{M}$ are functions representing the autoregressive and moving average terms and $\underline{\boldsymbol{\phi}}' = (\phi_1, \ldots, \phi_p)$ and $\underline{\boldsymbol{\theta}}' = (\theta_1, \ldots, \theta_q)$ are the autoregressive and moving average parameters. The moving average error terms, $\mathcal{M}$, could be, for example, deviance residuals, Pearson residuals, residuals measured on the original scale (i.e., $y_t - \mu_t$), or, as in what follows, residuals on the predictor scale [i.e., $g(y_t) - \eta_t$].

Model (2) is too general for practical application, so consider the following flexible and parsimonious submodel, which includes many well-known special cases:

$$g(\mu_t) = \eta_t = \underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} + \sum_{j=1}^{p} \phi_j\{g(y_{t-j}) - \underline{\mathbf{x}}_{t-j}'\underline{\boldsymbol{\beta}}\} + \sum_{j=1}^{q} \theta_j\{g(y_{t-j}) - \eta_{t-j}\}. \qquad (3)$$

Equations (1) and (3) together define the GARMA(p, q) model. The parameters $\underline{\boldsymbol{\beta}}, \underline{\boldsymbol{\phi}}$, and $\underline{\boldsymbol{\theta}}$ may be estimated by maximum likelihood, as discussed in Section 3. For certain functions g, it may be necessary to replace y_{t-j} with y_{t-j}^* in (3) to avoid the nonexistence of $g(y_{t-j})$ for certain values of y_{t-j}. The form of y_{t-j}^* depends on the particular function g and is defined for specific cases later.

A Box–Cox power transformation (Box and Cox 1964) may provide a suitable flexible form for g in (3), giving rise to a form of Box–Cox transformation-of-both-sides model. In this case the conditional mean, μ_t, and the lagged observations, y_{t-j}, in (3) are subject to the same power transformation $g(u)$, where $g(u) = (u^\lambda - 1)/u$ (if $\lambda \neq 0$) $+ \log(u)$ (if $\lambda = 0$). If $\lambda = 1$, the model is linear; if $\lambda = 0$, the model is log-linear. The optimal value for λ may be obtained by maximizing its profile likelihood.

2.2 Submodels of the GARMA Model and Related Models

Special cases of the GARMA model apart from the standard Gaussian ARMA(p, q) model give rise to well-known models, some of which are detailed here.

2.2.1 Poisson GARMA(p, q). A Poisson conditional distribution for y_t (given $\underline{\mathbf{H}}_t$) in (1) gives the Poisson GARMA(p, q) model. If, in addition, g is the log function in (3), then,

$$\log(\mu_t) = \underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} + \sum_{j=1}^{p} \phi_j\{\log(y_{t-j}^*) - \underline{\mathbf{x}}_{t-j}'\underline{\boldsymbol{\beta}}\} + \sum_{j=1}^{q} \theta_j\{\log(y_{t-j}^*/\mu_{t-j})\}, \qquad (4)$$

where $y_{t-j}^* = \max(y_{t-j}, c)$ and $0 < c < 1$; that is, any 0 values of y_{t-j} are replaced by c, the threshold parameter, in (4). If $\theta_j = 0$ for $j = 1, 2, \ldots, q$ in (4), then this gives an autoregressive model for counts from Zeger and Qaqish (1988). Feigin (1981) and Zeger and Qaqish (1988) discussed the Poisson GARMA(1, 0) with no time-dependent covariates in the context of branching processes. If $\phi_j = 0$ for $j = 1, 2, \ldots, p$ in (4), then this gives a pure moving average model for counts of the type considered by Li (1994).

Alternative forms of conditional Poisson models have been considered by Davis et al. (1999) and Brumback et al. (2000). Davis et al. (1999) discussed conditional Poisson time series models with predictors defined by

$$g(\mu_t) = \eta_t = \underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} + \sum_{j=1}^{\infty} \theta_j \varepsilon_{t-j},$$

where $\varepsilon_t = (y_t - \mu_t)/\mu_t^{\omega}$ for $\omega > 0$. They focused on the case when $\omega = .5$ (i.e., using Pearson residuals for the "moving average" terms). Davis et al. (1999) used a nonlinear optimization routine within S-PLUS to estimate their model parameters. The GLM iteratively reweighted least squares framework can in fact be used, because their model can be written in the format of our general observation-driven model (2) and fitted using an approach similar to that used in Section 3.1.

Brumback et al. (2000) discussed a conditional Poisson model similar in form to (2) that they call a "transitional GLM" (TGLM). In their example they reported four different forms of τ_t that correspond to our model (2) when $\mathcal{A}(\cdot) = y_{t-1}$, $\mathcal{A}(\cdot) = y_{t-1} - \exp(\underline{\mathbf{x}}'_{t-1}\underline{\boldsymbol{\beta}})$, $\mathcal{A}(\cdot) = \log(y^*_{t-1}) - \underline{\mathbf{x}}'_{t-1}\underline{\boldsymbol{\beta}}$, and a "moving average model" $\mathcal{M}(\cdot) = y_{t-1} - \mu_{t-1}$.

2.2.2 Binomial Logistic GARMA(p, q). If y_t has a conditional binomial $B(N_t, \mu_t)$ distribution (given $\underline{\mathbf{H}}_t$) in (1) and g is the logit function, that is, $g(u_t) = \text{logit}(u_t) = \log[u_t/(N_t - u_t)]$, in (3), then this gives a binomial logistic GARMA(p, q) model. As in Section 2.2.1, a threshold parameter $0 < c < 1$ is used to define y_t^*, where $y_t^* = \min[\max(y_t, c), (N_t - c)]$. If $N_t = 1$ for $t = 1, 2, \ldots, n$, this gives the binary logistic GARMA(p, q) model.

2.2.3 Gamma GARMA(p, q). A gamma conditional distribution for y_t (given $\underline{\mathbf{H}}_t$) in (1) gives the gamma GARMA(p, q) model. If in addition g is the reciprocal function [i.e., $g(u) = 1/u$] and $\theta_j = 0$ for $j = 1, 2, \ldots, q$, then this gives a model used by Zeger and Qaqish (1988) to model interspike times from neurons in the motor cortex of a monkey.

2.2.4 GARMA-GARCH(p, q). An important special case of the gamma GARMA(p, q) model is a reparameterised form of the GARCH model introduced by Bollerslev (1986). Assume that $\{\epsilon_t, t = 1, 2, \ldots, n\}$ is a process with a normal conditional distribution, that is, $\epsilon_t|\underline{\mathbf{H}}_t \sim N(0, h_t)$, where

$$h_t = \underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} + \sum_{j=1}^{p} \phi_j\{\epsilon_{t-j}^2 - \underline{\mathbf{x}}'_{t-j}\underline{\boldsymbol{\beta}}\} + \sum_{j=1}^{q} \theta_j\{\epsilon_{t-j}^2 - h_{t-j}\}. \quad (5)$$

Now let $y_t = \epsilon_t^2$, then y_t has a conditional gamma distribution; that is,

$$y_t|\underline{\mathbf{H}}_t \sim h_t \chi_1^2 \equiv Ga(\mu_t, \varphi), \quad (6)$$

with mean $\mu_t = E(y_t|\underline{\mathbf{H}}_t) = h_t$ and fixed scale parameter $\varphi = 2$. Hence from (5) and (6), the model for y_t is a gamma GARMA(p, q) model with identity function g and scale $\varphi = 2$, called here a GARMA-GARCH(p, q) model.

