Journal of Modern Applied Statistical Methods

Volume 5 | Issue 1 Article 13

5-1-2006

Properties Of The GAR(1) Model For Time Series Of Counts

Vasiliki Karioti
National Technical University of Athens, ccar@math.ntua.gr

Chrys Caroni
National Technical University of Athens

Journal of Modern Applied Statistical Methods
May, 2006, Vol. 5, No. 1, 140-151

1538 – 9472/06/$95.00

Properties Of The GAR(1) Model For Time Series Of Counts

Vasiliki Karioti Chrys Caroni
Department of Mathematics
National Technical University of Athens

Models for time series count data include several proposed by Zeger and Qaqish (1988), subsequently generalized into the *GARMA* family. The *GAR(1)* model is examined in detail. The maximum likelihood estimation of the parameters will be discussed and the properties of Pearson and randomized residuals will be examined.

Key words: time series, count data, GARMA, GAR(1), residuals

Introduction

Many of the time series recorded in practice consist of count data, in which each observation represents the number of events occurring at a point in time or in a given time interval. Examples include the number of cases of a particular disease reported each month. Especially when the counts are low, standard Gaussian time series models may need to be replaced by other models more suitable for count data, based on the Poisson distribution or another discrete distribution on the non-negative integers.

A number of models of this type have been developed. In this article, regression models for time series count data will be examined. These models, originally proposed by Zeger and Qaqish (1988), have been considered subsequently by several other authors (see, in particular, Kedem and Fokianos, 2002) and extended by Benjamin, Rigby, and Stasinopoulos (2003). In these models, each observation y_t in the series is represented as a Poisson variate which is conditionally independent of previous observations, given its mean, but whose mean depends on the previous observations $y_{t-1},...,y_1$ and possibly on covariates. These are examples of observation-driven models for time-dependent data in the terminology introduced by Cox (1981). In the simplest case, with first-order dependence and no covariates:

$$y_t \mid y_{t-1} \sim Poisson\ (\mu_t)$$

where

$$\mu_t = \mu_t(y_{t-1})\ .$$

In this article, the basic model is examined from several points of view relevant to its practical application to data. Principally, the performance of maximum likelihood estimation of the parameters and the properties of the residuals from the models are examined.

Vasiliki Karioti's main interest is in time series analysis. Email: vaskar@otenet.gr. Chrys Caroni is Associate Professor in the Department of Mathematics. Her research work is mainly on multivariate outliers and reliability. E-mail: ccar@math.ntua.gr

Models

Following Zeger and Qaqish (1988), let y_t be an outcome random variable and x_t an mx1 vector of covariates at time t. Define $\mu_t = E(y_t \mid D_t)$ where $D_t = \{x_t, x_{t-1},\ldots, y_{t-1}, \ldots, y_1\}$ includes past outcomes and the past and present covariates. It is assumed that

$$g(\mu_t) = x_t'\beta + \sum_{i=1}^{p} \theta_i f_i(D_t)$$

where g is a link function, the f_i are functions of the past data and the parameters β and $\theta = (\theta_1, \ldots, \theta_p)'$ are to be estimated. Because the link function is applied to the lagged observations y_{t-j}, this model goes beyond standard generalized linear models (GLM) with independent data (McCullagh and Nelder, 1989). A general model for μ_t is:

$$\begin{aligned} g(\mu_t) &= \eta_t \\ &= x_t'\beta + \sum_{j=1}^{p} \theta_j \left\{ g(y_{t-j}) - x_{t-j}'\beta \right\} \\ &\quad + \sum_{j=1}^{q} \phi_j \left\{ g(y_{t-j}) - \eta_{t-j} \right\} \end{aligned} \tag{1}$$

This defines a class of models called generalized autoregressive moving average models (GARMA: Benjamin, Rigby, and Stasinopoulos, 2003). A special case of GARMA arises when the conditional distribution for y_t (given D_t) is Poisson and g the canonical link function as in standard GLM, that is, the logarithm. Equation (1) becomes:

$$\begin{aligned} g(\mu_t) &= \log(\mu_t) \\ &= x_t'\beta + \sum_{j=1}^{p} \theta_j \left\{ \log(y^*_{t-j}) - x_{t-j}'\beta \right\} \\ &\quad + \sum_{j=1}^{q} \phi_j \left\{ \log(y^*_{t-j}) - \eta_{t-j} \right\} \end{aligned} \tag{2}$$

