Modeling time series with zero observations

Andrew Harvey and Ryoko Ito*

Faculty of Economics, Cambridge University and

Department of Economics and Nuffield College,

Oxford University

February 21, 2017

Abstract

We consider situations in which a significant proportion of observations in a time series are zero, but the remaining observations are positive and measured on a continuous scale. We propose a new dynamic model in which the conditional distribution of the observations is constructed by shifting a distribution for non-zero observations to the left and censoring negative values. The key to generalizing the censoring approach to the dynamic case is to have (the logarithm of) the location/scale parameter driven by a filter that depends on the score of the conditional distribution. An exponential link function

*Corresponding author. Email: ryoko.ito@economics.ox.ac.uk.

means that seasonal effects can be incorporated into the model and this is done by means of a cubic spline (which can potentially be time-varying). The model is fitted to daily rainfall in northern Australia and compared with a dynamic zero-augmented model.

KEYWORDS: Censored distributions; dynamic conditional score model; generalized beta distribution; rainfall; seasonality, zero augmented model.

JEL: C22

1 Introduction

Situations arise in which time series observations cannot be negative, but a significant proportion of them are identically zero (or have been set to zero because they fall below a certain threshold). The remaining observations are positive and measured on a continuous scale. The usual way of dealing with non-negative observations is to fit a location/scale model. When the occurence of zeroes is too frequent to be compatible with such a distribution, there are two ways of dealing with the problem. The first is by shifting a continuous location/scale distribution to the left and censoring all the negative observations so that they are assigned a value of zero. The second is by the introduction of a binary mechanism which assigns probabilities to zero and positive observations and then draws positive observations from a continuous location/scale distribution. Hautsch et al (2014) - hereafter HMS - adopt this zero-augmented approach for modeling high frequency stock returns.

The key to generalizing the censoring approach to the dynamic case is to set up an observation driven model in which (the logarithm of) the location/scale parameter driven by a filter that depends on the score of the conditional distribution. It can also be used to modify, and we think improve, the zero-augmented approach. However, the censoring model is our main focus because the model is a new one, with a potentially wide range of applications, and its properties are of considerable interest. The score has the important feature of automatically solving the problem of how to weight the zeroes. (It is not correct to assign them a weight of zero or treat them as missing because they are not independent of positive observations.) Furthermore for a fat-tailed distribution, such as generalized beta distribution of the second kind (GB2), the form of the score for location/scale is such that extreme observations are downweighted. Hence the path of the filter is not adversely affected by outliers. Such score-driven models, known as Dynamic Conditional Score (DCS) or Generalized Autoregressive Score (GAS) models, have already proved to be highly effective in a wide range of situations; see, for example, Harvey (2013) and Creal *et al* (2011, 2015).

One reason for setting up a dynamic equation for the logarithm of the location/scale parameter is to ensure that it remains positive. This is of considerable importance in our example, which concerns daily rainfall data with a strong seasonal pattern that we capture by a cubic spline. More generally it allows explanatory variables to be included in the model without the risk of predicting negative values. However, a second reason for modeling the logarithm of location/scale is that the theoretical properties of the model may be derived more easily. The score has the convenient property of having a zero mean and for many continuous (uncensored) distributions its distribution has a known form. For example the score for a GB2 distribution follows a beta distribution (of the first kind).

In an observation-driven model the one-step ahead distribution, and hence the likelihood function, is available by construction. Corresponding parameter-driven models, based on unobserved components, can also be set up but their statistical treatment requires computationally intensive techniques, such as MCMC or particle filtering; see the comments in Allik et al (2016), where the practical importance of simple recursive techniques is stressed.

The censoring approach is closely connected to the dynamic Tobit model, where location changes over time. In the Tobit case, the underlying variable may be defined over the range $-\infty$ to ∞ and is subject to censoring below a certain value, which is usually known (and could be zero). A number of researchers, including Zeger and Brookmayer (1986) and Park *et al* (2007), have addressed this problem for autoregressive and autoregressive moving average (ARMA) models when the underlying (uncensored) observations are Gaussian. Allik et al (2016) generalize to state space models. Our approach could be adapted to such situations by letting the conditional distribution be generalized-t; see McDonald and Newey (1988) and Harvey and Lange (2016). The generalized-t is derived from the GB2 and it includes Student's t and the general error distribution (GED) as special cases. The normal distribution is, of course, a limiting case of Student's t and a special case of GED.

The plan of the paper is as follows. Section 2 examines the score for a censored distribution, with particular reference to GB2 and generalized gamma distributions. The score is then brought into a dynamic censored model in Section 3 and into a dynamic zero-augmented model in Section 4. Models are fitted to daily data on rainfall in northern Australia in Section 5. Diagnostics are discussed and it is shown how the full multi-step predictive distribution can be estimated by simulation. Although such pure time series models cannot outperform a short-term weather forecast, they may be useful for describing and simulating weather patterns. They may also be extended to include explanatory variables, such as air pressure, and so may have a role to play in improving the quality of weather forecasts.

2 Shifting and Censoring

Let x_t be a continuous non-negative variable with scale $\varphi = \exp(\lambda)$ and probability density function (PDF) f_x and define a new variable as

$$y_t = \begin{cases} x_t - c, & x_t > c > 0 \\ 0, & x_t \leq c \end{cases} \tag{1}$$

where c is a constant. The original distribution is therefore shifted to the left[1] and negative values are censored with $\Pr(y_t = 0) = F_x(c)$, where F_x is again the cumulative density function (CDF) of x_t. In this case c is treated as though it were an unknown shape parameter. When c is fixed, $\Pr(y_t = 0)$ will tend to decrease as λ gets bigger.

More generally let $c(\lambda) = \exp(\alpha_0 + \alpha_1\lambda)$, where α_0 and α_1 are parameters. The value of F_x depends on the standardized variable, $x_t \exp(-\lambda)$, so $c(\lambda)\exp(-\lambda) = \exp(\alpha_0 + (\alpha_1 - 1)\lambda)$. Thus $\Pr(y_t = 0) \to 0$ as $\lambda \to \infty$ provided $\alpha_1 < 1$. As a rule, α_1 will be negative.

Let $I(y > 0)$ be an indicator that is zero when $y = 0$ and one when $y > 0$. The distribution of y_t is a discrete-continuous mixture with a point mass at zero, that is

$$\ln f(y_t; \lambda, \boldsymbol{\theta}, c) = I(y_t > 0) \ln f_x(y_t + c) + (1 - I(y_t > 0)) \ln F_x(c), \tag{2}$$

where $\boldsymbol{\theta}$ is a vector of shape parameters and c does not depend on λ. The corresponding score with respect to λ is

$$\frac{\partial \ln f(y_t)}{\partial \lambda} = I(y_t > 0)\frac{\partial \ln f_x(y_t + c)}{\partial \lambda} + (1 - I(y_t > 0))\frac{\partial \ln F_x(c)}{\partial \lambda}. \tag{3}$$

When the score of the uncensored distribution is monotonically increas-

[1] Letting c be negative shifts the distribution to the right. Such distributions are not uncommon in statistics; an application in hydrology can be found in Ahmad et al (1988).

ing,

$$\frac{\partial \ln f_x(0)}{\partial \lambda} \leq \frac{\partial \ln F_x(c)}{\partial \lambda} \leq \frac{\partial \ln f_x(c)}{\partial \lambda}; \tag{4}$$

see Appendix A. In words, the weight attached to a zero observation lies between the weight given to a very small positive observation and the weight given to zero in an uncensored model.

2.1 GB2 distribution

The generalized beta distribution of the second kind, denoted GB2, contains a wide range of distributions for non-negative variables as special cases; see Kleiber and Kotz (2003). The GB2 distribution has PDF

$$f_x(x) = \frac{\upsilon (x/\varphi)^{\upsilon\xi - 1}}{\varphi B(\xi, \varsigma) \left[(x/\varphi)^{\upsilon} + 1\right]^{\xi+\varsigma}}, \qquad x \geq 0, \qquad \varphi, \upsilon, \xi, \varsigma > 0, \tag{5}$$

where φ is the scale parameter, υ, ξ and ς are shape parameters and $B(\xi, \varsigma)$ is the beta function. An exponential link function will be used for the scale parameter, so $\varphi = \exp(\lambda)$. In uncensored models this leads to explicit expressions for the unconditional moments and the information matrix when scale is time-varying; see Harvey (2013, ch 5).

The CDF of x_t, $F(x_t; \upsilon, \xi, \varsigma)$, is a (regularized) incomplete beta function,

$$\beta(z_t; \xi, \varsigma) = B(z_t; \xi, \varsigma)/B(\xi, \varsigma),$$

where $B(z_t; \xi, \varsigma)$ is the incomplete beta function and $z_t = (x_t e^{-\lambda})^{\upsilon}$; see

Kleiber and Kotz (2003, p 184). The incomplete beta function can be written in closed form when ς and/or ξ is one. Note that in many packages, the argument in $\beta(.;\xi,\varsigma)$ is $z_t/(1+z_t)$ rather than z_t.

The tail index is $\eta = \upsilon\varsigma$. The lower this index, the fatter the tail. The $m-th$ moment of a distribution only exists for $m < \eta$. It can be more convenient to replace the parameter ς by the tail index. Redefining scale by replacing φ by $\varphi\eta^{1/\upsilon}$ then gives a reparameterized GB2 with PDF

$$f(y) = \frac{\upsilon(y/\varphi)^{\upsilon\xi-1}}{\varphi\eta^{\xi}B(\xi,\eta/\upsilon)\left[(y/\varphi)^{\upsilon}/\eta+1\right]^{\xi+\eta/\upsilon}}, \qquad \varphi,\upsilon,\xi,\eta > 0. \tag{6}$$

The generalized gamma (GG) is a limiting case of (6), obtained by letting the inverse tail index $\overline{\eta} = 1/\eta \to 0$; see Kleiber and Kotz (2003, p.187) and Harvey and Lange (2016). The incomplete beta functions are replaced by incomplete gamma functions. Note that the $\overline{\eta}$ parametization with $0 \leq \overline{\eta} \leq 1$ tends to be more computationally stable.

