

ORIGINAL ARTICLE

REGIME SWITCHING MODELS FOR CIRCULAR AND LINEAR TIME SERIES

ANDREW HARVEY^a  AND DARIO PALUMBO^{b,c} 

^a*Faculty of Economics, University of Cambridge, Cambridge, UK*

^b*Department of Economics, Ca' Foscari University of Venice, Venice, Italy*

^c*Homerton College, University of Cambridge, Cambridge, UK*

The score-driven approach to time series modelling is able to handle circular data and switching regimes with intra-regime dynamics. Furthermore it enables a dynamic model to be fitted to a linear and a circular variable when their joint distribution is a cylinder. The viability of the new method is illustrated by estimating models for hourly data on wind direction and speed in Galicia, north-west Spain. The modelling of intra-regime dynamics is shown to be of critical importance.

Received 25 January 2022; Revised 22 December 2022; Accepted 22 December 2022

Keywords: circular data; conditional score; cylinder; hidden Markov model; von Mises distribution; wind.

JEL. C22; C32; C46.

MOS subject classification: 62M10.

1. INTRODUCTION

Many areas of environmental statistics involve applications where circular data are collected and statistically analysed. For example, modelling and forecasting wind direction is relevant for tracking pollution and wildfires. Harvey *et al.* (2022) show how a score-driven approach¹ provides a solution to time series modelling of circular data. The present article extends this approach to handle data from time series where there is switching between different regimes and shows how dynamic bivariate models for speed and direction can be constructed.

The viability and effectiveness of the new methods is illustrated with data on wind direction and speed at a wind farm site in Galicia, North-West Spain. The observations were taken every minute over the month of January 2004. These data are used by García-Portugués *et al.* (2013) in a study of pollution. Figure 1 shows the time series of observations, measured in radians from 0 to 2π , obtained by taking the last observation in each hour. Zero radians corresponds to due east with the subsequent coding being anti-clockwise so $\pi/2$ is north, π is west and $3\pi/2$ is south. Because of circularity some of the measurements in the North East (NE) orbit appear at the top, close to 2π , rather than near the bottom. Serious distortions can arise if circularity is not taken into account and standard linear time series methods are used.

García-Portugués *et al.* (2013) note that the prevailing wind comes from two directions: SW and NE. The two dominant directions are apparent in Figure 1 with NE being a little less than one radian and SW around four. Hence a regime switching model may be appropriate. Holtzmann *et al.* (2006) propose a switching regime model for wind

* Correspondence to: Department of Economics, Ca' Foscari University of Venice, Fondamenta San Giobbe, Cannaregio 873, Venice 30121, Italy.
Email: dario.palumbo@unive.it

¹ Score-driven time series models were developed by Harvey (2013) and Creal *et al.* (2013), where they were called DCS and GAS models respectively.

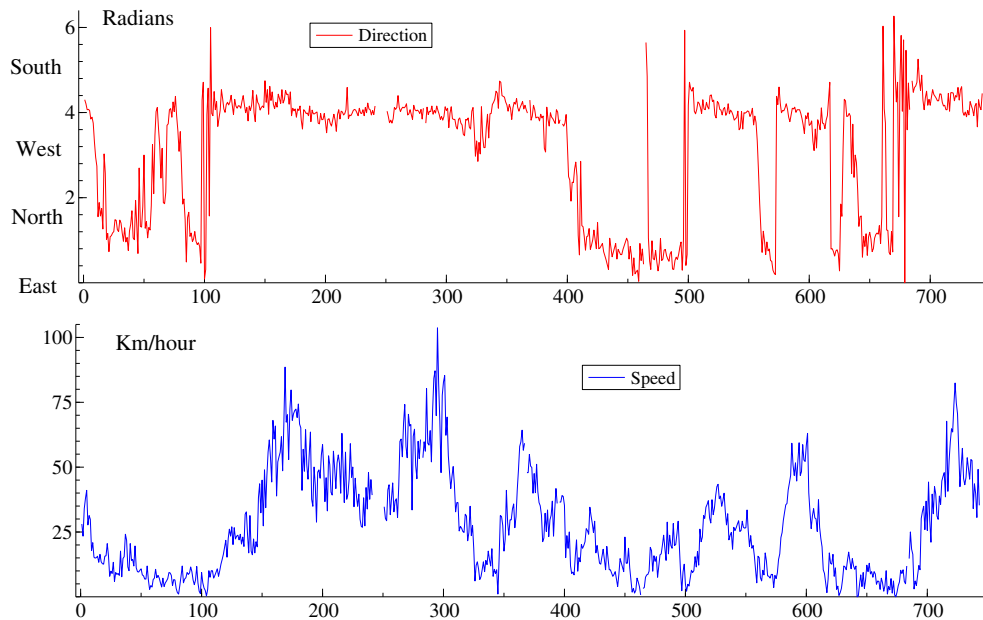


Figure 1. Hourly wind direction and speed in January 2004 at a site in Galicia

direction and the same approach is used by Zucchini *et al.* (2016, pp. 228–35), for modeling the change in direction of flight for fruit flies. The basic formulation assumes a finite number of regimes and introduces dynamics by a Markov chain in which there is a fixed probability of staying in the current regime or moving to another; see Hamilton (1989, 1994). The regime is not observed: hence the term hidden Markov model (HMM) as in Zucchini *et al.* (2016). The probability of being in a particular regime is given by a filter that depends on past observations. These probabilities yield a conditional mixture distribution for the current observation and hence a likelihood function. Catania (2021) proposes a different approach that by-passes the hidden Markov chain and instead sets up filters for the regime probabilities in the conditional distribution directly by using their scores. He calls these dynamic adaptive mixture models (DAMMs). The score-driven approach leads naturally to a solution of how to model intra-regime dynamics. The key point is that the forcing variables in dynamic equations are a function of past observations weighted according to the probability that they are in a given regime. Intra-day dynamics have not been modelled in this way in standard HMMs. Yet in many environmental applications there are marked intra-regime dynamics. The evidence presented here shows that neglecting these dynamics can be a major drawback.

Figure 1 also shows wind speed. Wind speed is a (non-negative) linear variable and a joint model distribution of a linear and a circular variable takes the form of a cylinder. Abe and Ley (2017) proposes a general distribution for cylindrical data. A key feature of their distribution is that the circular concentration is allowed to increase with the linear component, a phenomenon first identified in the seminal article by Fisher and Lee (1992, p. 666). The challenge is to move from the static to the dynamic case. We show here that a score-driven approach facilitates the construction of a model that allows the location of the circular variable and the level (scale) of the linear variable to change over time. The concentration may also change.

The bivariate score-driven model can be incorporated in a regime switching model. Switching bivariate models have been used by Lagona *et al.* (2015) for modeling the joint distribution of wave height and direction in the Adriatic; they employ an HMM for fitting a cylindrical distribution when there are three distinct regimes. Other applications involving speed and direction include tracking and forecasting the movements of animals, boats and wildfires, as in García-Portugués *et al.* (2014).

