UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Cambridge Working Papers in Economics

Cambridge Working Papers in Economics: 1971

# MODELING DIRECTIONAL (CIRCULAR) TIME SERIES

Andrew Harvey    Stan Hurn    Stephen Thiele

25 July 2019

Circular observations pose special problems for time series modeling. This article shows how the score-driven approach, developed primarily in econometrics, provides a natural solution to the difficulties and leads to a coherent and unified methodology for estimation, model selection and testing. The new methods are illustrated with hourly data on wind direction.

# Modeling directional (circular) time series

Andrew Harvey*, Stan Hurn and Stephen Thiele
Faculty of Economics, Cambridge University and QUT, Brisbane

July 25, 2019

**Abstract**

Circular observations pose special problems for time series modeling. This article shows how the score-driven approach, developed primarily in econometrics, provides a natural solution to the difficulties and leads to a coherent and unified methodology for estimation, model selection and testing. The new methods are illustrated with hourly data on wind direction.

KEYWORDS: Autoregression; circular data; dynamic conditional score model; von Mises distribution; wind direction.

JEL: C22

**Corresponding author**.
Andrew Harvey ach34@cam.ac.uk.

*We would like to thank participants at the INET workshop on Score-driven and Nonlinear Time Series Models held at Trinity College, Cambridge, in March 2019 for helpful comments. We are particularly grateful to Richard Davis and Neil Shephard for their suggestions. We are also grateful to Howell Tong and others for comments made a seminar at the University of Bologna.

# 1 Introduction

Directional variables are circular. If the starting point is due south then moving through 180 degrees ends up at due north. The same point is reached by moving 180 degrees in the opposite direction. In terms of radians the points $-\pi$ and $\pi$ meet up and this poses a challenge for directional time series modelling.

A number of ways for modelling circular time series have been proposed in the literature. The most widely used is based on transformations, the aim of which is to try to put the data in a form that lends itself to conventional autoregressive (AR) or autoregressive moving average (ARMA) modeling. A second method uses transformations but to formulate an autoregressive model in which the conditional distribution of the next observation is circular. By contrast, the approach proposed is also based a circular conditional distribution, but the dynamics are formulated so as to be consistent with the circularity of the data. It draws on recently developed procedures for dealing with non-Gaussian conditional distributions in a wide variety of situations, primarily in economics and finance. The defining feature of the new class of circular time series models, which turns out to be crucial for performance as well as theoretical properties, is that the dynamics of the time-varying parameter are driven by the score of the conditional distribution. Score-driven models are known as Dynamic Conditional Score (DCS) or Generalized Autoregressive Score (GAS) models; see, for example, Harvey (2013) and Creal, Koopmans and Lucas (2013).

Harvey and Luati (2014) show how the score-driven model may be used for modeling changing location when the conditional distribution is Student's $t$. The score automatically handles observations that would be classed as outliers for a normal distribution by making them less influential. The same methodology applied to directional data deals with the problem of circularity. The asymptotic distribution of the maximum likelihood (ML) estimator can be derived for a first-order model and general principles can be used for testing and model selection. A nonstationary model is also feasible. Therefore although circular data have special features that need to be explicitly recognized, their overall treatment follows from a well-developed time series approach. A score-driven autoregressive model is also proposed. A model of this kind has not yet appeared in the dynamic score literature but it turns out to be particularly attractive here.

Many scientific fields have applications in which directions are collected and statistically analysed. In particular, modeling wind direction is becoming increasingly important as energy generation by means of wind power increases. The score driven models developed in this paper are applied to wind direction data from the Black Mountain in Canberra. This is a relatively short time series and is chosen mainly for comparative purposes with previous studies. Our models will, however, generalize to handle many of the issues raised by longer time series.

The plan of the paper is as follows. Sections 2 and 3 review the von Mises distribution and existing methods for modeling circular time series. Score-driven models are described in Section 4. The small sample properties of these models are investigated by Monte Carlo experiments in Section 5. Model selection methods are discussed in Section 6, while Section 7 applies the new models to data on wind direction and highlights their advantages over existing methods. The last section concludes and points to future developments.

## 2 Circular data and the von Mises distribution

A (continuous) circular probability distribution (PDF) which depends on a vector of parameters $\boldsymbol{\theta}$, denoted $f(y;\boldsymbol{\theta})$, must satisfy the following conditions:

$$
\begin{aligned}
&\text{(i)} && f(y;\boldsymbol{\theta}) \geq 0 \\
&\text{(ii)} && \int_{-\pi}^{\pi} f(y;\boldsymbol{\theta})dy = 1 \\
&\text{(iii)} && f(y \pm 2\pi k;\boldsymbol{\theta}) = f(y;\boldsymbol{\theta}),
\end{aligned}
$$

where $k$ is an integer. General classes of distributions are proposed in Jones and Pewsey (2005) and Fernandez-Duran (2004). The latter is able to capture multi-modality as well as skewness. Provided the derivatives of the log-density with respect to elements of $\boldsymbol{\theta}$ are continuous, they are circular in the sense that the periodicity condition (iii) is satisfied.

Given circular data measured in radians, a common assumption is that the data have a von Mises (vM) distribution (also called the circular normal or Tikhonov distribution) with PDF given by

$$f(y;\mu,\upsilon) = \frac{1}{2\pi I_0(\upsilon)} \exp\{\upsilon \cos(y-\mu)\}, \qquad -\pi \leq y,\mu < \pi, \quad \upsilon \geq 0, \tag{1}$$

where $I_k(\upsilon)$ denotes a modified Bessel function of order $k$, $\mu$ denotes location (directional mean) and $\upsilon$ is a non-negative concentration parameter that is inversely related to scale. When $\upsilon = 0$ the distribution is uniform and when $\upsilon$ is large, $y$ is approximately $N(\mu, 1/\upsilon)$. The normal distribution is generally considered a good approximation for $\upsilon > 2$.

The (circular) variance of the von Mises distribution is

$$1 - I_1(\upsilon)/I_0(\upsilon) = 1 - A(\upsilon). \tag{2}$$

Note that

$$A(\upsilon) = \mathbb{E}\cos(y - \mu),$$

which defines the mean resultant length and that $A(\upsilon) \to 1$ as $\upsilon \to \infty$, whereas $A(0) = 0$.

The score with respect to the location parameter is

$$\frac{\partial \ln f}{\partial \mu} = \upsilon \sin(y - \mu), \tag{3}$$

with variance $\upsilon A(\upsilon)$. The score with respect to the concentration parameter is

$$\frac{\partial \ln f}{\partial \upsilon} = \cos(y - \mu) - A(\upsilon). \tag{4}$$

# 3 Existing time series models

Data generated by a Gaussian time series model over the real line, that is $-\infty < x_t < \infty$, can be converted into wrapped circular time series observations in the range $[-\pi, \pi)$ by letting

$$y_t = x_t \bmod(2\pi) - \pi, \qquad t = 1, ..., T. \tag{5}$$

Fitting such models can be seen as a missing data problem because the unwrapped observations can be decomposed as

$$x_t = y_t + 2\pi k_t, \qquad t = 1, ..., T,$$

where $k_t$ is an integer that needs to be estimated; see Breckling (1989) and Fisher and Lee (1994, p 329). The difficulty with this approach is computational. Estimation is usually by the EM algorithm, but, as Fisher and Lee (1994, p 333-4) observe, this can become complicated for all but the simplest models. Coles (1998) shows how Markov chain Monte Carlo can be used to fit wrapped autoregressive models, denoted $WAR(p)$.

