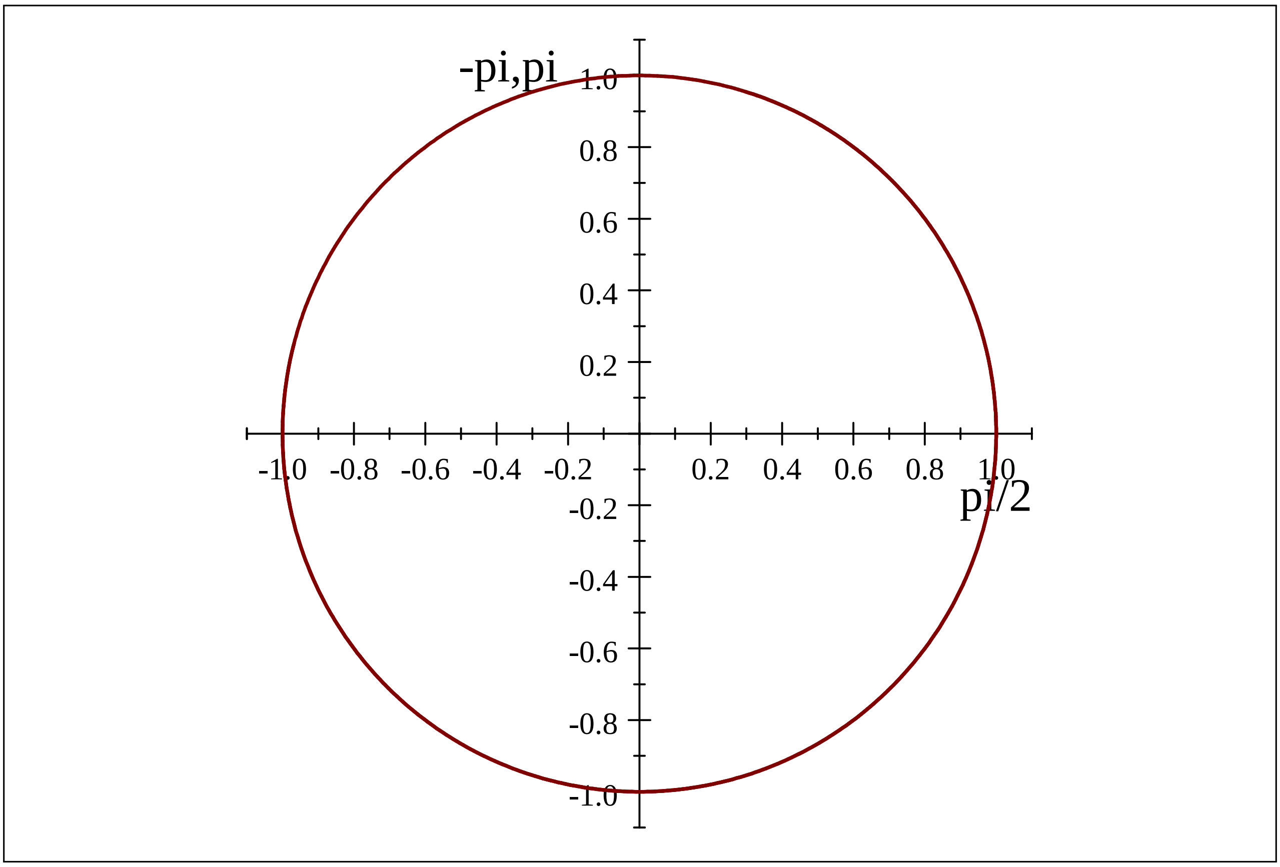
Modeling directional (circular) time series

Andrew Harvey (ach34@cam.ac.uk)
Faculty of Economics, Cambridge University
with Stan Hurn and Stephen Thiele (QUT, Brisbane)

March 2019



Due south is zero, West is negative, east is positive and north is $-\pi, \pi$.