where $y^*_{t-1} = \max(y_{t-1}, c)$, $0 < c < 1$ (Zeger and Qaqish, 1988; Benjamin, Rigby, and Stasinopoulos, 2003). The effect of using y^*_t in place of y_t is that zero values of y_t are replaced by *c*. This device is adopted in order to avoid an absorbing state at *y = 0*. If $\phi_j = 0$, for $j = 1, \ldots, q$, the model is autoregressive order *p*, *GAR(p)*. If $\theta_j = 0$, for $j = 1, \ldots, p$, it is a moving average model of order *q*, *GMA(q)* (Li, 1994). In the special case of $\phi_j = 0$, and *p* = 1, the model (2) is *GAR(1)* with the form:

$$\log(g(\mu_t)) = x_t'\beta + \theta_1 \left\{ \log(y^*_{t-1}) - x_{t-1}'\beta \right\} \tag{3}$$

If there are no covariates *x*, then writing $x_t'\beta = \mu = \text{constant}$, equation (3) becomes:

$$\mu_t = \exp(\mu) \left[\frac{y^*_{t-1}}{\exp(\mu)} \right]^{\theta_1} \tag{4}$$

Positive values of θ_1 represent positive autocorrelation within the series and negative values represent negative autocorrelation. Zeger and Qaqish (1988) also proposed another way of solving the problem of the absorbing state. Instead of introducing y^*_t, this model defines:

$$\mu_t = \exp(\mu) \left[\frac{y_{t-1} + c}{\exp(\mu) + c} \right]^{\theta_1} \tag{5}$$

where *c* is a constant added to each observation rather than only to zero outcomes. In some situations it might be interpreted as an immigration rate. This model is not part of the GARMA family.

Maximum Likelihood Estimation

The likelihood function conditional on the first term of the series is given by

$$L(y_2, y_3, \ldots, y_m \mid y_1) = \prod_{t=2}^{m} P[Y_t = y_t \mid y_{t-1}] = \prod_{t=2}^{m} \left(\frac{e^{-\mu_t} \mu_t^{y_t}}{y_t!} \right)$$

with log-likelihood

$$\ell = \ln L = \sum_{t \geq 2} \{ -\mu_t + y_t \ln \mu_t \} - \sum_{t \geq 2} \ln y_t!$$

Let the vector of model parameters to be estimated be denoted by η. Then

$$\frac{\partial \ell}{\partial \eta_i} = \sum_{t \geq 2} \left(\frac{y_t}{\mu_t} - 1 \right) \frac{\partial \mu_t}{\partial \eta_i}$$

$$\frac{\partial^2 \ell}{\partial \eta_i \partial \eta_j} = \sum_{t \geq 2} \left(\frac{y_t}{\mu_t} - 1 \right) \frac{\partial^2 \mu_t}{\partial \eta_i \partial \eta_j} - \sum_{t \geq 2} \left(\frac{y_t}{\mu_t^2} \right) \left(\frac{\partial \mu_t}{\partial \eta_i} \right) \left(\frac{\partial \mu_t}{\partial \eta_j} \right)$$

Closed-form expressions are not available for the estimation of η. Consequently, the likelihood must be maximised numerically. The BCOAH subroutine was used from the IMSL library to minimize the negative of the log-likelihood. This employs a modified Newton method and a user-supplied Hessian. Zeger and Qaqish (1988) fitted their models by quasi-likelihood estimation. Benjamin, Rigby, and Stasinopoulos (2003) fitted GARMA models by maximum likelihood using iteratively weighted least squares.

Simulation study

To examine the *GAR(1)* model from several points of view relevant to its practical application to data, a numerical study of simulated data was carried out. The limitation to first-order autoregression is common throughout the time series literature, chiefly for practical reasons (Greene, 2000). Because there is only one autoregressive parameter θ_1, its subscript will be dropped from this point on. To generate a realization of a time series of length *m* for selected values of μ, θ and *c*, the *GAR(1)* model (4) was used to generate a sequence of *m* + 50 counts, starting from a Poisson deviate. The pseudorandom number generator *RNPOI* from the IMSL library was used to generate Poisson deviates. The first 50 counts were discarded and the remaining *m* values were retained for analysis. A relatively short series of *m* = 50 observations and longer series of *m* = 150 observations were examined.