Section 2 reviews the score-driven model for circular data proposed by Harvey *et al.* (2022) and fits it to wind direction in Galicia. The theory underlying regime switching models is described in section 3 and it is shown how

the score-driven approach set out in Section 2 is able to model dynamic location and concentration within regimes. The methods are applied to the Galicia data on wind direction in Section 4. Dynamic cylindrical distributions are described in Section 5 and extended to handle switching regimes. These time series models are fitted to the Galicia data on direction and speed in Section 6. Section 7 concludes.

2. SCORE-DRIVEN MODELS FOR CIRCULAR DATA

Circular observations measured in radians are usually taken to have a von Mises (vM) distribution with probability density function (PDF)

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

where $I_k(v)$ denotes a modified Bessel function of order k , μ is mean direction and v is a non-negative concentration parameter that is inversely related to dispersion. When $v = 0$ the distribution is uniform whereas y is approximately $N(\mu, 1/v)$ for large v . The maximum likelihood (ML) estimator of location, μ , is the sample mean direction, \bar{y}_d ; see Mardia and Jupp (2000). A class of general circular distributions, in which the cardioid and wrapped Cauchy are special cases, is described in Jones and Pewsey (2005).

Circular time series models currently in use are almost all based on the autoregressive-moving average (ARMA) type models proposed in Fisher and Lee (1994) or are regime switching models; see Pewsey and García-Portugués (2021, section 10.1). Score-driven models offer considerable advantages and allow regime switching to be combined with intra-regime dynamics.

2.1. Dynamic Location

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

$$y_t = z_t \bmod(2\pi) - \pi, \quad t = 1, \dots, T; \quad (2)$$

see Breckling (1989) and Fisher and Lee (1994). The score-driven model for circular data is

$$z_t = \mu_{t|t-1} + \varepsilon_t, \quad -\pi \leq \varepsilon_t < \pi, \quad t = 1, \dots, T, \quad (3)$$

where the ε_t 's are independent and identically distributed (IID) random variables from a circular distribution with location zero and $\mu_{t|t-1}$ is a filter for location at time t based on information available at time $t - 1$. The basic stationary first-order filter is

$$\mu_{t+1|t} = (1 - \phi)\omega + \phi\mu_{t|t-1} + \kappa u_t, \quad |\phi| < 1, \quad t = 1, \dots, T, \quad (4)$$

where $\mu_{1|0} = \omega$ is the unconditional location of $\mu_{t|t-1}$ and the forcing variable, u_t , is defined as being (proportional to) the conditional score for location, that is $\partial \ln f(y_t; \mu_{t|t-1}, v) / \partial \mu_{t|t-1}$, $t = 1, \dots, T$.

A defining property of a (continuous) circular distribution is that it satisfies the periodicity condition $f(z \pm 2\pi k; \boldsymbol{\psi}) = f(z; \boldsymbol{\psi})$, where k is an integer and $\boldsymbol{\psi}$ denotes parameters. Provided the derivatives of the log-density with respect to the elements of $\boldsymbol{\psi}$ are continuous, they too are circular in that the periodicity condition is satisfied. Thus the path of $\mu_{t|t-1}$ is the same irrespective of whether u_t is defined in terms of z_t or y_t . The conditional distribution of y_t in a model defined by (3), (4) and (2) is therefore the same as that of z_t and so the likelihood function of the wrapped observations, the y_t 's, is the same as that of the unobserved variables, the z_t 's.

In the case of the von Mises distribution, that is $\varepsilon_t \sim vM(0, v)$ in (3), the score is

$$u_t = v \sin(z_t - \mu_{t|t-1}) = v \sin(y_t - \mu_{t|t-1}), \quad u_t \sim IID(0, A(v)/v). \quad (5)$$

The general continuous circular distribution of Jones and Pewsey (2005) has a score that, like (5), is invariant to wrapping as well as being IID.

Harvey *et al.* (2022) provide further details and derive the asymptotic distribution of the ML estimators of ϕ , κ and ω in (4) for the stationary vM model.

2.2. Tests

The Lagrange multiplier, or score, 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, section 2.5). For a vM distribution with $v > 0$, the scores under the null hypothesis of constant location are proportional to $\sin(y_t - \bar{y}_d)$. Hence the sample autocorrelations correspond to the circular autocorrelations (CACFs) in Jammalamadaka and SenGupta (2001, pp. 176–9).

2.3. Heteroscedasticity

Score-driven models can be extended to allow for dynamic heteroscedasticity by setting up a filter for the conditional concentration. Thus ε_t in (3) is distributed as $vM(0, v_{t|t-1})$ with the dynamics dependent on the score with respect to $v_{t|t-1}$, that is

$$u_t^v = \cos(y_t - \mu_{t|t-1}) - A(v_{t|t-1}).$$

The scores are a martingale difference sequence with mean zero and variance $1 - A(v_{t|t-1})^2 - A(v_{t|t-1})/v_{t|t-1}$. An exponential link function can be used to ensure the concentration remains positive. Thus $v_{t|t-1} = \exp(\zeta_{t|t-1})$. The first-order dynamic model for $\zeta_{t|t-1}$ is then

$$\zeta_{t+1|t} = (1 - \phi_\zeta)\omega_\zeta + \phi\zeta_{t|t-1} + \kappa u_t^\zeta, \quad t = 1, \dots, T, \quad (6)$$

where $u_t^\zeta = v_{t|t-1} u_t^v$ and $\zeta_{1|0} = \omega_\zeta$. The likelihood function is

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

where $\boldsymbol{\psi}$ denotes the parameters in the dynamic equations for $v_{t|t-1}$ as well as $\mu_{t|t-1}$. The forcing variable for location is now $u_t = v_{t|t-1} \sin(y_t - \mu_{t|t-1})$.

2.4. Application to Circular Data from Galicia

When a basic first-order dynamic model, (4) is fitted to the Galicia data,² the result is $\tilde{\phi} = 1.0$, $\tilde{\kappa} = 0.19$ and $\tilde{v} = \exp(\tilde{\omega}_\zeta) = 4.78$. The maximized log-likelihood is $\ln L = -510.62$. The residual CACF shows there is considerable serial correlation remaining and the fit, as measured by dispersion (circular variance), is no better than that of a random walk in which $\mu_{t|t-1} = y_{t-1}$; see Mardia and Jupp (2000, pp. 18–19 and 30). Adding heteroscedasticity gives a better fit, but although this reduces the serial correlation in dispersion, the serial correlation in location increases.

² The fact that the range of the observations is $[0, 2\pi)$ rather than $[-\pi, \pi)$ makes no substantive difference to the results.

3. SWITCHING REGIMES

The plot of wind direction in Galicia provides a clear motivation for a switching model. The first subsection below gives a short description of the static mixture model before moving on to review the dynamic adaptive mixture model (DAMM) of Catania (2021). The third subsection then shows how the score-driven approach can handle intra-regime dynamics and in the subsection following the associated diagnostic tests are set out. The last subsection discusses circular observations.