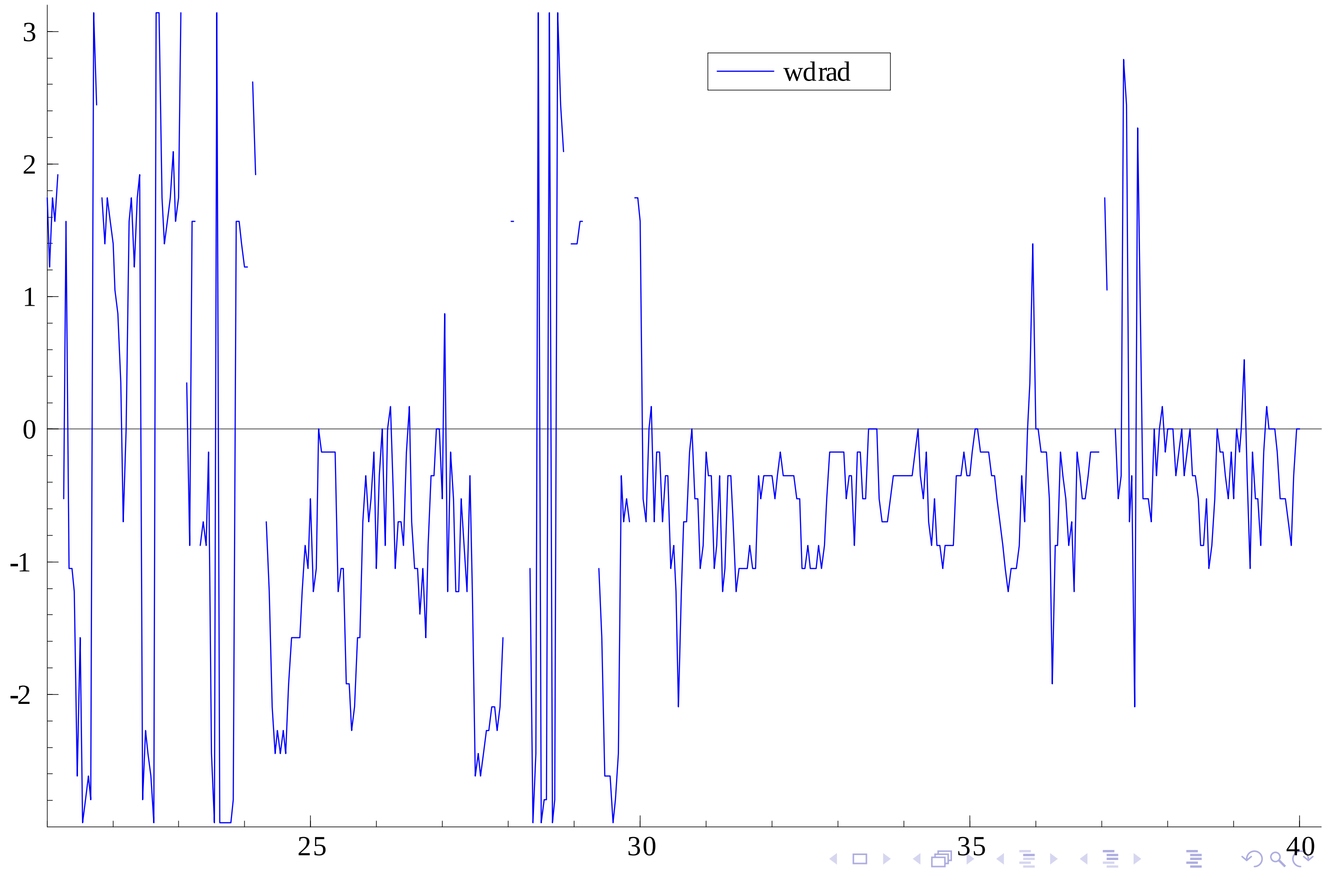


Figure: Hourly wind direction in Cairns

Directional variables and circularity

The points $-\pi$ and π meet up. There is a discontinuity at $\pm\pi$ and this poses a challenge for directional time series modelling. A number of ways of dealing with the problem 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 AR or ARMA modeling. (CARMA). A second approach uses tranformation to model the conditional mean as an AR.

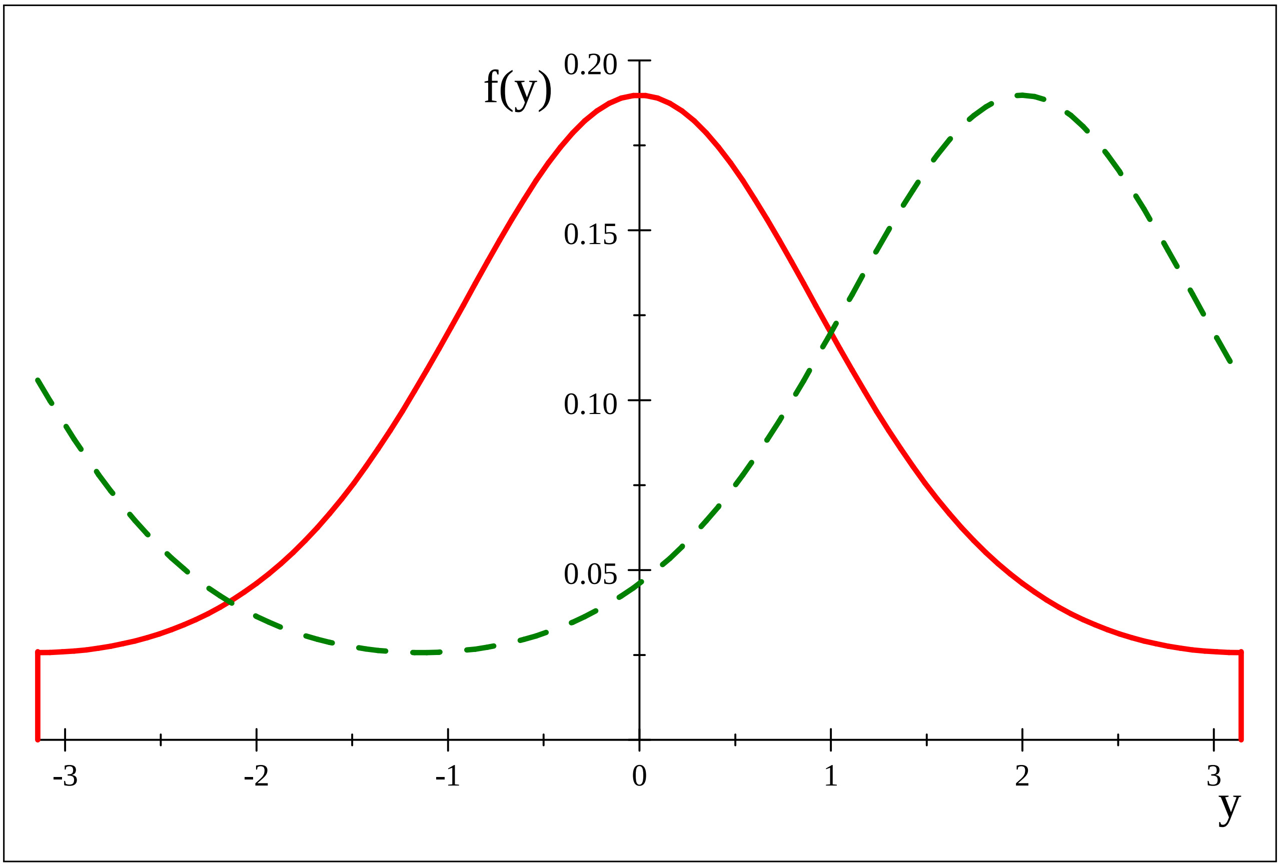
Dynamic Conditional Score models provide a much better solution.

von Mises (vM) distribution

Circular data measured in radians is usually taken to have a von Mises (vM) distribution with PDF

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

where $I_k(v)$ denotes a modified Bessel function of order k, μ is location and v is a non-negative concentration parameter that is inversely related to scale. When $v = 0$ the distribution is uniform whereas y is approximately $N(\mu, 1/v)$ for large v; the normal distribution is generally considered a good approximation for $v > 2$.
The (circular) variance of the von Mises distribution is $1- A(v)$, where $A(v) = I_1(v)/I_0(v)$.