From (4), the parameter *c* appears in the likelihood only in the terms, if any, that immediately follow a zero. If there are few zeros in the series, then there is very little information available for the estimation of *c*. If desired, its estimation can be avoided in order to simplify the likelihood equations. As well, a very flat likelihood surface (with respect to *c*) can be avoided by dividing the series into blocks. A block ends when a zero occurs, and the following block starts with the next non-zero outcome. The overall likelihood is the product of the likelihoods of the separate blocks, each of which is conditional on the first member of the block, and it is a function of θ and μ only. The minor drawback of this procedure is that some information is lost, because the overall likelihood consists not of *m* - 1 but *m* - *1* - m_0 terms, where m_0 is the number of zeros occurring within the series.

Results

Table 1 shows summary statistics for the estimates of θ in the *GAR(1)* model. Difficulties with the numerical fitting procedure prevented the use of the larger values of $|\theta|$ when μ was small. It appears that the maximum likelihood estimate of μ is effectively unbiased, although a minor downward bias appears as θ increases to large positive values. The precision of the estimate of μ increases as μ increases, and appears to be a decreasing function of θ being lowest when θ takes large positive values. Table 2 shows results for the estimation of θ. There is some downwards bias in $|\theta|$, larger when $\theta > 0$ than when $\theta \le 0$, and larger for series of length 50 than ones of length 150. The precision of estimation of θ is also a decreasing function of θ but depends less heavily on the value of μ. Comparison of mean squared errors between Tables 1 and 2 shows that μ is estimated relatively much more precisely than θ.

Table 3 shows the correlation between estimates of θ and μ. Correlations appear to be a decreasing function of θ and also of μ, but do not depend heavily on the length of the series. For the larger values of μ (= 4, 6) and for θ positive or moderately negative, the estimates of the two parameters are virtually uncorrelated.

Table 1. Average and mean squared error of maximum likelihood estimate of μ in the GAR(1) model. Each entry is based on 2,000 simulated sets of data.

		Length 50		Length 150	
μ	θ	mean	m.s.e.	mean	m.s.e.
2	-0.6	2.000	0.0011	2.000	0.0004
	-0.3	1.998	0.0017	2.001	0.0057
	0	1.996	0.0028	1.999	0.0010
	0.3	1.996	0.0057	1.998	0.0021
	0.6	1.976	0.0215	1.990	0.0068
4	-0.8	4.000	0.0001	4.000	0.00004
	-0.6	4.000	0.0001	4.000	0.00005
	-0.3	3.999	0.0002	4.000	0.0001
	0	3.999	0.0004	4.000	0.0001
	0.3	3.999	0.0007	4.000	0.0003
	0.6	3.997	0.0023	3.999	0.0008
	0.8	3.988	0.0102	3.995	0.0032
6	-0.8	6.000	0.00002	6.000	0.00001
	-0.6	6.000	0.00002	6.000	0.00001
	-0.3	6.000	0.00003	6.000	0.00001
	0	6.000	0.00005	6.000	0.00002
	0.3	6.000	0.0001	6.000	0.00004
	0.6	5.999	0.0003	6.000	0.00011
	0.8	5.998	0.0013	5.999	0.0018

Table 2. Average and mean squared error of maximum likelihood estimate of θ in the GAR(1) model. Each entry is based on 2,000 simulated sets of data.

		Length 50		Length 150	
μ	θ	mean	m.s.e.	mean	m.s.e.
2	-0.6	-0.586	0.0087	-0.596	0.0026
	-0.3	-0.300	0.0129	-0.298	0.0044
	0	-0.014	0.0162	-0.007	0.0053
	0.3	0.267	0.0173	0.288	0.0054
	0.6	0.549	0.0165	0.584	0.0046
4	-0.8	-0.776	0.0086	-0.790	0.0025
	-0.6	-0.586	0.0127	-0.596	0.0036
	-0.3	-0.303	0.0168	-0.301	0.0056
	0	-0.021	0.0186	-0.007	0.0065
	0.3	0.258	0.0206	0.285	0.0064
	0.6	0.539	0.0186	0.580	0.0049
	0.8	0.727	0.0168	0.777	0.0033
6	-0.8	-0.777	0.0094	-0.791	0.0027
	-0.6	-0.587	0.0130	-0.597	0.0041
	-0.3	-0.307	0.0177	-0.302	0.0059
	0	-0.021	0.0202	-0.009	0.0069
	0.3	0.258	0.0202	0.285	0.0064
	0.6	0.542	0.0184	0.578	0.0053
	0.8	0.729	0.0162	0.775	0.0034

Table 3. Correlations between maximum likelihood estimates of μ and θ.