The practical relevance of the shifted GB2 can be seen by considering the density at the origin, that is when $y = 0$ and $x = c$. Only when $\upsilon\xi = 1$, is $f_x(0)$ positive and finite, taking the value $f_x(0) = \upsilon/(\varphi B(\xi,\varsigma))$; this is the mode, ie $f_x(0) > f_x(x)$ for $x > 0$. For $\upsilon\xi > 1$, $f_x(0) = 0$ and for $\upsilon\xi < 1$, $f_x(0) = \infty$. In a shifted distribution the ordinate of the continuous part of the y distribution is positive and finite at the origin (for $c > 0$). The potential importance of having a distribution with this property is clear from Figure 1, which shows the histogram of days with positive rainfall in Darwin in

January; the percentage of days with no rainfall was 26.1% (81 in total). The IBM graph of Figure 2 in HMS (2014, p 95) also displays a histogram where the continuous part of the distribution appears to be positive and finite at the origin.

[Figure 1 about here.]

Remark 1 *The F-distribution is a special case of GB2 when* $\upsilon = 1$ *and the degrees of freedom,* ν_1 *and* ν_2*, are the same in the numerator and the denominator. However, even when these restrictions do not hold, the theory for a censored F-distribution is similar to that of a GB2; for example the incomplete beta function for* $F(\nu_1, \nu_2)$ *is* $B(z_c; \nu_1/2, \nu_2/2)$*.*

2.2 Likelihood and score

The first term in (2), the continuous part, is

$$\begin{aligned}\ln f_x(y_t + c) &= \ln \upsilon - \upsilon\xi\lambda + (\upsilon\xi - 1)\ln(y_t + c) - (\xi + \varsigma)\ln(\{(y_t + c)e^{-\lambda}\}^{\upsilon} + 1) - \ln B(\xi, \varsigma) \\ &= \ln \upsilon - \upsilon\xi\lambda + (\upsilon\xi - 1)\ln x_t - (\xi + \varsigma)\ln((x_t e^{-\lambda})^{\upsilon} + 1) - \ln B(\xi, \varsigma).\end{aligned}$$

The score with respect to λ is

$$\frac{\partial \ln f(y_t)}{\partial \lambda} = I(y_t > 0)[\upsilon(\xi + \varsigma)b_t - \upsilon\xi] - (1 - I(y_t > 0))\frac{\upsilon b_c^{\xi}(1 - b_c)^{\varsigma}}{B(z_c; \xi, \varsigma)}, \quad (7)$$

where

$$b_t = b_t(\xi, \varsigma) = \frac{(x_t e^{-\lambda})^{\upsilon}}{(x_t e^{-\lambda})^{\upsilon} + 1} = \frac{z_t}{z_t + 1}, \qquad b_c \leq b_t(\xi, \varsigma) \leq 1. \tag{8}$$

The expression for the second term comes about because $F_x(x_t) = \beta(z_t; \xi, \varsigma)$, which is also the CDF of $b_t(\xi, \varsigma)$, a beta distribution of the first kind. Thus

$$\begin{aligned} \frac{\partial \ln F_x}{\partial \lambda} &= \frac{\partial \ln \beta(z; \xi, \varsigma)}{\partial b} \frac{\partial b}{\partial \lambda} = \frac{1}{\beta(\xi, \varsigma)} \frac{\partial \beta(\xi, \varsigma)}{\partial b} \frac{\partial b}{\partial z} \frac{\partial z}{\partial \lambda} \\ &= \frac{-\upsilon}{\beta(z_c; \xi, \varsigma)} f_b b_c (1 - b_c) = \frac{-\upsilon b_c^{\xi} (1 - b_c)^{\varsigma}}{\beta(z_c; \xi, \varsigma) B(\xi, \varsigma)} \leq 0 \end{aligned} \tag{9}$$

where f_b is the PDF of the $beta(\xi, \varsigma)$ distribution and b_c is b_t defined with $x_t = c$.

Confirmation that the expectation of the score is zero is given in Appendix A. It is also shown that the variance of the score is

$$\begin{aligned} E\left(\frac{\partial \ln f(y_t)}{\partial \lambda}\right)^2 &= \frac{\upsilon^2 b_c^{2\xi} (1 - b_c)^{2\varsigma}}{B(z_c; \xi, \varsigma) B(\xi, \varsigma)} + \upsilon^2 \frac{(\xi + \varsigma)(\xi + 1)\xi}{(\xi + \varsigma + 1)} (1 - \beta(z_c; \xi + 2, \varsigma)) \\ &\quad - 2\upsilon^2 \xi^2 (1 - \beta(z_c; \xi + 1, \varsigma)) + \upsilon^2 \xi^2 (1 - \beta(z_c; \xi, \varsigma)) \end{aligned} \tag{10}$$

Thus the variance depends on λ, which is not the case in the model without censoring.

Remark 2 *The result in (4) implies (when multiplied by minus one) that*

$$\upsilon \xi - \upsilon(\xi + \varsigma) \frac{(ce^{-\lambda})^{\upsilon}}{(ce^{-\lambda})^{\upsilon} + 1} \leq \frac{1}{B(z_c; \xi, \varsigma)} \frac{\upsilon (ce^{-\lambda})^{\upsilon \xi}}{((ce^{-\lambda})^{\upsilon} + 1)^{\xi + \varsigma}} \leq \upsilon \xi.$$

The second inequality is consistent with 26.5.4 in Abramovitz and Stegun (1964, p 944).

As in an uncensored fat-tailed distribution, the score is bounded as $y_t \to \infty$; it tends towards the tail index $\upsilon\varsigma$. On the other hand, as $y_t \to 0$ it tends to $\upsilon(\xi+\varsigma)(ce^{-\lambda})^{\upsilon}/((ce^{-\lambda})^{\upsilon}+1) - \upsilon\xi$ and when $y_t = 0$ it is

$$\frac{-1}{B(z_c;\xi,\varsigma)}\frac{\upsilon(ce^{-\lambda})^{\upsilon\xi}}{((ce^{-\lambda})^{\upsilon}+1)^{\xi+\varsigma}}.$$

When ξ and/or ς is unity the incomplete beta function is available in closed form and the score is easier to compute. For $\varsigma = 1$, the Dagum distribution, $B(b_c;\xi,1) = b_c^{\xi}B(\xi,1) = b_c^{\xi}/\xi$, whereas for $\xi = 1$, the Burr distribution, $B(b_c;1,\varsigma) = (1-(1-b_c)^{\varsigma})B(1,\varsigma) = (1-(1-b_c)^{\varsigma})/\varsigma$; see also Kleiber and Kotz (2003, p. 198, 213). For the log-logistic distribution, when $\xi = \varsigma = 1$, $b_t(1,1)$ is uniform and so $\beta(b_t;\xi,\varsigma) = b_t$. The score is just

$$\frac{\partial \ln f(y_t)}{\partial\lambda} = I(y_t>0)\frac{\upsilon((y_t+c)e^{-\lambda})^{\upsilon}-\upsilon}{((y_t+c)e^{-\lambda})^{\upsilon}+1} - (1-I(y_t>0))\frac{\upsilon}{(ce^{-\lambda})^{\upsilon}+1}.$$

Figure 2 shows the score for the log-logistic distribution with $\upsilon = 2$, unit scale, that is $\lambda = 0$, and c set to 0.5. The value of the score at the origin is -1.6.

[Figure 2 about here.]

The score for the more general formulation, $c(\lambda) = \exp(\alpha_0 + \alpha_1\lambda)$, has

$$\frac{\partial \ln F_x}{\partial \lambda} = \frac{\upsilon(\alpha_1 - 1)b^{\xi}(1-b)^{\varsigma}}{B(z_c; \xi, \varsigma)} \tag{11}$$

because $z_t = (e^{\alpha_0 + (\alpha_1 - 1)\lambda})^{\upsilon}$.

2.3 Generalized gamma

The generalized gamma (GG) distribution is

$$f(y; \varphi, \gamma, \upsilon) = \frac{\upsilon}{e^{\lambda}\Gamma(\gamma)} \left(\frac{y}{e^{\lambda}}\right)^{\upsilon\gamma - 1} \exp\left(-(ye^{-\lambda})^{\upsilon}\right), \quad 0 \leq y < \infty,$$

with $\gamma, \upsilon > 0$ and $-\infty < \lambda < \infty$. The gamma distribution is obtained by setting $\upsilon = 1$, whereas the Weibull sets $\gamma = 1$. The exponential distribution sets $\upsilon = \gamma = 1$. The CDF of the GG is the regularized incomplete gamma function, $\gamma(z; \gamma)$. If $\gamma \to \infty$ in the GG then we get the lognormal distribution, provided additional conditions are put on the other parameters; see Kleiber and Kotz (2003, p.149).

Remark 3 *A slightly different form of the GG is obtained by letting the inverse tail index in the GB2 of (6) go to zero.*

The score for the censored GG distribution is

$$\frac{\partial \ln f(y_t)}{\partial \lambda} = -(1 - I(y_t > 0))\frac{\upsilon z_c^{\gamma} \exp(z_c)}{\Gamma(z_c; \gamma)} + I(y_t > 0)(((y_t + c)e^{-\lambda})^{\upsilon} - \upsilon\gamma),$$

where $\Gamma(z_t;\gamma) = \gamma(z;\gamma)\Gamma(\gamma)$ is the incomplete gamma function and $z_t = g_t = (x_t e^{-\lambda})^{\upsilon}$ is distributed as $gamma(\upsilon, \gamma)$, with PDF f_g at the true parameter values. The first term follows because at $x_t = c$

$$\frac{\partial \ln F_x}{\partial \lambda} = \frac{f_{g,c}}{\gamma(z_c;\gamma)} \upsilon (ce^{-\lambda})^{\upsilon}. \tag{12}$$

Remark 4 *For the Weibull distribution, direct evaluation of the derivative of* $\ln F_x$ *is possible because* $F_x = 1 - \exp(-x^{\upsilon} e^{-\lambda\upsilon})$. *Thus*

$$\frac{\partial \ln f(y_t)}{\partial \lambda} = -(1 - I(y_t > 0)) \frac{\upsilon c^{\upsilon} e^{-\upsilon\lambda} \exp\left(-c^{\upsilon} e^{-\upsilon\lambda}\right)}{1 - \exp\left(-c^{\upsilon} e^{-\upsilon\lambda}\right)} + I(y_t > 0)\upsilon((y_t + c)^{\upsilon} e^{-\lambda\upsilon} - 1).$$

The exponential distribution is a special case in which $\upsilon = 1$.