Von Mises with $\mu = 0$ (red) $\mu = 2$ (green). Both have $v = 2$.

von Mises (vM) distribution

The score wrt μ is

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

Its variance is $\upsilon A(\upsilon)$.
The ML estimator of μ is the sample mean direction, $\overline{y}_d$. (The sample mean is not well-defined as it depends on where the circle is cut.)
The score wrt υ is

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

Existing time series models - CARMA

When observations are in radians, the range is 2π, so, for example, $-\pi < y \leq \pi$. The two standard approaches, described by Fisher and Lee (1994), transform y to a variable x in the range $-\infty < x < \infty$. The aim is have x close to being Gaussian.
The first method is to transform $y - \mu$, $-\pi < y - \mu \leq \pi$, and to fit a linear time series model to x. When the time series model is an ARMA it is called $CARMA(p, q)$, where the C denotes 'circular'. The link function is written $x = g^{-1}(y - \mu)$. Fisher and Lee (1994) prefer the inverse 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.
The probit gives a normal if the circular distribution is uniform. If $\upsilon > 2$, there may be no need for a transformation

Existing time series models- IAR

In the second class, the inverse form is an autoregression, $IAR(p)$, whereby the conditional mean, $\mu_{t|t-1}$, of a circular distribution is

$$\mu_{t|t-1} = \mu + g\{\alpha_1 g^{-1}(y_{t-1} - \mu) + ... + \alpha_p g^{-1}(y_{t-p} - \mu)\}, \qquad -\pi < y - \mu \leq \pi.$$

The $IAR(1)$ model can be approximated without the transformation when α is not too far from one, so

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

There is no $IARMA(p, q)$ model.
The transformations in the CARMA and IAR approaches do not actually address the issue of circularity because they set up a barrier for $y_t - \mu$ at π, $-\pi$ and so ignore the proximity of observations on either side.
A score-driven model produces a filter that respects the circularity of the distribution.

Score-driven models

When the conditional distribution is vM, letting the score drive a dynamic equation for $\mu_{t|t-1}$ solves the circularity problem because the sine curve of angular observations is continuous at π and $-\pi$ and $\sin(y_t \pm 2\pi) = \sin(y_t)$. Unlike the *CARMA* and *IAR* approaches which set up a barrier for $y_t - \mu$ at π, $-\pi$, a value of y_t slightly bigger than $-\pi$ is treated in the same way as if it were slightly bigger than π.
Dividing the score by its information quantity and dropping $A(\upsilon)$ means the filter is not dependent on υ. The first-order model is

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

where the conditional distribution of y_t is $vM(\mu_{t|t-1}, \upsilon)$, $\mu_{1|0} = \mu$ and

$$u_t = \sin(y_t - \mu_{t|t-1}).$$

Score-driven models

The variable u_t is a MD with mean zero and variance $A(\upsilon)/\upsilon$. Hence the model

$$y_t^u = \mu_{t|t-1} + \varepsilon_t, \qquad t = 1, ..., T,$$

where ε_t is $vM(0, \upsilon)$, is stationary with an unconditional mean of μ. Unlike in the *CARMA* and *IAR* models, the observations are not constrained to lie in the range $\mu \pm \pi$. An observation outside this range can be re-set by adding or subtracting 2π, that is

$$y_t = y_t^u \bmod 2\pi - \pi$$

if y_t is to be in the range $[-\pi, \pi)$.

Score-driven models

The DGP is invariant to such '**wrapping**' of the observations.
It is this invariance property which in our view defines a genuine circular time series model.
Figure contrasts the score variable with a linear response when $\mu = 0$. The score starts to downweight observations beyond $\pi/2$ and $-\pi/2$, reflecting the fact that, like the score for a $t-$distribution, it is a redescending function.
For small deviations from the mean, the score is approximately linear as $\sin(y_t - \mu_{t|t-1}) \simeq y_t - \mu_{t|t-1}$.