	Length 50			Length 150		
θ	μ=2	4	6	μ=2	4	6
-0.8		0.179	0.082		0.201	0.081
-0.6	0.309	0.145	0.040	0.351	0.125	0.049
-0.3	0.246	0.084	0.033	0.246	0.043	0.035
0	0.168	0.060	0.054	0.168	0.069	0.001
0.3	0.108	0.065	0.025	0.150	0.055	0.059
0.6	0.021	0.070	0.006	0.055	-0.015	-0.027
0.8		0.063	0.015		0.008	0.028

Residuals

In any regression model, it is important to examine residuals in order to assess the model's adequacy. Our ability to do this depends quite heavily on whether or not the residuals follow the normal distribution; otherwise it may be difficult to draw conclusions from their behavior. Benjamin, Rigby, and Stasinopoulos (2003) advocated using Dunn and Smyth's (1996) randomized quantile residuals for this purpose, because they expected Pearson or deviance residuals to be highly non-normally distributed for count data, at least when the mean count is low. Randomized quantile residuals are defined by

$$r_t = \Phi^{-1}(u_t) \qquad (6)$$

where Φ^{-1} is the inverse standard normal cumulative distribution function, u_t is a random variable uniformly distributed on the interval $\left[F(y_t\text{-}1;\hat{\mu}_t),\ F(y_t;\hat{\mu}_t)\right]$ and $F(y_t;\hat{\mu}_t)$ is the fitted Poisson cumulative distribution function.

Figures 1-4 show examples of the behavior of ordinary Pearson residuals $(y_t-\hat{\mu}_t)/\hat{\mu}_t^{1/2}$ and randomized quantile residuals in series of length 50, first within a series (all residuals from one simulated series) and then across series (the residual for t=20 examined across all 2000 simulations of the same set of parameter values). Figure 1 shows that, even though the counts are quite low (μ=2), the Pearson residuals within a series do not depart from normality as much as might be expected, so although the randomized quantile residuals (Figure 2) give an improvement, this does not seem to be important. However, across series the Pearson residuals depart markedly from a normal distribution (Figure 3) in the extreme tails whereas the randomized quantile residuals have much better behavior (Figure 4).

In the corresponding Figures 5-8 for series of length 150, it can be seen that the Pearson residuals are quite satisfactory; therefore there is little scope for the randomized quantile residuals to offer any improvement.

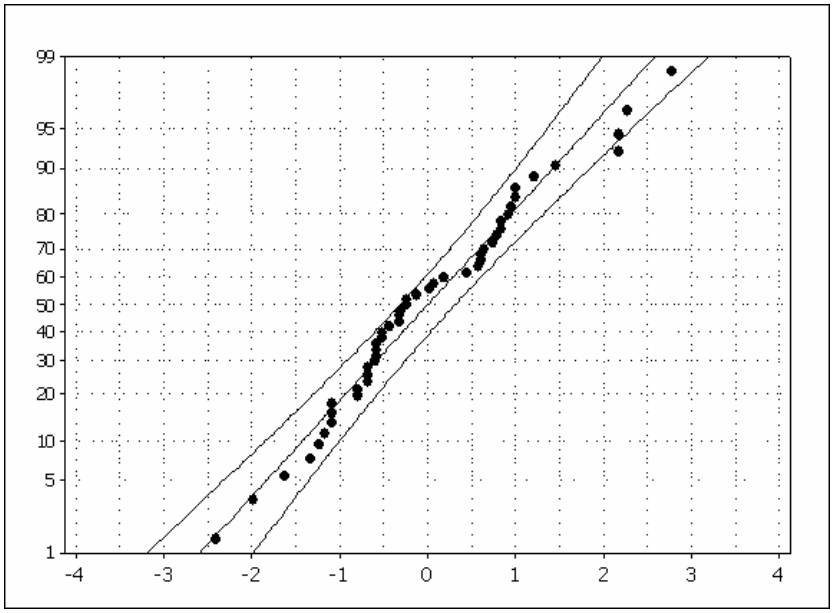

Figure 1. Normal probability plot of Pearson residuals from one realization of *GAR(1)* with m = 50, μ = 2, θ = 0.3.