3 Dynamic model for the censored distribution model

In a parameter-driven model, dynamics would be introduced into λ by letting it follow a stochastic process. In other words it is an unobserved component. Estimation then requires a computer-intensive approach, such as MCMC. By contrast the DCS model is observation-driven, with the predictive distribution defined conditional on a filtered value of λ, denoted $\lambda_{t|t-1}$.

The DCS model for observations generated by the shifted location/scale model, (1) with $x_t = \varepsilon_t \exp(\lambda_{t|t-1})$, is easily implemented. The dynamic equation for the logarithm of location/scale is driven by the score, $\partial \ln f(y_t \mid Y_{t-1}; \boldsymbol{\theta}, c)/\partial \lambda_{t|t-1}$, and it is straightforward to add explanatory variables, contained in a vector $\mathbf{z}_t$. Thus

$$\lambda_{t|t-1} = \omega + \lambda^{\dagger}_{t|t-1} + \mathbf{z}_t' \boldsymbol{\beta} \tag{13}$$

with

$$\lambda^{\dagger}_{t+1|t} = \phi \lambda^{\dagger}_{t|t-1} + \kappa u_t, \tag{14}$$

where u_t denotes the conditional score and $\boldsymbol{\beta}$ denotes a vector of parameters. Clearly $\lambda^{\dagger}_{t|t-1}$ is stationary when $|\phi| < 1$. For a GB2 the score is defined as in (7). The dynamics can be extended by adding lags of $\lambda^{\dagger}_{t|t-1}$ and/or u_t.

The variance of the score depends on $\lambda_{t|t-1}$, unlike in the location/scale model without censoring. Here there is the question as to whether to redefine u_t as the score divided by the information quantity, (10), or its square root; see Creal et al (2013). In our application, u_t is taken to be the raw score. Whichever course of action is taken, the fact remains that the $u_t' s$ are not identically distributed. This point needs to be borne in mind for the development of an asymptotic theory for the ML estimator.

The one-step ahead predictive distribution gives the probability of a zero. The conditional mean and quantiles can also be found. The $\tau - th$ conditional quantile is defined as $\Pr(y_{T+1} \geq q_\tau \mid Y_T)$. Hence $q_\tau = F_x^{-1}(\tau) - c$ for $\tau > F_x(c)$ (that is $y > 0$). For a GB2, $F_x^{-1}(\tau)$ is the inverse regularized incomplete beta function; see Abramovitz and Stegun (1964, p 944-5). When ξ and/or ς

is one, the quantile function is relatively simple. For example, for a Burr distribution, $F^{-1}(\tau) = \exp \lambda_{t|t-1}[(1-\tau)^{-1/\varsigma} - 1]^{1/\upsilon}$.

The conditional mean is

$$E(y_{T+1} \mid Y_T) = \int_0^\infty y f_x(y+c)dy = \int_c^\infty x f_x(x)dx$$

and for a GB2

$$E(y_{T+1} \mid Y_T) = \exp(\lambda_{T+1|T}) \frac{\Gamma(\xi + 1/\upsilon)\Gamma(\varsigma - 1/\upsilon)}{\Gamma(\xi)\Gamma(\varsigma)} (1 - \beta(z_c; \xi + 1/\upsilon, \varsigma - 1/\upsilon)).$$

When $c = 0$ there is no censoring, $\beta(z_c; \xi + 1/\upsilon, \varsigma - 1/\upsilon) = 0$ giving the formula in Kleiber and Kotz (2003, p 188). For the GG distribution

$$E(y_{T+1} \mid Y_T) = \exp(\lambda_{T+1|T}) \frac{\Gamma(\gamma + 1/\upsilon)}{\Gamma(\gamma)} (1 - \gamma(z_c; \gamma + 1/\upsilon)).$$

Multi-step forecasts can be made by simulation. Values of $\lambda_{T+\ell|T}$ are obtained by simulating beta variates, and hence the score, ℓ times. A value of $y_{T+\ell}$ is then simulated, again from a beta (which converts into GB2). This process is then repeated to build up a predictive distribution for $y_{T+\ell}$.

4 Zero-augmented distributions

A zero-augmented point-mass mixture distribution assumes that positive observations occur with probability π and zeroes occur with probability $1 - \pi$.

The positive observations are drawn from a continuous distribution. For the $t-th$ observation

$$\ln f(y_t;\pi) = I(y_t > 0)\ln\pi + I(y_t > 0)\ln f(y_t) + (1 - I(y_t > 0))\ln(1-\pi).$$

The extra parameter is now π, $0 \leq \pi \leq 1$, rather than c. HRS (2014, p 96-7) adopt this formulation with a GB2 distribution (which they call generalized F) for non-zero observations. Their concern is with trading volume and so $I(.)$ is a 'trade indicator' and π denotes the 'trading probability'.

The difference between the zero-augmented and censored distributions is illustrated in Figure 3. The continuous part of the zero-augmented distribution is the original PDF multiplied by π and the probability of a zero is $1-\pi$. By contrast, in the case of censoring, the original distribution - for x - is shifted to the left by c units and the area below the curve and to the left of the vertical axis is the probability of a zero observation. If the scale is changed the probability of a zero in the censored distribution changes accordingly.

[Figure 3 about here.]

When $\upsilon\xi > 1$ for GB2, $f(0) = 0$ and the issue then is whether it is plausible to have zeroes occuring when the probability of values close to zero is very small. On the other hand, when $\upsilon\xi < 1$, $f(0) = \infty$ which is also unappealing. When $\upsilon\xi = 1$ the zero ordinate is positive and finite but this has the limitation that it is the mode: thus it cuts out distributions where the continuous part is away from zero.

The introduction of dynamics for π typically involves involves a logistic transformation. When π_t depends on $\lambda_{t|t-1}$,

$$\pi_{t|t-1} = \frac{\exp(\delta_0 + \delta_1 \lambda_{t|t-1})}{1 + \exp(\delta_0 + \delta_1 \lambda_{t|t-1})}. \tag{15}$$

Thus the π parameter is replaced by two new parameters, δ_0 and δ_1. We expect δ_1 to be positive because as location/scale increases, the probability of a zero, $1 - \pi_{t|t-1}$, is likely to fall.

Remark 5 *Another possibility considered by HMS (2014, p 106-10) is a dynamic equation for* $\ln(\pi_{t|t-1}/(1 - \pi_{t|t-1}))$, *which is unconnected with the equation for* $\lambda_{t|t-1}$; *compare the related autologistic model of Rydberg and Shephard (2003). This model is a plausible one for capturing a trading probability but it is less appealing for modeling rainfall.*

The score, which drives the dynamics for $\lambda_{t|t-1}$ in (14), is

$$\frac{\partial \ln f(y_t)}{\partial \lambda_{t|t-1}} = I(y_t > 0) \left[(1 - \pi_{t|t-1})\delta_1 + \frac{\partial \ln f(y_t)}{\partial \lambda_{t|t-1}} \right] - (1 - I(y_t > 0))\pi_{t|t-1}\delta_1.$$

For GB2,

$$\frac{\partial \ln f(y_t)}{\partial \lambda_{t|t-1}} = I(y_t > 0)[(1-\pi_{t|t-1})\delta_1 + \upsilon(\xi+\varsigma)b_t - \upsilon\xi - (1-I(y_t > 0))\pi_{t|t-1}\delta_1, \tag{16}$$

with $x = y$ in (8).

When $\delta_1 = 0$, so $\pi_{t|t-1}$ is constant, the score wrt $\lambda_{t|t-1}$ is zero[2] for $y_t = 0$. This makes sense because the probability of a zero does not depend on $\lambda_{t|t-1}$. In contrast, HMS (2014, p 101) follow researchers in the MEM literature, such as Bauwens *et al* (2004), and adopt a model in which κu_t in (14) is replaced by

$$\kappa I(y_t > 0)\ln(y_t/\exp(\lambda_{t|t-1})) + \kappa^0(1 - I(y_t > 0)) \tag{17}$$

so there is an additional parameter, κ^0, for zero observations. Note also that whereas the score is bounded for a fat-tailed distribution, as in (16), the response in (17) is unbounded.