3.1. Static Mixture Model

The PDF of a mixture of K distributions is

$$f(y_t; \boldsymbol{\psi}) = \sum_{i=1}^K \xi_i f_i(y_t; \boldsymbol{\psi}_i), \quad \sum_{i=1}^K \xi_i = 1, \quad t = 1, \dots, T,$$

where $0 \leq \xi_i \leq 1$, $i = 1, 2, \dots, K$, with ξ_i , denoting the probability of being in i th regime; the parameters in the i th regime are contained in the vector $\boldsymbol{\psi}_i$ and $\boldsymbol{\psi}$ includes all these parameters, together with ξ_i , $i = 1, 2, \dots, K-1$. When the observations are independent, the probability of being in regime i given observation y_t is

$$\xi_i(y_t) = \xi_i f_i(y_t; \boldsymbol{\psi}_i) / f(y_t; \boldsymbol{\psi}), \quad i = 1, 2, \dots, K. \quad (7)$$

The maximum likelihood (ML) estimates for the parameters in each of the $\boldsymbol{\psi}_i$'s, together with the unconditional probabilities, ξ_i , $i = 1, \dots, K$, can be computed iteratively by what turns out to be a special case of the EM algorithm; see Hamilton (1994, pp. 688–9), or Zucchini *et al.* (2016).

3.2. Dynamic Mixture Model

In the dynamic mixture model, the PDF of y_t , conditional on information at time $t-1$, is

$$f_{t|t-1}(y_t; \boldsymbol{\psi}) = \sum_{i=1}^K \xi_{i,t|t-1} f_{i,t|t-1}(y_t; \boldsymbol{\psi}_i), \quad t = 1, \dots, T, \quad (8)$$

where $0 \leq \xi_{i,t|t-1} \leq 1$, $i = 1, 2, \dots, K$, and $\sum_{i=1}^K \xi_{i,t|t-1} = 1$. The probabilities, $\xi_{i,t|t-1}$, are given by filters constructed from past observations. There are two ways in which this may be done. The first is implicitly, with the filters derived as a consequence of setting up an HMM for the unobserved probabilities, $\xi_{i,t}$; see Hamilton (1994, pp. 690–3). The second is by explicitly writing down a filter as in DAMM. Here we concentrate on the DAMM option because it is relatively simple and transparent. It also embodies a score-driven approach and as such it leads naturally to the formulation of dynamic nonlinear models for time-varying parameters within each regime.

The DAMM can be described most easily by restricting attention to the case where there are only two regimes. For simplicity of notation, the parameter vector $\boldsymbol{\psi}$ (which now includes parameters associated with the switching filters) will be dropped from $f_{t|t-1}(y_t; \boldsymbol{\psi})$ and similarly for the regime conditional distributions, $f_{i,t|t-1}(y_t; \boldsymbol{\psi}_i)$, $i = 1, \dots, K$. The conditional distribution can now be written

$$f_{t|t-1}(y_t) = \xi_{1,t|t-1} f_{1,t|t-1}(y_t) + (1 - \xi_{1,t|t-1}) f_{2,t|t-1}(y_t), \quad t = 1, \dots, T, \quad (9)$$

where $\xi_{1,t|t-1} = \xi_{1,t|t-1}$ is the probability of being in the first regime, implying that $1 - \xi_{1,t|t-1} = \xi_{2,t|t-1}$ is the probability of being in the second. The probability, $\xi_{1,t|t-1}$, can be confined to the range $0 < \xi_{1,t|t-1} < 1$ by a logistic link function,

$\xi_{t|t-1} = \exp \gamma_{t|t-1} / (1 + \exp \gamma_{t|t-1})$, $-\infty < \gamma_{t|t-1} < \infty$. The dynamics of $\gamma_{t|t-1}$ are then driven by the score with respect to $\gamma_{t|t-1}$ which is

$$w_t = \frac{\partial \ln f_{t|t-1}}{\partial \gamma_{t|t-1}} = \frac{\partial \ln f_{t|t-1}}{\partial \xi_{t|t-1}} \frac{\partial \xi_{t|t-1}}{\partial \gamma_{t|t-1}} = \frac{f_{1,t|t-1} - f_{2,t|t-1}}{f_{t|t-1}} \frac{\exp \gamma_{t|t-1}}{(1 + \exp \gamma_{t|t-1})^2},$$

or equivalently

$$w_t = \frac{f_{1,t|t-1} - f_{2,t|t-1}}{f_{t|t-1}} \xi_{t|t-1} (1 - \xi_{t|t-1}). \quad (10)$$

The basic first-order dynamic equation is

$$\gamma_{t+1|t} = (1 - \phi_\gamma) \omega_\gamma + \phi_\gamma \gamma_{t|t-1} + \kappa_\gamma w_t, \quad t = 1, \dots, T, \quad (11)$$

where $\gamma_{1|0}$ is set to the unconditional mean, ω_γ , and the condition $|\phi_\gamma| < 1$ is all that is required to ensure that $\gamma_{t+1|t}$, and hence, $\xi_{t+1|t}$, is stationary. No restrictions are imposed on κ_γ . As $\kappa_\gamma \rightarrow \infty$ there will be an abrupt change in regime when w_t changes sign, whereas when κ_γ is close to zero any change will be gradual.

3.3. Dynamics Within Regimes

When the location parameter within the i th regime is time-varying, its dynamics can be captured by a filter that depends on the score. The PDFs in (8) now depend on $\mu_{i,t|t-1}$ so $f_{i,t|t-1}(y_t; \boldsymbol{\psi}_i)$ becomes $f_{i,t|t-1}(y_t; \mu_{i,t|t-1}, \boldsymbol{\psi}_i)$ and $\boldsymbol{\psi}_i$ is redefined accordingly. Differentiating the logarithm of (8) gives

$$\begin{aligned} \frac{\partial \ln f_{t|t-1}}{\partial \mu_{i,t|t-1}} &= \frac{\partial \ln f_{t|t-1}}{\partial f_{t|t-1}} \frac{\partial f_{t|t-1}}{\partial f_{i,t|t-1}} \frac{\partial f_{i,t|t-1}}{\partial \ln f_{i,t|t-1}} \frac{\partial \ln f_{i,t|t-1}}{\partial \mu_{i,t|t-1}} \\ &= \xi_{i,t|t} \frac{\partial \ln f_{i,t|t-1}}{\partial \mu_{i,t|t-1}} = u_{i,t}, \quad i = 1, \dots, K, \end{aligned} \quad (12)$$

where, following on from (7),

$$\xi_{i,t|t} = \xi_{i,t|t-1} f_{i,t|t-1} / f_{t|t-1}, \quad i = 1, 2, \dots, K, \quad (13)$$

is the estimated probability of being in the i th regime given y_t and $\xi_{i,t|t-1}$. The smaller is $\xi_{i,t|t}$, the more the contribution of y_t to the corresponding score is downweighted. These scores drive dynamic filters such as

$$\mu_{i,t+1|t} = (1 - \phi_i) \omega_i + \phi_i \mu_{i,t|t-1} + \kappa_i u_{i,t}, \quad |\phi_i| < 1, \quad i = 1, \dots, K, \quad (14)$$

with $\mu_{i,1|0} = \omega_i$, $i = 1, \dots, K$.