An alternative approach is to transform a circular variable $y$, where $-\pi < y - \mu < \pi$, to a variable $x$ in the range $-\infty < x < \infty$ by means of a link function, $x = g^{-1}(y - \mu)$. There are then two ways to proceed. The first is to fit a linear time series model to $x$ and then transform back to $y$. The model for $x$ is sometimes called a direct or linked linear process. When the time series model is an $ARMA(p, q)$ process it is called $LARMA(p, q)$ or, more commonly, $CARMA(p, q)$, where the $C$ denotes 'circular' as opposed to the $L$ for 'linked'. In the second class of models, which are nonlinear, the inverse form is an autoregression, denoted $IAR(p)$, whereby the conditional mean, $\mu_{t|t-1}$, of a circular distribution, such as $vM$, is specified as

$$\mu_{t|t-1} = \mu + g\{\phi_1 g^{-1}(y_{t-1} - \mu) + ... + \phi_p g^{-1}(y_{t-p} - \mu)\}, \quad -\pi < y - \mu < \pi. \tag{6}$$

The $IAR(1)$ model with $\alpha$ close to one can be approximated without the transformation so

$$\mu_{t|t-1} \simeq \mu + \alpha(y_{t-1} - \mu). \tag{7}$$

When $\alpha$ is equal to one, $\mu_{t|t-1} = y_{t-1}$, so $y_t$ is a random walk. There seems to be no $IARMA(p, q)$ model.

The aim of the transformations is have $x$ close to being Gaussian. Fisher and Lee (1994) prefer the probit transformation to the commonly used $\tan(y/2)$ transformation as they argue that the latter can give rise to fat tails whereas the former is closer to a normal distribution. More precisely, the probit transformation yields a normal distribution when the circular observations are uniform. On the other hand, if the variance is small, the untransformed observations can be treated as normally distributed.

# 4 Score-driven models

When the conditional distribution is continuous and circular, letting the score drive a dynamic equation for the conditional mean, $\mu_{t|t-1}$, solves the circularity problem because the score function is also circular and continuous. Unlike the $CARMA$ and $IAR$ approaches, which set up a barrier for $y_t - \mu$ at $\pi$ and $-\pi$ and thus ignore the proximity of observations on either side, in the proposed approach a value of $y_t$ slightly bigger than $-\pi$ is treated in the same way as if it were slightly bigger than $\pi$.

The conditional score, $u_t$, enters into a dynamic equation such as

$$\mu_{t|t-1} = (1-\phi)\mu + \phi\mu_{t-1|t-2} + \kappa u_{t-1}, \qquad |\phi| < 1, \tag{8}$$

where the conditional distribution of a variable $x_t$ defined over the real line, has location $\mu_{t|t-1}$ so that

$$x_t = \mu_{t|t-1} + \varepsilon_t, \qquad t = 1, ...., T, \tag{9}$$

in which $\varepsilon_t$ has location zero. An observation falling outside the range $\mu \pm \pi$ can be wrapped, as in (5), so that $y_t$ lies in the range $[-\pi, \pi)$. The score-driven data generating process is invariant to this wrapping of the observations. Hence the question of how to estimate $k_t'$ defined below (5) does not arise. Note that there is no need to wrap $\mu_{t|t-1}$ but if it is reset neither the data generation process nor estimation is affected.

The conditional score is a martingale difference sequence with mean zero. In the case of the von Mises distribution, that is $\varepsilon_t \sim vM(0, \upsilon)$ in (9), dividing the score, (3) by its information quantity gives $\sin(y_t - \mu_{t|t-1})/A(\upsilon)$ but there is a good case for dropping $A(\upsilon)$ because then the filter is not dependent on $\upsilon$. Hence the forcing variable in an equation such as (8) is

$$u_t = \sin(y_t - \mu_{t|t-1}), \qquad u_t \sim i.i.d.(0, A(\upsilon)/\upsilon). \tag{10}$$

It follows that $x_t$ and $y_t$ are strictly stationary.

Figure 1 contrasts the score variable for a vM with a linear response when $\mu = 0$. The score starts to downweight observations beyond $\pi/2$ and $-\pi/2$,

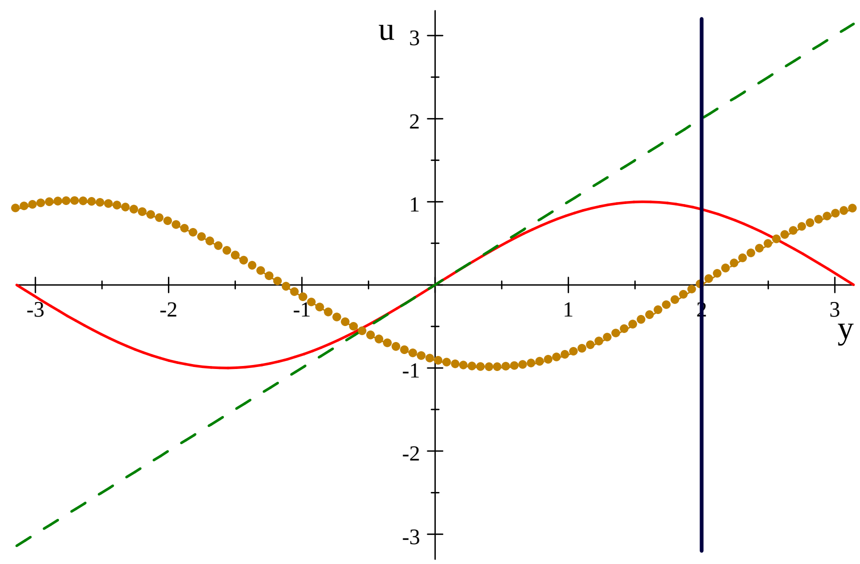


Figure 1: Score functions for vM and normal. Dots denote vM with $\mu = 2$.

reflecting the fact that, like the score for a $t-$distribution, it is a redescending function. However, for small deviations from the mean, the score is approximately linear; the MacLaurin expansion shows that $\sin(y_t - \mu_{t|t-1}) \simeq y_t - \mu_{t|t-1}$. If the concentration is large, so that a Gaussian conditional distribution is a reasonable approximation, then $y_t - \mu_{t|t-1}$ is the score and the model corresponds to the steady-state innovations form of the Kalman filter from a Gaussian unobserved components model made up of a first-order autoregressive process and white noise.

The dots in Figure 1 illustrate the critical role played by the score when $\mu_{t|t-1}$ is close to $\pi$. Suppose $\mu_{t|t-1} = \pi - a$, where $a$ is small and positive. (In the figure it is $\pi - 2$.) Suppose the next observation is negative at $-\pi + b$, where $b$ is small and positive. The distance between $\mu_{t|t-1}$ and $y_t$ is only $a+b$, but $y_t - \mu_{t|t-1} = -2\pi + b + a$. However, $\sin(y_t - \mu_{t|t-1}) = \sin(-2\pi + b + a) = \sin(b + a)$. Thus the impact of the negative observation on $\mu_{t|t-1}$ is positive.

## 4.1 Maximum likelihood estimation

The log-likelihood function when the observations have a von Mises conditional distribution is

$$\ln L(\upsilon, \boldsymbol{\psi}) = T \ln(2\pi I_0(\upsilon)) + \upsilon \sum_{t=1}^{T} \cos(y_t - \mu_{t|t-1}),$$

where $\boldsymbol{\psi}$ denotes the parameters in the dynamic equation for $\mu_{t|t-1}$. In a stationary model it is convenient to assume that $\mu_{1|0}$ is given by the unconditional mean $\mu$. When $\upsilon = 0$ the observations are uniformly distributed on the circle and so cannot be predicted because the probability of the next observation falling in any equal width interval on the circle is the same. It will therefore be assumed that $\upsilon > 0$.

Asymptotic results for the stationary first-order DCS model are as follows.

