GENERALIZED LINEAR AUTOREGRESSIONS

BY NEIL SHEPHARD
Nuffield College, Oxford OX1 1NF, UK

February 6, 1995

SUMMARY

This paper develops a class of autoregressive and moving average models which extend the generalized linear model. Likelihood and quasi-likelihood estimation procedures are developed which allow the models to be easily estimated and tested. Several examples are given which illustrate the usefulness and simplicity of the approach advocated in this paper.

Some key words: Autoregression, Diagnostic checking, Generalized linear model, Martingale difference, Overdispersion, Poisson.

1 INTRODUCTION

This paper develops autoregressive and moving average extensions of the generalized linear model. These allow serial dependence by specifying a one-step-ahead prediction distribution, so yielding the likelihood, which frees the predicted mean of the model to depend on lagged values of the observations.

The class of generalized linear models put forward here have

$$f(y_t \mid \mathcal{Y}_{t-1}) = \exp[\{y_t'\theta_t - b(\theta_t)\}/\phi + d(y_t)], \quad t = 1, ..., n, \tag{1}$$

where $\mathcal{Y}_t$ denotes the information available at time t. So θ_t, and consequently the first two moments of $y_t \mid \mathcal{Y}_{t-1}$

$$E(y_t \mid \mathcal{Y}_{t-1}) = \mu_t = b'(\theta_t), \quad \text{var}(y_t \mid \mathcal{Y}_{t-1}) = \phi b''(\theta_t) = \phi V(\mu_t) = \phi V_t. \tag{2}$$

are functions of $\mathcal{Y}_{t-1}$. Here $b'()$ denotes the first derivative of the $b()$ function.

The following linear structure is assumed for the conditional mean:

$$g(\mu_t) = g\{b'(\theta_t)\} = X_t\beta + \tau_t = \eta_t, \tag{3}$$

where $\tau_t = f(\mathcal{Y}_{t-1})$. In this setup, the regressor variables X_t are treated as being fixed. The $g(.)$ function is called the link.

One way of progressing to a useful parametric model is to follow Zeger and Qaqish (1988) and write

$$\tau_t = \sum_{j=1}^{p} \gamma_j f_j(\mathcal{Y}_{t-1})$$

where the f_j are some functions yet to be determined. In a further step Zeger and Qaqish (1988) suggest taking particular forms for the f_j for specific distributional problems. In the Bernoulli example, they suggest taking $f_j(Y_{t-1}) = y_{t-j}$. This gives the Markov chain model suggested by Cox and Snell (1989, pg. 101) for this problem. In the Poisson model they argue for taking $f_j(Y_{t-1}) = \log\{\max(y_{t-j}, c)\}$ where $c \in (0,1)$. In the $p = 1$ case, c and γ_1 control $Pr(y_t = 0 | y_{t-1} = 0)$. In more complicated modelling problems the interpretation of c is less clear. A similar constant is exploited for the gamma distribution.

In this paper I will suggest a new form for f_j, built out of the choice of link function rather than distributional model. I will argue for the building of models based on the adjusted dependant variate:

$$\tau_t = \sum_{j=1}^{p} \gamma_j z(y_{t-j}) \quad \text{where} \quad z(y_t) = \eta_t + g'(\mu_t)(y_t - \mu_t),$$

the Taylor expanded version of $g(y_t)$. This leads to non-linear autoregression models with martingale difference errors:

$$z(y_t) = X_t\beta + \sum_{j=1}^{p} \gamma_j z(y_{t-j}) + g'(\mu_t)(y_t - \mu_t).$$

I will show that these models have attractive properties, are extendible to allow moving average terms and are simple to fit using the usual generalized linear models approach. Section 4 details two examples of the use of these time series models.

2 AUTOREGRESSIVE MODELS

The approach I suggest is based on the fact that

$$E(y_t \mid \mathcal{Y}_{t-1}) = \mu_t = g^{-1}(\eta_t), \tag{4}$$

that is, η_t is mapped into the expectation of the series via $g^{-1}(\eta_t)$. This suggests that y_t might be related back to η_t via the inverse of this function, $g(y_t)$. This transformation would allow the lagged observations to enter **additively on the scale of** η_t.