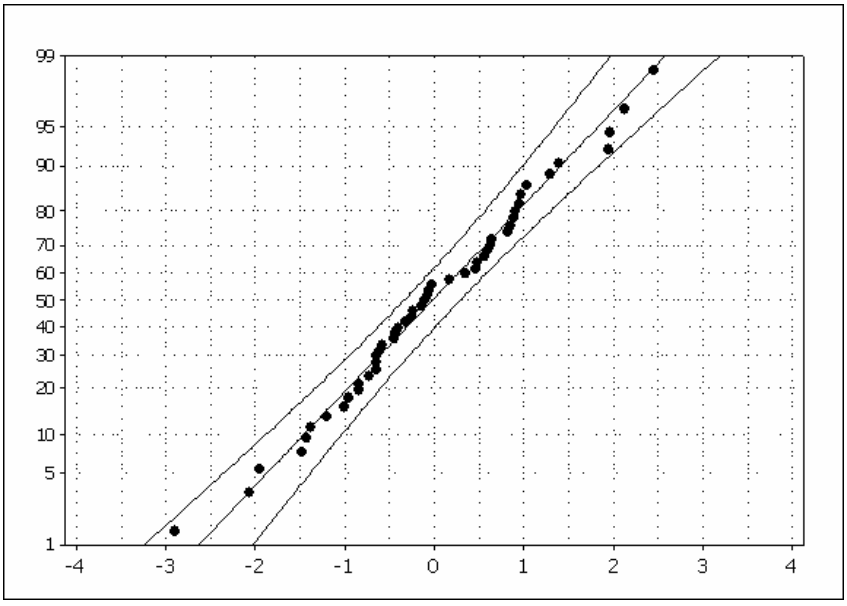

Figure 2. Normal probability plot of randomized residuals from one realization of *GAR(1)* with m = 50, μ = 2, θ = 0.3.

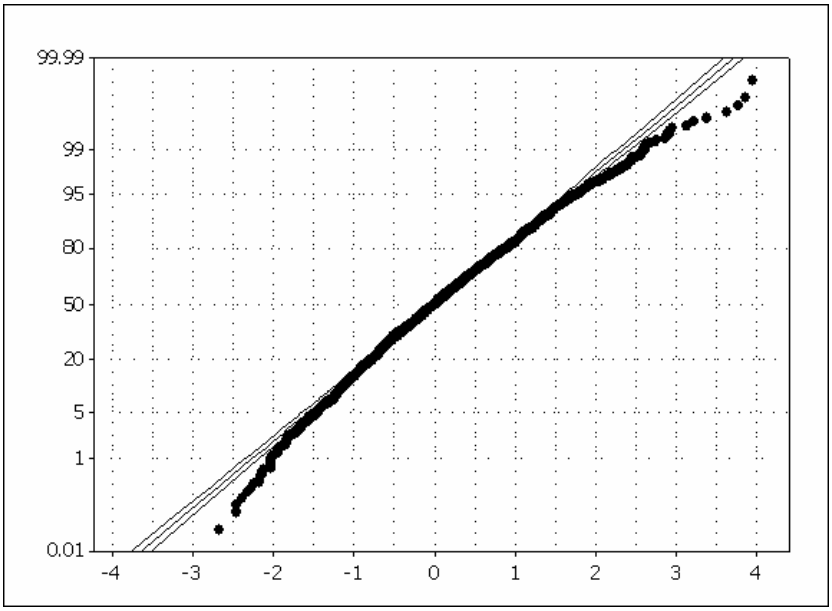

Figure 3. Normal probability plot of Pearson residuals at t = 20 from 2000 realizations of *GAR(1)* with m = 50, μ = 2, θ = 0.3.

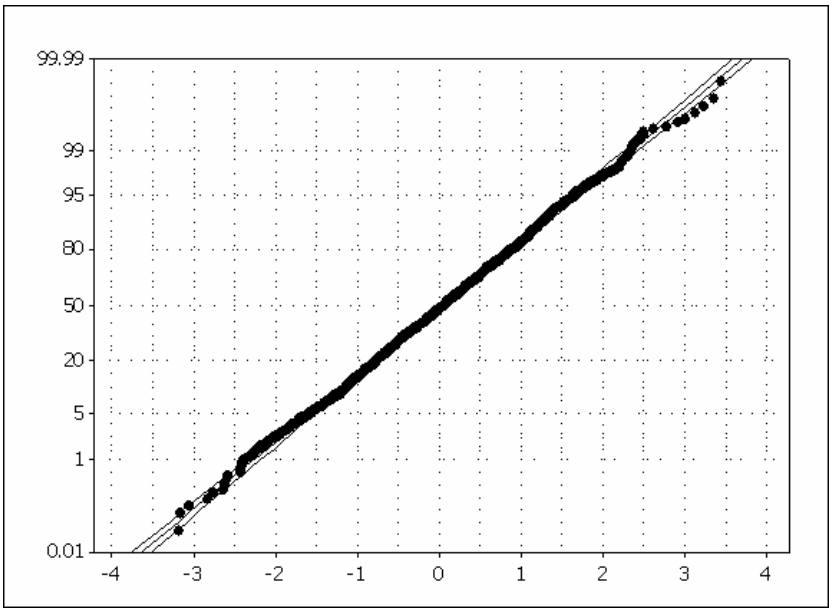

Figure 4. Normal probability plot of randomized residuals at t = 20 from 2000 realizations of *GAR(1)* with m = 50, μ = 2, θ = 0.3.