Remark 6 *From the properties of an exponential GB2 (EGB2) distribution,* $E[\ln(y_t/\exp(\lambda_{t|t-1}))] = \upsilon^{-1}\pi[\psi(\xi) - \psi(\varsigma)]$, *where* ψ *is the digamma function. Thus the (conditional) expectation of (17) is* $\kappa\upsilon^{-1}\pi[\psi(\xi) - \psi(\varsigma)] + \kappa^0(1-\pi)$ *rather than zero, as for the score.*

Remark 7 *A zero-augmented censored distribution may be constructed to*

[2] It is assumed that the probability of a zero from the continuous distribution is effectively zero.

give

$$\ln f(y_t; c, \pi) = (1 - I(y_t > 0)) \ln\{\pi F_x(c) + 1 - \pi\} + I(y_t > 0) \ln \pi$$
$$+ I(y_t > 0) \ln f_x(y_t + c).$$

Here $\Pr(y_t = 0) = \pi F_x(c) + 1 - \pi$ *which becomes* $F_x(c)$ *when* $\pi = 1$ *(pure censoring) and* $1 - \pi$ *when* $c = 0$ *(pure zero-augmention). When* $\pi = 0$*, positive observations cannot occur, so it is effectively ruled out, that is* $0 < \pi \leq 1$*. Given a sufficiently large sample, fitting the zero-augmented censored model could be the starting point. The censored and zero-augmented models then emerge as special cases.*

5 Daily Rainfall in Northern Australia

Daily rainfall as recorded by the Bureau of Meteorology of the Australian Government (http://www.bom.gov.au/climate/data/) is the total amount, in millimeters (mm), of precipitation that reaches the ground (in a rain gauge) in the 24 hours preceding 9am of each day. Our chosen locations (identified by the associated station number in brackets) are Darwin Airport, Darwin, Northern Territory (014015) and Kuranda Railway Station, Cairns, Queensland (031036). The sample period is a 10-year window from 1 January 2006 to 31 December 2015, which includes[3] 3652 days and two leap

[3] There were a few missing observations; ee http://www.bom.gov.au/climate/cdo/about/about-rain-data.shtml. These were replaced by the average of the two adjacent observations.

years. As can be seen from Table 1, it fails to rain on more than half the days, but when there is rainfall it can be very heavy. The occasional day of very heavy rain is apparent in Figure 1 and the case for using a fat-tailed distribution is a strong one.

[Table 1 about here.]

Figure 4 shows the strong seasonal pattern in rainfall; it tends to be dry during the (Australian) winter, with a high proportion of days when it fails to rain. Figure 5 indicates that the volume of rain is inversely related to the probability of no rain.

[Figure 4 about here.]

[Figure 5 about here.]

Figure 6 shows the empirical distribution of non-zero observations in November in Darwin. The rainfall for January, shown earlier in Figure 1, is much higher. The hope is that when the time-varying location/scale changes, the shape of the censored distribution will adapt appropriately.

[Figure 6 about here.]

5.1 Model

The dynamic censored model has a conditional GB2 distribution in the form (6) and with scale dependent censoring, that is $c(\lambda) = \exp(\alpha_0 + \alpha_1\lambda)$. Thus the score is

$$\frac{\partial \ln f(y_t)}{\partial \lambda} = I(y_t > 0)[\upsilon(\xi + \varsigma)b_t - \upsilon\xi] - (1 - I(y_t > 0))\frac{\upsilon(\alpha_1 - 1)b^{\xi}(1-b)^{\varsigma}}{B(z_c; \xi, \varsigma)}$$

where $z_t = (e^{\alpha_0 + (\alpha_1 - 1)\lambda})^{\upsilon}$. The dynamic equation for the logarithm of location/scale is as in (13) with explanatory variables used to model the seasonal pattern, γ_t, as a deterministic cubic spline function, as in Harvey and Koopman (1992) and Ito (2016). We assume that one seasonal cycle is complete in 365 days, that the pattern of seasonality is fixed and that continuity is maintained from the end of one year to the beginning of the next. A little experimentation indicated that the pattern was well captured by placing the knots of the spline on the 50th day, the 100th day, the 160th day, the 240th day, and the 300th day of the calendar year. It was also found that the GB2 worked best when parameterized in terms of $\overline{\eta}$ and it is these results that are reported.

The zero-augmented model is also based on the GB2, but with a logistic link function, as in (15). The seasonal is modeled by a spline as in the censored model. Table 2 shows goodness of fit statistics, with AIC denoting Akaike Information Criterion and BIC denoting Bayes (Schwartz) Information Criterion. For Darwin the censored model gives the best fit whereas for

Cairns the zero-aumented model is better. Table 3 gives the ML estimates for the preferred models. The first column of numbers is for the censored model with $\alpha_1 = 0$ and the third column of numbers is for the zero-aumented model with $\delta_1 = 0$. In both cases the unrestricted model gives the better fit. LR tests[4] lead to the same conclusion. For Darwin, the LR statistic for the restricted and unrestricted censored models is 17.8 and for Cairns the LR statistic for the zero-augmented models is 413.8.

[Table 2 about here.]

The unrestricted models reported in Table 2 were also fitted without the dynamics of (14). The fit of the dynamic models was much better than the fit of the corresponding static models. For the censored model fitted to Darwin, the static AIC and BICs were 3.24 and 3.27 respectively (as opposed to 3.17 and 3.20). For the zero-augmented model fitted to Cairns the static AIC and BICs were 4.18 and 4.22. respectively.

[Table 3 about here.]

Table 3 shows the estimated parameter values of the preferred model specifications. The parameters of the spline component were estimated simultaneously with these parameters, but they are omitted from the table.

[4]For large sample sizes, Terasvirta and Mellin (1986) argue that the BIC (SIC) can give a better indication of statistical significance than a standard test at a conventional level of significance.

The value of $\widehat{\phi}$ indicates that the non-periodic component, $\lambda_{t|t-1}^{\dagger}$, is comfortably stationary. A negative α_1 means that the probability of no rain increases when heavy rain is unlikely. Although ξ close to one, the null hypothesis that the distribution is Burr is rejected by an LR test at the 5% level of significance - the statistic is 4.09 with a $p-$value of 0.04. The estimated tail-indices are small, reflecting the fatness of the tails apparent in the histograms of Figures 1 and 6.

5.2 Diagnostics

The PITs (probability integral transforms) for the positive observations are

$$PIT_{y>0}(y_t; c_{t|t-1}, \lambda_{t|t-1}) = \frac{F_x(y_t e^{-\lambda_{t|t-1}} + c_{t|t-1}e^{-\lambda_{t|t-1}}) - F_x(c_{t|t-1}e^{-\lambda_{t|t-1}})}{1 - F_x(c_{t|t-1}e^{-\lambda_{t|t-1}})}$$

for the censored model. For the zero-augmented model the PITs are as above but multiplied by $\pi_{t|t-1}$. The GB2 distribution appears to capture the empirical distribution of the data reasonably well in both cases. Figure 7 plots the empirical CDF against the PITs for Darwin.

[Figure 7 about here.]

The correlogram of scores for Darwin is shown in Figure 8. The use of scores for model checking follows from the principles underlying Lagrange multiplier tests[5]. There is no indication of significant residual serial correlation. The diagnostics for Cairns were also satisfactory.

[Figure 8 about here.]

The probability of the next observation being non-zero is given by $F_x(c_{t|t-1}e^{-\lambda_{t|t-1}})$. Figure 9 shows the estimated dynamic probability of no rain tomorrow in 2011; compare Figure 5. The small circles mark days with more than 100mm of rain. Such days roughly coincide with a high probability of rain. As with the scores, the correlogram of the binary variables, shown in Figure 10, gives no indication of serial correlation.

[Figure 9 about here.]

[Figure 10 about here.]

5.3 Out-Of-Sample Performance

The out-of-sample period is from 1 January 2016 to 31 August 2016, which is 244 days. The correlogram of post-sample scores in Figure 11 shows no serial correlation. The same is true for the binary variables (not shown).

[5] Tests against time-variation that has not been captured by the model can, in principle, be constructed using score-based (Lagrange multiplier) tests. Such tests have been shown to be very effective; see Harvey (2013, section 2.5), Harvey and Thiele (2016) and Calvori et al (2016). When the focus of attention is on λ in a location/scale model, score tests differ from tests based on the raw residuals; see, for example, HMS (2014, p 95-6). LM test statistics have not been computed here but the residual correlogram shown here is informative and re-assuring.

The post-sample goodness of fit can be assessed by the predictive likelihood, that is the sum of the logarithms of the log-likelihoods. We can also look at the ability of the model to forecast the binary outcome of whether or not it rains tomorrow. The overall forecasting performance may be measured by the Brier probability score

$$PS = \{F_x(c_{t|t-1}e^{-\lambda_{t|t-1}}) - (1 - I(y_t > 0)\}^2; \tag{18}$$

see Gneiting and Ranjan (2011, p 412). Table 4 gives the average predictive likelihood and the Brier probability score, computed using a uniform weighting scheme. The associated t-statistics are for comparing one-step ahead density forecasts. The t-statistics are negative and statistically significantly showing that the dynamic model specification is preferred.

[Table 4 about here.]

[Figure 11 about here.]

Finally the viability of obtaining multi-step ahead density forecasts by simulation is illustrated in Figure 12. This shows the median and the 75% quantile for Darwin obtained at the start of 2016.

[Figure 12 about here.]

6 Conclusion

Two ways of constructing a point-mass distribution were considered. One, which is new, deals with zeroes by shifting a continuous distribution for non-negative variables to the left and then censoring it at zero. The other augments the continuous distribution with a binary mechanism for zeroes and positive observations. In the dynamic case, the scale changes and is driven by the score. The scale in turn feeds into a mechanism that determines the degree of censoring or the probability of a zero. The application to daily rainfall illustrates the use of these models and shows that both are viable. The censored distribution model appears to be better for the Darwin data whereas for Cairns the zero-augmented model is superior.

The data are highly seasonal and a large part of the change in scale is determined by a seasonal pattern that is parsimoniously modeled by a cubic spline. The seasonal pattern was assumed to be fixed, but for a longer time series it might be possible to capture a changing seasonal pattern, as in Proietti and Hillebrand (2017). Ito (2016) shows how dynamics may be introduced into the seasonal spline. Another unexplored issue concerns the use the use of periodic models, where parameters are different in different seasons; see Hipel and McLeod (1994). The parameters in periodic models are typically the dynamic parameters, such as ϕ and κ, but shape parameters may also change as distributions change with the seasons. For daily data a viable way of including such effects would be to relate the dynamic and/or shape parameters to the seasonal variation captured by the spline. For example, we might use a logistic transformation for the autoregressive parameter. We did not investigate periodic models given our relatively short time series, but it may be a topic for future research.

The model may be extended by including explanatory variables other than seasonals. For example, rainfall depends on air pressure and hence predicted air pressure could be used in conjunction with our models to forecast the probability of rain and give a distribution of rainfall.

Finally it should be noted that our treatment of dynamics for a censored distribution has a wide range of applications in economics and finance.