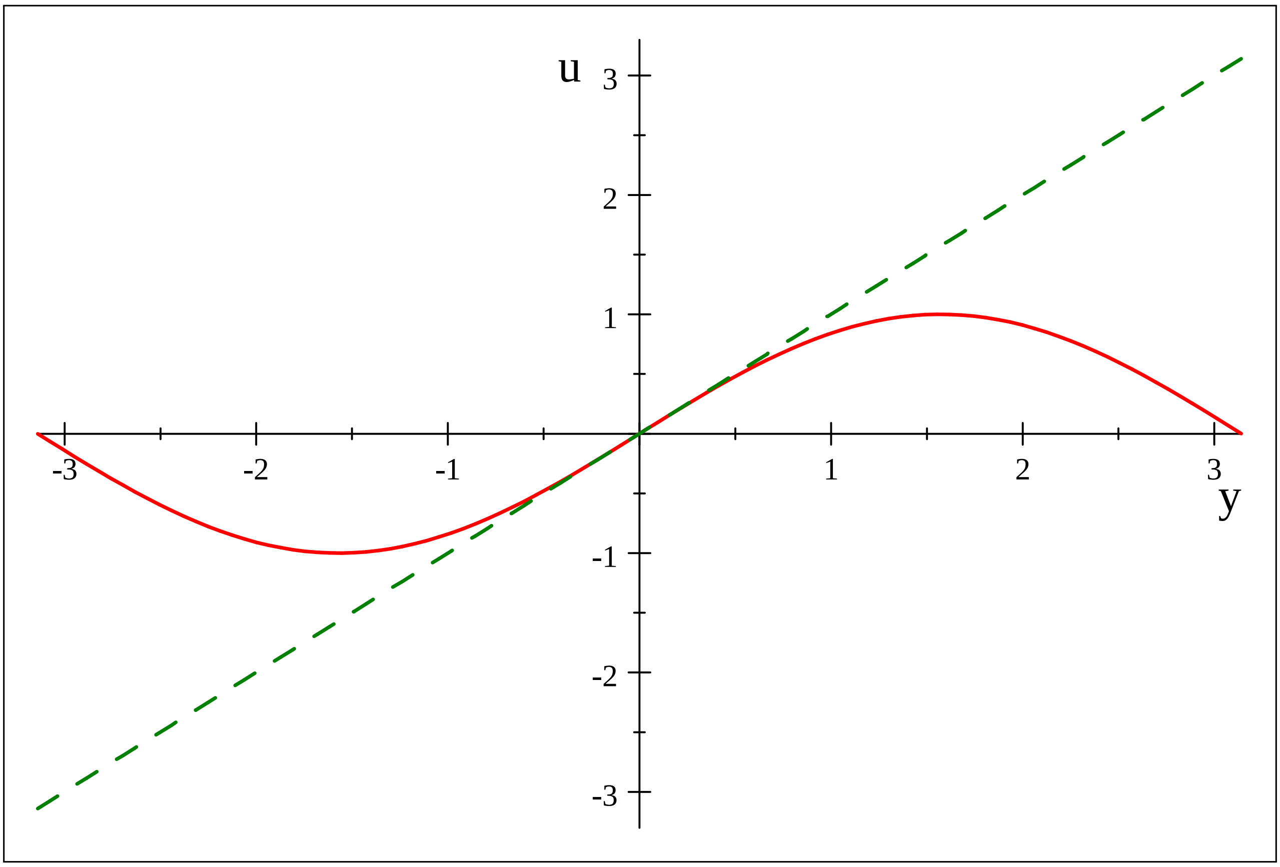
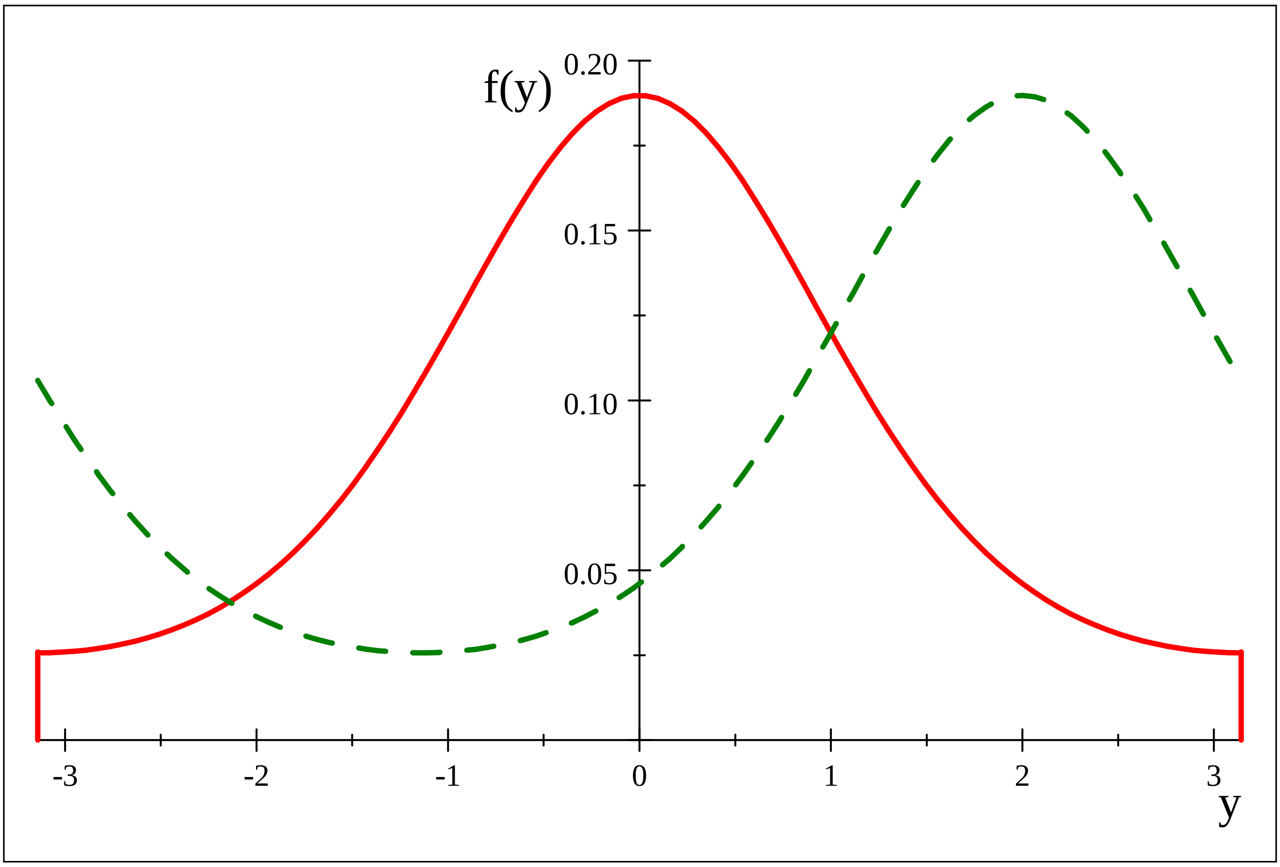


Figure: Sine and linear response for deviations from the conditional mean, set to zero for solid and dashed

Score-driven models

The critical role played by the score becomes clear when $\mu_{t|t-1}$ is close to π, say $\pi - a$, where a is small and positive. 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 is positive. There is no problem if its effect is to move the location beyond π.



Von Mises with $\mu = 0$ (red) $\mu = 2$ (green). Both have $\upsilon = 2$.