For some models, such as Bernoulli trials, $g(y_t)$ cannot be used, for it leads to singularities. Instead, I propose to generally work with a Taylor expanded version of $g(y_t)$, the adjusted dependent variate (e.g. Firth (1990, pg. 63)), to construct the model. In particular, I define the function

$$\begin{aligned} z(y_t) &= g(\mu_t) + g'(\mu_t)(y_t - \mu_t) \\ &= \eta_t + g'(\mu_t)(y_t - \mu_t). \end{aligned} \tag{5}$$

This gives the time series model studied in this paper $\tau_t = \sum_{j=1}^{p} \gamma_j z(y_{t-j})$. This will be written as an generalized linear autoregressive model of order p, denoted GLAR(p). Notice this part of the model is selected by the choice of link function, not the properties of the relevant observation model.

2.1 Properties

If focus is put on (5) and the special case of the GLAR(1) model $\eta_t = \beta + \gamma z(y_{t-1})$ is considered, then to establish the broad-brush properties of the model it is useful to write $c_t = g'(\mu_t)(y_t - \mu_t)$, and note that c_t is a martingale difference (MD) sequence. For then

$$z(y_t) \quad = \quad \eta_t + c_t = \beta + \gamma z(y_{t-1}) + c_t, \tag{6}$$

an autoregression in $z(y_t)$, with MD errors.

To illustrate the working of this consider the Bernoulli case, where the following 3 links are standard:

	$g(z)$	c_t
Logistic	$\log\{z/(1-z)\}$,	$\dfrac{y_t - \mu_t}{\mu_t(1-\mu_t)}$,
Probit	$\Phi^{-1}(z)$,	$\dfrac{(y_t - \mu_t)}{\phi\{\Phi^{-1}(\mu_t)\}}$,
Log-Log	$\log\{-\log(1-z)\}$,	$(y_t - \mu_t)\dfrac{\mu_t}{(1-\mu_t)}\dfrac{-1}{\log(1-\mu_t)}$,

where $\Phi()$ and $\phi()$ denote the distribution and density of the standard normal distribution. If μ_t does not change very much then these will be close to the Cox and Snell (1989, pg. 101) model.

If y_t has an infinite history then if $|\gamma| < 1$

$$z(y_t) = \frac{\beta}{1-\gamma} + c_t + \gamma c_{t-1} + \gamma^2 c_{t-2} + ..., \tag{7}$$

which implies

$$\begin{aligned} E\{z(y_t)\} &= E[E\{z(y_t) \mid \mathcal{Y}_{t-1}\}] &= E(\eta_t) \\ &= \beta + \gamma E\{z(y_{t-1})\} &= \tfrac{\beta}{1-\gamma}. \end{aligned} \tag{8}$$

It is difficult to show that the c_t process is white noise for all link functions and models, and so in general I cannot directly express the conditions needed for the model to be covariance stationary. This is because

$$\text{var}(c_t) = E\{Var(c_t|\mathcal{Y}_{t-1})\} = \phi E\{g'(\mu_t)^2 V(\mu_t)\}. \tag{9}$$

2.2 Adding moving average terms

The GLAR model, with regressors, has $z(y_t) = X_t\beta + \sum_j \gamma_j z(y_{t-j}) + c_t$. An obvious generalization is the GLARMA model

$$z(y_t) = X_t\beta + \sum_{j=1}^{p} \gamma_j z(y_{t-j}) + \sum_{j=0}^{q} \delta_j c_{t-j}, \text{ where } \delta_0 = 1. \tag{10}$$

3 INFERENCE

One motivation for the design of autoregressive models is to allow the log-likelihood to be built out of predictions. In particular, conditioning on some initial values $\mathcal{Y}_0$ means that

$$\log f(y_1, ..., y_n \mid \mathcal{Y}_0) = \sum_{t=1}^{n} \log f(y_t \mid \mathcal{Y}_{t-1}), \tag{11}$$

can straightforwardly be computed. The score is, writing ω as the $p \times 1$ vector of parameters indexing the model,