APPENDIX

A Score inequality

The result is proven by noting that the expectation of the score is zero. In the censored model

$$E\left[\frac{\partial \ln f(y_t)}{\partial \lambda}\right] = F_x(c)\frac{\partial \ln F_x(c)}{\partial \lambda} + \int_c^\infty \frac{\partial \ln f_x}{\partial \lambda} f_x dx = 0,$$

whereas in the uncensored case

$$\int_0^c \frac{\partial \ln f_x}{\partial \lambda} f_x dx + \int_c^\infty \frac{\partial \ln f_x}{\partial \lambda} f_x dx = 0$$

Thus

$$\int_0^c \frac{\partial \ln f_x}{\partial \lambda} f_x dx = F_x(c) \frac{\partial \ln F_x(c)}{\partial \lambda}$$

Because the score is monotonically increasing

$$\frac{\partial \ln f_x(0)}{\partial \lambda} \int_0^c f_x dx \leq \int_0^c \frac{\partial \ln f_x}{\partial \lambda} f_x dx \leq \frac{\partial \ln f_x(c)}{\partial \lambda} \int_0^c f_x dx$$

and (4) follows.

B Expectation and variance of score of a censored GB2

$$\begin{aligned}
E\frac{\partial \ln f(y_t)}{\partial \lambda} &= -F_x(c)\frac{\upsilon b_c^\xi (1-b_c)^\varsigma}{\beta(z_c;\xi,\varsigma)B(\xi,\varsigma)} + \int_{b_c}^1 [\upsilon(\xi+\varsigma)b_t - \upsilon\xi] f_b db_t \\
&= -\frac{\upsilon b_c^\xi (1-b_c)^\varsigma}{B(\xi,\varsigma)} + \int_{b_c}^1 [\upsilon(\xi+\varsigma)b_t] f_b db_t - \upsilon\xi(1-\beta(z_c;\xi,\varsigma)) \\
&= -\frac{\upsilon b_c^\xi (1-b_c)^\varsigma}{B(\xi,\varsigma)} + \upsilon(\xi+\varsigma)\frac{B(\xi+1,\varsigma)}{B(\xi,\varsigma)}(1-\beta(z_c;\xi+1,\varsigma)) - \upsilon\xi(1-\beta(z_c;\xi,\varsigma)) \\
&= -\frac{\upsilon b_c^\xi (1-b_c)^\varsigma}{B(\xi,\varsigma)} + \upsilon\xi(1-\beta(z_c;\xi+1,\varsigma) - \upsilon\xi(1-\beta(z_c;\xi,\varsigma)) \\
&= -\frac{\upsilon b_c^\xi (1-b_c)^\varsigma}{B(\xi,\varsigma)} + \upsilon\xi(\beta(z_c;\xi,\varsigma) - \beta(z_c;\xi+1,\varsigma))
\end{aligned}$$

Note that $z_c = (ce^{-\lambda})^{\upsilon}$ and $b_c = z_c/(z_c + 1)$. Then, using (26.5.16) in Abramovitz and Stegun (1964, p 944),

$$\beta(z_c; \xi, \varsigma) - \beta(z_c; \xi + 1, \varsigma) = \frac{1}{\xi B(\xi, \varsigma)} b_c^{\xi}(1 - b_c)^{\varsigma}.$$

As regards the variance of the score

$$\begin{aligned} E\left(\frac{\partial \ln f(y_t)}{\partial \lambda}\right)^2 &= \frac{\upsilon^2 b_c^{2\xi}(1-b_c)^{2\varsigma}}{\beta(z_c;\xi,\varsigma)B^2(\xi,\varsigma)} + \int_{b_c}^{1} [\upsilon(\xi+\varsigma)b_t - \upsilon\xi]^2 f_b db_t \\ &= \frac{\upsilon^2 b_c^{2\xi}(1-b_c)^{2\varsigma}}{B(z_c;\xi,\varsigma)B(\xi,\varsigma)} + \upsilon^2(\xi+\varsigma)^2 \frac{B(\xi+2,\varsigma)}{B(\xi,\varsigma)}(1-\beta(z_c;\xi+2,\varsigma)) \\ &\quad -2\upsilon^2(\xi+\varsigma)\xi\frac{B(\xi+1,\varsigma)}{B(\xi,\varsigma)}(1-\beta(z_c;\xi+1,\varsigma)) + \upsilon^2\xi^2(1-\beta(z_c;\xi,\varsigma)) \end{aligned}$$

which then gives (10).

ACKNOWLEDGEMENT

Ryoko Ito thanks Nuffield College and the Institute for New Economic Thinking at the Oxford Martin School of Oxford University for their financial support.

REFERENCES

Abramowitz, M. and I. A. Stegun, (eds.), 1964, *Handbook of Mathematical Functions*, New York: Dover Publications Inc.

Ahmad, M. I., Sinclair, C. D. and A. Werritty, 1988, Log-Logistic Flood Frequency Analysis, *Journal of Hydrology*, 98, 205–224.

Allik, B., Miller, C., Piovoso, M. J. and R. Zurakowski, 2016, The Tobit Kalman Filter: An Estimator for Censored Measurements. *IEEE Transactions on Control Systems Technology*, 24, 365-71.

Bauwens, L., Giot, P., Grammig, J. and D. Veredas, 2004, A Comparison of Financial Duration Models via Density Forecasts. *International Journal of Forecasting,* 20, 589-609.

Blasques, F., Koopman, S. J. and A. Lucas, 2014, Maximum Likelihood Estimation for Generalized Autoregressive Score Models. Tinbergen Institute Discussion Paper, TI 2014-029/III, Amsterdam.

Calvori, F., Creal, D.W., Koopman, S.J. and A. Lucas, 2016, Testing for Parameter Instability in Competing Modeling Frameworks, *Journal of Financial Econometrics* (forthcoming).

Creal, D., Koopman, S. J. and A. Lucas, 2011, A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations, *Journal of Business and Economic Statistics* 29, 552-63.

Creal, D., Koopman, S. J. and A. Lucas, 2013, Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics,* 28, 777-95.

Gneiting, T. and R. Ranjan, 2011, Comparing Density Forecasts using Threshold- and Quantile-Weighted Scoring Rules. *Journal of Business and Economic Statistics,* 29, 411-22.

Hautsch, N., Malec, P. and M. Schienle, 2014, Capturing the Zero: New Class of Zero-Augmented Distributions and Multiplicative Error Processes, Journal of Financial Econometrics, 12, 89-121.

Harvey, A.C. , 2013, *Dynamic Models for Volatility and Heavy Tails: with Applications to Financial and Economic Time Series.* Econometric Society Monograph, Cambridge University Press.

Harvey, A.C. and R-J. Lange, 2016, Volatility Modeling with a Generalized t-Distribution, *Journal of Time Series Analysis* (forthcoming).

Harvey, A.C. and S. Thiele, 2016, Testing Against Changing Correlation. Journal of Empirical Finance, 38, 575-89.

Hipel, R. W., and A. I. McLeod, 1994, *Time Series Modelling of Water Resources and Environmental Systems.* Developments in Water Science, 45, Elsevier, Amsterdam.

Ito, R. , 2016, Spline-DCS for Forecasting Trade Volume in High-Frequency Finance, Cambridge Working Papers in Economics (CWPE1606), University of Cambridge.

Kleiber, C. and S. Kotz, 2003, *Statistical Size Distributions in Economics and Actuarial Sciences.* New York: Wiley.

McDonald, J.B. and W.K. Newey, 1988, Partially Adaptive Estimation of Regression Models via the Generalized t Distribution, *Econometric Theory,* 4, 428-457.

Park, J. W., Genton, M. G., and Ghosh, S. K., 2007, Censored Time Series Analysis with Autoregressive Moving Average Models, *Canadian Journal of Statistics*, 35, 151–168.

Proietti, T. and E. Hillebrand, 2017, Seasonal Changes in Central England Temperatures. *Journal of the Royal Statistical Society, Series A (forthcoming)*

Rydberg, T., and N. Shephard, 2003, Dynamics of Trade-by-Trade Price Movements: Decomposition and Models. *Journal of Financial Econometrics* 1, 2–25.

Teräsvirta, T. and I. Mellin, 1986, Model Selection Criteria and Model Selection Tests in Regression Models. Scandinavian Journal of Statistics, 13, 159-171.

Zeger, S. L. and R. Brookmeyer, 1986, Regression Analysis with Censored Autocorrelated Data, *Journal of the American Statistical Association*, 81, 722–729.

TABLES

Town	Fraction of zeroes	75%-quantile	Mean	Max.
Darwin	0.665	1.8	5.2	367.6
Cairns	0.594	4.0	6.9	381.0

Table 1: Sample statistics of rainfall, in millimeters, for the period 1 January 2006 and 31 December 2015.

		Censor		Zero-augmented	
		$\alpha_1 = 0$	$\alpha_1 \neq 0$	$\delta_1 = 0$	$\delta_1 \neq 0$
Darwin	AIC	3.175	3.170	3.206	3.172
	BIC	3.204	3.199	3.235	3.200
Cairns	AIC	3.955	3.951	4.059	3.945
	BIC	3.984	3.980	4.088	3.974

Table 2 Goodness of fit statistics

Location	Darwin		Cairns	
Model	Censored		Zero-aug.	
	Estimate	S.E.	Estimate	S.E.
ω	0.43	0.55	0.54	0.13
ϕ	0.68	0.05	0.59	0.02
κ	0.41	0.06	0.29	0.04
δ_0	-	-	-1.59	0.42
δ_1	-	-	2.34	0.18
α_0	1.51	0.25	-	-
α_1	-0.38	0.20	-	-
ν	0.63	0.09	0.81	0.03
ξ	1.75	0.15	3.12	0.04
$\overline{\eta}$	0.31	0.16	0.44	0.10
η	3.26		2.29	
ln L	-5767.9		-7183.3	

Table 3 ML estimates of preferred models with GB2 as in the inverse tail index parameterization of (6).