The standard family of GARCH models is defined by replacing (5) with

$$h_t = \beta_0 + \sum_{j=1}^{p} \gamma_j \epsilon_{t-j}^2 + \sum_{j=1}^{q} \delta_j h_{t-j}.$$

Model (5) allows time-dependent explanatory variables to influence h_t, the conditional variance of ϵ_t. The standard GARCH model, however, sets $\underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} = \beta_0$ for all t.

Bollerslev (1986) considered the GARMA-GARCH(p, q) form of parameterization as an alternative to the usual GARCH parameterization and described it as an "autoregressive moving average process in ϵ_t^2." Note that for $\underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} = \beta_0$ for all t, the families of GARCH and GARMA-GARCH models are identical as a whole.

Note also that if the identity function g used in (3) to obtain (5) is replaced by the Box–Cox power transformation, then for the case of $\theta_j = 0$ for $j = 1, 2, \ldots, q$ and $\underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} = \beta_0$ for all t, this gives a reparameterized form of the nonlinear ARCH (NARCH) model developed by Higgens and Bera (1992).

3. MAXIMUM LIKELIHOOD ESTIMATION AND INFERENCE

3.1 Estimation

The GARMA model-fitting procedure described herein performs maximum likelihood estimation using iteratively reweighted least squares (IRLS). The estimation method is based on that for the standard GLM.

Denote the model parameters to be estimated by $\underline{\boldsymbol{\gamma}} = (\underline{\boldsymbol{\beta}}', \underline{\boldsymbol{\phi}}', \underline{\boldsymbol{\theta}}')'$. These parameters are estimated by maximum likelihood. Strictly speaking, if the covariates, $\underline{\mathbf{x}}_t$, are stochastic, then estimation of $\underline{\boldsymbol{\gamma}}$ is done by maximizing the partial likelihood (see Fahrmeir and Tutz 1994, p. 194). The term "partial likelihood" thus should be used instead of likelihood in such situations. The log-likelihood of the data $\{y_{m+1}, \ldots, y_n\}$ conditional on the first m observations $\{y_1, \ldots, y_m\}$ and on $\eta_t = g(y_t^*)$ for $t = 1, 2, \ldots, i$, where $i = \max(p, q)$ and $m \geq i$, is given by $l = \sum_{t=m+1}^{n} \log f(y_t|\underline{\mathbf{H}}_t)$. Expectations are conditional on the same conditions.

The Fisher scoring algorithm is used to maximize the conditional log-likelihood function l leading to an IRLS procedure using at the kth iteration–adjusted dependent variable $\underline{\mathbf{z}}^{(k)}$ and weights $\underline{\mathbf{W}}^{(k)}$ constructed from the equations

$$z_t = \frac{\partial \eta_t}{\partial \underline{\boldsymbol{\gamma}}'}\underline{\boldsymbol{\gamma}} + \alpha(y_t - \mu_t)g'(\mu_t)$$

and

$$w_t^{-1} = [g'(\mu_t)]^2 v(\mu_t)$$

for $t = m+1, \ldots, n$, where $g'(\mu_t) = \partial\eta_t/\partial\mu_t$ and $0 < \alpha \leq 1$ is the step length used to aid convergence of the estimation procedure. Updated parameter estimates $\underline{\boldsymbol{\gamma}}^{(k+1)}$ are obtained by regressing $\underline{\mathbf{z}}^{(k)}$ on $(\partial\underline{\boldsymbol{\eta}}/\partial\underline{\boldsymbol{\gamma}}')^{(k)}$ with weights $\underline{\mathbf{W}}^{(k)} = \text{diag}(w_t^{(k)})$. (For full details of the algorithm, including the use of recursion when calculating the derivatives $(\partial\underline{\boldsymbol{\eta}}/\partial\underline{\boldsymbol{\gamma}}')^{(k)}$, see Benjamin, Rigby, and Stasinopoulos 1998.) Brumback et al. (2000) have advocated a regression procedure similar to the algorithm described earlier from Benjamin et al. (1998) for their TGLMs. Their working variable and iterative weights are scaled by dividing z_t by $g'(\mu_t)$ with fixed step length $\alpha = 1$ and multiplying w_t by $[g'(\mu_t)]^2/\varphi$.

3.2 Model Inference and Prediction

Consider two nested GARMA models, H_0 and H_1, with fitted log-likelihoods $\hat{l}_0$ and $\hat{l}_1$ and total number of fitted parameters κ_0 and κ_1. The likelihood ratio statistic for testing between the models, $\Lambda = (\widehat{D}_0 - \widehat{D}_1)$, has an approximate chi-squared distribution with $(\kappa_1 - \kappa_0)$ degrees of freedom under H_0, where $\widehat{D}_j = -2\hat{l}_j$ is called the *global deviance* for $j = 0, 1$ and where the exponential family dispersion parameter, φ, if unknown, is estimated by its maximum likelihood estimator (MLE) $\hat{\varphi}_j$ for each model $j = 0, 1$.

Evaluating $\underline{\gamma}$ using IRLS allows us to obtain the approximate large-sample conditional variance of $\hat{\underline{\gamma}}$ as a byproduct of the iterative process (Green 1984; Kaufmann 1987). It can be shown that $\sqrt{(n-m)}(\hat{\underline{\gamma}} - \underline{\gamma}) \sim N(0, \underline{\mathbf{I}}(\underline{\gamma})^{-1})$, where

$$\underline{\mathbf{I}}(\underline{\gamma}) = \lim_{n\to\infty} \frac{\hat{\varphi}}{n-m} \left\{ \sum_{t=m}^{n} w_t \left(\frac{\partial \eta_t}{\partial \underline{\gamma}}\right) \left(\frac{\partial \eta_t}{\partial \underline{\gamma}}\right)' \right\}^{-1}.$$

Simulation techniques can be used for short-range forecasting. Fixing $\hat{\underline{\gamma}}$ and the known and estimated information set (i.e., $\widetilde{\underline{\mathbf{H}}} = \{\underline{\mathbf{x}}_t, \ldots, \underline{\mathbf{x}}_1, y_n, \ldots, y_1, \hat{\mu}_n, \ldots, \hat{\mu}_1\}$), N sample paths of simulated values to y_t $(t > n)$ give an empirical predictive distribution for Y_t given $\hat{\underline{\gamma}}$ and $\widetilde{\underline{\mathbf{H}}}$ [i.e., $f(y_t|\hat{\underline{\gamma}}, \widetilde{\underline{\mathbf{H}}})$], from which intervals for predicted values can be constructed (see, e.g., Hyndman 1995). Note that any stochastic explanatory time series will also need to be forecasted. Also, a check for nonstationarity is obtained by comparing the empirical distribution of Y_{t_1}, Y_{t_2}, and Y_{t_3}, where $n < t_1 < t_2 < t_3 < m$ for t_1, t_2, and t_3 sufficiently far apart.