$$\frac{\partial \log L}{\partial \omega} = \sum_{t=1}^{n} \frac{\partial \mu_t'}{\partial \omega} \frac{\partial \theta_t}{\partial \mu_t'} \frac{\partial \log f(y_t \mid \mathcal{Y}_{t-1}; w)}{\partial \theta_t} = \frac{1}{\phi} \sum_{t=1}^{n} D_t' V_t^{-1} (y_t - \mu_t), \tag{12}$$

where $V_t = V(\mu_t) = \partial \mu_t' / \partial \theta_t$ and $D_t' = \partial \mu_t' / \partial \omega$ which is a $p \times k$ matrix. Further, ignoring terms with zero expectation (under correct specification of the conditional mean) gives:

$$\frac{\partial^2 \log L}{\partial w \partial w'} = -\frac{1}{\phi} \sum_{t=1}^{n} D_t' V_t^{-1} D_t. \tag{13}$$

The derivatives D_t can be computed recursively. For ease of exposition let $\eta_t = x_t'\beta + z(y_{t-1})\gamma$ be univariate. Then

$$D_t' = \frac{\partial \eta_t}{\partial \omega} \frac{\partial \mu_t}{\partial \eta_t}, \frac{\partial \eta_t}{\partial \omega} = \left[\left\{ \begin{array}{c} x_t \\ z(y_{t-1}) \end{array} \right\} + \frac{\partial z(y_{t-1})}{\partial \omega} \gamma \right] \quad \text{and} \quad \frac{\partial z(y_t)}{\partial \omega} = \frac{\partial \eta_t}{\partial \omega} + \frac{\partial c_t}{\partial \mu_t} D_t'.$$

The model can be estimated using a Newton-Raphson technique until convergence. Unfortunately, unlike the generalized linear model with log concave links (Wedderburn (1976)), there seems to be no guarantee that the log-likelihood has a unique maximum. Indeed parallels with the Gaussian model suggest that this may be a particular problem when the model includes moving average terms.

3.1 Estimating the dispersion

In generalized linear models there are some difficulties associated with simultaneously estimating β and ϕ and so the conventional moment based estimator

$$\widehat{\phi} = \frac{1}{n-p} \sum_{t=1}^{n} \frac{(y_t - \mu_t)^2}{V_t}, \tag{14}$$

will be used. Using this estimated value within the second derivative matrix allows a consistent estimate of the standard errors of the regression parameters under overdispersion.

3.2 Dispersion stability

Movements of ϕ away from one can originate from many sources. However, it can be argued that although a model can be very useful even when overdispersion is present, the degree of overdispersion should normally be relatively constant throughout the time series sample.

To assess the stability of the parameter the sample can be split into three, and estimates of ϕ can be made in the first and last of those samples. Here ω is estimated using the whole of the data. The ratio of these two estimates form the statistic

$$H = \sum_{t=1}^{[n/3]} \frac{(y_t - \mu_t)^2}{V_t} / \sum_{t=[2n/3]}^{n} \frac{(y_t - \mu_t)^2}{V_t}, \tag{15}$$

which will be scaled by reference to a $F_{[n/3],[n/3]}$ distribution under the null of no change in the dispersion parameter. Significantly large or small values of H will be taken as evidence of parameter instability.

3.3 Quasi-likelihood inference

If the model is misspecified then the objective function will be called a quasi-likelihood function. As D_t and V_t are determined by lagged information the score is the sum of martingale differences if the conditional mean is correctly specified. This implies the estimating equation maintains its zero mean if the mean is correctly specified.

Using the martingale difference feature of $(y_t - \mu_t)$, then for $t > s$ the covariance between $y_t - \mu_t$ and $y_s - \mu_s$ is zero, which means that a consistent estimate of the covariance of the score can be computed using

$$C_{n1} = \frac{1}{n\phi^2} \sum_{t=1}^{n} D_t' V_t^{-1} (y_t - \mu_t)(y_t - \mu_t)' V_t^{-1} D_t. \tag{16}$$

If the additional assumption that the model correctly specifies $Var(y_t|Y_{t-1})$ is made, then this simplifies to

$$C_{n2} = \frac{1}{n\phi} \sum_{t=1}^{n} D_t' V_t^{-1} D_t, \tag{17}$$

which is also the expectation of the second derivative of the quasi-likelihood under just the correct specification of the conditional mean. Thus the information equality holds if the $\mathrm{var}(y_t|Y_{t-1})$ is correctly specified and so the maximum quasi-likelihood estimator's asymptotic properties are the same as that of the maximum likelihood's to the first order; see McCullagh (1983).