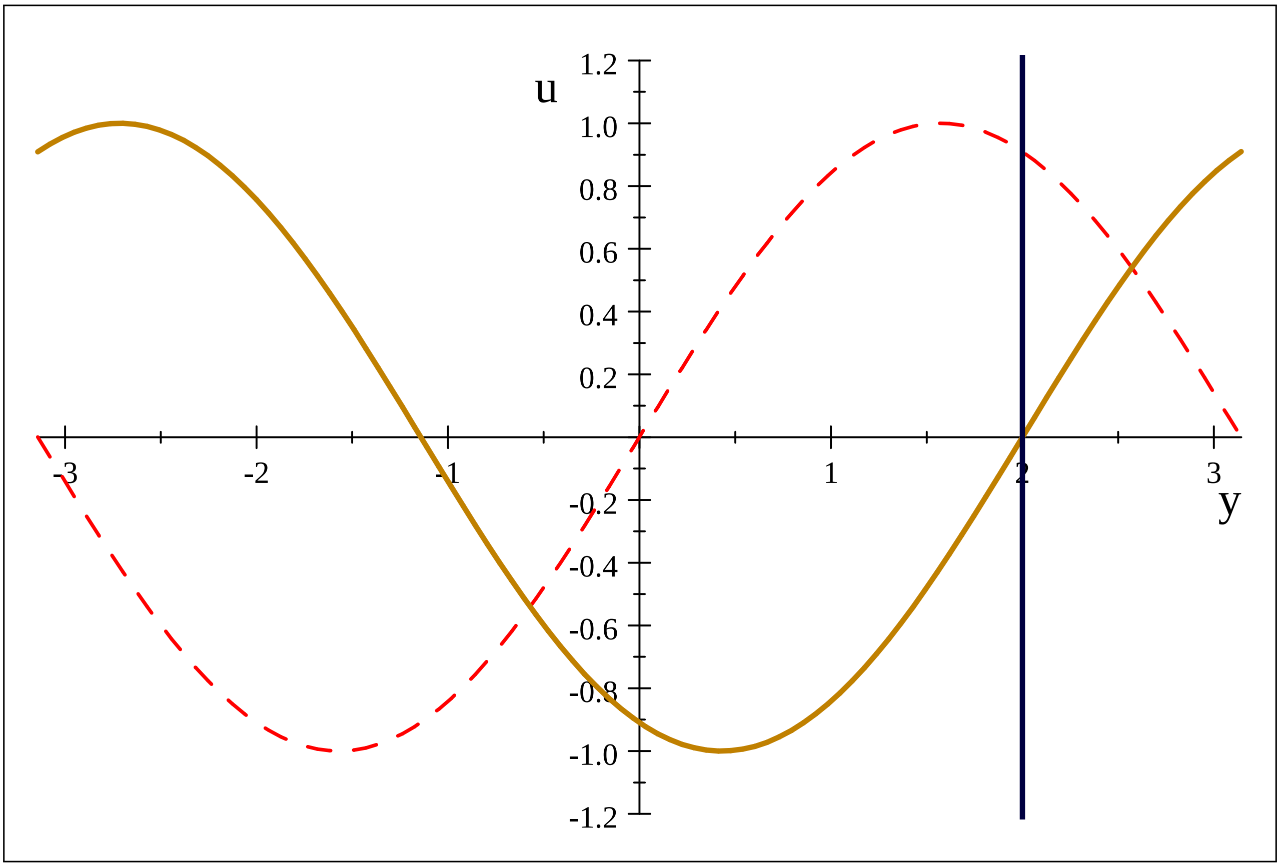


Figure: Sine response for deviations from the conditional mean, set to zero for dashed and to two for continuous.

Asymptotic distribution and nonstationarity

Asymptotic distribution of ML estimator can be derived as in Harvey (2013).
The non-stationary first-order model is

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

where $\mu_{1|0}$ is fixed. The mean can, in principle, travel all the way round the circle. When an observation falls outside the range $[-\pi, \pi)$, it is re-set so as to be in the admissible range by adding or subtracting 2π.
There is no need to reset $\mu_{t|t-1}$ but if it is reset - either in simulation or estimation - neither DGP nor estimation is not affected.
The unconditional distribution is uniform.

Score-driven autoregression - SCAR

Score-driven circular autoregressive (SCAR) model is

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

In other words the forcing variables are defined with the conditional mean, $\mu_{t|t-1}$, replaced by the **unconditional** mean, μ. Unlike the IAR models this model is invariant to translation.
Maximizing $S(\boldsymbol{\phi}) = \sum_{t=1}^{T} \cos(y_t - \mu_{t|t-1})$ for SCAR(p) model maximizes the log-likelihood and yields a set of normal equations

$$\frac{\partial S(\boldsymbol{\phi})}{\partial \phi_j} = \frac{\partial \ln L}{\partial \phi_j} = \upsilon \sum_{t=1}^{T} \sin(y_t - \mu_{t|t-1}) \sin(y_{t-j} - \mu) = 0, \quad j = 1, ..., p$$

Score-driven autoregression - SCAR

Differentiating again gives

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

Taking conditional expectations at time $t-1$ yields

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

at the true parameter values.

Score-driven autoregression - SCAR

The ML estimates could be computed by a (modified) Newton-Raphson algorithm. The υ parameter can be dropped. If $E\cos(y_t - \mu_{t|t-1})$ is treated as constant, the computations reduce to repeated regressions of $\sin(y_t - \widehat{\mu}_{t|t-1})$ on $\sin(y_{t-j} - \widehat{\mu})$, $j = 1, ..., p$. On convergence the asymptotic information matrix is estimated as

$$avar(\widetilde{\boldsymbol{\phi}}) = \frac{1}{\widetilde{\upsilon} A(\widetilde{\upsilon})} \left[\sum_{t=p+1}^{T} \mathbf{s}_t \mathbf{s}_t' \right]^{-1},$$

where the *k-th* element of the $p \times 1$ vector $\mathbf{s}_t$ is $\sin(y_{t-k} - \widehat{\mu})$.
Also

$$avar(\widetilde{\mu}) = \frac{1}{\widetilde{\upsilon} A(\widetilde{\upsilon})(T - p - \{\sum_k \phi_k\} \sum_{t=p+1}^{T} \cos(y_t - \widetilde{\mu}))^2}.$$

Forecasting

When a forecast of an observation, $\widetilde{y}^u_{t+1|t} = \mu_{t+1|t}$, falls outside the range it can be reset, that is

$$\widetilde{y}_{t+1|t} = \widetilde{y}^u_{t+1|t} \bmod 2\pi - \pi$$

gives $\widetilde{y}_{t+1|t}$ in the range $[-\pi, \pi)$.
The conditional distribution for y_{t+1} is $vM(\widetilde{y}_{t+1|t}, \widetilde{\upsilon})$ with the accuracy measured by $D = 1 - A(\widetilde{\upsilon})$.
The conditional distribution of $y_{t+\ell}$ may be obtained by simulation with the accuracy measured by

$$D(\ell) = 1 - \sum_{j=1}^{\ell} \cos(y_{t+j} - \mu_{t+j|t})/\ell.$$

Black mountain

Fisher and Lee (1994) consider hourly measurements of wind direction taken over a period of four days on Black mountain, ACT, Australia. They estimate a $CAR(1)$ model with parameter $\widetilde{\phi} = 0.52$ after an inverse probit transformation. They also fit the $IAR(1)$ model with a conditional vM distribution and obtain ML estimates of $\widetilde{\alpha} = 0.68$ and $\widetilde{\upsilon} = 2.47$.
Circular sample autocorrelations

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

don't have exactly the same interpretation as usual. Diminished.