**Proposition 1** *Define $\boldsymbol{\psi} = (\kappa, \phi, \mu)'$ in the stationary first-order dynamic equation (8). For a single observation, the information matrix for $\boldsymbol{\psi}$ and $\upsilon$ is*

$$\mathbf{I}\begin{pmatrix} \boldsymbol{\psi} \\ \upsilon \end{pmatrix} = \begin{bmatrix} \upsilon A(\upsilon)\mathbf{D}(\boldsymbol{\psi}) & \mathbf{0} \\ \mathbf{0}' & 1 - A(\upsilon)^2 - A(\upsilon)/\upsilon \end{bmatrix} \tag{11}$$

*where*

$$\mathbf{D}(\boldsymbol{\psi}) = \mathbf{D}\begin{pmatrix} \kappa \\ \phi \\ \mu \end{pmatrix} = \frac{1}{1-b}\begin{bmatrix} A(\upsilon)/\upsilon & \dfrac{a\kappa A(\upsilon)/\upsilon}{1-a\phi} & 0 \\ \dfrac{a\kappa A(\upsilon)/\upsilon}{1-a\phi} & \dfrac{\kappa^2 A(\upsilon)(1+a\phi)/\upsilon}{(1-\phi^2)(1-a\phi)} & 0 \\ 0 & 0 & \dfrac{(1-\phi)^2(1+a)}{1-a} \end{bmatrix}$$

*with*

$$a = \phi - \kappa A(\upsilon) \tag{12}$$

$$b = \phi^2 - 2\phi\kappa A(\upsilon) + \kappa^2(1 - A(\upsilon)/\upsilon) < 1. \tag{13}$$

The information matrix for a vM distribution with parameters $\mu$ and $\upsilon$ is given in Mardia and Jupp (2000, p 86, 350). The derivation of $\mathbf{D}(\boldsymbol{\psi})$ is in Appendix A.

**Proposition 2** *Provided* $b < 1$ *and* $\kappa \neq 0$*, the ML estimator,* $(\widetilde{\boldsymbol{\psi}}' \; \widetilde{\upsilon})'$*, is consistent and asymptotically normal with mean* $(\boldsymbol{\psi}' \; \upsilon)'$ *and covariance matrix given by the inverse of (11).*

The proof follows from Lemma 1 in Jensen and Rahbek (2004) and Harvey (2013). The conditions $b < 1$ and $\kappa \neq 0$ are needed for the information matrix to be positive definite. Third derivatives associated with the mean are bounded because they depend only on sines and cosines. Derivatives with respect to concentration are also bounded because

$$\frac{\partial A(\upsilon)}{\partial \upsilon} = 1 - A(\upsilon)^2 - \frac{A(\upsilon)}{\upsilon}.$$

When $\upsilon \to \infty$ the condition $b < 1$ leads to $(\phi - \kappa)^2 < 1$ which corresponds to the familiar condition for invertibility in an ARMA(1,1) model and the asymptotic distribution for estimators in a Gaussian model is obtained; see Harvey (2013, p 67-8). On the other hand, as $\upsilon \to 0$ the information on $\mu$ tends to zero, as do all elements in the $\mathbf{I}(\boldsymbol{\psi})$ matrix.

The ML estimates, $\widetilde{\boldsymbol{\psi}}$ need to satisfy the condition

$$\sum_{t=1}^{T} \sin(y_t - \mu_{t|t-1}) = 0,$$

which is achieved by maximizing

$$S(\boldsymbol{\psi}, \mu) = \sum_{t=1}^{T} \cos(y_t - \mu_{t|t-1})$$

with respect to $\boldsymbol{\psi}$ and $\mu$. This may be done independently of $\upsilon$. Once $\widetilde{\boldsymbol{\psi}}$ has been computed, the ML estimate of $\upsilon$ may be obtained by solving

$$A(\widetilde{\upsilon}) = S(\widetilde{\boldsymbol{\psi}}, \widetilde{\mu})/T.$$

Unfortunately there is no exact solution; see Mardia and Jupp (2000, pp 85-6).

**Remark 1** *An initial estimate of $\mu$ is given by the sample mean direction, defined as $\overline{y}_d = \arctan(\overline{S}/\overline{C})$, where $\overline{S} = \sum \sin y_t/T$ and $\overline{C} = \sum \cos y_t/T$; if $\overline{C} < 0$ then $\pi$ is added. The ML estimator of $\mu$ in the DCS model is, like $\overline{y}_d$, equivariant under rotation and it does not matter whether the data given by a particular cut are over $[-\pi, \pi)$ or $[0, 2\pi)$; see Mardia and Jupp (2000, p 17). Furthermore the estimates of the other parameters are unchanged[1]. Fitting a $CARMA$ or $IAR$ model, on the other hand, requires that the data be adjusted so as to be in the range $\overline{y}_d \pm \pi$ and if observations are not re-categorized for updated estimates of $\mu$, there is the potential for a large positive transformed observation switching to becoming a correspondingly extreme negative observation.*

**Remark 2** *Blasques, Gorgi, Koopman and Wintenberger (2018) draw attention to the importance of ensuring the invertibility of a nonlinear time series model. They show that a sufficient condition for invertibility of a stationary and ergodic model is $E \ln \Lambda_t(\boldsymbol{\psi}) < 0$, where $\Lambda_t(\boldsymbol{\psi}) := \sup |z_t|$, with $z_t = d\mu_{t+1|t}/d\mu_{t|t-1} = \phi + \kappa(\partial u_t/\partial \mu_{t|t-1})$ and the supremum is over all admissible $\boldsymbol{\psi}$. Thus a sufficient condition for invertibility of the first-order DCS circular model is $\kappa < 1 - \phi$ and $\kappa < 1 + \phi$. The result is obtained by noting that $\partial u_t/\partial \mu_{t|t-1} = -\cos(y - \mu_{t|t-1})$ which lies between -1 and 1. The first of these conditions is almost certainly too restrictive; compare the situation for a t-distribution as analysed in Blasques et al. (2018, p 1041). There is the option of computing an estimate of $\mathbb{E} \ln \Lambda_t(\boldsymbol{\psi})$. However, in the usual case*

[1]If a constant is added to all the observations, the likelihood is maximized simply by adding the same constant to the estimate of $\mu$.

*when both $\kappa$ and $\phi$ are positive and, in addition, $\kappa \leq \phi$, it follows that $z_t \geq 0$ and so $|z_t| = z_t$. Since $\mathbb{E}(z_t) = \phi - \kappa A(\upsilon)$, Jensen's inequality shows that* $\mathbb{E} \ln \Lambda_t(\boldsymbol{\psi}) < 0$.

## 4.2 Non-stationarity

The non-stationary first-order DCS model is

$$\mu_{t|t-1} = \delta + \mu_{t-1|t-2} + \kappa u_{t-1}, \tag{14}$$

where $\delta$ is a drift term and $\mu_{1|0}$ is fixed. The conditional mean can, in principle, travel all the way round the circle; see the footnote below (5). Such situations can arise in practice. For example, Fisher (1993, p 249) gives a data set of weekly observations at a location in England where the wind direction moves round the full circle every quarter.

Because $\text{var}(\mu_{t|t-1}) \to \infty$ as $t \to \infty$ in (14), we have the following property.

**Proposition 3** *The unconditional distribution for wrapped observations, (5), generated by (14) is uniform.*

We cannot initialize $\mu_{1|0}$ with the directional mean because it does not exist. The best option is to start off the recursion in (14) with $\mu_{2|1} = y_1$ and compute estimates of the other parameters. These provide starting values for full ML estimation with $\mu_{1|0}$ treated as a fixed parameter. The transformation $\mu_{1|0} = 2\arctan(\omega)$, where $\omega$ is unconstrained, may be employed to ensure $\left|\mu_{1|0}\right| < \pi$.

The asymptotics still hold for (14), as in Harvey (2013, p 45-6), so for a conditional vM distribution, $\widetilde{\kappa}$ is asymptotically normal with mean $\kappa$ and

$$\text{avar}(\widetilde{\kappa}) = \frac{1}{T}\frac{2\kappa A(\upsilon) - \kappa^2(1 - A(\upsilon)/\upsilon))}{A(\upsilon)^2}. \tag{15}$$

Estimating the initial value, $\mu_{1|0}$, makes no difference to avar$(\widetilde{\kappa})$ because the asymptotic variance is $O(1)$. The equation for $b$, that is (13), now implies $\kappa > 0$ and $\kappa < 2A(\upsilon)\upsilon/(\upsilon - A(\upsilon))$. Note that $\kappa \to 0$ as $\upsilon \to 0$ whereas for $\upsilon \to \infty$, $0 < \kappa < 2$. However, for $\upsilon = 2$, $\kappa < 2.16$. The result in Remark 4 suggests[2] that invertibility is guaranteed by $\kappa \leq 1$.