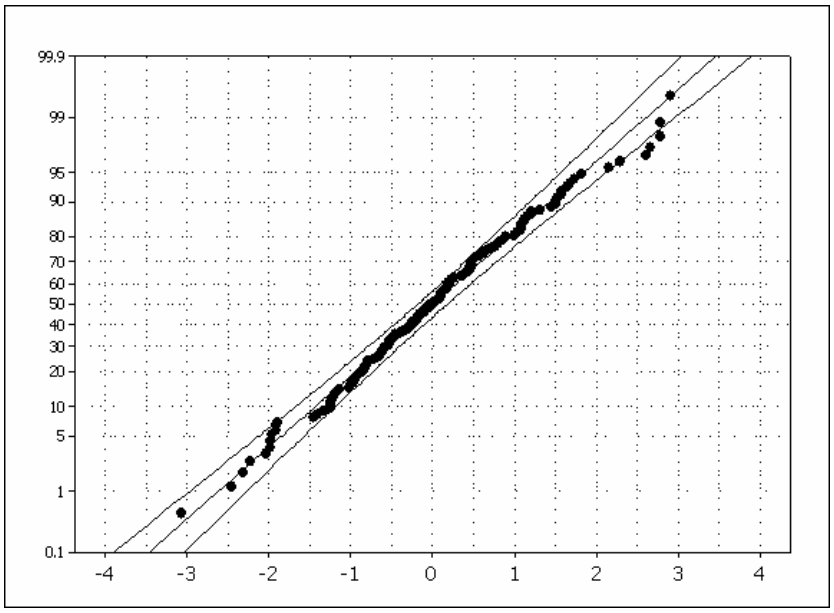

Figure 5. Normal probability plot of Pearson residuals from one realization of *GAR(1)* with m = 150, μ = 4, θ = -0.6.

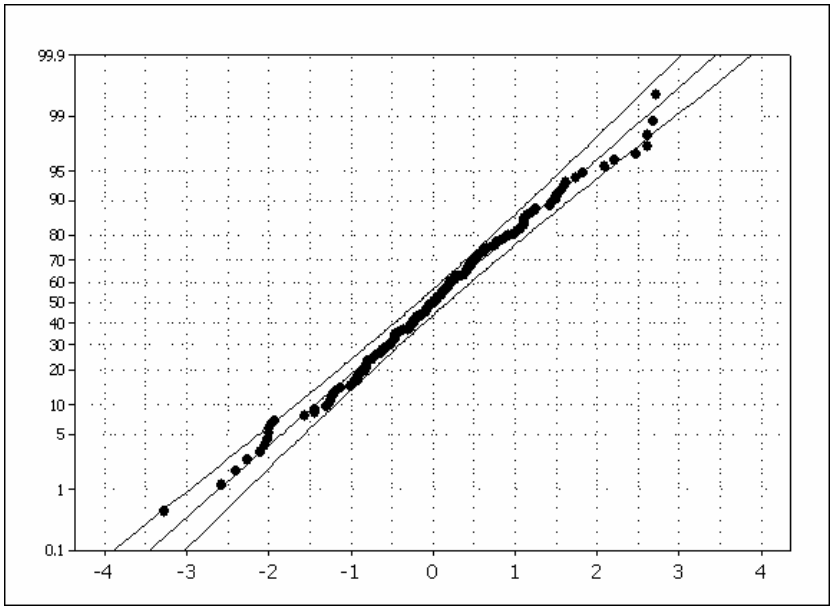

Figure 6. Normal probability plot of randomized residuals from one realization of *GAR(1)* with m = 150, μ = 4, θ = -0.6.