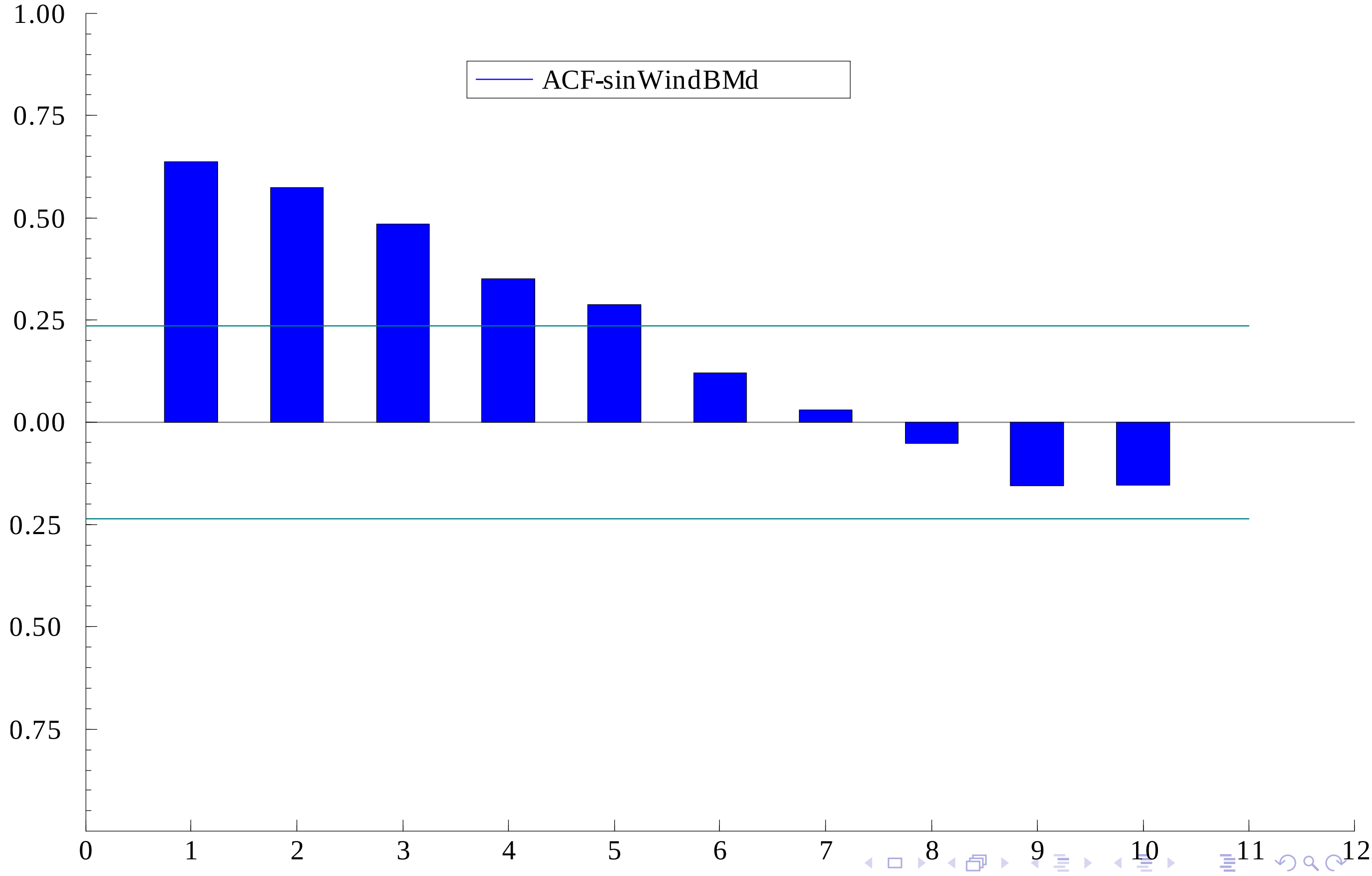
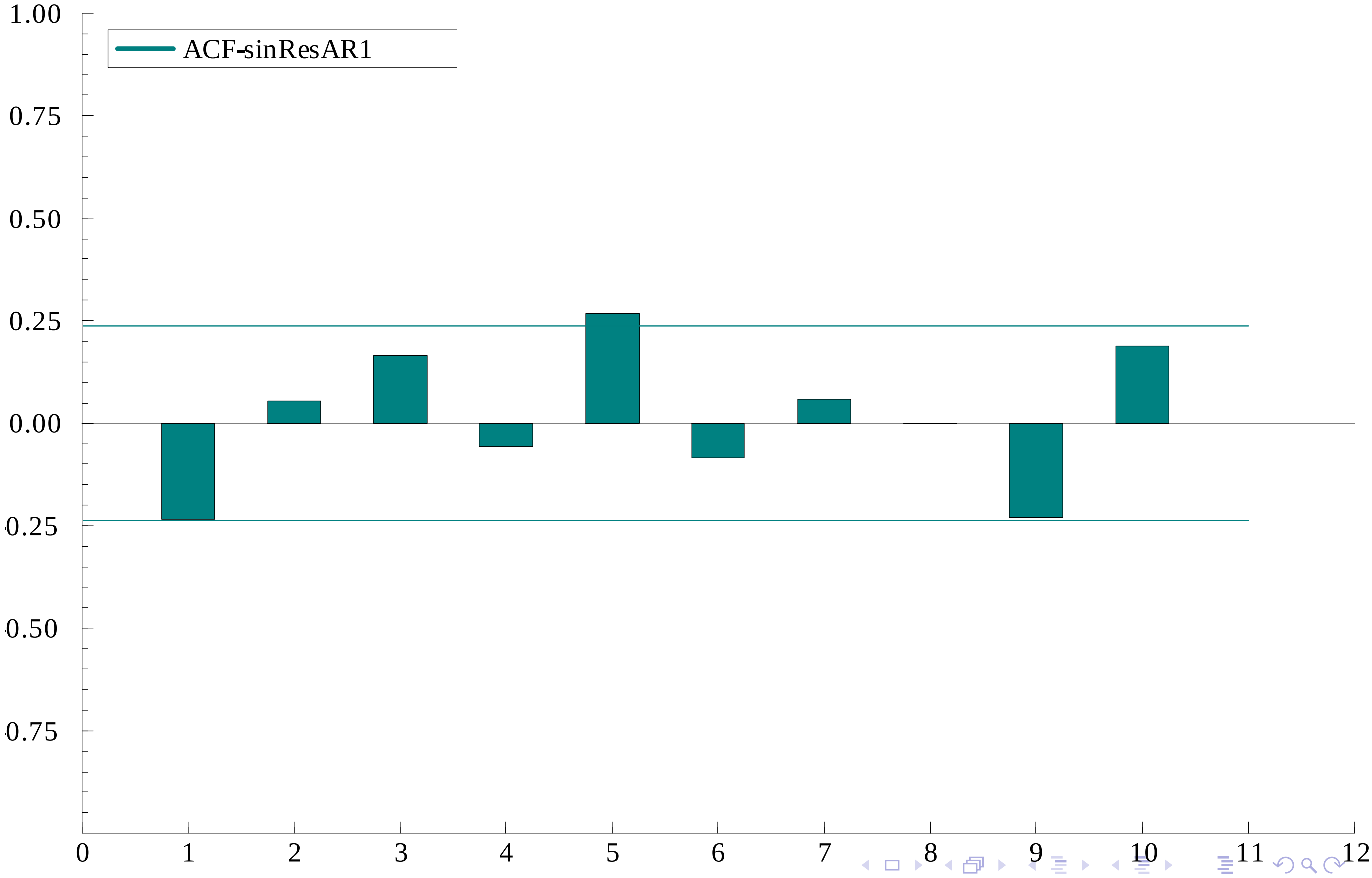