Town	Darwin		Cairns	
Model	Censored		Zero-aug.	
	Static	Dynamic	Static	Dynamic
Predictive likelihood	1.23	1.21	2.15	1.96
t-stat	-3.70		-9.39	
Brier PS	0.884	0.876	0.86	0.78
t-stat	-2.66		-10.28	

Table 4 Average predictive likelihood and Brier probability scores, together with the associated t-statistics for comparing one-step ahead density forecasts.

FIGURES

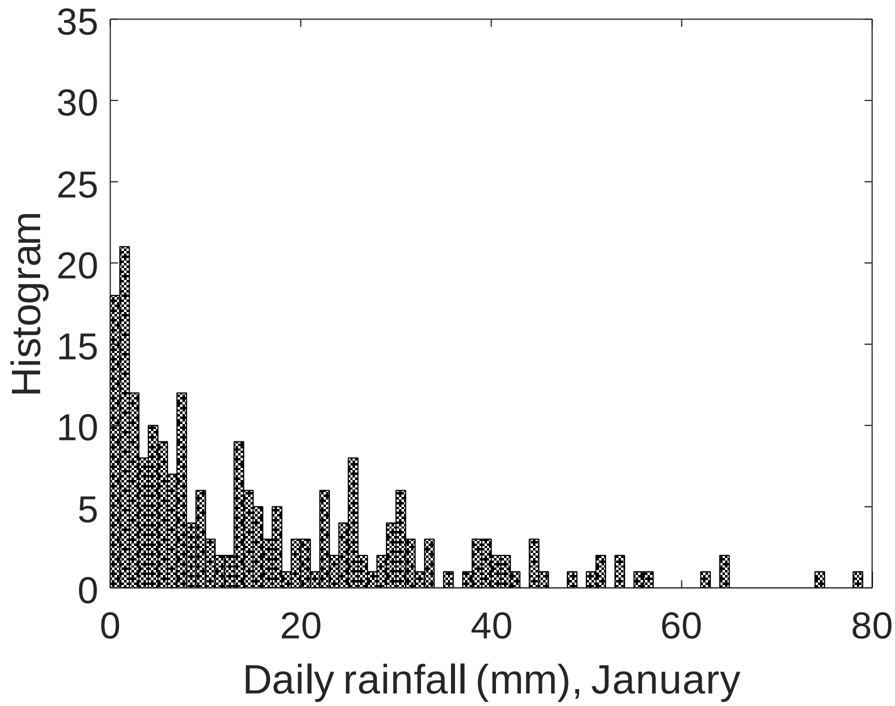

Figure 1: Daily rainfall in Darwin in December. Positive observations only arranged in bins of width 1mm

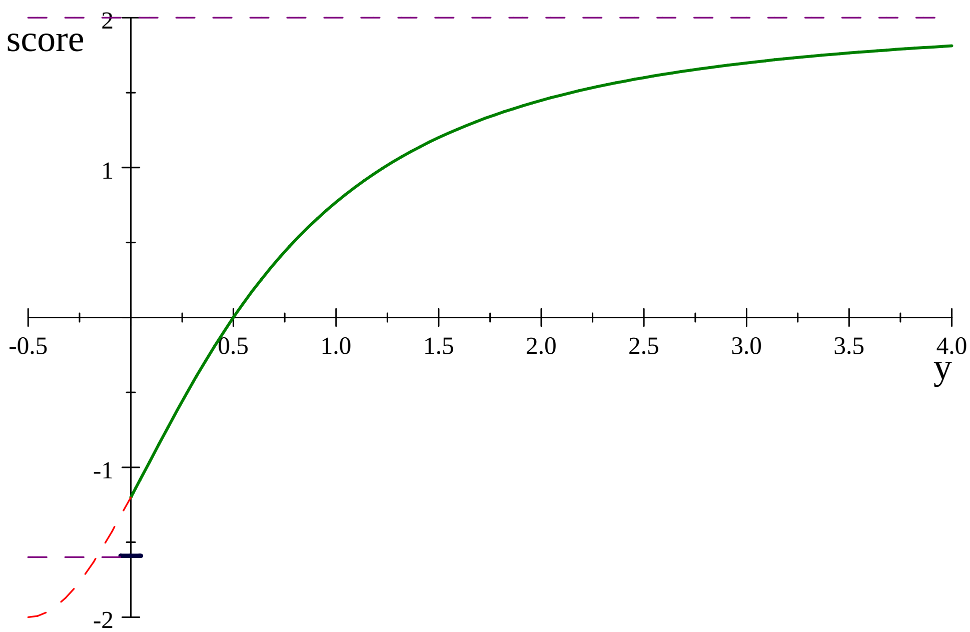

Figure 2: Log-logistic score with $\upsilon = 2$ and $c = 0.5$.

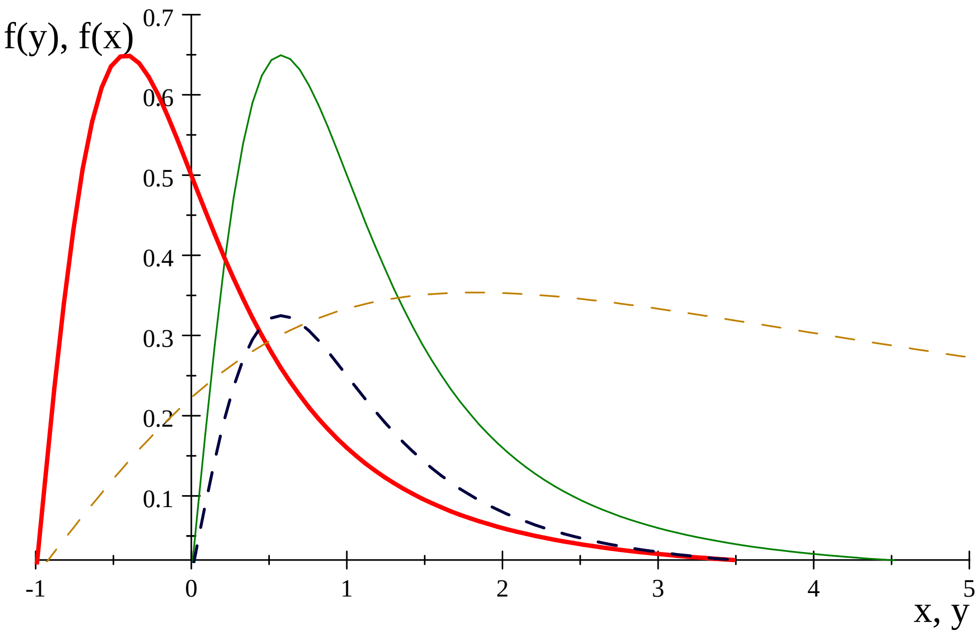

Figure 3: Log-logistic (thin line) with unit scale and $\upsilon = 2$. Censored with $c = 1$ is thick line. Thin dash is censored with scale of 4. Thick dash is distribution for positive observations for point-mass mixture with $\pi = 0.5$.

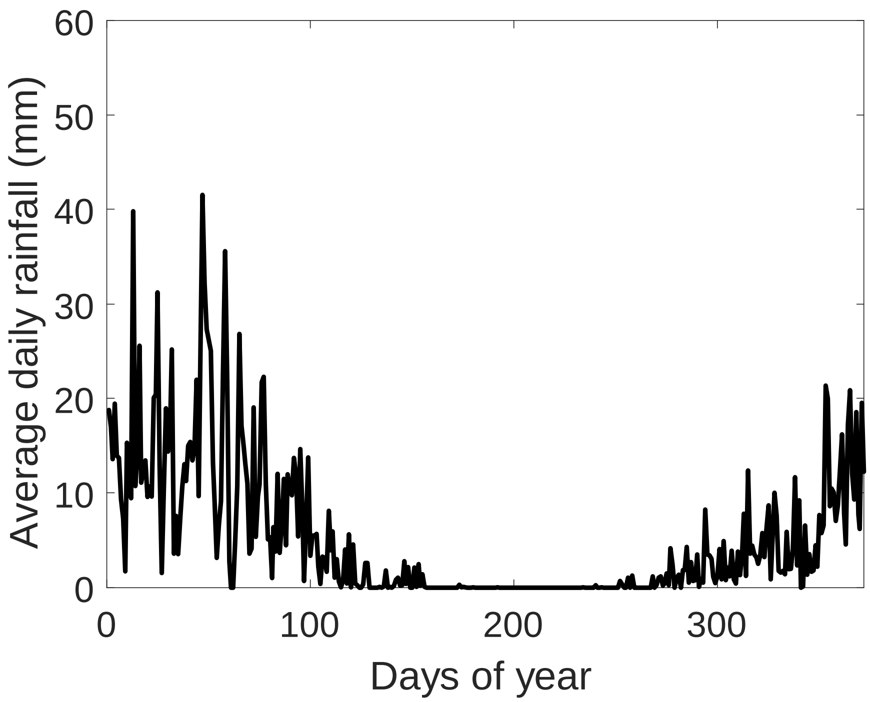

Figure 4: Average daily rainfall in Darwin throughout the year, for 1 January 2006 to 31 December 2015.