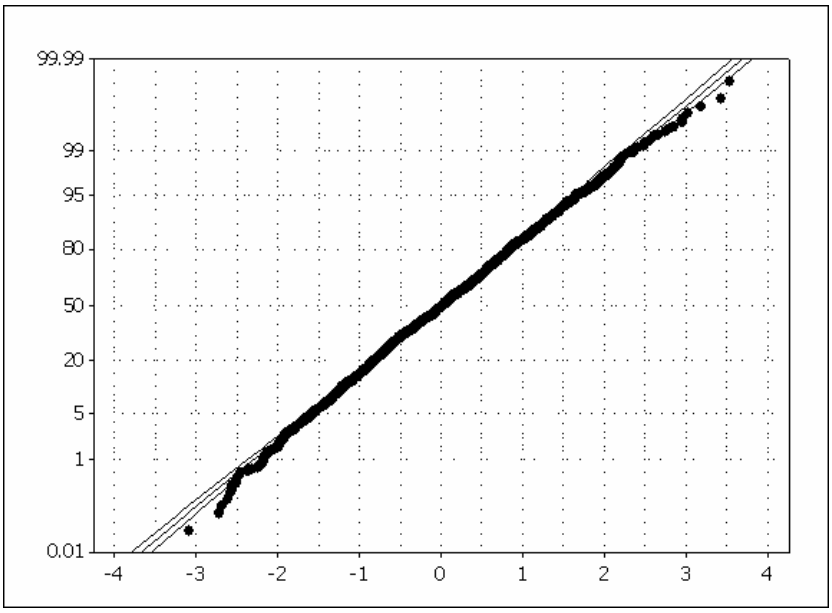

Figure 7. Normal probability plot of Pearson residuals at t = 20 from 2000 realizations of *GAR(1)* with m = 150, μ = 4, θ = -0.6.

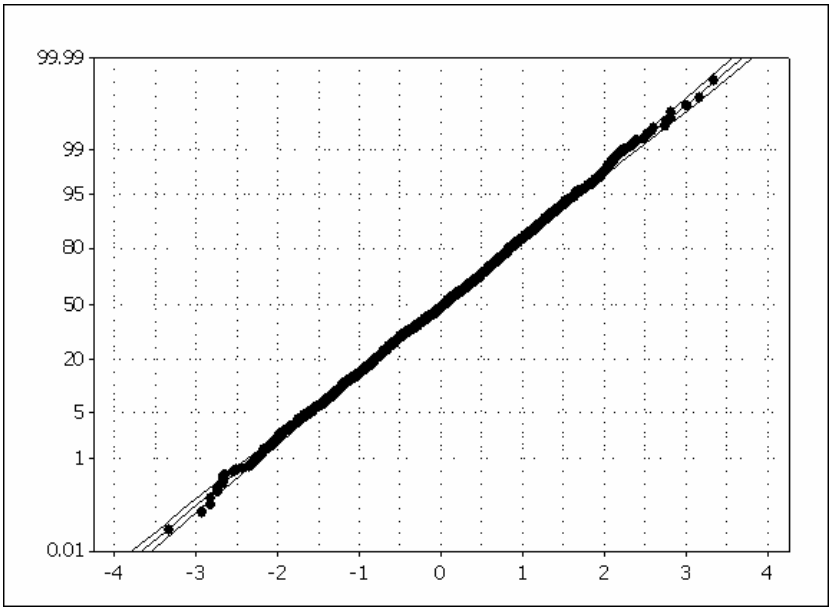

Figure 8. Normal probability plot of randomized residuals at t = 20 from 2000 realizations of *GAR(1)* with m = 150, μ = 4, θ = -0.6.

Table 4 presents results on the distribution of the residuals in relation to the 5% and 1% critical values of the standard normal distribution. Binomial standard errors of these simulated exceedance probabilities with n = 2000 are about 0.5% for the 5% point and about 0.2% for the 1% point. There is a moderate tendency for the exceedance probabilities to be lower than the nominal values, which would lead to conservative tests based on the normal distribution. Fitting logistic regression models with factors μ, θ and the type of residual (Pearson or randomized) confirmed a difference between the exceedance probabilities of the two residuals for m = 50 at the 5% point (logistic regression coefficient for randomized versus Pearson = 0.154 with standard error 0.029) but not at the 1% point (-0.067, s.e. 0.067).

Table 4. Simulated exceedance probabilities (x1000) of normal 5% and 1% critical values of a randomly selected Pearson residual (P) and randomized residual (R). Each entry is based on 2,000 simulations of the GAR(1) model.