When the conditional mean is correctly specified, but the variance is modelled incorrectly, then this simplification cannot be used. Instead the covariance of the estimator has to be robustified. One way of achieving this is to use the sandwich estimator.

Extending this type of argument to pseudo-likelihood methods (Carroll and Ruppert (1988)) is straightforward and useful for problems where it is important to directly model the variance function.

3.4 Diagnostic checking

Diagnostics for the time series models will be built out of the so called deviance residuals. These are discussed at some length by McCullagh and Nelder (1989, pg. 39) and Pregibon (1981). They take on the form of the individual observations contribution to the total deviance $r_t = sign(y_t - \mu_t)d_t^{1/2}$ where $D(y;\widehat{\mu}) = \sum r_t^2$. For many distributions the r_t will be close to being normally distributed and can be used as inputs into the correlogram and Box-Ljung statistic (Ljung and Box (1978)).

3.5 Conditioning on initial values

Throughout this section inferences have been made conditionally on some initial values $\mathcal{Y}_0$. In practice it is not immediately clear how to construct $\mathcal{Y}_0$. In the Gaussian case there are two basic strategies for initialising ARMA models. The first assumes stationarity and completes the model by adding the unconditional distribution of $\mathcal{Y}_0$ to the likelihood (Harvey (1993, Ch. 3)). This does not seem feasible for the non-Gaussian case. The second conditions on some initial observations (inducing a 'marginal' likelihood) and sets the lags of the moving average, c_t, to be zero. As a result for GLARMA(p,q) models the first p observations would be used to construct the initial values while the remaining $n-p$ observations determine the likelihood given $z(y_p), ..., z(y_1)$.

A difficulty with this approach is that it is not possible to construct $z(y_{t-1})$ immediately from raw data without past values to compute η_{t-1}. The best solution I have found to this problem is to fall back on the Zeger and Qaqish (1988) approach of using ad hoc rules to approximate $z(y_p)$ by an altered $g(y_p)$. The suggested forms are:

Identity	y,	Logistic	y,
Logarithm	$\log\{\max(0.5, y)\}$,	Probit	y.
Square-Root	$y^{1/2}$,		

As a result the initial values are parameter free. Although not entirely satisfactory, this approach does yield stable solutions.

3.6 Reparameterisation

For simplicity of exposition I will assume that the series is univariate and that the model is driven by the variable $\eta_t = X_t\beta + \sum_{j=1}^{p} \gamma_j z(y_{t-j})$. This parameterisation of the process may not be the most useful. First, if the series is changing slowly over time, then the $z(y_{t-j})$ are likely to be quite highly correlated with one another, leading to the usual multicollinearity problems. Second, it may be useful to reparameterise the model to allow a direct estimate of the long-run effect, $\beta/(1-\gamma)$, of the lagged transformed observations.

Both of these problems can be tackled by working with the parameterization

$$\eta_t = X_t\beta + \gamma z(y_{t-1}) + \sum_{j=1}^{p-1} \gamma_j^* \triangle z(y_{t-j}), \quad \text{where} \quad \gamma = \sum_{j=1}^{p} \gamma_j \text{ and } \gamma_j^* = -\sum_{s=j+1}^{p} \gamma_s, \tag{18}$$

a trick which is used in Gaussian autoregressions: see Harvey (1993, pg 159).

4 EXAMPLES

4.1 Sudden Infant Death Syndrome (SIDS)

Campbell (1994) used the Zeger (1988) regression model for counts to analyse the relationship between SIDS and temperature. The hypothesis of interest is that a short-term drop in temperature leads to a rise in the rate of SIDS. In this paper, I analyse a very slightly updated version of the UK data taken from 8th January 1979 to 31st December 1983. In total there are 1813 daily observations. The basic model exploited here comes from Table 2 in Campbell (1994), which relates SIDS to the logarithm of the number of babies at risk, the mean temperature for lags 2-5 days and a dummy weekend effect which takes on the value one for a weekend and zero elsewhere.