4. PROPERTIES OF THE GENERALIZED AUTOREGRESSIVE MOVING AVERAGE PROCESS

In Section 4.1 we derive stationarity conditions for the marginal mean and variance and the resulting stationary mean and variance of the dependent variable y in the GARMA model for the identity link for some key exponential family distributions. Some of these results are generalizations of results for first-order autoregressive models given by Grunwald et al. (2000). In Section 4.2 we discuss how simulation can be used to investigate models with a nonidentity link function. In Section 4.3 we consider parameter restrictions.

4.1 Identity Link

We derive theoretical formulas for the marginal mean and variance for a GARMA model with identity link function g. The marginal variance is found explicitly for certain specific exponential family error distributions.

Theorem 1. The marginal mean of y_t of the GARMA model defined by (1) and (3) with identity link function g is

$$E(y_t) = \underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} \tag{7}$$

provided that $\Phi(B)$ (as defined in Theorem 2) is invertible. The proof of Theorem 1 is given in the Appendix. The marginal mean is stationary provided also that $\underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} = \beta_0$ for all t.

Theorem 2. The marginal variance of y_t for the GARMA model with identity link g is

$$V(y_t) = \varphi E[\Psi^{(2)}(B)v(\mu_t)], \tag{8}$$

where $\Psi^{(2)}(B) = 1 + \psi_1^2 B + \psi_2^2 B^2 + \cdots$ and $\Psi(B) = \Phi(B)^{-1}\Theta(B) = 1 + \psi_1 B + \psi_2 B^2 + \cdots$ [assuming that $\Phi(B)$ is invertible], $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, and $\Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$.

Equation (8) can be used to provide the conditions for a stationary variance for specific exponential family distributions. Note also that for independent observations with mean μ_t, the marginal variance of y_t is given by $\varphi v(\mu_t)$. Hence the expectation operator and the term $\Psi^{(2)}(B)$ in (8) represent overdispersion factors that have arisen due to the time series nature of the data. Because the term $v(\mu_t)$ is distribution dependent, the explicit stationary marginal variance of y_t is derived separately for the Poisson, negative binomial, and gamma distributions in the following corollaries, setting $\underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} = \beta_0$ for all t. Proofs are given in the Appendix.

Corollary 1. The stationary marginal variance of y_t for the Poisson GARMA model with identity link is

$$V(y_t) = \Psi^{(2)}(1)\beta_0, \tag{9}$$

where $\Psi^{(2)}(1) = 1 + \sum_{j=1}^{\infty} \psi_j^2$.

Corollary 2. The stationary marginal variance of y_t for the negative binomial GARMA model with identity link is

$$V(y_t) = \Psi^{(2)}(1)[\beta_0 + E(\mu_t^2)/k], \tag{10}$$

where k is a distribution parameter defined later in (12), and $E[\mu_t^2] = [1 + (1/k) - (1/k)\Psi^{(2)}(1)]^{-1}\{\beta_0^2 + [\Psi^{(2)}(1) - 1]\beta_0\}$ provided that $[1 + (1/k) - (1/k)\Psi^2(B)]$ is invertible, giving the extra conditions for a stationary finite variance.

Corollary 3. The stationary marginal variance y_t of the gamma GARMA model with identity link is

$$V(y_t) = \varphi\Psi^{(2)}(1)[1 + \varphi - \varphi\Psi^{(2)}(1)]^{-1}\beta_0^2, \tag{11}$$

provided that $[1 + \varphi - \varphi\Psi^{(2)}(B)]$ is invertible.

Note the similarities between the marginal variance for the negative binomial and gamma GARMA models when φ is replaced by $1/k$. This is due to their conditional variances of $\varphi\mu^2$ and $\mu + \mu^2/k$. The above corollaries hold within model-specific parameter restrictions to ensure that $\mu_t > 0$ for all t.

4.2 Nonidentity Link

In general, for link functions g other than the identity, the first two moments of the marginal distribution appear to be intractable, as do the parameter restrictions required to ensure stationarity. We carried out a Monte Carlo simulation to investigate the region of the parameter space for which a GARMA process was nonstationary and to ascertain how the marginal moments are affected by different values of the parameters in the stationarity region of the parameter space. For each parameter combination, we used a Monte Carlo simulation with 1,000 realizations of length 200 to check for nonstationarity by comparing the empirical distributions at times 150, 175, and 200 using a chi-squared goodness-of-fit test.

For parameter combinations that exhibited no evidence of nonstationarity, we carried out a further Monte Carlo simulation using 10 realizations of length 20,150. We found that this provided reliable marginal moment estimates having small standard errors.

We calculated sample statistics over times 150 to 20,150 for each of the 10 realizations. These provided estimates (and standard errors) for the corresponding marginal moments of dependent variable y.

We illustrate the approach by simulation for the important special case of a Poisson GARMA(1, 1) model with the log link function, as defined by (4) with threshold parameter $c = .1$ and $\underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} = \beta_0 = \log_e 2$ for all t. We considered each of the 16 combinations of ϕ and $\theta \in (-.4, 0, .4, .8)$. Entries in the table for which $\phi = -\theta$ correspond to an independent Poisson series with mean 2, and hence are the same as those for $(\phi, \theta) = (0, 0)$. We found no evidence of nonstationarity for any of the parameter combinations considered. Table 1 provides the estimates of the stationary marginal moments of y. Specifically, it gives the marginal mean, variance, skewness, and kurtosis of y for each parameter combination.

Table 1 shows that the marginal mean, variance, skewness, and kurtosis of y increase as $(\phi, \theta) \to (1, 1)$ or $(-1, -1)$. In comparison, an independent Poisson series with mean 2 has variance 2, skewness .71, and kurtosis 3.5. Hence the time series structure can lead to an extremely long-tailed marginal distribution. In contrast, the Gaussian ARMA(1, 1) model has marginal mean, skewness, and kurtosis all constant over the stationary parameter region $(0 < \phi < 1, 0 < \theta < 1)$, whereas its marginal variance increases as $(\phi, \theta) \to (1, 1)$ or $(-1, -1)$, but more slowly than in Table 1.

4.3 Parameter Restrictions

GARMA models with conditional distributions with range restrictions $y_t \geq 0$ require the constraint $\mu_t \geq 0$, which results in restrictions on the parameter space, depending on the particular link function. Additional parameter restrictions may also be needed for a stationary finite mean and variance, depending on the particular conditional exponential family distribution and link function.