Figure: Circular correlogram (sines) for Black mountain observations.

Black mountain

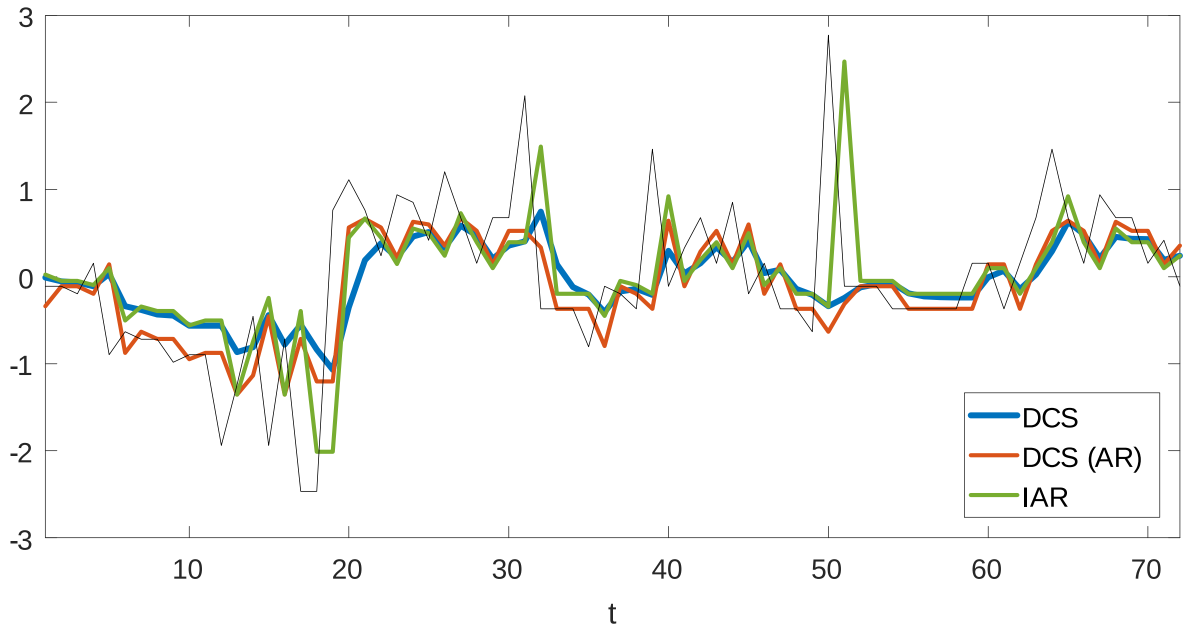
Model \ Parameter	ϕ	κ	υ	D	s^2
Gaussian $AR(1)$	0.46	-	-	0.250	0.575
$AR(1) + N$	0.81	0.194	-	0.232	0.528
$IAR(1)$ *probit*	0.68	-	2.47	0.238	0.544
$IAR(1)$ *tan*	0.55	-	2.23	0.268	0.624
$SCAR(1)$	1.01 (0.14)	-	2.86 (0.39)	0.201	0.449
$DCS(1)$	0.74 (0.16)	0.453 (0.144)	2.62 (0.37)	0.223	0.505

The benchmark given by the circular variance for the random walk, that is first differences, is $D_\Delta = 0.282$.