The results from Campbell's basic poisson regression (the GLAR(0)) model are given in Table 1, which shows all three variables are 'significant'. The correlogram for the deviance residuals is given in Fig 1, while the Box-Ljung statistic is very large. Campbell corrects for the serial dependence using the Zeger (1988) method by employing an AR(7) template for the serial dependence. Campbell's results (using his original data) are given in Table 2 (so the poisson result nearly replicates the GLAR(0) result of Table 1). The Zeger method has increased the standard error on the coefficient of interest by a factor of 9, reducing its t statistic down to around 3. By adding deterministic trends and trigonometric seasonals, Campbell reversed this result and obtained a very large t statistic for temperature.

The first time series model I fit is the GLAR(7) model; moving average models were not very helpful for this data set. I selected this order to give a baseline for the Campbell model selection. I have chosen to present the results in a way which focuses on the long-run formulation. Table 1 holds most of the results from the fit, but the short-run dynamics are also reported here: the coefficients on $v_{t-1} = z(y_{t-1}) - z(y_{t-2})$ and lagged versions of this. The first line gives the estimates, the second standard errors. All are highly significant.

$$\begin{array}{ccccccc} z(y_{t-1}) & v_{t-1} & v_{t-2} & v_{t-3} & v_{t-4} & v_{t-5} & v_{t-6} \\ .416 & -.376 & -.319 & -.274 & -.226 & -.163 & -.079\ . \\ .045 & .044 & .042 & .039 & .035 & .030 & .022 \end{array} \tag{19}$$

The main results are interesting. The temperature variable is quite stable, with a t statistic of close to 15. The weekend variable has increased, but so has its standard error. The Box-Ljung statistic and overdispersion parameter have both fallen. The fall in the deviance is highly significant.

The highly significant coefficient at lag 7 prompts the fitting of higher order models to this data. The results from fitting an GLAR(15) are given in Table 1 and show a further increase in the estimated weekend effect, but the temperature effect is unchanged. The Box-Ljung statistic is now moderate while the deviance has fallen by 37 compared to the GLAR(7) model. The lags now take on the form:

$$\begin{array}{cccccccc} z(y_{t-1}) & v_{t-1} & v_{t-2} & v_{t-3} & v_{t-4} & v_{t-5} & v_{t-6} & v_{t-7} \\ .489 & -.457 & -.419 & -.389 & -.357 & -.315 & -.251 & -.190 \\ .050 & .051 & .051 & .051 & .050 & .050 & .050 & .049 \end{array}, \qquad (20)$$

$$\begin{array}{ccccccc} v_{t-8} & v_{t-9} & v_{t-10} & v_{t-11} & v_{t-12} & v_{t-13} & v_{t-14} \\ -.170 & -.142 & -.115 & -.061 & -.023 & -.019 & .007 \\ .048 & .046 & .044 & .040 & .036 & .031 & .023 \end{array}.$$

This strongly suggests that this model has dealt with most of the serious autocorrelation structure in the data: an GLAR(10) would probably deal with the data reasonably satisfactorily. The correlogram of the deviance residuals is given in Fig 1.

4.2 Weekly returns on the S&P 500 index: 1962 to 1992

The second series considered is the Wednesday-to-Wednesday 7/11/1962 to 12/30/1992 (n=1591) returns series ($100 \times \Delta \log y_t$) for the US's value weighted S&P 500 market index constructed using the daily CRSP tape. As is typical for financial time series, the squared returns exhibit higher autocorrelations than the returns themselves and these autocorrelations decay slowly, indicating that the variance process is persistent.

To whiten the observations an AR(1) model is fitted to this series. The squares of the resulting residuals are then analysed using gamma based models exploiting a log-link. The first of these is a baseline constant mean. The results can be seen in Table 3 and Fig 2 and again show the persistence in the volatility, now the mean, process. Notice the figure presents results using the absolute residuals from the AR(1) fit, rather than the squares, as this displays more of the details.