### 4.3 Score-driven circular autoregression

A score-driven circular autoregression (SCAR) for a conditional vM distribution can be formulated, not with the conditional mean, defined as in (10), but rather as

$$\mu_{t|t-1} = \mu + \phi_1 \sin(y_{t-1} - \mu) + ... + \phi_p \sin(y_{t-p} - \mu), \quad -\pi \leq y, \mu < \pi, \quad \upsilon > 0, \tag{16}$$

where $\mu$ is the unconditional mean. Unlike the $IAR(p)$ model in (6), the $SCAR(p)$ model is invariant to translation and like the DCS model it is invariant to wrapping: hence (16) is written in terms of $y_t$ rather than $x_t$ to simplify the discussion. It is assumed that the SCAR model is stationary, although it is not clear what conditions[3], if any, need to be imposed on the autoregressive parameters to ensure that this is the case.

Estimation is best carried out by starting at $t = p+1$ rather than setting pre-sample values equal to zero. Some restrictions on the parameters may be desirable. For example, when $p = 1$, the fact that $|\sin(y_{t-1} - \mu)| \leq 1$, means that $\left|\mu_{t|t-1} - \mu\right| \leq \pi$ for $|\phi| \leq \pi$. In any event, the model satisifies the conditions of Lemma 1 in Jensen and Rahbek (20014) and so the following result holds; see Appendix B.

**Proposition 4** *Assuming that in the $SCAR(p)$ model, (16), the ML estimator, $\widetilde{\phi}$, is consistent, the limiting distribution of $\sqrt{T}(\widetilde{\phi} - \phi)$ is multivariate*

[2] Blasques *et al.* (2018) assume stationarity when stating their result.

[3] The variance is finite for finite values of the parameters because $|\sin(y_t - \mu)| \leq 1$.The process can, in principle, be initialized by drawing the pre-sample observations from the unconditional distribution.

*normal with mean* 0 *and covariance matrix,* $(1/\upsilon A(\upsilon))Q^{-1}$, *where the* $ij-th$ *element of* $Q$ *is the circular autocovariance of order* $|i-j|$, *as defined in the numerator of (21). Given that* $\sin(y_t-\mu)$ *is stationary, the matrix* $Q$ *will be positive definite. The ML estimators of* $\mu$ *and* $\upsilon$ *are similarly asymptotically normal, being distributed independently of* $\widetilde{\phi}$ *and of each other. The limiting distribution of* $\sqrt{T}(\widetilde{\mu}-\mu)$ *is normal with mean* 0 *and variance* $1/[\upsilon A(\upsilon)(1-\sum_k \phi_k \mathbb{E}\cos(y_t-\mu))^2/T]$. *The asymptotic distribution of* $\widetilde{\upsilon}$ *is as implied by (11).*

**Corollary 1** *The large-sample covariance matrix of* $\widetilde{\phi}$ *can be estimated as*

$$\mathrm{avar}(\widetilde{\phi}) = \frac{1}{\widetilde{\upsilon}A(\widetilde{\upsilon})}\left[\sum_{t=p+1}^{T} s_t s_t'\right]^{-1}, \tag{17}$$

*where the* $k$*-th element of the* $p\times 1$ *vector* $s_t$ *is* $\sin(y_{t-k}-\widetilde{\mu})$, $k=1,...,p$; *see (28). As regards* $\mu$,

$$\begin{aligned}\mathrm{avar}(\widetilde{\mu}) &= \frac{1}{\widetilde{\upsilon}A(\widetilde{\upsilon})\sum_{t=p+1}^{T}(1-\sum_k \phi_k\cos(y_{t-k}-\widetilde{\mu}))^2} \\ &\simeq \frac{1}{\widetilde{\upsilon}A(\widetilde{\upsilon})(T-p-\{\sum_k \phi_k\}\sum_{t=p+1}^{T}\cos(y_t-\widetilde{\mu}))^2}.\end{aligned} \tag{18}$$

**Remark 3** *The ML estimates could be computed by a (modified) Newton-Raphson algorithm*[4], *independent of* $\upsilon$*; see Appendix B. Upon convergence the ML estimate of* $\upsilon$ *is obtained as* $\widetilde{\upsilon}=A^{-1}(S(\widetilde{\phi},\mu)/T)$. *The initial estimates are given by regressing* $\sin(y_t-\overline{y}_d)$ *on its lags, but these estimates are also obtained from the circular autocorrelations using the Yule-Walker equations.*

The SCAR(p) model, (16), can be extended so as to become a SCARMA$(p,q)$

[4] Fisher and Lee (1994, p 331) observe that for $IAR(p)$, ML estimation with exact derivatives becomes complicated for $p$ greater than 2 or 3.

by adding lagged scores as defined in (10). Thus

$$\mu_{t|t-1} = \mu + \phi_1 \sin(y_{t-1} - \mu) + ... + \phi_p \sin(y_{t-p} - \mu) + \theta_1 u_{t-1} + ... + \theta_q u_{t-q}, \quad (19)$$

where $-\pi < y, \mu \leq \pi$. The DCS component model is a parsimonious way of modeling SCMA($\infty$).

# 5 Monte Carlo experiments

A series Monte Carlo experiments were carried out in order to evaluate the finite sample properties of the estimated score driven circular model. The data were generated from the model summarised in equations (8), (9) and (10) with $\varepsilon_t \sim vM(0, \upsilon)$. The first fifty observations were discarded in order to remove the effect of initialisation and time series of length $T = \{250, 500, 1000, 2000\}$ were used for estimation. The parameter values are as follows:

$$\mu = \pi/4, \qquad \phi = \{0.7, 0.9, 0.98\}, \qquad \kappa = \{0.2, 0.5, 1\}, \qquad \upsilon = \{0.5, 2, 4\}.$$

Although the value of $\mu$ is set at $\pi/4$, experimentation with other values for $\mu$ indicates that, as theory suggests, the value of $\mu$ has no bearing on the results. The parameter values cover an empirically plausible range and include values similar to those reported below in the study of wind direction for Black Mountain. After generating the data, the circular DCS model was estimated by ML with the conditional mean initialised at $\mu_0 = \mu$. The whole process was replicated 10,000 times and the resulting mean square errors (MSEs) of the estimates reported in Table 1.

For $\phi = 0.7$ and $\phi = 0.9$, the asymptotic MSEs are quite close to the simulated MSEs when $T = 1000$ or more. For smaller samples, the MSEs for $\phi$ are not always close to the asymptotic values. In particular, when $\phi = 0.98$ the asymptotic MSEs for $\phi$ and $\mu$ are not a good guide to the sample MSEs, even for $T = 2000$. Given the proximity to the unit root this may not be surprising. Overall, the sample MSEs for $\kappa$ and $\upsilon$ are much closer to the asymptotic MSEs. Finally, note that, in accordance with the theory, a lower $\upsilon$ means a higher MSE.

Table 2 shows the results of Monte Carlo experiments, again based on 10,000 replications, for the nonstationary model (14) with no intercept. The MSEs are close to the values given by the asymptotic theory.

# 6 Model selection and forecasting

Circular sample autocorrelations can be used to suggest possible dynamic specifications and to check on their effectiveness in handling dependence. Goodness of fit criteria can be used to choose the best model. Section 7 provides an illustration.

## 6.1 Testing for uniformity

The Rayleigh test of the null hypothesis of a uniform distribution is based on the square of the mean resultant length, that is

$$R^2 = \left(T^{-1}\sum_{t=1}^{T}\cos y_t\right)^2 + \left(T^{-1}\sum_{t=1}^{T}\sin y_t\right)^2.$$

The asymptotic distribution of $2TR^2$ is $\chi_2^2$ under the null hypothesis of a uniform distribution; see Mardia and Jupp (2000, p 94-5). In the present context, failure to reject the null hypothesis indicates no serial correlation or, as was shown in sub-section 4.3, a unit root. A test of independence based on the circular autocorrelations of (20) below should indicate the possibility of a nonstationary time series; see the discussion in Fisher (1993, p 184).