Black mountain

Figure shows the DCS filters and compares them with the *IAR*(1) filter.
The DCS filter is much less variable.
If the data are not centred by subtracting the directional mean, the IAR filter behaves differently. The DCS remains the essentially the same
***
The ML estimation of μ is, like the directional mean, **equivariant under rotation**. Furthermore the estimates of the other parameters are also unchanged. It doesn't matter whether the data given by a particular cut are over $(-\pi, \pi]$ or $(0, 2\pi]$.
Fitting a *CARMA* or *IAR* model, on the other hand, requires that the data be adjusted so as to be in the range $\overline{y}_d \pm \pi$.



A general class of directional models

Jones and Pewsey (JASA, 2005) give a distribution that includes vM, Cardioid and wrapped Cauchy as special cases. The PDF is

$$f(y) = K[1 + \tanh(\varkappa \upsilon)\cos(y_t - \mu)]^{1/\varkappa}, \quad -\pi < y, \mu \leq \pi,$$

where $\upsilon \geq 0$, $-\infty \leq \varkappa \leq \infty$, and K is a normalizing constant. The value of $\varkappa$ is 1 for the cardioid distribution,, -1 for wrapped Cauchy, and 0 for vM. The cardioid and wrapped Cauchy distributions depend on a scale parameter which can be related to υ. For the cardioid the parameter is $\rho = \tanh(\upsilon)/2$, whereas for wrapped Cauchy it is $\rho = \tanh(\upsilon/2)$.

A general class of directional models

The location score is

$$\frac{1}{\varkappa}\frac{\tanh(\varkappa \upsilon)\sin(y_t - \mu_{t|t-1})}{1+\tanh(\varkappa \upsilon)\cos(y_t - \mu_{t|t-1})}$$

Figure contrasts the score of a vM, with $\upsilon = 1$, with that of a Cardioid and wrapped Cauchy. All three are unchanged if a multiple of 2π is added to or substracted from the angle. The wrapped Cauchy score reaches a turning point for $|y| < \pi/2$ and does so more rapidly as υ increases. The cardioid score, on the other hand, reaches its turning point when $|y| > \pi/2$. The scores become similar as $\upsilon \to 0$.
The use of $u_t = \sin(y_t - \mu_{t|t-1})$ in the dynamic equation lends itself to QML when the true distribution is from the JP class because the JP score is circular and the sine ensures this is the case.

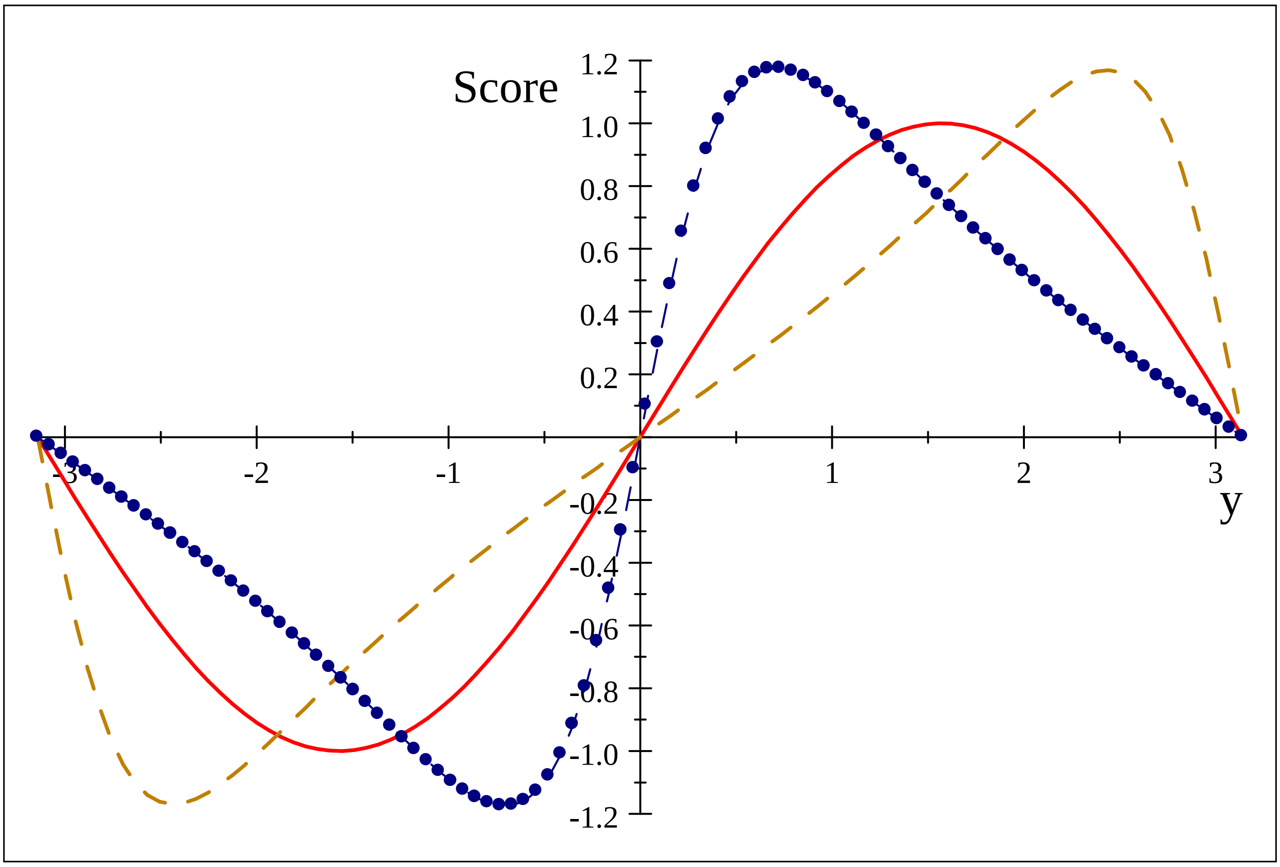


Figure: Sine, Cardioid (- -) and wrapped Cauchy (dots) scores for location with $\upsilon = 1$.

Conclusion and future research

Score-driven approach solves the problem of how to model circular time series.
New AR score model works well for wind direction.
Generalized models
1) Heteroscedasticity - based on cosines
2) JP family of distributions
3) Diurnality and seasonality can be handled.
4) Missing observation eg no wind.