4.3.1 GARMA(p, q) Model With Identity Link. The parameter restrictions for a GARMA(p, q) model with an identity link to have a stationary finite marginal mean are that $\underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} = \beta_0$ for all t and that $\Phi(B)$ is invertible, that is, has roots outside the unit circle. The additional parameter restrictions for the GARMA models with identity link also to have stationary finite variance depend on the particular conditional distribution. For example, for the Poisson conditional distribution, no additional parameter restrictions are required in (9), whereas for the gamma conditional distribution, the invertibility of $[1 + \varphi - \varphi\Psi^{(2)}(B)]$ is required in (11).

Because the GARCH model is a reparameterized form of the gamma GARMA model with identity link function and scale $\varphi = 2$, (11) can be used to obtain the fourth moment of ϵ_t in the GARCH model defined in Section 2.2.4. The results agree with those of Bollerslev (1986, p. 313) for the GARCH(1, 1) model, but (11) can also provide the fourth moments of ϵ_t for higher-order GARCH models not previously derived explicitly.

4.3.2 GARMA(1, 1) Model With Identity Link. For a GARMA(1, 1) model with identity link function, the constraint $\mu_t \geq 0$ results in the parameter restrictions to the triangle of the parameter space bounded by $\phi = -\theta$, $\phi = 1$, and $\theta = 0$ shown in Figure 1. The additional parameter restrictions for a stationary finite marginal mean are that $\underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} = \beta_0$ for all

Table 1. Marginal Moments for Poisson GARMA Models: Marginal Mean, Variance, Skewness and Kurtosis

θ	ϕ = −.4	0	.4	.8
.8	2	2	2.1	3.1
	2.7	4.1	6.6[b]	17.9[c]
	1.1	1.7	2.2[a]	2.1[a]
	4.7[a]	8[f]	12.9[g]	9.9[g]
.4		1.8	1.7	1.9
		2.3	3.2	7.3[c]
		.9	1.3	2
		3.9	4.7[a]	7.9[e]
0	2.6	2	1.6	1.2
	5.4	2	2.1	3.2[a]
	1.5	.7	1	2
	5.9[a]	3.5	4	8.3[e]
−.4	4.8	2.6		1
	45.6[f]	5		1.6
	1.9	1.5		1.7
	5.7[a]	6.2[a]		6.4[b]

NOTE: Standard errors are indicated by letters next to the relevant terms using the key below. Terms with no key have standard errors less than .1.
[a] .1 → .2.
[b] .2 → .3.
[c] .3 → .4.
[d] .4 → .5.
[e] .5 → .6.
[f] .6 → .8.
[g] 2 → 3.

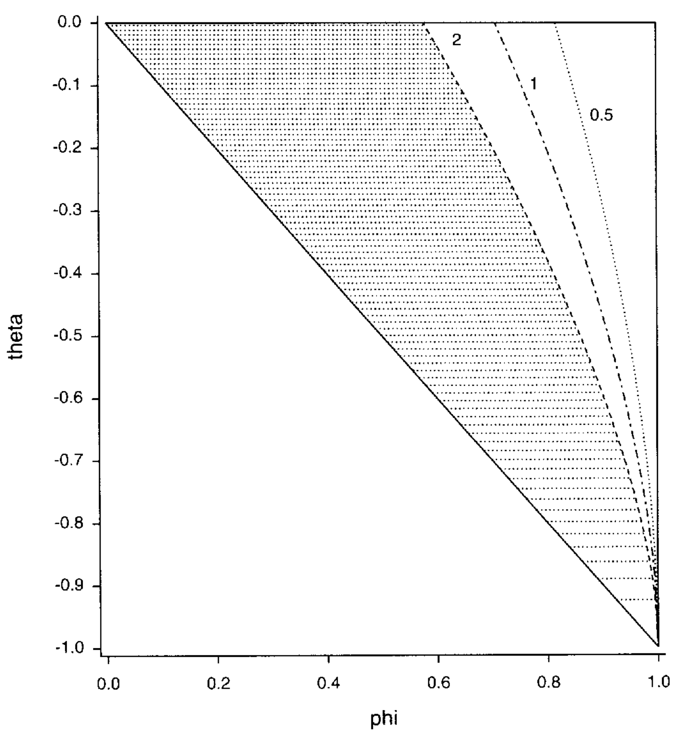

Figure 1. Stationarity Regions of the Gamma GARMA(1,1) Model With Identity Link for Scale Parameter $\varphi = .5, 1, 2$ (with shaded region for $\varphi = 2$).

t and that $-1 \leq \phi \leq 1$, giving marginal mean $E(y_t) = \beta_0$. For a Poisson GARMA(1, 1) model with identity link, no additional parameter restrictions are required for a stationary finite marginal variance, so the resulting parameter region is still the triangle in Figure 1.

However, for the gamma GARMA(1, 1) model with identity link, the additional parameter restrictions are that $|\phi^2 + \varphi(\phi + \theta)^2| < 1$, giving marginal variance $V(y_t) = \beta_0^2\varphi[(1-\phi^2) - (\phi+\theta)^2]/[(1-\phi^2) - \varphi(\phi+\theta)^2]$. The resulting parameter region is shown in Figure 1 as a shaded region for $\varphi = 2$ that extends to give larger regions for $\varphi = 1$ and $\varphi = .5$. The ϕ-axis intercept is $(1+\varphi)^{-.5}$, with intercept moving from $(\phi, \theta) = (1, 0)$ to the origin $(0, 0)$ along the horizontal axis as φ changes from 0 to ∞. The corresponding parameter restrictions for the negative binomial GARMA model are identical, but with φ replaced by $1/k$ with the limiting Poisson GARMA case as $k \to \infty$ given by the complete triangle.

5. ILLUSTRATIVE EXAMPLE

This section provides a practical demonstration of the capabilities of the GARMA model-fitting procedure and diagnostic facilities to model a univariate non-Gaussian time series. In this example, the GARMA model (4) with negative binomial conditional distribution is used to model the monthly number, y_t, of poliomyelitis cases reported to the U.S. Centers for Disease Control for the years 1970 to 1983, that is, 168 observations. The GARMA model is shown to compare favorably with previous models. The data were originally modeled by Zeger (1988), who used a parameter-driven approach in which a first-order autoregressive model was used for the latent process to conclude that there is evidence of a decreased polio infection rate. Extending this to assume different forms for the latent process leads to particularly complicated estimating procedures. Using an alternative observation-driven approach, Li (1994) compared a second-order moving average log-linear Poisson process to a second-order Markov model of Zeger and Qaqish (1988), although he did not take into account any seasonality, which is apparent from a plot of the series shown in Figure 2. These data were also modeled by Davis et al. (1999), who concluded that there was additional variation in y_t that was not accounted for using their observation-driven model (described earlier in Sec. 2.2).