ML estimates are obtained by maximizing the log-likelihood function

$$\ln L(\boldsymbol{\psi}) = \sum_{t=1}^T \ln f_{t|t-1} = \sum_{t=1}^T \ln \left[\sum_{i=1}^K \xi_{i,t|t-1} f_{i,t|t-1} \right], \quad (15)$$

where $\boldsymbol{\psi}$ now includes $\omega_i, \phi_i, \kappa_i$, $i = 1, \dots, K$, as well as the corresponding parameters for any dynamic dispersion models, the parameters in (11) and any constant shape parameters.

Remark 1. In the classic Markov switching model, dynamics are introduced into the location and/or scale of each regime by letting them depend directly on past observations. For example the conditional mean in the i th regime is often given by $\mu_i + \phi_i(y_{t-1} - \mu_i)$, $t = 1, \dots, T$, $i = 1, 2$; see Hamilton (1994, p. 691). By contrast, the score-driven approach leads to a filters driven by a forcing variable that is weighted by the probability of being in the i th regime. This defining feature of the score-driven model also distinguishes it from the treatment of switching conditional heteroscedasticity in the financial literature; see, for example, Haas *et al.* (2004) and the discussion in Catania (2021, sections 2 and 3).

3.4. Model Selection and Diagnostics

Following on from the work of Hamilton (1996), Smith (2008) finds LM tests to have the best size and power properties for Markov switching models. The structure of the DAMM is such that LM diagnostic tests are easily formulated. When a static mixture model has been fitted, evidence for serial correlation in the regime dynamics and the intra-regime dynamics is separated out. Formal tests against dynamics can be constructed and the pattern of serial correlation displayed in correlograms. As shown by the application in Section 4.1, this can be of great benefit for model specification. When dynamics have been estimated, diagnostics designed to assess the possibility of omitted dynamic effects can be formulated using the same principles.

Under the null hypothesis that the model is a static mixture with no dynamics, LM tests against dynamics in regime switching and in the parameters within each regime may be constructed. In a two state model, the test is against dynamic switching of the form $\gamma_{t|t-1} = \omega_\gamma + \kappa_1 w_{t-1} + \dots + \kappa_p w_{t-p}$, where w_t is as in (10). When the model is static $w_t = \xi(y_t) - \xi$, $t = 1, \dots, T$, because, from (7), $f_1(y_t) = \xi(y_t)f(y_t)/\xi$, where, as before, the subscript is dropped from ξ_1 . Thus, following Harvey (2013, section 2.5), the LM test of the null hypothesis that the model is static, that is $H_0 : \kappa_1 = \dots = \kappa_p = 0$ against $H_1 : \kappa_i \neq 0$ for some $i = 1, \dots, P$, is equivalent to a portmanteau Q -test based on the correlogram of the estimated probabilities, $\xi(y_t)$, $t = 1, \dots, T$. The critical values are taken from a χ_p^2 distribution.

An LM test against level dynamics in the i th regime can be based on the scores in expression (12) with $\xi_{i,t|t}$ and $\mu_{i,t|t-1}$ fixed under the null hypothesis at $\xi_i(y_t)$ and μ_i respectively. When the test is against dynamics in the i th location only, and μ_j , $j \neq i$ is fixed, the LM statistic is equivalent to the Q -statistic formed from sample autocorrelations, $r_i(\tau) = c_i(\tau)/c_i(0)$, where $c_i(\tau) = \sum_{t=\tau+1}^T u_{i,t}u_{i,t-\tau}/T$, $i = 1, \dots, K$, $\tau = 1, \dots, P$, that is

$$Q_i(P) = T \sum_{\tau=1}^P r_i^2(\tau), \quad i = 1, 2, \dots \quad (16)$$

As noted by Harvey and Thiele (2016, pp. 578–9), estimating fixed parameters makes no difference to the distribution of $Q_i(P)$, which is asymptotically distributed as χ_p^2 under the null hypothesis.

When dynamic models have been fitted to the regime probabilities, as in the basic DAMM, LM test statistics for location dynamics in individual regimes can be constructed. Similarly, when dynamics have been estimated within regimes, LM tests for omitted dynamics can be set up. Tests for heteroscedasticity can be similarly formed; see, for example, Calvori *et al.* (2017). However, if the effect of fitting dynamics to the ξ'_i s and to location and/or scale, is ignored, simple $Q_i(P)$ tests may be used to give an indication of serial correlation in each regime. Harvey and Thiele (2016) show that this can often be a good strategy and it is the one we adopt here.

3.5. Circular Mixture Models

Static and dynamic circular mixture models can be estimated as outlined in Sections 3.1–3.3. The variable w_t driving the switching equation, (11), is obviously circular so invariance to wrapping is retained. The scores driving the intra-regime dynamics, the u'_i s of (12), are as in (5). The filtered location is given by the mean direction of the

filtered locations in the individual regimes, computed as

$$\mu_{i,t-1} = \text{atan2} \left[\sum_{i=1}^K \xi_{i,t|t-1} \sin \mu_{i,t|t-1}, \sum_{i=1}^K \xi_{i,t|t-1} \cos \mu_{i,t|t-1} \right], \quad i = 1, 2, \dots, \quad (17)$$

rather than by a simple average, $\sum_{i=1}^K \xi_{i,t|t-1} \mu_{i,t|t-1}$, $t = 1, \dots, T$. Multi-step forecasts can be computed from recursions for the $\gamma'_{i,t|t-1}$ s and $\mu'_{i,t|t-1}$ s and the distributions of future observations can be simulated.

A portmanteau test, (16), against dynamics in the level of the i th regime is based on the (circular) correlogram of the residuals $\xi_i(y_t) \sin(y_t - \tilde{\mu}_i)$, $i = 1, \dots, K$. Note that although the score is $v_i \xi_i(y_t) \sin(y_t - \tilde{\mu}_i)$, the concentration parameter, v_i , cancels out.

4. SWITCHING MODELS FOR GALICIA

Tables I and II show ML estimates for switching models applied to the Galicia data. The numerical maximization was carried out in Matlab using an interior-point method that follows a barrier approach to solve the subproblems occurring in each Newton–Raphson iteration; see, for example, Waltz *et al.* (2006). Estimates of the asymptotic standard errors, obtained from the numerical Hessian, are shown in brackets. The parameter ω_ζ is the logarithm of the concentration, v . When heteroscedastic models are fitted, it is ω_ζ in (6).

Table I. Goodness of fit of the two regime vM model with the estimated dynamic parameters of the switching probability from fitting: (1) the static mixture model, (2) the pure DAMM model, (3) the DAMM model with intra regime dynamic location and (4) the DAMM model with intra regime dynamic location and concentration

Model	AIC	BIC	Logl	ω_γ	ϕ_γ	κ_γ	ξ
(1)	1,643.580	1,666.640	−816.790	0.680 (0.087)	–	–	0.664
(2)	980.255	1,012.539	−483.127	2.194 (0.808)	0.959 (0.012)	4.766 (0.631)	0.900
(3)	490.554	541.286	−234.277	1.192 (0.838)	0.959 (0.013)	5.774 (0.690)	0.767
(4)	459.400	528.581	−214.700	1.724 (0.840)	0.944 (0.021)	5.360 (0.895)	0.849

Note: The switching parameters are shown in the last four columns, with ξ denoting the logistic transformation of ω_γ .