Table 1:

Mean square errors of the maximum likelihood estimates of the score driven model for circular data based on the von Mises distribution. Asymptotic standard errors are shown in brackets.

| $\phi$ | $\kappa$ | $\upsilon$ | $T$ | $\mu$ | $\phi$ | $\kappa$ | $\upsilon$ |
|---|---|---|---|---|---|---|---|
| 0.9 | 0.5 | 2 | 250 | 0.916 | 0.033 | 0.062 | 0.264 |
| | | | 500 | 0.309 (0.272) | 0.011 (0.008) | 0.031 (0.029) | 0.125 (0.122) |
| | | | 1000 | 0.139 | 0.004 | 0.015 | 0.062 |
| | | | 2000 | 0.068 (0.068) | 0.002 (0.002) | 0.008 (0.007) | 0.031 (0.030) |
| 0.98 | 0.5 | 2 | 250 | 14.83 | 0.012 | 0.071 | 0.330 |
| | | | 500 | 11.45 (4.543) | 0.004 (0.001) | 0.036 (0.024) | 0.170 (0.122) |
| | | | 1000 | 7.752 | 0.002 | 0.018 | 0.083 |
| | | | 2000 | 3.390 (1.136) | 0.001 (0.0002) | 0.007 (0.006) | 0.039 (0.030) |
| 0.7 | 0.5 | 2 | 250 | 0.127 | 0.132 | 0.077 | 0.260 |
| | | | 500 | 0.064 (0.064) | 0.056 (0.045) | 0.038 (0.037) | 0.128 (0.122) |
| | | | 1000 | 0.032 | 0.024 | 0.018 | 0.062 |
| | | | 2000 | 0.016 (0.016) | 0.012 (0.011) | 0.009 (0.009) | 0.030 (0.030) |
| 0.9 | 1 | 2 | 250 | 3.432 | 0.017 | 0.075 | 0.278 |
| | | | 500 | 1.134 (0.756) | 0.007 (0.004) | 0.035 (0.033) | 0.131 (0.122) |
| | | | 1000 | 0.477 | 0.003 | 0.017 | 0.064 |
| | | | 2000 | 0.222 (0.189) | 0.001 (0.001) | 0.008 (0.008) | 0.032 (0.030) |
| 0.9 | 0.2 | 2 | 250 | 0.196 | 0.211 | 0.048 | 0.262 |
| | | | 500 | 0.082 (0.081) | 0.051 (0.019) | 0.023 (0.022) | 0.130 (0.122) |
| | | | 1000 | 0.040 | 0.014 | 0.011 | 0.062 |
| | | | 2000 | 0.020 (0.020) | 0.006 (0.005) | 0.006 (0.005) | 0.031 (0.030) |
| 0.9 | 0.5 | 0.5 | 250 | 1.470 | 0.328 | 0.374 | 0.094 |
| | | | 500 | 0.656 (0.574) | 0.068 (0.016) | 0.158 (0.123) | 0.046 (0.044) |
| | | | 1000 | 0.306 | 0.014 | 0.070 | 0.022 |
| | | | 2000 | 0.151 (0.143) | 0.005 (0.004) | 0.033 (0.031) | 0.011 (0.011) |
| 0.9 | 0.5 | 4 | 250 | 0.603 | 0.026 | 0.045 | 1.137 |
| | | | 500 | 0.204 (0.162) | 0.009 (0.007) | 0.022 (0.022) | 0.541 (0.520) |
| | | | 1000 | 0.084 | 0.004 | 0.011 | 0.265 |
| | | | 2000 | 0.041 (0.040) | 0.002 (0.002) | 0.005 (0.005) | 0.131 (0.130) |

Table 2:

Scaled Mean square errors of the maximum likelihood estimates of the score driven model for the nonstationary model (14). Asymptotic standard errors are shown in brackets.

| $\kappa$ | $\upsilon$ | $T$ | $\kappa$ | $\upsilon$ |
|---|---|---|---|---|
| 0.5 | 2 | 250 | 0.052 (0.044) | 0.260 (0.244) |
| | | 500 | 0.026 (0.022) | 0.128 (0.122) |
| | | 1000 | 0.017 (0.011) | 0.062 (0.062) |
| | | 2000 | 0.010 (0.005) | 0.032 (0.030) |
| 1 | 2 | 250 | 0.066 (0.061) | 0.255 (0.244) |
| | | 500 | 0.032 (0.031) | 0.127 (0.122) |
| | | 1000 | 0.017 (0.015) | 0.062 (0.062) |
| | | 2000 | 0.009 (0.008) | 0.030 (0.030) |

## 6.2 Circular autocorrelation functions

The circular ACF for a uniform distribution is defined as

$$\rho_c^*(\tau) = \frac{\gamma_\tau^{CC}\gamma_\tau^{SS} - \gamma_\tau^{CS}\gamma_\tau^{SC}}{\gamma_0^{SS}\gamma_0^{CC} - (\gamma_0^{SC})^2}, \qquad \tau = 1, 2, ... \tag{20}$$

where $\gamma_\tau^{CC} = \mathbb{E}[\cos y_t \cos y_{t-\tau}]$ and similarly for $\gamma_\tau^{SS}, \gamma_\tau^{CS}$ and $\gamma_\tau^{SC}$. Both $\sin y_t$ and $\cos y_t$ have zero means because of a uniformity assumption; see Holzmann, Munk, Suster and Zucchini (2006). An alternative form is in (6.36) of Fisher (1993, p 151). Fisher and Lee (1994, p 333) write down the corresponding correlogram.

When the distribution is not uniform, the directional mean needs to be subtracted; see Fisher (1993, p151-2). The circular correlation coefficient proposed by Jammalamadaka and SenGupta (2001, p176-9) is formulated somewhat differently and it implies a circular ACF given by

$$\rho_c(\tau) = \gamma_\tau^{SS}/\gamma_0^{SS}, \qquad \tau = 0, 1, 2, ... \tag{21}$$

where $\gamma_\tau^{SS} = \mathbb{E}[\sin(y_t - \mu)(\sin(y_{t-\tau} - \mu)]$, $\tau = 0, 1, 2, ...$

The sample[5] circular ACF corresponding to (21) is

$$r_c(\tau) = \frac{\sum \sin(y_t - \overline{y}_d)\sin(y_{t-\tau} - \overline{y}_d)}{\sum \sin^2(y_t - \overline{y}_d)}, \qquad \tau = 1, 2, ... \tag{22}$$

The limiting distribution when the observations are independent and identically distributed (IID) is standard normal, that is $\sqrt{T} r_c(\tau) \rightarrow N(0,1)$; see Brockwell and Davis (1991, Theorem 7.7.2).

The Lagrange multiplier (LM) test against serial correlation in location is based on the portmanteau or Box-Ljung statistic constructed from the autocorrelations of the scores; see Harvey (2013, p 52-4) and Harvey and Thiele (2016). For a vM distribution with $\upsilon > 0$, the scores are proportional to the sines of the angular observations measured as deviations from their directional mean, so the autocorrelations are the circular autocorrelations as defined in (21). The derivation can be based on the SCAR or SCMA models, the latter being a special case of (19) with no lagged unconditional scores. When the Q-statistic in the portmanteau test is based on the first $P$ sample autocorrelations, it is asymptotically distributed as $\chi^2_P$ under the null hypothesis of serial independence. Once a dynamic model has been fitted, a formal test requires that the degrees of freedom be adjusted by subtracting the number of estimated dynamic parameters from $P$. This is the Box-Pierce test. An alternative is to carry out an LM test; see the discussion in Harvey and Thiele (2016).

For the purposes of initial model identification it is helpful to know something about the behaviour of the CACF in (21) for wrapped models. From Jammalamadaka and SenGupta (2001, p 180), the circular ACF for a wrapped Gaussian model, constructed as in (5), is

$$\rho_c(\tau) = \frac{\sinh(2\gamma_x(\tau))}{\sinh(2\gamma_x(0))}, \qquad \tau = 0, 1, 2... \tag{23}$$

[5] There appears to be a typographical error in the sample correlation given in (8.2.5) of Jammalamadaka and SenGupta (2001, p178) because it is not consistent with the theoretical definition.