Here additional variation in y_t is accounted for by fitting a negative binomial GARMA model with conditional distribution

$$f(y_t|\underline{\mathbf{H}}_t) = \frac{\Gamma(y_t+k)}{\Gamma(k)\Gamma(y_t+1)}\left\{\frac{\mu_t}{\mu_t+k}\right\}^{y_t}\left\{\frac{k}{\mu_t+k}\right\}^{k} \qquad (12)$$

for $y_t = 0, 1, 2, \ldots$, where k is a dispersion parameter. The conditional mean and variance of y_t are given by

$$E(y_t|\underline{\mathbf{H}}_t) = \mu_t \qquad \text{and} \qquad V(y_t|\underline{\mathbf{H}}_t) = \mu_t + \frac{\mu_t^2}{k}.$$

We jointly fitted the parameters $(\underline{\boldsymbol{\gamma}}, k)$ by adopting the approach of Lawless (1987), maximizing the likelihood $l(\underline{\boldsymbol{\gamma}}, k)$ with respect to $\underline{\boldsymbol{\gamma}}$ for selected values of k, using the general algorithm outlined in Section 3. This gives a maximum likelihood estimate of $\hat{\underline{\boldsymbol{\gamma}}}(k)$ and hence the profile likelihood $l(\hat{\underline{\boldsymbol{\gamma}}}(k), k)$, from which the maximum likelihood estimate, $\hat{k}$, can be determined. As $k \to \infty$, the limiting form of the negative binomial distribution is the Poisson distribution.

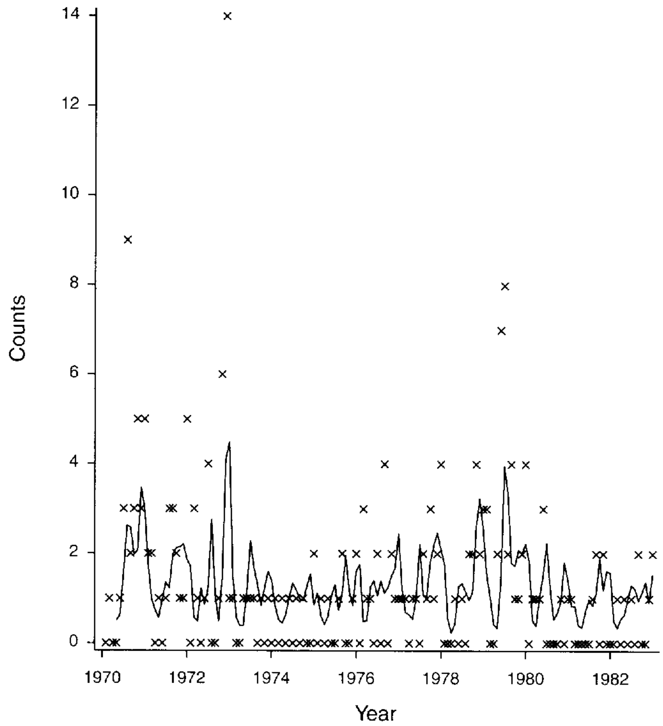

Figure 2. Poliomyelitis Data (x) and Fitted Mean Values (—) From the Negative Binomial GARMA(0, 2) Model.

To investigate whether the incidence of polio has been decreasing since 1970, we fit models with and without the trend effect. We modeled the seasonal effect by sine and cosine pairs with both annual and semiannual cycles. The predictor model is thus given by (4), where $y_t^* = \max(y_t, .1)$ and $\underline{\mathbf{x}}_t'$ comprises $1, t, \cos(2\pi t/12), \sin(2\pi t/12), \cos(2\pi t/6), \sin(2\pi t/6)$.

We fit the negative binomial GARMA model with log link function defined by (12) and (4) to the data for a range of values p and q to model the autocorrelation structure in y_t. Fitting the additional parameter k caused a significant improvement in the fit relative to the Poisson model for all of the models. The resulting fitted global deviances (i.e., minus twice the fitted log-likelihood conditional on the first three observations), together with the maximum likelihood estimate of k, are shown in Table 2 for the negative binomial GARMA models with and without a trend. For models with a trend, the estimate of $\hat{\beta}$ is also given.

We carried out testing between the models using the likelihood ratio test statistic based on the difference in global deviances as described in Section 3.2. Global deviances are essential for comparing GARMA models with different variance functions (as here with models with different fitted values of k). Alternatively, the generalized Akaike information criterion, $GAIC = -2\hat{l} + \zeta.\kappa$, which penalizes the global deviance $(-2\hat{l})$ by penalty ζ for each of the κ parameters fitted in the model, can be used to compare models. Using a penalty $\zeta = 3.8$ (equal to $\chi^2_{1,.05}$, a 5% critical value of a chi-squared distribution) in the GAIC and considering models with and

Table 2. Fitted Global Deviances for Negative Binomial GARMA Models

	Trend			No trend	
Order (p, q)	Deviance	$\hat{k}$	$\hat{\beta} \times 10^3$	Deviance	$\hat{k}$
(0, 0)	501.6	1.79	−4.69	507.8	1.60
(1, 0)	494.9	2.14	−4.46	499.0	1.98
(0, 1)	496.5	2.04	−4.51	501.4	1.86
(2, 0)	490.7	2.28	−4.40	493.6	2.16
(1, 1)	492.8	2.19	−4.67	495.9	2.07
(0, 2)	487.8	2.52	−4.43	490.9	2.37
(3, 0)	486.3	2.56	−4.48	490.3	2.36
(2, 1)	486.2	2.45	−4.85	489.7	2.31
(1, 2)	484.4	2.65	−4.65	488.3	2.48
(0, 3)	485.3	2.73	−4.47	489.4	2.50

without a trend separately leads to the choice of $p = 0$ and $q = 2$ in both cases. Comparing these GARMA(0, 2) models (i.e., $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$), the trend effect was not significant at a 5% level. Hence there is no significant evidence of a decreasing trend in polio infection after accounting for both the seasonality and the autocorrelation structure exhibited by the data. This is in agreement with the results of Zeger (1988), Fahrmeir and Tutz (1994), Kuk and Cheng (1997), and Davis et al. (1999). Note also that Davis et al. (1999) did not consider fitting a model without a trend term, although its t statistic (-1.69) would indicate that this term is not significant. They used a Poisson conditional distribution model; however, the negative binomial model used here provides a significantly better fit.

The fitted negative binomial and Poisson GARMA(0, 2) models without a trend are detailed in Table 3, which gives for each model the parameter estimates and their standard errors together with their global deviance. Note that the negative binomial model produces a clearly better fit than the Poisson model and has similar parameter estimates but higher standard errors. Considering the differences in the case contributions to the global deviance between the two models (Fig. 3), it is clear that observations at July 1970 ($y_t = 9$), November 1972 ($y_t = 14$), and May 1979 ($y_t = 7$) with deviance differences of 3.26, 4.61, and 3.05, are fitted better by the negative binomial GARMA model.

The fitted (conditional) mean μ_t from the negative binomial GARMA(0, 2) model is given in Figure 2, and the resulting residual analysis is shown in Figure 4. As the distribution

Table 3. Fitted Negative Binomial and Poisson GARMA(0,2) models

	Negative binomial		Poisson	
Parameter	$\hat{\gamma}$	$SE(\hat{\gamma})$	$\hat{\gamma}$	$SE(\hat{\gamma})$
Intercept	.406	.135	.414	.114
$\cos(2\pi t/12)$	.139	.155	.149	.126
$\sin(2\pi t/12)$	−.482	.181	−.533	.162
$\cos(2\pi t/6)$	.404	.136	.454	.112
$\sin(2\pi t/6)$	−.000159	.133	−.020	.109
MA(1)	.214	.063	.265	.050
MA(2)	.203	.063	.242	.047
Deviance	490.9		513.1	

NOTE: SE = standard error.