Table II. Estimated dynamic parameters for the two regime vM model from fitting: (1) the static mixture model, (2) the pure DAMM model, (3) the DAMM model with intra regime dynamic location and (4) the DAMM model with intra regime dynamic location and concentration

Model	$\omega_{\mu 1}$	$\phi_{\mu 1}$	$\kappa_{\mu 1}$	$\omega_{\zeta 1}$	$\phi_{\zeta 1}$	$\kappa_{\zeta 1}$	$\omega_{\mu 2}$	$\phi_{\mu 2}$	$\kappa_{\mu 2}$	$\omega_{\zeta 2}$	$\phi_{\zeta 2}$	$\kappa_{\zeta 2}$
(1)	4.051 (0.014)	–	–	2.540 (0.101)	–	–	1.055 (0.056)	–	–	0.735 (0.130)	–	–
(2)	4.041 (0.014)	–	–	2.437 (0.066)	–	–	1.032 (0.049)	–	–	0.848 (0.090)	–	–
(3)	4.136 (0.040)	0.914 (0.026)	0.013 (0.002)	3.318 (0.067)	–	–	3.165 (0.403)	0.989 (0.005)	0.257 (0.026)	1.243 (0.073)	–	–
(4)	4.107 (0.038)	0.917 (0.023)	0.015 (0.002)	3.060 (0.104)	0.856 (0.069)	0.255 (0.068)	2.305 (0.156)	0.983 (0.004)	0.173 (0.023)	1.426 (0.125)	0.309 (0.068)	0.774 (0.143)

Table III. Box–Ljung test with $P = 5$, $Q(5)$, for residual correlation on the fitted scores from fitting: (1) the static mixture model, (2) the pure DAMM model, (3) the DAMM model with intra regime dynamic location and (4) the DAMM model with intra regime dynamic location and concentration

Model	μ_1	ζ_1	μ_2	ζ_2	γ
(1)	726.152***	184.370***	794.447***	371.265***	2392.15***
(2)	852.972***	302.547***	807.591***	226.344***	13.588**
(3)	4.176	36.536***	10.931*	71.257***	43.851***
(4)	7.776	19.369***	27.266***	25.941***	26.291***

Note: (Nominal) significance levels: 0.01 ‘***’, 0.05 ‘**’, 0.1 ‘*’.

Diagnostic test statistics for assessing residual serial correlation in different components are in Table III. We originally calculated Q -statistics for $P = 1, 5$ and 20 but the message in all three is essentially the same, so only $P = 5$ is given here. When dynamics are fitted, the tests should not be treated formally because, as indicated earlier, the distribution is no longer χ_p^2 under the null hypothesis. Furthermore, because the sample size is large, with $T = 744$, Q -statistics based on relatively small sample autocorrelations can still be big. Their main value is to convey a strong message about which models are most effective.

4.1. Static Mixture

Although the static mixture model can be ruled out from the correlogram of the raw data, it is nevertheless informative about the regimes. Numerically optimizing the log-likelihood function gave the values in the first row of Table II, that is $\tilde{\mu}_1 = 4.051$ (0.014), $\tilde{\mu}_2 = 1.055$ (0.056), $\tilde{v}_1 = 12.68$, $\tilde{v}_2 = 2.08$ and $\tilde{\xi} = 0.66$; note that $v = \ln \omega_\zeta$. The log-likelihood, in Table I, is -816.8 , which is far lower than we obtained for the single regime dynamic model reported in Section 2.4, and the Q -statistics are huge.

The plot of $\xi(y_t)$ in Figure 2 shows how the contrast between the distributions in the two regimes gives a clear indication of which regime is operative at any one time. The regimes are obviously not determined randomly and the ACF of the $\xi_t(y_t)$'s indicates that a fairly persistent first-order filter, as in (11), is likely to give a good fit. The circular ACFs (CACFs) for the individual regimes in the lower panels indicate persistent dynamics in the location. The correlations between the three scores are not far from zero.

4.2. Dynamic Mixtures

Although the tests indicate dynamics within each regime, it is useful to begin by fitting a pure DAMM, that is one without intra-regime dynamics but with a dynamic equation for $\gamma_{t|t-1}$ as in (11). The estimates of ω_{μ_1} , ω_{μ_2} , ω_{ζ_1} and ω_{ζ_2} are similar to the estimates found for the static mixture model except that the estimate of ξ is, at 0.90, somewhat higher than the static estimate of 0.66. The log-likelihood is -483.1 so there is a clear improvement over the static mixture model.

The diagnostic test based on the switching residuals, shown in the column headed γ in Table III, indicates that most of the dynamic movements have been captured by the regime-switching equation. However, the Q -statistics for location dynamics remain very high in both regimes; indeed the correlograms are not dissimilar to those in Figure 2. Both are consistent with a first-order dynamic model, (14), for each regime.

Fitting an HMM gives a result very close to that of the DAMM. The log-likelihood is slightly lower, at -490.5 , but the AIC and BIC are slightly bit smaller, reflecting the fact that the HMM has one fewer parameter. The finding that the pure switching model is inadequate is important because most, if not all, of the research in this area has been restricted to pure Markov switching models. Indeed only three pages are devoted to intra-regime dynamics in the book by Zucchini *et al.* (2016, pp. 150–2).

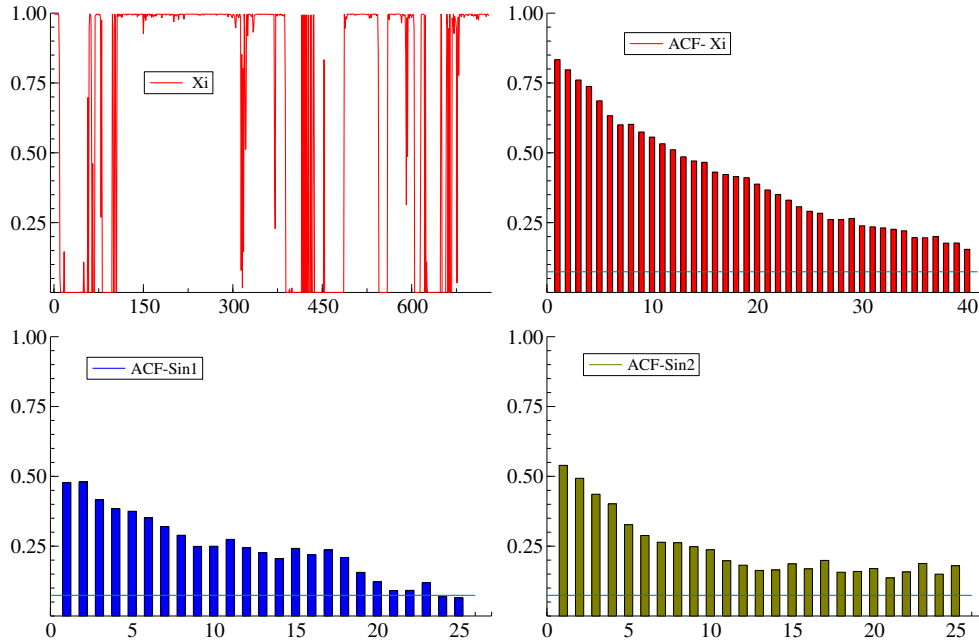


Figure 2. Regime probabilities and ACFs for a static mixture model of wind direction in Galicia