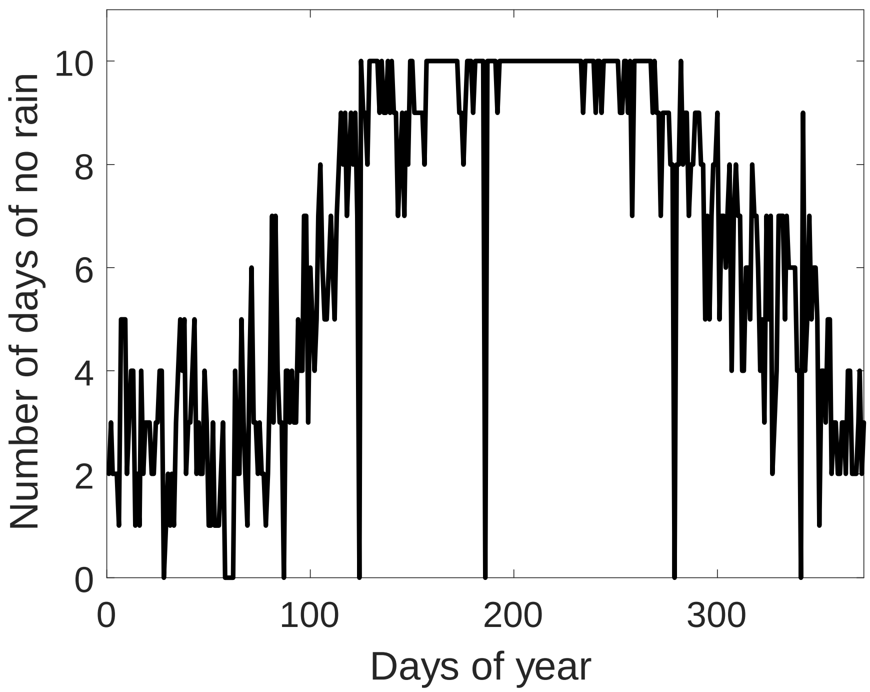

Figure 5: Darwin: Number of days in the sample period with no rain.

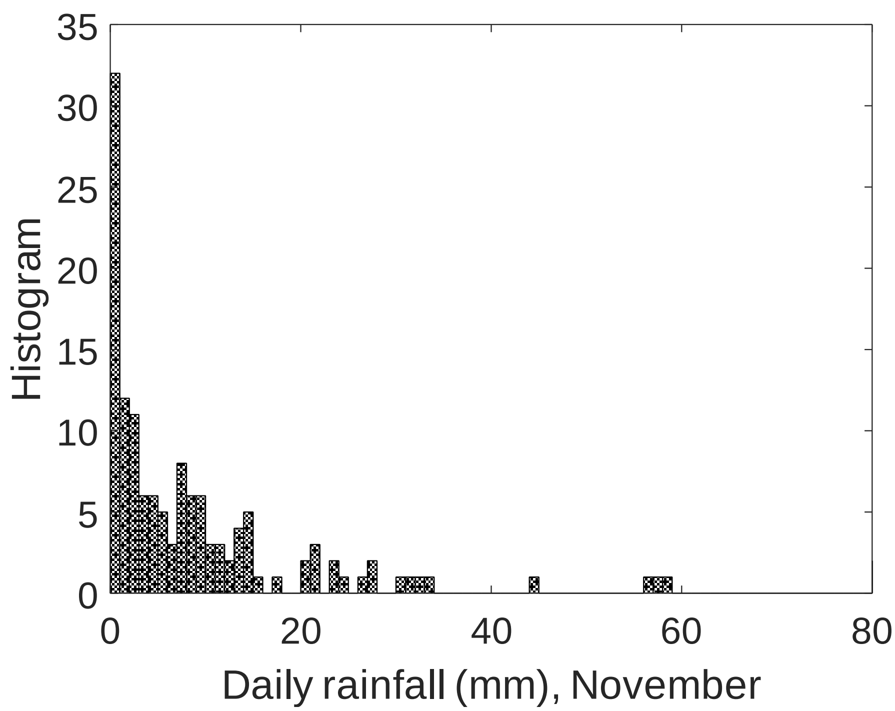

Figure 6: Daily rainfall in Darwin in November. Positive observations only arranged in bins of width 1mm