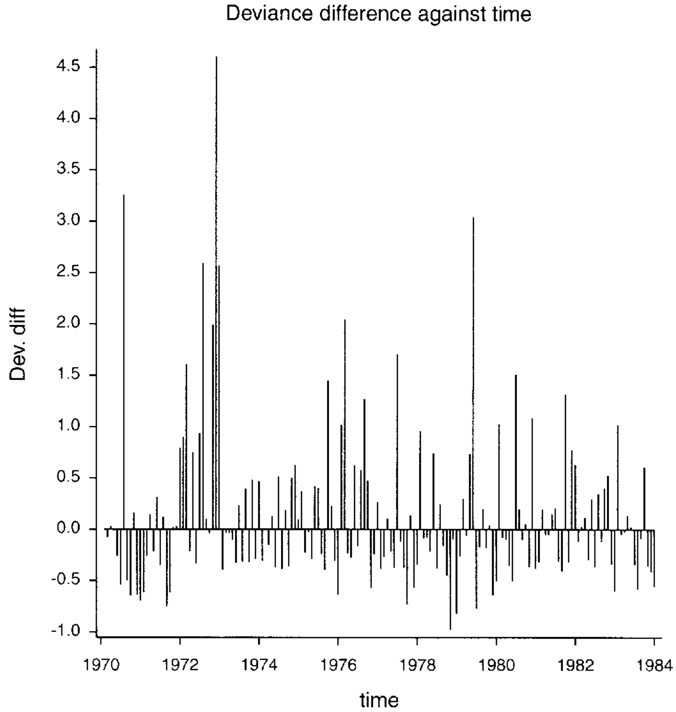

Figure 3. Differences in the Case Contributions to the Global Deviance Between the Fitted Poisson and Negative Binomial Models.

of the (conditional) deviance (and also Pearson) residuals are highly nonnormally distributed for count data with low fitted means, we advocate using the normalized conditional (randomized) quantile residuals of Dunn and Smyth (1996) for (discrete) GARMA models. The normalized randomized quantile residuals are given by $r_t = \Phi^{-1}(u_t)$, where Φ^{-1} is the inverse cumulative distribution function of a standard normal variate and u_t is a random value from the uniform distribution in the interval $[F(y_t - 1, \hat{\mu}_t, \hat{k}), F(y_t, \hat{\mu}_t, \hat{k})]$, where $F(y_t; \hat{\mu}, \hat{k})$ is the fitted conditional negative binomial cumulative distribution function. Figures 4(a) and 4(b) plot the autocorrelation and partial autocorrelation functions of the randomized quantile residuals. There is no evidence of any correlation within the residuals, a finding is supported by the Box–Ljung statistic of 12.9 based on 15 lags (because $\chi^2_{13,.05} = 22.36$). Figures 4(c) and 4(d) show kernel density and normal Q-Q plots for the normalized randomized quantile residuals, which appear to be roughly normally distributed as expected. A plot of the normalized randomized quantile residuals against time (Fig. 5) shows a random scatter. We conclude that the chosen model provides an adequate fit to the data.

No fitted observations had extreme lower tail probabilities [i.e., $F(y_t; \hat{\mu}, \hat{k}) \leq .02$], although the indication of this in Figure 5 is a result of the residual randomization process. Table 4 gives the observations y_t for which the upper-tail probability $p_t = P(Y_t \geq y_t | \underline{\mathbf{H}}_t) = 1 - F(y_t - 1; \hat{\mu}, \hat{k}) \leq .02$. This shows that the observation at November 1972 ($y_t = 14$), often considered an outlier, is not as extreme as other modelers have indicated. This is due to the fatter tails of the negative binomial distribution causing the tail probability to increase from .000782 to .016 when a Poisson conditional distribution assumption is replaced by a negative binomial.

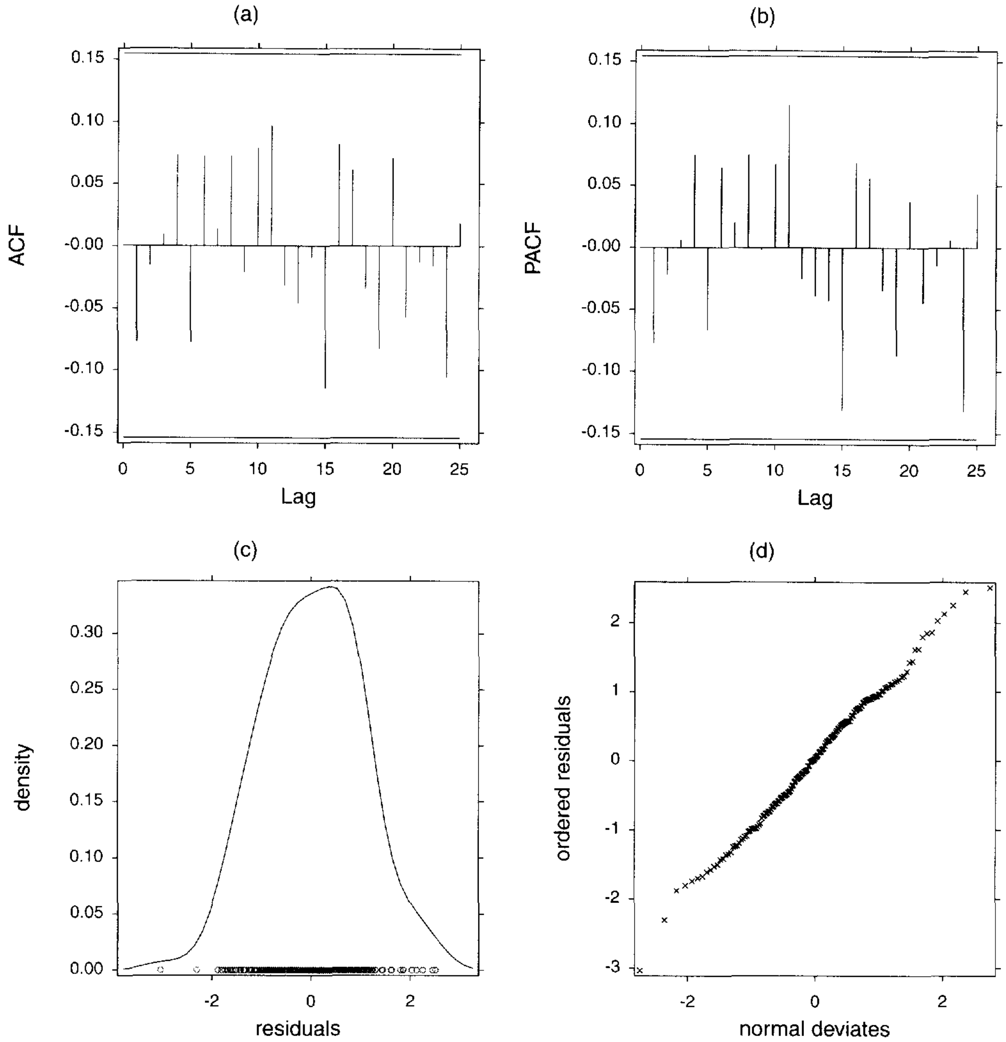

Figure 4. Randomized Quantile Residual Analysis. (a) The ACF of residuals (b) The PACF of residuals (c) density estimation of the residuals (d) QPLOT of the residuals.