4.3. Dynamic Mixture Model with Intra-regime Dynamics

The first-order score-driven models for the location in each regime are as in (14) with

$$u_{i,t} = \xi_{i,t} v_{i,t|t-1} \sin(y_t - \mu_{i,t|t-1}), \quad i = 1, 2,$$

where $v_{i,t|t-1}$ is the filtered concentration when the model takes account of heteroscedasticity. The results, for the models labelled (3) and (4), show the dynamics to be fairly persistent in both regimes, with ϕ_1 and ϕ_2 of (14) always above 0.9. The mean of location in the second regime has increased to 3.165 but if it is constrained to its value for the static model, that is 1.05, the log-likelihood is much lower at -279.3 as opposed to -234.3 for the unconstrained model. It seems that the parameters in the second regime are more difficult to estimate than those in the first, perhaps because the concentration is lower. Nevertheless there is a huge increase in the likelihood as compared with the pure DAMM.

The diagnostics show that serial correlation in location has been eliminated. However, the scores for concentration indicate dynamics. When, in the last line, the model is extended to allow for heteroscedasticity, there is a further improvement in goodness of fit and the level in regime 2 falls to 2.31. On the other hand the underlying probability of being in regime 1, that is ξ , rises from 0.767 to 0.849.

Finally the likelihoods for the two switching models with intra-regime dynamics are, as expected, much bigger than those of the corresponding single regime models and, despite the extra parameters, the AICs and BICs are much smaller.

5. MODELING THE CYLINDER

A bivariate distribution for a circular and a linear variable takes the form of a cylinder. This section shows how a dynamic model can be constructed. The last subsection makes the extension to a bivariate regime switching model.

5.1. Weibull–von Mises (Abe-Ley) Distribution

The Weibull–Sine Skewed–von Mises (WeiSSVM) proposed³ by Abe and Ley (2017) combines a von Mises circular distribution for a circular variable, y , with a Weibull distribution for a non-negative linear variable, x . The skewing term will be dropped here to simplify the exposition, giving the Weibull–von Mises (WeiVM) distribution

$$f(y, x) = \frac{\alpha \exp(-\alpha \lambda)}{2\pi \cosh v} x^{\alpha-1} \exp\{-(x/\exp \lambda)^\alpha (1 - \tanh v \cdot \cos(y - \mu))\},$$

$$-\pi \leq y, \mu < \pi, \quad x \geq 0, \quad \alpha > 0, v \geq 0,$$

where $\exp(\lambda)$ is the scale, φ , for the linear variable and v is a parameter that determines concentration for y .

5.2. Dynamic Model

In the dynamic score-driven cylinder model the parameters μ and λ change over time and the logarithm of the joint density of the conditional WeiVM distribution is

$$\ln f(y_t, x_t; \mu_{t|t-1}, \lambda_{t|t-1}, \boldsymbol{\psi}) = \ln(\alpha/2\pi) - \ln \cosh v - \alpha \lambda_{t|t-1} + (\alpha - 1) \ln x_t$$

$$- (x_t / \exp \lambda_{t|t-1})^\alpha (1 - \tanh v \cos(y_t - \mu_{t|t-1})), \quad t = 1, \dots, T, \quad (18)$$

with $-\pi \leq y_t < \pi$, but with no corresponding restriction on $\mu_{t|t-1}$; $\boldsymbol{\psi}$ denotes the parameters v and α and those in the dynamic equations. The conditional scores are

$$\frac{\partial \ln f_{t|t-1}}{\partial \mu_{t|t-1}} = u_t^\mu = \tanh(v)(x_t / \exp \lambda_{t|t-1})^\alpha \sin(y_t - \mu_{t|t-1}) \quad (19)$$

and

$$\frac{\partial \ln f_{t|t-1}}{\partial \lambda_{t|t-1}} = u_t^\lambda = \alpha(x_t / \exp \lambda_{t|t-1})^\alpha (1 - \tanh(v) \cos(y_t - \mu_{t|t-1})) - \alpha. \quad (20)$$

The filters for $\mu_{t|t-1}$ and $\lambda_{t|t-1}$ are driven by u_t^μ and u_t^λ and so for first-order dynamics

$$\mu_{t+1|t} = (1 - \phi_\mu)\omega_\mu + \phi_\mu \mu_{t|t-1} + \kappa_\mu u_t^\mu,$$

$$\lambda_{t+1|t} = (1 - \phi_\lambda)\omega_\lambda + \phi_\lambda \lambda_{t|t-1} + \kappa_\lambda u_t^\lambda. \quad (21)$$

with $\mu_{1|0} = \omega_\mu$ and $\lambda_{1|0} = \omega_\lambda$.

Both u_t^μ and u_t^λ retain the univariate circularity property of being unchanged when multiples of 2π are added or subtracted from y_t . The circularity of the scores confirms that when a dynamic score-driven model for a WeiVM distribution, $f(z_t, x_t)$, allows z_t to range over the whole real line, it may be wrapped, as in (2), to give a likelihood function, based on (18), that is the same as that of the (infeasible) likelihood function for $f(z_t, x_t)$.

Basing the dynamics for $\mu_{t|t-1}$ and $\lambda_{t|t-1}$ on scores means that their movements interact with each other in a way that makes sense given the structure of the WeiVM bivariate distribution. It follows from Abe and Ley (2017) that the distribution of y_t conditional on x_t , together with all the information at time $t - 1$, is vM with mean $\mu_{t|t-1}$ and concentration

$$v(x_t) = (\tanh v) \cdot (x_t / \exp \lambda_{t|t-1})^\alpha, \quad (22)$$

³ Abe and Ley (2017) have $\beta = \exp(-\lambda)$ and $\kappa = v$.

so the more x_t exceeds its expected value, the higher the concentration. Thus (19) can be written as $u_t^\mu = v(x_t) \sin(y_t - \mu_{t|t-1})$. When x_t is close to zero, there is no clear direction so the concentration is low. The conditional distribution of x_t given y_t together with all the information at time $t - 1$ is Weibull with scale

$$\varphi(y_t) = (1 - \tanh v \cdot \cos(y_t - \mu_{t|t-1}))^{-1/\alpha} \varphi_{t|t-1}, \quad (23)$$

where $\varphi_{t|t-1} = \exp(\lambda_{t|t-1})$. Substituting in (20) gives $u_t^\lambda = \alpha[(x_t/\varphi(y_t))^\alpha - 1]$. When y_t is close to $\mu_{t|t-1}$ it will boost the effect of x_t .