The first time series model considered is an GLAR(7); the fit is reported in Table 3. Again the table uses a long-run parameterisation: the short-term dynamics are estimated to be:

$$\begin{array}{ccccccc} z(y_{t-1}) & v_{t-1} & v_{t-2} & v_{t-3} & v_{t-4} & v_{t-5} & v_{t-6} \\ .613 & -.479 & -.374 & -.354 & -.271 & -.208 & -.125 \\ .040 & .039 & .039 & .038 & .036 & .031 & .023 \end{array}. \qquad (21)$$

This shows that the serial correlation is quite persistent, requiring a model which deals with longer degrees of dependence. To attempt to produce a reasonably parsimonious model, a GLARMA(1,1) model is estimated. Again the results are reported in Table 3 and Fig 2. It has a much better fit than the GLAR(7) model, even though it has fewer parameters. It achieves this by having much more persistence as the parameter on $z(y_{t-1})$ is close to one. The moving average parameter is close to minus one and so there is a near-cancellation of unit roots, which is required for the autocorrelation function to have the low correlation/persistent feature which is vital for this data set. This model analogous the very successful GARCH(1,1) model originally suggested by Bollerslev (1986).

One difficulty with modelling financial data is the occasional presence of a large outlier. This is a problem in this data set, for the crash of 19th October 1987 and its aftermath produce an observation of 277, while for the GLARMA(1,1) model the predicted mean and standard error are both 5.5. As a result, the observation is 50 standard deviations away from its predicted value and has a residual deviance of -9.5! It is important to check that the inferences are robust to this outlier. This is done by reducing this aberrant observation from 277 down to 50. The results are given in Table 4 and suggest that the model is little affected by this behaviour, although the overdispersion parameter falls quite significantly. The H statistic is still small, but this is caused by the volatile period following the crash and does not dramatically influence the parameter estimates.

5 Conclusion

This paper I have proposed an autoregressive extension of the generalized linear model. The resulting GLAR models are easy to fit and check. As the are non-linear autoregressions with martingale difference errors, their properties are straightforward to find.

Acknowledgements

I would like to thank David Firth for his encouragement and helpful advice on various aspects of this paper and Rob Engle and Sir David Cox for useful conversations. I would also like to thank Maxine Brant for her keyboarding skills and ESRC for their financial support. Finally I thank Mike Campbell for allowing me to use his SIDS data and Sangjoon Kim for constructing the weekly returns on the S&P500 index used in this paper .

List of Figures

Figure 1. 1a: Campbell's(1994) daily count of the number of UK Sudden Infant Deaths. 1b: Correlogram for the deviances from the poisson GLIM fit, X-axis scale is years. 1c: Correlogram for the deviances from the GLAR(7) fit, X-axis scale is years. 1d: Correlogram for the deviances from the GLAR(15) fit, X-axis scale is years.

Figure 2. 2a: S&P500 Absolute values of the daily returns ($100 \times \Delta \log y_t$) on the S&P500 index. 2b: Correlogram for the deviances from a gamma GLIM fit (constant mean). 2c: One-step-ahead predicted mean from a GLARMA(1,1) fit. 2d: Correlogram for the deviances from a GLARMA(1,1) fit.

References

Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *J. Econometrics 51*, 307–327.

Campbell, M. J. (1994). Time series regression for counts: an investigation into the relationship between Sudden Infant Death Syndrome and environmental temperature. *J. R. Statist. Soc. A 157*, 191–208.

Carroll, R. J. and D. Ruppert (1988). *Transformations and Weighting in Regression.* London: Chapman and Hall.

Cox, D. R. and E. J. Snell (1989). *The Analysis of Binary Data* (2 ed.). London: Chapman and Hall.

Firth, D. (1990). Generalized linear models. In D. V. Hinkley, N. Reid, and E. J. Snell (Eds.), *Statistical theory and Modelling*, pp. 55–82. London: Chapman and Hall.

Harvey, A. C. (1993). *Time Series Models* (2 ed.). New York: Philip Allan.

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models. *Biometrika 66*, 67–72.

McCullagh, P. (1983). Quasi-likelihood functions. *Ann. Statist. 11*, 59–67.

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

Pregibon, D. (1981). Logistic regression diagnostics. *Ann. Statist. 9*, 705–724.

Wedderburn, R. W. M. (1976). On the existence and uniqueness of the maximum likelihood estimates of certain generalized linear models. *Biometrika 63*, 27–32.