6. DISCUSSION

In this article we have proposed a class of GARMA models that extend univariate ARMA models to a non-Gaussian situation. We used the GLM framework to fit these models. The maximum likelihood fitting algorithm described is within the IRLS framework. We used the fitted likelihood for model comparisons. We have developed GLIM programs to carry out menu-driven model specification, estimation, and diagnostics outlined in the article.

This work extends earlier work on observation driven models by providing a parsimonious representation of a process in which the mean of the conditional exponential family distribution depends on the past history of the process. We have derived the stationary conditions and resulting stationary mean and variance for the important cases of the Poisson, negative binomial and Gamma models with identity link, providing as a byproduct the conditions and formula for the fourth moment of the GARCH model. This model has potentially wide applications. In particular, the Poisson GARMA model with log link function considered in the Section 2.2.1 has potential uses for modeling time series count data with a highly kurtotic marginal distribution, and negative binomial GARMA models provide a useful model for overdispersed time series count data repeatedly observed in real datasets, which has not previously been modeled adequately.

APPENDIX: PROOFS

Proof of Theorem 1

For the identity link, (3) becomes

$$\mu_t = \underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} + \sum_{j=1}^{p} \phi_j\{y_{t-j} - \underline{\mathbf{x}}_{t-j}'\underline{\boldsymbol{\beta}}\} + \sum_{j=1}^{q} \theta_j\{y_{t-j} - \mu_{t-j}\}. \tag{A.1}$$

Let $y_t = \mu_t + \nu_t$. Then the ν_t are martingale errors (i.e., marginally mean 0 and uncorrelated). Substituting μ_t from (A.1) gives

$$w_t = \sum_{j=1}^{p} \phi_j w_{t-j} + \sum_{j=1}^{q} \theta_j \nu_{t-j} + \nu_t,$$

where $w_t = y_t - \underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}}$. Hence

$$w_t = \Psi(B)\nu_t, \tag{A.2}$$

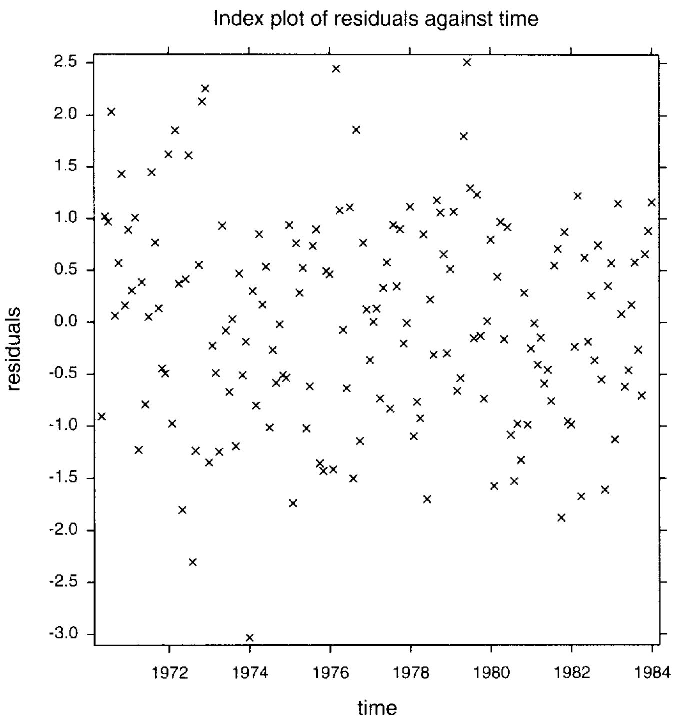

Figure 5. Randomized Quantile Residuals Against Time for the Negative Binomial GARMA(0, 2) Model.

where $\Psi(B) = \Phi(B)^{-1}\Theta(B)$, provided that $\Phi(B)$ is invertible, and $\Psi(B)$, $\Phi(B)$, and $\Theta(B)$ are defined as in Section 4.1. The marginal mean $E(y_t)$ of y_t is given by

$$E(y_t) = E(\underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} + w_t) = \underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} + E(w_t) = \underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}},$$

as $E(w_t) = 0$ from (A.2).

Proof of Theorem 2

Applying the law of iterated expectations gives the marginal variance, $V(\nu_t)$, of the martingale errors, ν_t,

$$V(\nu_t) = E(\nu_t^2) = E[E(\nu_t^2|\underline{\mathbf{H}}_t)] = \varphi E[v(\mu_t)], \tag{A.3}$$

because $v(\mu_t) = V(y_t|\underline{\mathbf{H}}_t) = V(\nu_t|\underline{\mathbf{H}}_t)$. Hence the marginal variance $V(y_t)$ of y_t is given by

$$\begin{aligned} V(y_t) &= V(\underline{\mathbf{x}}'\underline{\boldsymbol{\beta}} + w_t) = V(w_t) = E(w_t^2) \\ &= E[\Psi(B)\nu_t \Psi(B)\nu_t] \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \Psi_i \Psi_j E(\nu_{t-i}\nu_{t-j}) \\ &= \sum_{i=0}^{\infty} \Psi_i^2 E(\nu_{t-i}^2) \\ &= \varphi E[\Psi^{(2)}(B) v(\mu_t)], \end{aligned} \tag{A.4}$$

using (A.3) and where $\Psi^{(2)}(B)$ is as defined in Section 4.1.

Table 4. Tail Probabilities for Extreme Residuals

Date	y_t	$\hat{\mu}_t$	p_t
October 1972	6	1.423	.020
November 1972	14	4.077	.016
May 1979	7	1.378	.008

The marginal variances of y_t for the GARMA model with identity link and Poisson, gamma, and negative binomial conditional distributions are derived in Corollaries 1–3, which use the following lemma.

Lemma A.1. For any GARMA(p, q) model with identity link function, the marginal variance of μ_t is given by

$$V(\mu_t) = \varphi E\{[\Psi^{(2)}(B) - 1] v(\mu_t)\} \tag{A.5}$$

Proof of Lemma A.1

$$\begin{aligned} y_t &= \mu_t + \nu_t \\ V(y_t) &= V(\mu_t) + V(\nu_t) \end{aligned}$$

because ν_t is uncorrelated with μ_t. Hence, using (A.3) and (A.4), the Lemma A.1 result (A.5) is obtained.

In the proofs of the corollaries that follow, let $\underline{\mathbf{x}}_t'\underline{\boldsymbol{\beta}} = \beta_0$ for all t for simplicity of notation. Hence $E(\mu_t) = \beta_0$ for all t.