The wrapping diminishes the autocorrelations, the more so the bigger is the variance of the unwrapped series, $\gamma_x(0)$. On the other hand, as $\gamma_x(0) \to 0$, $\rho_c(\tau) \to \rho_x(\tau)$, that is the ACF of $\rho_c(\tau)$ is close to that of $x_t$. Thus whereas the ACF of unwrapped observations, were they available, could be interpreted in the usual way for linear data, this is no longer true for the wrapped observations unless the variance is small. For a score-driven model, the issues are somewhat different because $\sin(y_t - \mu_d) = \sin(x_t - \mu)$ and so, since $\mu_d = \mu$, the dynamic properties of the wrapped and unwrapped series are the same. The challenge is therefore to determine the properties of the unwrapped series.

## 6.3 Goodness of fit of the distribution

The residuals are given by

$$\nu_t = \min \left| y_t - \mu_{t|t-1}, \; y_t - \mu_{t|t-1} \pm 2\pi \right| \tag{24}$$

There is no closed form CDF for the vM distribution, but probability integral transforms (PITs) can be computed by approximations as detailed in Mardia and Jupp (2000, p 41). An LM test against a class of exponential distributions, given in Mardia and Jupp (2000, p 142-3), can be carried out. A rejection of the vM distribution may lead one to consider a more general class of distributions, such as the one proposed by Jones and Pewsey (2005).

When a model has been fitted, the most informative diagnostic plot is one where $y_t$ is adjusted, by adding or subtracting $2\pi$, so as to be in the range $\mu_{t|t-1} \pm \pi$. In this way observations close to $\pm\pi$ no longer appear at both the top and bottom of the graph.

Goodness of fit may be assessed by the dispersion (circular variance)

$$D = 1 - \sum_{t=1}^{T} \cos(y_t - \mu_{t|t-1})/T \tag{25}$$

or the circular standard deviation, $s = \sqrt{-2\ln(1-D)}$, a measure whose square is most comparable to the prediction error variance; see Mardia and Jupp (2000, pp 18-19, 30). In time series forecasting the random walk often provides a useful benchmark. The equivalent benchmark for dispersion is $D_\Delta = 1 - \sum_{t=2}^{T}\cos(y_t - y_{t-1})/(T-1)$. Hence goodness of fit for a particular model might be characterized by $A_\Delta = 1 - D/D_\Delta$; with a perfect fit $A_\Delta = 1$. Alternatively we could use $B_\Delta = 1 - s^2/s_\Delta^2$, where $s_\Delta^2 = -2\ln(1-D_\Delta)$. A negative value for $A_\Delta$ or $B_\Delta$ indicates that the model is worse than a forecast given by the last observation.

### 6.4 Forecasts

When a forecast of the next observation, that is $\mu_{t+1|t}$, falls outside the range it can be reset, as in (5), to give $\widetilde{y}_{t+1|t}$ in the range $[-\pi, \pi)$. The conditional distribution for $y_{t+1}$ is $vM(\widetilde{y}_{t+1|t}, \upsilon)$.

The conditional distribution of $y_{t+\ell}$ may be obtained by simulation with the accuracy measured by $D(\ell) = 1 - \sum_{j=1}^{\ell}\cos(y_{t+j} - \mu_{t+j|t})/\ell$.

## 7 Wind direction on Black Mountain

Fisher and Lee (1994) consider $T = 72$ hourly measurements of wind direction taken over a period of four days on Black Mountain, ACT, Australia. The data can be found in Fisher (1993) and are in degrees from 0 to 360. We converted to radians, subtracted the directional mean, $\overline{y}_d = 5.083$ and added $2\pi$ to some observations [6] so that they are all in the range $[-\pi, \pi)$.

Figure 2 shows the circular correlogram based on sines. This is very similar to the circular correlogram in Figure 2a of Fisher and Lee (1994, p 336). At first sight the pattern casts doubt on the AR(1) specification in that

[6] This means that once a model has been fitted, the observations need to be transformed back to what they were originally. In the DCS model no transformations are needed prior to model fitting.

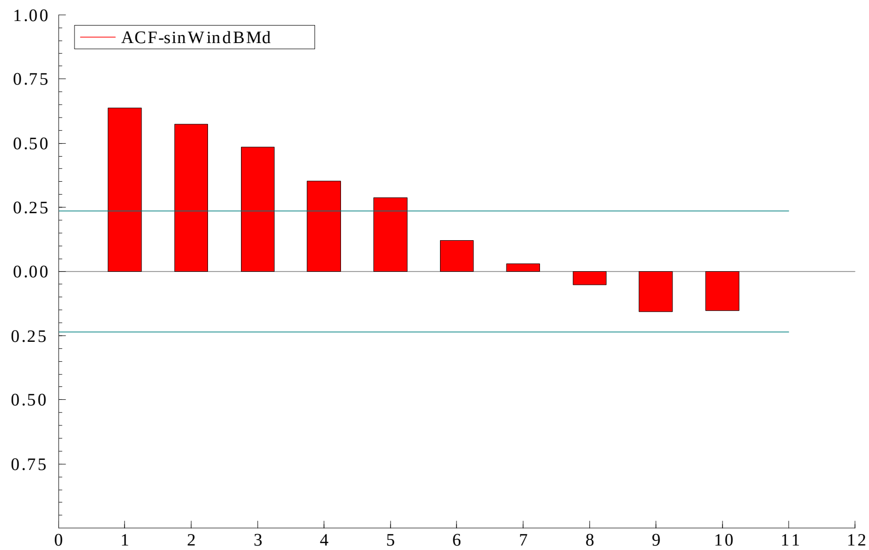


Figure 2: Circular correlogram (sines) for Black mountain observations. The horizontal lines are $\pm 2/\sqrt{T}$.

the structure is more indicative of $ARMA(1,1)$. However, given the damping affect on circular autocorrelations highlighted by (23), an $AR(1)$ model may not be unreasonable. The ambiguity shows that circular correlograms need to be interpreted with care.

The histogram of the circular observations, which is shown in Figure 3, suggests that a transformation may be neither necessary nor desirable. A probit would produce a normal distribution if the original distribution were uniform - which it clearly is not. As it is, the excess kurtosis for a probit transformation is 3.02. The $\tan(y/2)$ transformation is even more extreme in this respect with excess kurtosis of 13.51; it is perhaps not surprising that the estimate of $\phi$ in the CAR(1) model is only 0.35.

Table 3 shows the results of fitting various models. The first model ignores circularity by assuming a conditional Gaussian distribution. The other models take it to be von Mises. The benchmark given by the circular variance for first differences is $D_{\Delta} = 0.258$ implying $s_{\Delta}^2 = 0.597$. Standard errors are

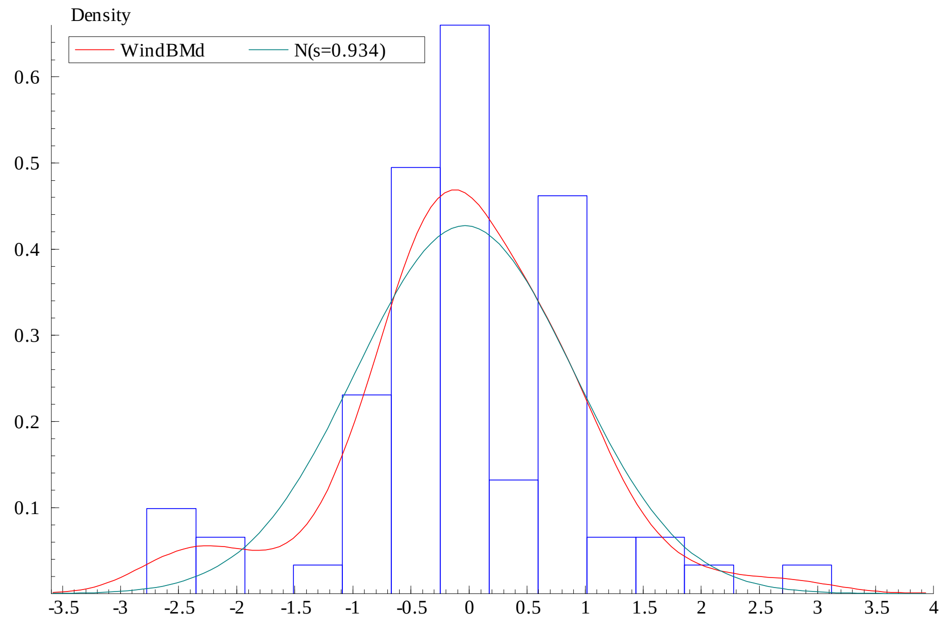


Figure 3: Histogram of Black mountain data

shown for the DCS(1) model - the first-order filter of (8) - and the SCAR(1) model. For DCS(1) these were obtained from (11).