		Length 50				Length 150			
		5%		1%		5%		1%	
μ	θ	P	R	P	R	P	R	P	R
2	-0.6	430	445	90	90	475	485	135	80
	-0.3	430	440	60	80	425	470	100	85
	0	500	550	130	130	510	470	65	95
	0.3	460	510	105	95	430	500	90	80
	0.6	370	435	65	65	435	465	80	75
4	-0.8	435	420	110	105	550	565	110	110
	-0.6	415	410	80	75	480	510	75	95
	-0.3	490	495	85	105	445	455	85	100
	0	450	485	115	85	445	435	90	65
	0.3	440	460	110	100	570	560	70	90
	0.6	465	465	130	125	525	505	95	90
	0.8	440	420	85	85	460	485	85	75
6	-0.8	490	485	120	100	505	510	80	90
	-0.6	435	420	80	70	490	495	100	100
	-0.3	415	410	35	40	485	490	95	105
	0	500	505	125	120	515	525	105	115
	0.3	520	545	140	140	545	550	110	130
	0.6	460	470	55	55	505	515	105	115
	0.8	500	490	110	110	535	525	130	135

Conclusion

These results suggest that the *GAR(1)* model without covariates is numerically well behaved, except in the case of the combination of small μ and large $|\theta|$. Restricting the study to *GAR(1)* is not unreasonable, because this is likely to be the most important practical case. According to Greene (2000), "The first-order autoregression has withstood the test of time and experimentation as a reasonable model for underlying processes that probably, in truth, are impenetrably complex" (p.531).

The results also show that the Pearson residuals do not depart from normality as much as might have been expected. However, the randomized residuals are available for use, if preferred, and their distribution seems to be very close to normal. Sometimes there are objections to using randomization within statistical analysis but, as Dunn and Smyth (1996) pointed out, these do not apply when the aim is to look at the overall pattern of residuals, which is what happens when all the residuals within one run are being considered. On the other hand, the random element does become an issue when specific residuals are being examined. This is the case when, for instance, extreme values are under consideration as potential outliers.

Although the simulation results show that the normal distribution applies quite well even at the 1% points, outlier detection may be based on much more extreme values than this (for example, when Bonferroni adjustments are used). Figure 4 compared to Figure 3 and to a lesser extent, Figure 8 compared to Figure 7, show that the randomized residuals would work far better than the Pearson residuals for this purpose. One way of obtaining the advantage of adjusting the residuals, but avoiding randomization, is as follows. Instead of definition (6), define adjusted residuals by

$$r_t^* = \Phi^{-1}\left(u_t^*\right)$$

where u_t^* is the mid-point of the interval $\left[F\left(y_t - 1; \hat{\mu}_t\right), F\left(y_t; \hat{\mu}_t\right)\right]$. In other words, the random variable u_t in (6) is replaced by its expected value. The distribution of these adjusted residuals across series in the simulations was very close to the distribution of the randomized residuals shown in Figures 4 and 8.

One unsatisfactory feature of the model (2) or (4) is the necessity for introducing y_t^*. This is an artificial device to enable the series to restart from zero, which otherwise would be an absorbing state. As remarked above, the amount of information available on the parameter *c* is very small and it is preferred to ignore it entirely by dividing the series up into blocks. This is only an issue when μ is small, because otherwise the chances of reaching zero are negligible. On the other hand, this case may be the most interesting for the application of these models. It is noted that Benjamin, Rigby, and Stasinopoulos (2003) did not discuss this problem and in their example (which includes many zeroes) they appear simply to have used *c* = 0.1 without estimation. Kedem and Fokianos (2002) used examples without zeroes.

During the course of the investigations, the alternative model (5) was also examined. It was found that the likelihood surface tends to be very flat with respect to *c*. Because of this practical problem, but especially because of the dislike of the unrealistic device of adding a constant to every observation, this work has not been pursued and was not reported in this article. Another model, replacing both (4) and (5), could allow a random quantity (independent of other parts of the model and other time periods) to be added to each observation. This could be a much more satisfactory physical model of immigration from elsewhere than is offered by the existing proposals.

References

Benjamin, M. A., Rigby, R. A. & Stasinopoulos, M. D. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association, 98,* 214-223.

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

Dunn, P. K. & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics, 5*, 236-244.

Greene, W. H. (2000). *Econometric analysis* (4th ed.). Upper Saddle River, N.J.: Prentice Hall.

Kedem, B. & Fokianos, K. (2002). *Regression models for time series analysis.* New Jersey: John Wiley.

Li, W. K. (1994). Time series models based on generalized linear models: Some further results. *Biometrics, 50*, 506-511.

McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

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

Zeger, S. L. & Qaqish, B. (1988). Markov regression models for time series: A quasi-likelihood approach. *Biometrics, 44*, 1019-1031.