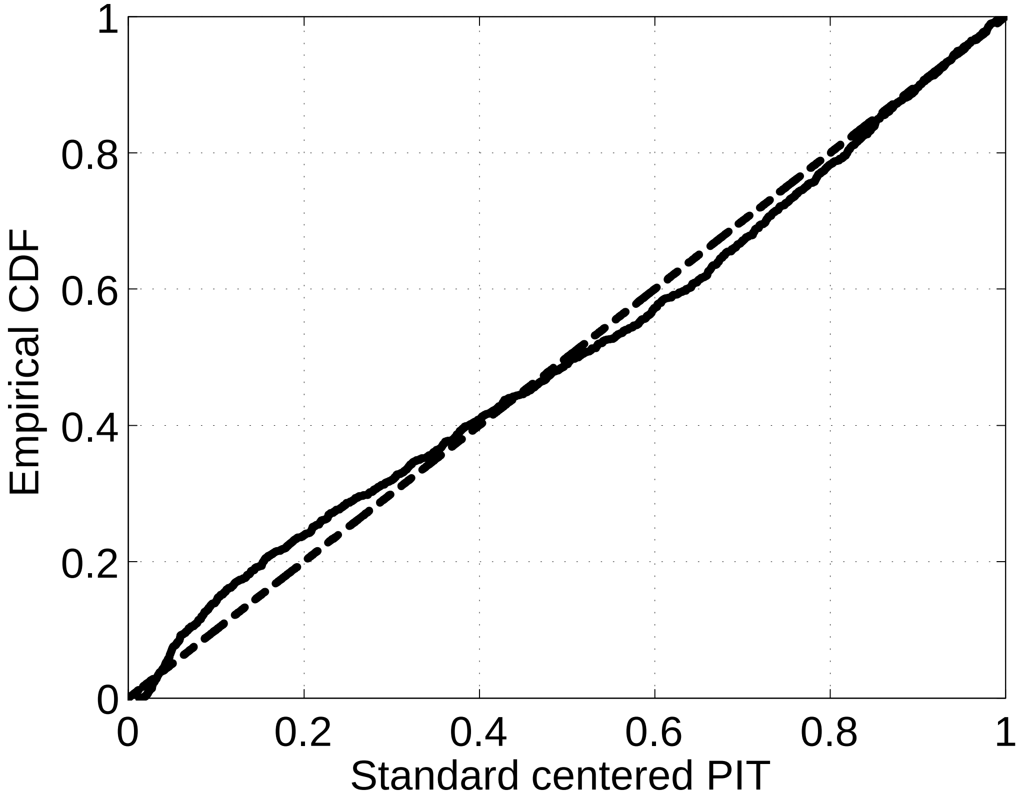

Figure 7: PITs for Darwin

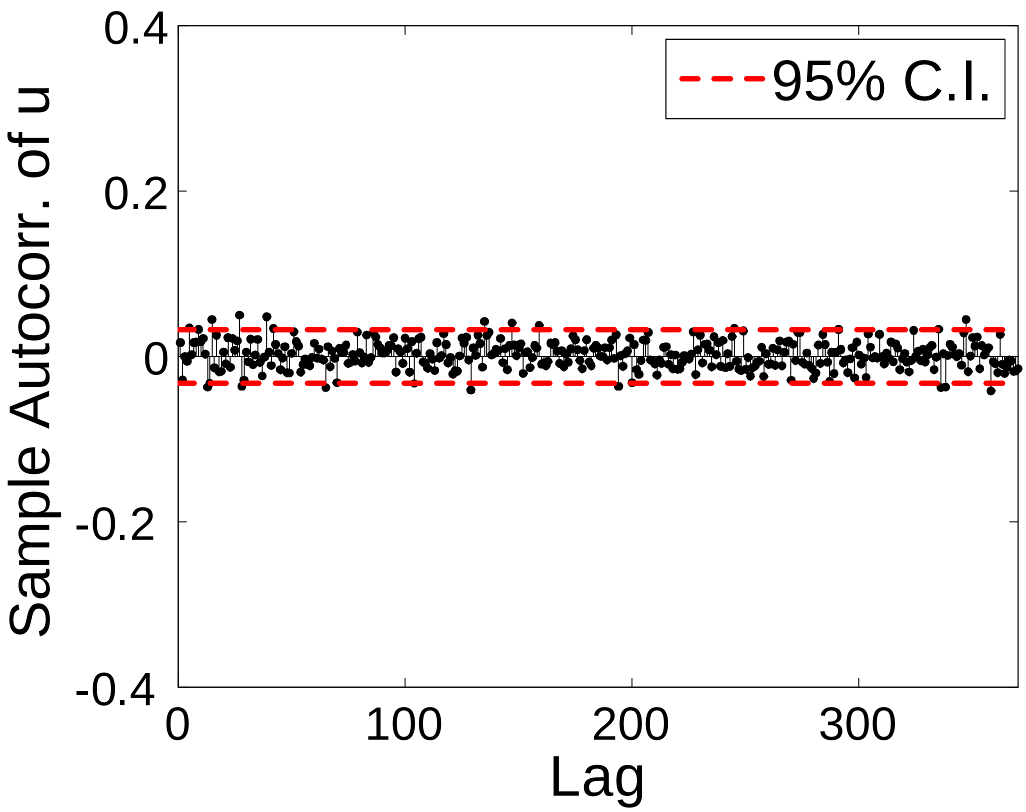

Figure 8: Correlogram of scores from dynamic model fitted to Darwin

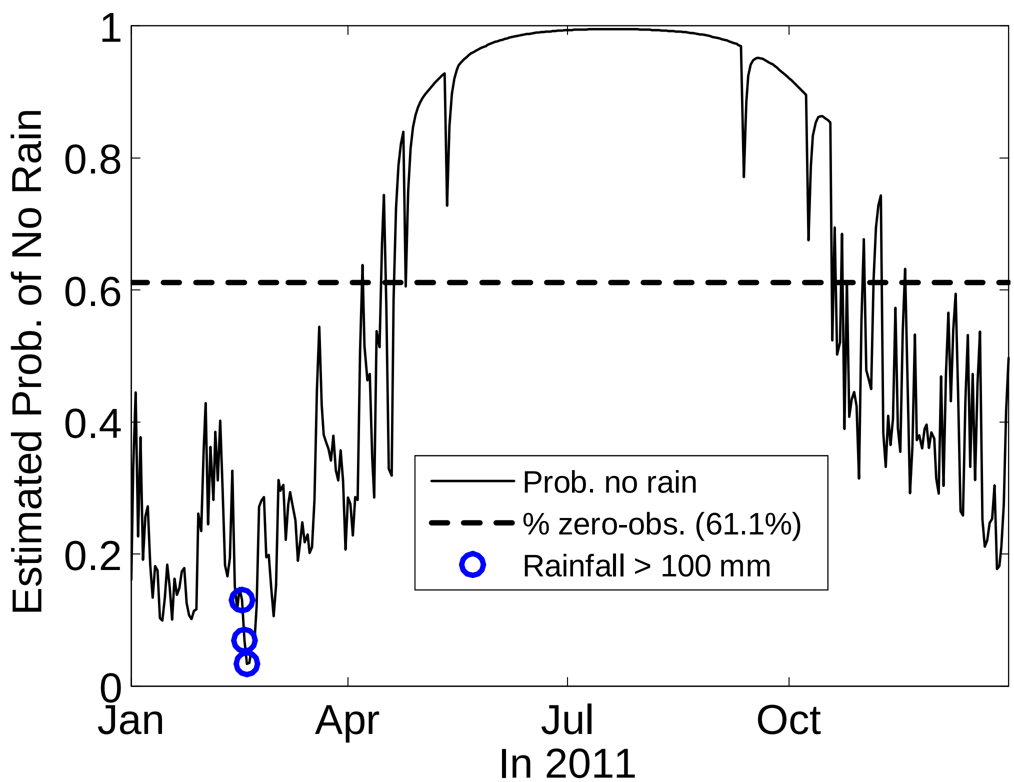

Figure 9: Estimated probability of no rain next day in Darwin in 2011.

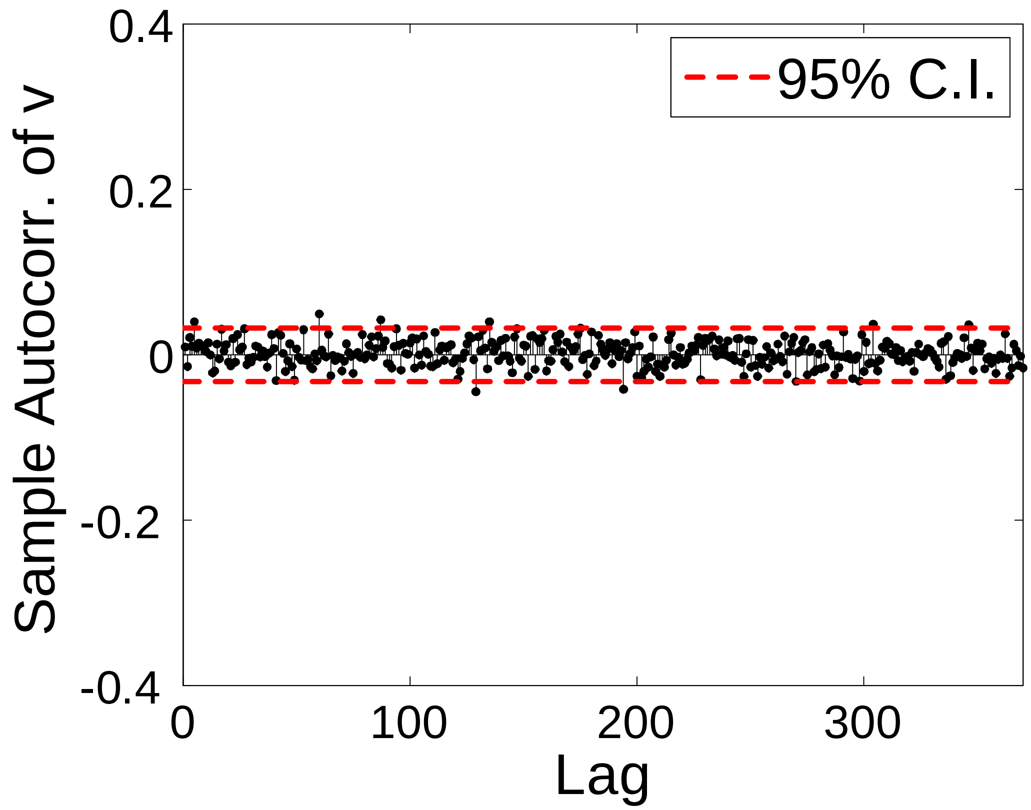

Figure 10: Correlogram of binary variables from dynamic model fitted to Darwin

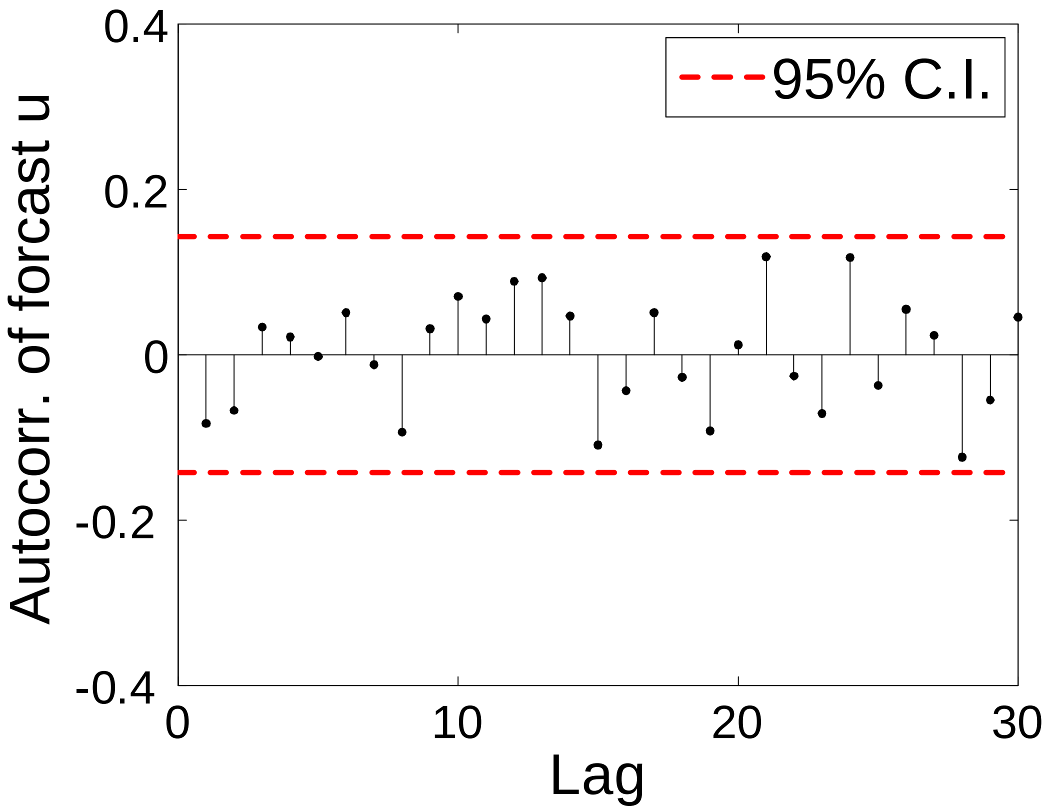

Figure 11: Correlogram of out of sample scores for dynamic model fitted to Darwin

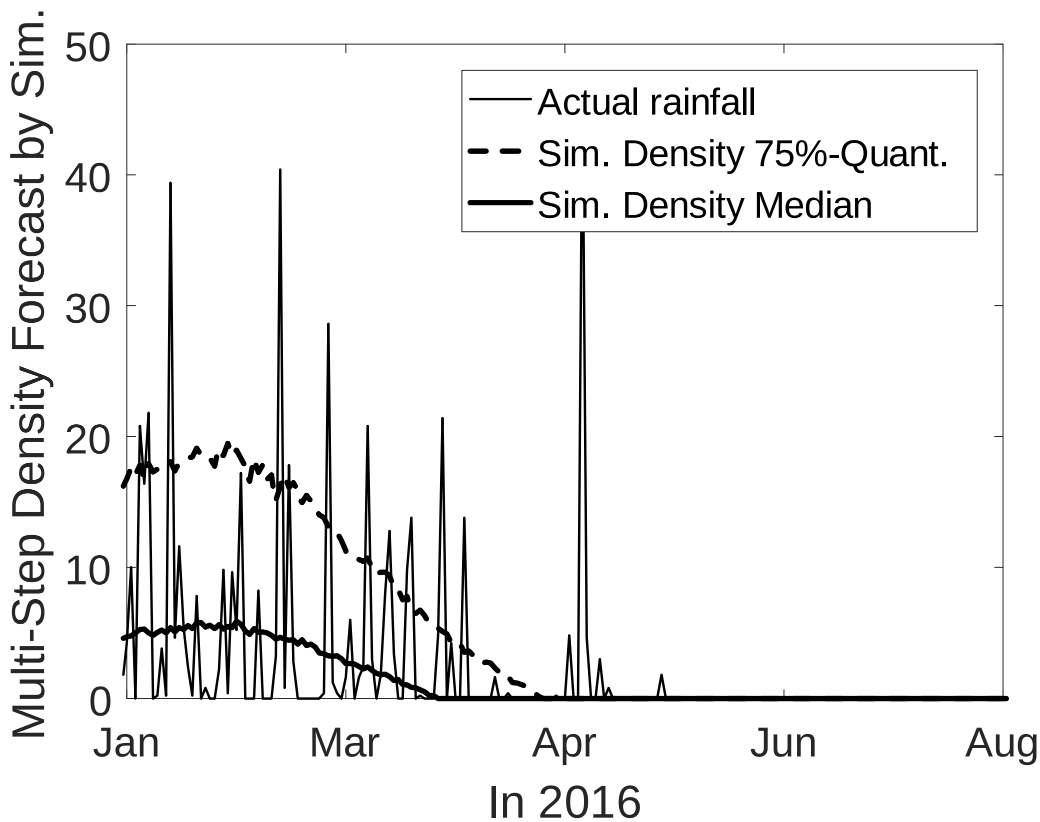

Figure 12: Multi-step predictive density for Darwin at the start of 2016.