The score-driven models give the best fit. Furthermore the circular correlogram of the best-fitting model, SCAR(1), shows very little evidence of residual serial correlation; see Figure 4. Note however, that the superior fit of the SCAR model over the DCS is due to the first 18 observations which lie mainly below the others. If these observations are dropped, DCS(1) is the better model.

Figure 5 shows the filtered conditional mean for the DCS model and compares it with the $IAR(1)$ filter given by the $\tan(y/2)$ transformation. The DCS filter is much less variable; the SCAR filter behaves in a similar way. A plot for the probit $IAR(1)$ lies between the $IAR$ $\tan(y/2)$ and the DCS. If the data are not centred by subtracting the directional mean, the IAR filters behave differently whereas the score filters are basically unaffected.

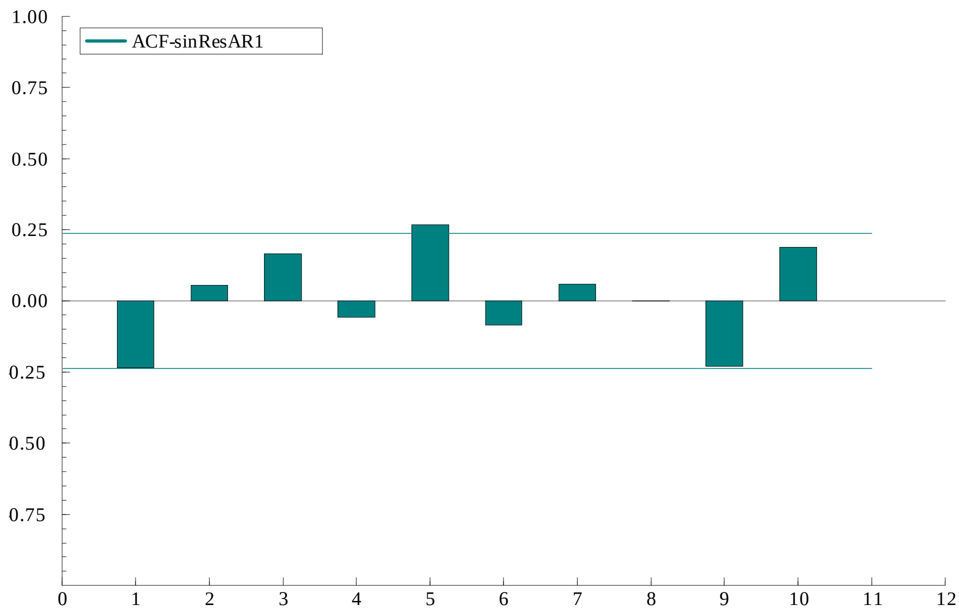


Figure 4: Circular correlogram of residuals from SCAR(1) model.

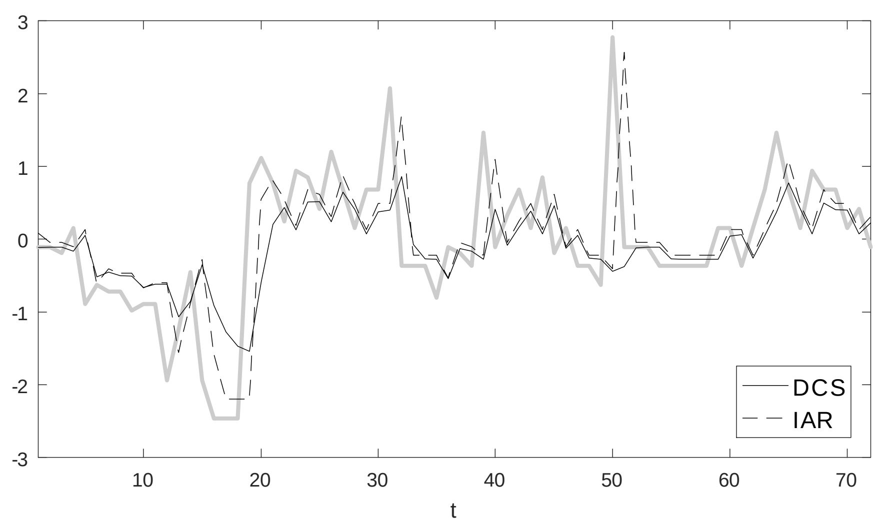


Figure 5: Filters for DCS and IAR tan models

Table 3:
Estimates and goodness of fit measures for Black Mountain data.
Standard errors are shown in parentheses.

| Model | $\mu$ | $\phi$ | $\kappa$ | $\upsilon$ | $D$ | $s^2$ | $A_{\Delta}$ |
|---|---|---|---|---|---|---|---|
| Gaussian AR(1) | $-0.02$ | 0.52 | - | - | 0.243 | 0.557 | 0.057 |
| | (0.10) | (0.14) | | | | | |
| IAR(1) Probit | $-0.03$ | 0.68 | - | 2.46 | 0.239 | 0.547 | 0.071 |
| | (0.26) | (0.14) | - | (0.35) | | | |
| IAR(1) tan | 0.08 | 0.67 | - | 2.44 | 0.242 | 0.555 | 0.060 |
| | (0.27) | (0.15) | - | (0.35) | | | |
| SCAR(1) | $-0.60$ | 1.24 | - | 3.00 | 0.190 | 0.421 | 0.263 |
| | (0.13) | (0.13) | - | (0.43) | | | |
| DCS(1) | $-0.11$ | 0.66 | 0.64 | 2.54 | 0.231 | 0.526 | 0.103 |
| | (0.20) | (0.16) | (0.15) | (0.36) | | | |

Fisher and Lee (1994) also estimated a $CAR(1)$ model with parameter $\widetilde{\phi} = 0.52$ after a probit transformation. The CAR models do not give one-step ahead forecasts with a vM distribution and so are difficult to compare directly with IAR models. In any case they are a much less attractive option.

# 8 Conclusions

This article shows how the score-driven approach provides a natural solution to the difficulties posed by circular data and leads to a coherent and unified methodology for estimation, model selection and testing. The data generating process is unaffected by any wrapping of the observations and the models estimated by maximum likelihood are unaffected by the way the data is cut. Two classes of models are introduced, one based on a filtered component and the other taking an autoregressive form. An asymptotic theory is developed and Monte Carlo experiments examine small sample performance. Diagnostic checks for serial correlation follow straightforwardly. The new models are fitted to hourly data on wind direction and are shown to provide a better fit than existing methods.

The score-driven approach may be extended in a number of directions. Firstly conditional distributions other than von Mises are easily accommodated. Secondly heteroscedasticity can be modeled with dynamic equations driven by the score with respect to concentration. Thirdly dynamic seasonal and diurnal effects (for hourly data) can be handled and finally the approach can be used to formulate models for circular-linear data. These issues will be addressed in later work.