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

Zeger, S. L. and B. Qaqish (1988). Markov regression models for time series, a quasi likelihood approach. *Biometrics 44*, 1019–1032.

GLAR	(0)		(7)		(15)	
Variable	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
LAR	3.223	0.6390	3.448	1.082	2.950	1.2557
Weekend	0.104	0.0291	0.167	0.0489	0.205	0.0559
Meantemp	-0.067	0.0026	-0.070	0.0040	-0.075	0.0042
$z(y_{t-1})$			0.416	0.023	0.489	0.0504
ϕ	1.109		1.057		1.045	
Dev & Mean Dev	2174	1.202	2069	1.148	2032	1.133
BL(25) & H	248.1	0.726	37.6	0.826	21.7	0.824

Table 1: Fitting of the main GLAR models for the revised Campbell SIDS data. The numbers at the top of the Table indicate the order of GLAR model. LAR: log of the number of babies at risk. Weekend: weekend dummy, 1 at weekend. Meantemp: average temperature in last 2-5 days. BL(25): Box-Ljung statistic with 25 lags. Dev: Total deviance and Mean Dev: Mean Deviance. H is the dispersion stability statistic.

	Poisson		Zeger	
Variable	Estimate	S.E.	Estimate	S.E.
LAR	3.021	0.593	3.250	0.913
Weekend	0.106	0.028	0.104	0.016
Meantemp	-0.067	0.002	-0.056	0.018

Table 2: Part of Table 3 in Campbell(1994) SIDS data. LAR: log of number of babies at risk. Weekend: weekend dummy, 1 at weekend. Meantemp: average temperature in last 2-5 days.

GLARMA	(0,0)		(7,0)		(1,1)	
Variable	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Const	1.408	0.061	1.342	0.076	1.229	0.189
$z(y_{t-1})$			0.613	0.040	0.972	0.009
c_{t-1}					-0.896	0.019
ϕ	5.887		2.685		3.628	
Dev & Mean Dev	4612	2.902	4149.5	2.623	4118.0	2.595
BL(25) & H	511.8	0.109	32.6	0.360	21.6	0.223

Table 3: S&P500 Weekly data. The numbers at the top of the Table indicate the order of GLARMA model. Only explanatory variable used in this model is a constant. BL(25): Box-Ljung statistic with 25 lags. Dev: Total deviance and Mean Dev: Mean Deviance. H is the dispersion stability statistic.

GLARMA	(0,0)		(7,0)		(1,1)	
Variable	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Const	1.372	0.046	1.330	0.068	1.187	0.189
$z(y_{t-1})$			0.615	0.038	0.977	0.008
c_{t-1}					-0.890	0.017
ϕ	3.395		2.215		2.202	
Dev & Mean Dev	4502.1	2.833	4085.7	2.583	4011.2	2.527
BL(25) & H	515.1	0.340	30.7	0.522	20.9	0.605

Table 4: S&P500 Weekly data. Only explanatory variable used in this model is a constant. The models reported in this table are the same as in Table 3 except the 19th October 1987 observation has been changed from 277 to 50. BL(25): Box-Ljung statistic with 25 lags. Dev: Deviance and Mean Dev: Mean Deviance. H is the dispersion stability statistic.