5.3. Heteroscedasticity

As the model stands, concentration, $v(x_t)$, changes only with x_t , depending on whether x_t is higher or lower than expected given $\lambda_{t|t-1}$. Using a result in Abe and Ley (2017, p. 95), the expected value of $v(x_t)$ based on information at time $t - 1$ is

$$\begin{aligned} E_{t-1}v(x_t) &= \tanh(v)E_x(x_t/\exp \lambda_{t|t-1})^\alpha = \tanh v \cosh v \cdot P_1^0(\cosh v) \\ &= \tanh v \cosh^2 v = 0.5 \sinh(2v), \end{aligned}$$

where $P_1^0(\cdot)$ is the associated Legendre function of the first kind with degree v and order zero. Thus the prediction of $v(x_t)$ is constant. It is not dependent on $\lambda_{t|t-1}$ and so if, in the context of wind, speed has been high for some time, a value of x_t lower than its expectation will imply that concentration is suddenly lower than average. This seems implausible and it points to the need to introduce dynamic heteroscedasticity into the model by letting v be dynamic. The score with respect to this new dynamic parameter, denoted, $v_{t|t-1}$, is

$$u_t^v = (x_t/\exp \lambda_{t|t-1})^\alpha [1 - \tanh^2 v_{t|t-1}] \cos(y_t - \mu_{t|t-1}) - \tanh v_{t|t-1}. \quad (24)$$

The score u_t^v is very close to that of $\lambda_{t|t-1}$, in (20), but it differs in that when y_t is close to $\mu_{t|t-1}$ it increases whereas u_t^λ reacts in the opposite direction. Note that u_t^λ is now defined with $\tanh v_{t|t-1}$ replacing $\tanh v$. Using the filter for $v_{t|t-1}$ now gives $E_{t-1}v(x_t) = 0.5 \sinh(2v_{t|t-1})$.

The heteroscedastic dynamic model includes an equation for $\zeta_{t|t-1} = \ln v_{t|t-1}$ to complement those in (21). The information matrix for μ , λ and ζ is given in the Supporting information. Its availability raises the possibility of pre-multiplying the scores by its inverse, as is often done in the dynamic score literature.

Remark 2. Abe and Ley (2017, pp. 96–7), p 96-7, give a generalization,⁴ the GGSSVM, in which the generalized gamma (GG) distribution replaces the Weibull; the circular marginal distribution is then the Jones-Pewsey distribution. Imoto *et al.* (2019) propose a generalized Pareto-type cylindrical distribution that can handle heavier tails. In both cases a score driven model can again be formulated.

5.4. Forecasts

Forecasts are based on information at T so for $T + 1$ we plug $\mu_{T+1|T}$, $v_{T+1|T}$ and $\lambda_{T+1|T}$ into the joint distribution. The (marginal) distribution of y_{T+1} , conditional on information at time T , is wrapped Cauchy that is

$$f_T(y_{T+1}) = \frac{1}{2\pi} \frac{1 - \tanh^2(v_{T+1|T}/2)}{1 + \tanh^2(v_{T+1|T}/2) - 2 \tanh(v_{T+1|T}/2) \cos(y_{T+1} - \mu_{T+1|T})}, \quad (25)$$

⁴ Note that they have α replacing our $\alpha\gamma$ and γ replacing α .

where $-\pi \leq y_{T+1} < \pi$. The one-step ahead forecast for direction, $E_T(y_{T+1})$, is just the predicted location $\mu_{T+1|T}$. The marginal distribution for the linear variable is given in Abe and Ley (2017, p. 94) as

$$f_T(x_{T+1}) = V_{T+1|T}(x_{T+1}) \frac{\alpha}{e^{\lambda_{T+1|T}}} \left(\frac{x_{T+1}}{e^{\lambda_{T+1|T}}} \right)^{\alpha-1} \exp(-(x_{T+1}/e^{\lambda_{T+1|T}})^\alpha),$$

where $0 \leq x_{T+1} < \infty$ and

$$V_{T+1|T}(x_{T+1}) = \frac{I_0((x_{T+1}/e^{\lambda_{T+1|T}})^\alpha \tanh v_{T+1|T})}{\cosh v_{T+1|T}}.$$

The one-step ahead forecast of x_{T+1} is

$$E_T(x_{T+1}) = \exp(\lambda_{T+1|T}) \Gamma(1 + 1/\alpha) [(\cosh v_{T+1|T})^{1/\alpha} P_{1/\alpha}(\cosh v_{T+1|T})].$$

Except for the normalizing term $V_{T+1|T}(x_{T+1})$, the form of $f_T(x_{T+1})$ is that of a Weibull distribution and likewise $E_T(x_{T+1})$ is as for a Weibull distribution, apart from the term in square brackets. Multi-step forecasts can be obtained by simulation. Abe and Ley (2017, p. 94), provide details on how to simulate from the WeiSSVM distribution.

5.5. Switching Cylinders

DAMMs can be applied to multivariate series as in Catania (2021, eq. (3)). In a bivariate model the switching filter for $\xi_{i,t|t-1}$ depends on the joint PDF $f(y_t, x_t)$. All parameters, including those that are fixed, such as α , are regime dependent. Following on from (17), the score for location within a regime is then given by

$$u_{it}^\mu = \frac{\partial \ln f_{i,t|t-1}}{\partial \mu_{i,t|t-1}} = \xi_{i,t|t} (y_t, x_t) v_i(x_t) \sin(y_t - \mu_{i,t|t-1}), \quad i = 1, 2,$$

where $v_i(x_t) = \tanh(v_i)(x_t / \exp \lambda_{i,t|t-1})^{\alpha_i}$, and similarly for the other scores.

The graphs in Figure 1 suggest that the regimes for direction are more clearly defined than those for speed. Thus it is worth considering whether to model regime switching only in terms of the marginal distribution of direction, y_t . To implement such a regime switching mechanism, the PDF in the score with respect to the dynamic switching probability, that is (11), is taken to be wrapped Cauchy, as in (25), and the same density is used in the contemporaneous probability (13).

6. WIND DIRECTION AND SPEED IN GALICIA

The scatter plot of speed and direction for the Galicia data in Figure 3 highlights the regimes in direction and confirms the impression gained from Figure 1 that speed tends to be higher when the wind is coming from the SW.

As might be expected from the univariate results on wind direction, a static mixture model fares badly, with $\ln L = -3732.2$ as opposed to $\ln L = -3371.2$ for the single regime model without heteroscedasticity and $\ln L = -3355.7$ with heteroscedasticity. The pure DAMM model, shown in the second line of Table IV and labelled model (6), is much better, with $\ln L = -3413.7$ but it too fails to beat the single regime models. The score-based Q -statistics for residual serial correlation shown in Table VI are huge.