## Appendix A: Information matrix for the DCS model

The information matrix for the ML estimator of $\boldsymbol{\psi}$ is, from Harvey (2013, p 37),

$$\mathbf{D}(\boldsymbol{\psi}) = \mathbf{D}\begin{pmatrix} \kappa \\ \phi \\ \mu \end{pmatrix} = \frac{1}{1-b}\begin{bmatrix} A & D & E \\ D & B & F \\ E & F & C \end{bmatrix} \tag{26}$$

with

$$\begin{aligned} A &= \sigma_u^2 = A(\upsilon)/\upsilon, \qquad B = \frac{\kappa^2\sigma_u^2(1+a\phi)}{(1-\phi^2)(1-a\phi)}, \qquad C = \frac{(1-\phi)^2(1+a)}{1-a}, \\ D &= \frac{a\kappa\sigma_u^2}{1-a\phi}, \qquad E = \frac{c(1-\phi)}{1-a} \qquad \textit{and} \qquad F = \frac{ac\kappa(1-\phi)}{(1-a)(1-a\phi)}. \end{aligned}$$

Now

$$\begin{aligned} a &= \phi - \kappa A(\upsilon) \\ b &= \phi^2 - 2\phi\kappa A(\upsilon) + \kappa^2(1 - A(\upsilon)/\upsilon) \end{aligned}$$

because $\mathbb{E}\left(\partial u_t/\partial\mu\right)^2 = \mathbb{E}\left(\cos^2(y_t - \mu_{t|t-1})\right)$ and we know from the information quantity for $\upsilon$ that $\mathbb{E}[\left(\cos(y_t - \mu_{t|t-1}) - A(\upsilon)\right)^2] = 1 - A(\upsilon)^2 - A(\upsilon)/\upsilon$. Finally $c = -\kappa\mathbb{E}\left(\sin(y_t - \mu_{t|t-1})\cos(y_t - \mu_{t|t-1})\right) = -\kappa\mathbb{E}\left(\sin\{2(y_t - \mu_{t|t-1})\}\right)/2 = 0$. Thus $E = F = 0$.

There are no extra terms because the off-diagonals in the information matrix, (11), are zero and $u_t = \sin(y_t - \mu_{t|t-1})$ does not depend on $\upsilon$; see http://www.econ.cam.ac.uk/DCS/docs/Lemma10.pdf for further details on the issues involved.

## Appendix B: The SCAR(p) model and the Newton-Raphson algorithm

Note that the normal equations for $\boldsymbol{\phi}$ are

$$\frac{\partial \ln L}{\partial \phi_j} = \upsilon \sum_{t=p+1}^{T} \sin(y_t - \mu_{t|t-1}) \sin(y_{t-j} - \mu) = 0, \qquad j = 1, ..., p. \tag{27}$$

Differentiating again gives

$$\frac{\partial^2 \ln L}{\partial \phi_j \partial \phi_k} = -\upsilon \sum_{t=p+1}^{T} \cos(y_t - \mu_{t|t-1}) \sin(y_{t-j} - \mu) \sin(y_{t-k} - \mu), \qquad j, k = 1, ..., p$$

Taking conditional expectations at time $t-1$ yields

$$-\mathbb{E}_{t-1} \frac{\partial^2 \ln f}{\partial \phi_j \partial \phi_k} = \upsilon A(\upsilon) \sin(y_{t-j} - \mu) \sin(y_{t-k} - \mu), \qquad j, k = 1, ..., p \tag{28}$$

at the true parameter values. The unconditional expectation gives the circular autocovariances. Furthermore

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \phi_j \partial \mu} = & -\upsilon \sum_{t=p+1}^{T} \cos(y_t - \mu_{t|t-1})[1 - \sum_k \phi_k \cos(y_{t-k} - \mu)] \sin(y_{t-j} - \mu) \\ & -\upsilon \sum_{t=p+1}^{T} \sin(y_t - \mu_{t|t-1}) \cos(y_{t-j} - \mu), \qquad j = 1, ..., p \end{aligned}$$

so

$$\mathbb{E}_{t-1} \frac{\partial^2 \ln L}{\partial \phi_j \partial \mu} = -\upsilon A(\upsilon) \sum_{t=p+1}^{T} [1 - \sum_k \phi_k \cos(y_{t-k} - \mu)] \sin(y_{t-j} - \mu).$$

The unconditional expectation is zero because sine is odd and cosine is even and so $\cos(y_{t-k} - \mu) \sin(y_{t-j} - \mu)$ is odd and its (unconditional ) expectation is zero. As regards $\upsilon$, taking conditional expectations shows that $\mathbb{E}(\partial^2 \ln L / \partial \phi_j \partial \upsilon) = 0$, $j = 1, ..., p$ and $\mathbb{E}(\partial^2 \ln L / \partial \mu \partial \upsilon) = 0$.

The result for $\mu$ follows because

$$\frac{\partial \ln L}{\partial \mu} = \upsilon \sum_{t=p+1}^{T} \sin(y_t - \mu_{t|t-1})[1 - \sum_k \phi_k \cos(y_{t-k} - \mu)]$$

and

$$\begin{aligned}\frac{\partial^2 \ln L}{\partial \mu^2} =& -\upsilon \sum_{t=p+1}^{T} \cos(y_t - \mu_{t|t-1})[1 - \sum_k \phi_k \cos(y_{t-k} - \mu)]^2 \\ & -\upsilon \sum_{t=p+1}^{T} \sin(y_t - \mu_{t|t-1}) \sum_k \phi_k \sin(y_{t-k} - \mu)\end{aligned}$$

Taking conditional expectations removes the last term.

A Newton-Raphson algorithm is an option because if $\cos(y_t - \widehat{\mu}_{t|t-1})$ is dropped, the computations reduce to repeated regressions of $\sin(y_t - \widehat{\mu}_{t|t-1})$ on $\sin(y_{t-j} - \widehat{\mu})$, $j = 1, ..., p$, with the estimate of $\boldsymbol{\phi}$ updated by adding the latest regression coefficients to it. Alternatively, expression (28) suggests that it could be replaced by $\widehat{A}(\upsilon) = \sum_{t=p+1}^{T} \cos(y_t - \widehat{\mu}_{t|t-1})/(T-p)$ in what amounts to the method of scoring. By a similar argument, a regression of $\sin(y_t - \widehat{\mu}_{t|t-1})$ on $1 - \sum_k \widehat{\phi}_k \cos(y_{t-k} - \widehat{\mu})$ with $\widehat{A}(\upsilon)$ included as a divisor, could be used to update $\widehat{\mu}$. If $\cos(y_t - \widehat{\mu}_{t|t-1})$ is retained in the Hessian its role is to downweight observations far from the conditional mean (a heteroscedasticity correction).

# References

Blasques, F., Gorgi, P., Koopman, S.J. and O. Wintenberger (2018). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics*, **12**, 1019–1052.

Breckling, J. (1989). *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*. Berlin: Springer.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods*. Springer Series in Statistics. Berlin: Springer.

Coles, S. G. (1998). Inference for circular distributions and processes. *Statistical Computing*, **5**, 108–13.

Creal, D., Koopman, S. J. and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, **28**, 777–95.

Fernandez-Duran, J.J. (2004) Circular distributions based on nonnegative trigonometric sums. *Biometrics*, **60**, 499–503.

Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge: Cambridge University Press.

Fisher, N.I. and A.J. Lee (1992). Regression models for an angular response. *Biometrics*, **48**, 665–77.

Fisher, N.I. and A.J. Lee (1994). Time series analysis of circular data. *Journal of the Royal Statistical Society, B*, **70**, 327–332.

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

Harvey, A.C. and A. Luati. (2014). Filtering with heavy tails. *Journal of the American Statistical Association*, **109**, 1112–1122.

Harvey, A.C. and S. Thiele (2016). Testing against changing correlation. *Journal of Empirical Finance*, **38**, 575–89.

Holzmann, H., Munk, A., Suster, M. and W. Zucchini (2006). Hidden Markov models for circular and linear–circular time series. *Environmental and Ecological Statistics*, **13**, 325–347.

Jammalamadaka, S.A. and A.S. SenGupta (2001). *Topics in Circular Statistics*. New York: World Scientific Publishing Co.

Jensen, S. T. and A. Rahbek (2004). Asymptotic inference for nonstationary GARCH. *Econometric Theory*, **20**, 1203–26.

Jones, M.C. and A. Pewsey (2005) A family of symmetric distributions on the circle. *Journal of the American Statistical Association*, **100**,1422–1428.

Mardia, K. V. and P.E. Jupp (2000). *Directional Statistics*. Chichester: Wiley.