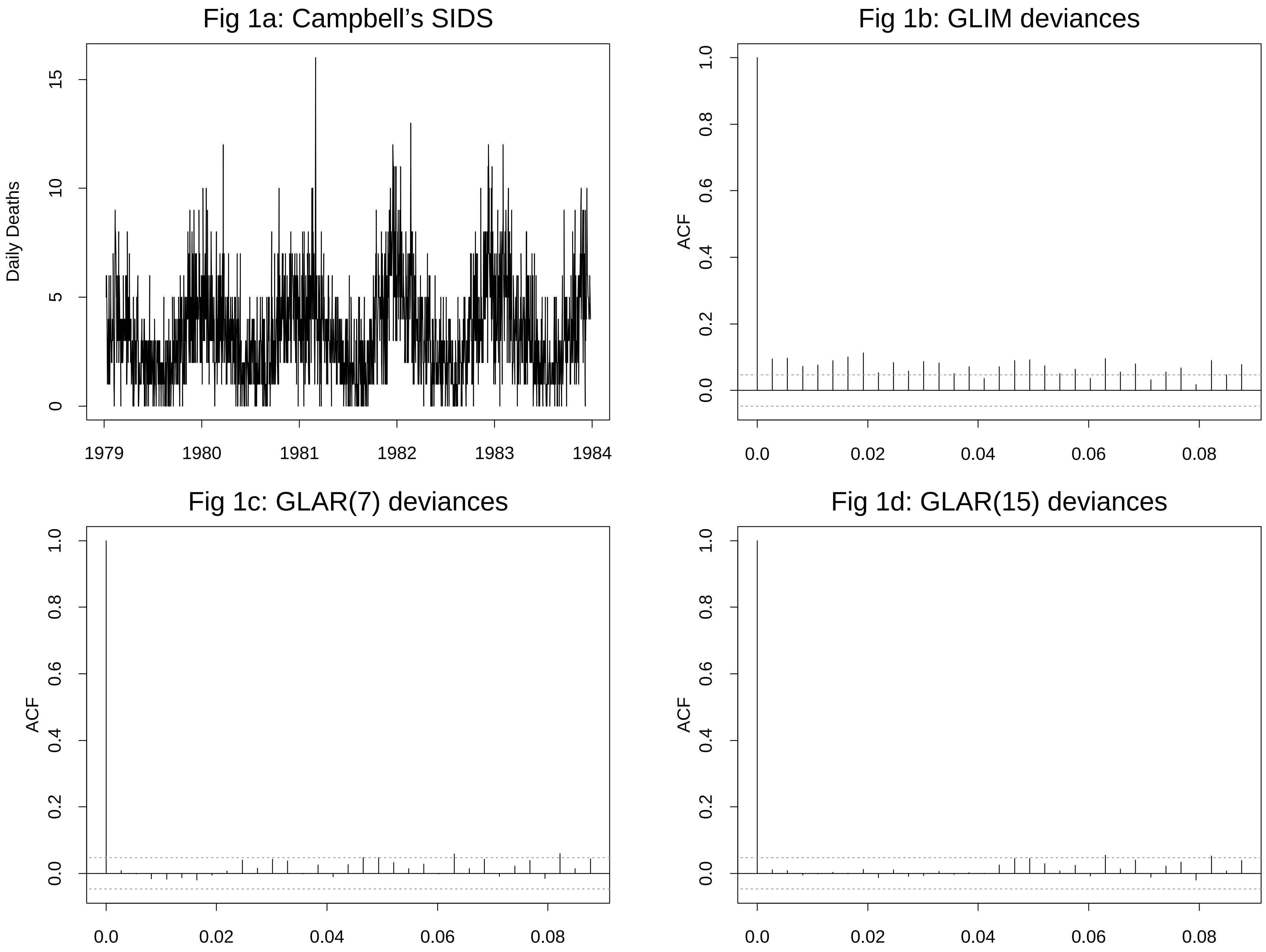
Fig 1a: Campbell's SIDS
Daily Deaths
0
5
10
15
1979
1980
1981
1982
1983
1984
Fig 1b: GLIM deviances
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.02
0.04
0.06
0.08
Fig 1c: GLAR(7) deviances
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.02
0.04
0.06
0.08
Fig 1d: GLAR(15) deviances
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.02
0.04
0.06
0.08

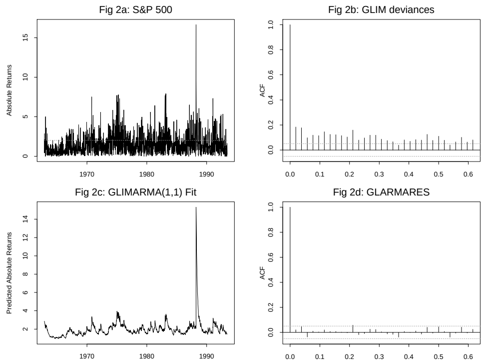
Fig 2a: S&P 500
Absolute Returns
Fig 2b: GLIM deviances
ACF
Fig 2c: GLIMARMA(1,1) Fit
Predicted Absolute Returns
Fig 2d: GLARMARES
ACF