Proof of Corollary 1

Applying (A.4) to the Poisson conditional distribution gives

$$\begin{aligned} V(y_t) &= \Psi^{(2)}(B) E(\mu_t) \\ &= \Psi^{(2)}(B)\beta_0 = \Psi^{(2)}(1)\beta_0, \end{aligned}$$

because $\varphi = 1$ and $v(\mu_t) = \mu_t$. Hence Corollary 1 is established.

Proof of Corollary 2

Applying (A.5) to the negative binomial conditional distribution gives

$$\begin{aligned} E(\mu_t^2) &= [\Psi^{(2)}(B) - 1] E(\mu_t + \mu_t^2/k) + \beta_0^2 \\ &= [\Psi^{(2)}(B) - 1]\beta_0 + \beta_0^2 + [\Psi^{(2)}(B) - 1] E(\mu_t^2)/k, \end{aligned}$$

because $\varphi = 1$ and $v(\mu_t) = \mu_t + \mu_t^2/k$. Hence

$$E(\mu_t^2) = [1 + (1/k) - (1/k)\Psi^{(2)}(B)]^{-1} \cdot \{\beta_0^2 + [\Psi^{(2)}(B) - 1)]\beta_0\}, \tag{A.6}$$

provided that $[1 + (1/k) - (1/k)\Psi^{(2)}(B)]$ is invertible. Hence, using (A.4),

$$\begin{aligned} V(y_t) &= \Psi^{(2)}(B) E(\mu_t + \mu_t^2/k) \\ &= \Psi^{(2)}(B)[\beta_0 + E(\mu_t^2)/k], \end{aligned}$$

where $E(\mu_t^2)$ is given by (A.6). Hence Corollary 2 is established.

Proof of Corollary 3

Applying (A.5) to the gamma conditional distribution gives

$$E(\mu_t^2) = \varphi[\Psi^{(2)}(B) - 1] E[\mu_t^2] + \beta_0^2,$$

because $v(\mu_t) = \mu_t^2$. Hence

$$E(\mu_t^2) = [1 + \varphi - \varphi\Psi^{(2)}(B)]^{-1}\beta_0^2,$$

provided that $\{1 + \varphi - \varphi\Psi^{(2)}(B)\}$ is invertible. Hence, using (A.4),

$$V(y_t) = \varphi\Psi^{(2)}(B)[1 + \varphi - \varphi\Psi^{(2)}(B)]^{-1}\beta_0^2.$$

Thus Corollary 3 is established.

[Received January 1999. Revised October 2001.]

REFERENCES

Benjamin, M. A., Rigby, R. A., and Stasinopoulos, D. M. (1998), "Fitting Non-Gaussian Time Series Models," in *COMPSTAT Proceedings in Computational Statistics*, eds. R. Payne and P. Green, Heidelburg: Physica-Verlag, pp. 191–196.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 304–327.

Box, G. E. P., and Cox, D. R. (1964), "An Analysis of Transformations," *Journal of the Royal Statistical Society*, 26, 211–243.

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*, 95, 16–27.

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

Davis, R. A., Dunsmuir, W. T. M., and Wang, Y. (1999), "Modelling Time Series of Count Data," in *Asymptotics, Nonparametrics and Time Series*, ed. S. Ghosh, New York: Marcel Dekker, pp. 63–113.

Dunn, P. K., and Smyth, G. K. (1996), "Randomized Quantile Residuals," *Journal of Computational and Graphical Statistics*, 5, 236–244.

Durbin, J., and Koopman, S. J. (1997), "Monte Carlo Maximum Likelihood Estimation for Non-Gaussian State-Space Models," *Biometrika*, 84, 669–684.

——— (2000), "Time Series Analysis of Non-Gaussian Observations Based on State-Space Models From Both Classical and Bayesian Perspectives" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 62, 3–56.

Fahrmeir, L., and Tutz, G. (1994), *Multivariate Statistical Modeling Based on Generalized Linear Models*, New York: Springer-Verlag.

Fahrmeir, L., and Wagenpfeil, S. (1997), "Penalized Likelihood Estimation and Iterative Kalman Filtering for non-Gaussian Dynamic Regression Models," *Computational Statistics and Data Analysis*, 24, 295–320.

Feigin, P. D. (1981), "Conditional Exponential Families and a Representation Theorem for Asymptotic Inference," *The Annals of Statistics*, 9, 597–603.

Francis, B., Green, M., and Payne, C. (eds.) (1993), *GLIM4: The Statistical System for Generalized Linear Interactive Modelling*, Oxford, U.K.: Clarendon Press.

Green, P. (1984), "Iterative Reweighted Least Squares for Maximum Likelihood Estimation, and Some Robust and Resistant Alternatives" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 46, 149–192.

Grunwald, G. K., Hyndman, R. J., Tedesco, L. M., and Tweedie, R. L. (2000), "Non-Gaussian Conditional Linear AR(1) Models," *Australian and New Zealand Journal of Statistics*, 42, 179–495.

Heyde, C. C. (1978), "On an Optimal Asymptotic Property of the Maximum Likelihood Estimator of a Parameter From a Stochastic Process," *Stochastic Processes and Their Applications*, 8, 1–9.

Heyde, C. C., and Feigin, P. D. (1975), "On Efficiency and Exponential Families in Stochastic Process Estimation," in *Statistical Distributions in Scientific Work*, Vol. 1, eds. G. P. Patil et al., Dordrecht: D. Riedel, 227–240.

Higgens, M. L., and Bera, A. K. (1992), "A Class of Nonlinear ARCH Models," *International Economic Review*, 33, 137–158.

Hyndman, R. (1995), "Highest Density Forecast Regions for Non-Linear Non-Normal Time Series Models," *Journal of Forecasting*, 14, 431–441.

Kaufmann, H. (1987), "Regression Models for Nonstationary Categorical Time Series: Asympototic Estimation and Theory," *The Annals of Statistics*, 15, 79–98.

Kuk, A. Y. C., and Cheng, Y. W. (1997), "The Monte Carlo Newton–Raphson algorithm," *Journal of Statistical Computation and Simulation*, 59, 233–250.

Lawless, J. F. (1987), "Negative Binomial and Mixed Poisson Regression," *The Canadian Journal of Statistics*, 15, 209–225.

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

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), London: Chapman and Hall.

Shephard, N., and Pitt, M. K. (1997), "Likelihood Analysis of Non-Gaussian Measurement Time Series," *Biometrika*, 84, 653–667.

West, M., Harrison, P. J., and Migon, H. S. (1985), "Dynamic Generalized Linear Models and Bayesian Forecasting," *Journal of the American Statistical Association*, 80, 73–96.

Zeger, S. L. (1988), "A Regression Model for Time Series of Counts," *Biometrika*, 75, 822–835.

Zeger, S. L., and Qaqish, B. (1988), "Markov Regression Models for Time Series: A Quasi-Likelihood Approach," *Biometrics*, 44, 1019–1032.