The inclusion of dynamics within regimes offers considerable improvement. As before the fit is better with heteroscedasticity dynamics. The main issue to resolve is whether the dynamics in the switching equation should depend on both direction and speed or on direction only. The results favour the second possibility, especially when the dynamics include heteroscedasticity. Thus the estimates reported in the last lines of Tables IV–VI are for the preferred model. As can be seen, $\ln L = -2946.8$. There is still some residual serial correlation in some of the

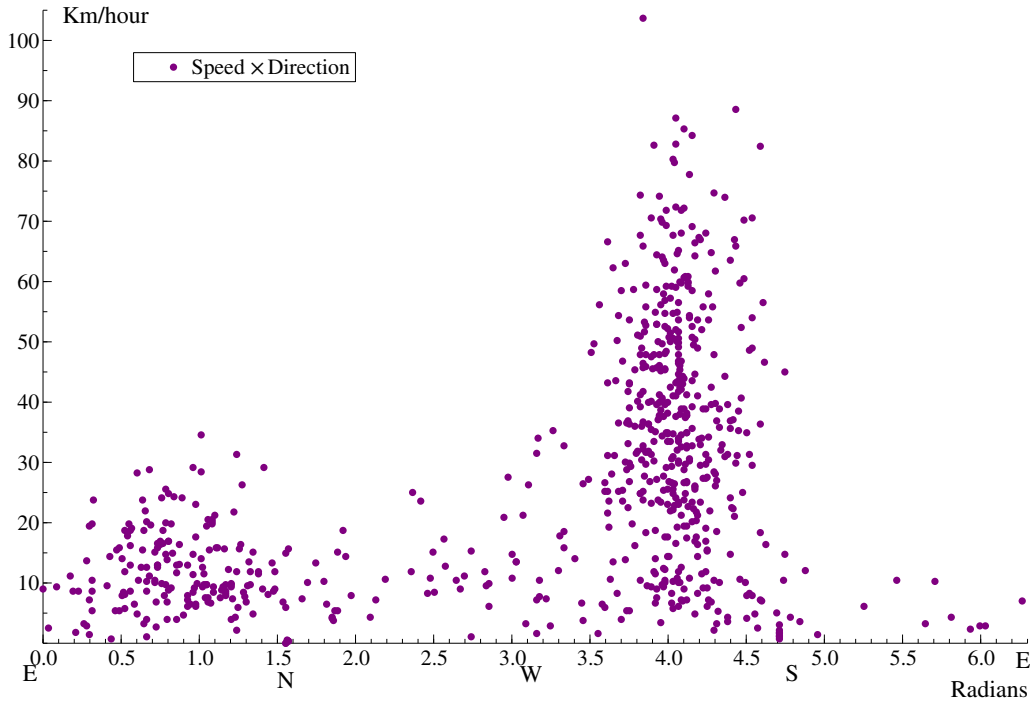


Figure 3. Scatter diagram of wind speed and direction in Galicia

Table IV. Goodness of fit of the two regime WeiVM model with the estimated dynamic parameters of the switching probability from fitting, (5), the static mixture model, (6), the pure DAMM model, (7), the DAMM model with intra regime dynamic location and scale and (8), DAMM model with intra regime dynamic location, scale and concentration

Model	AIC	BIC	Logl	ω_γ	ϕ_γ	κ_γ	ξ
(5)	7,482.446	7,523.954	-3,732.223	0.994 (0.097)	-	-	0.730
(6)	6,849.460	6,900.193	-3,413.730	1.084 (0.552)	0.846 (0.022)	11.079 (1.224)	0.747
(7)	6,034.761	6,122.389	-2,998.380	1.151 (0.664)	0.924 (0.015)	7.729 (1.114)	0.760
(8)	5,992.685	6,098.762	-2,973.343	1.669 (0.561)	0.870 (0.020)	27.487 (0.833)	0.842
(9)	6,018.408	6,106.037	-2,990.204	1.336 (0.828)	0.966 (0.010)	5.594 (1.076)	0.792
(10)	5,939.650	6,045.727	-2,946.825	0.032 (0.552)	0.879 (0.016)	15.945 (2.524)	0.508

Note: Models (9) and (10) are specified in the same way as (7) and (8) except that the switching probability is driven by the marginal distribution with respect to location. The switching parameters are shown in the last four columns, with ξ denoting the logistic transformation of ω_γ .

Table V. Estimated dynamic parameters for the two regime WeivM model from fitting, (5), the static mixture model, (6), the pure DAMM model, (7), the DAMM model with intra regime dynamic location and scale and (8), DAMM model with intra regime dynamic location, scale and concentration

Model	$\omega_{\mu 1}$	$\phi_{\mu 1}$	$\kappa_{\mu 1}$	$\omega_{\lambda 1}$	$\phi_{\lambda 1}$	$\kappa_{\lambda 1}$	$\omega_{\gamma 1}$	$\phi_{\gamma 1}$	$\kappa_{\gamma 1}$	α_1	$\omega_{\mu 2}$	$\phi_{\mu 2}$	$\kappa_{\mu 2}$	$\omega_{\lambda 2}$	$\phi_{\lambda 2}$	$\kappa_{\lambda 2}$	$\omega_{\gamma 2}$	$\phi_{\gamma 2}$	$\kappa_{\gamma 2}$	α_2	
(5)	4.046 (0.012)	-	-	2.418 (0.054)	-	-	0.695 (0.024)	-	-	2.115 (0.033)	0.886 (0.028)	-	-	1.561 (0.096)	-	-	0.543 (0.051)	-	-	-	2.021 (0.057)
(6)	4.048 (0.011)	-	-	2.422 (0.045)	-	-	0.735 (0.022)	-	-	2.171 (0.029)	0.917 (0.032)	-	-	1.732 (0.025)	-	-	0.381 (0.047)	-	-	-	2.002 (0.041)
(7)	4.002 (0.032)	0.927 (0.025)	0.010 (0.002)	2.218 (0.080)	0.978 (0.005)	0.021 (0.002)	0.748 (0.021)	-	-	3.575 (0.029)	1.694 (0.110)	0.992 (0.003)	0.043 (0.007)	1.622 (0.016)	0.811 (0.047)	0.089 (0.010)	0.489 (0.041)	-	-	-	2.329 (0.031)
(8)	4.048 (0.001)	0.941 (0.024)	0.008 (0.001)	2.161 (0.124)	0.976 (0.000)	0.023 (0.002)	0.592 (0.027)	0.912 (0.027)	0.015 (0.003)	3.398 (0.054)	0.854 (0.071)	0.982 (0.019)	0.013 (0.004)	1.798 (0.002)	0.802 (0.068)	0.046 (0.006)	1.261 (0.133)	0.998 (0.000)	0.067 (0.014)	0.077 (0.033)	2.342 (0.033)
(9)	3.998 (0.030)	0.919 (0.025)	0.010 (0.002)	2.274 (0.110)	0.972 (0.007)	0.027 (0.003)	0.744 (0.020)	-	-	3.478 (0.031)	1.408 (0.082)	0.967 (0.013)	0.052 (0.011)	1.482 (0.166)	0.924 (0.038)	0.051 (0.010)	0.485 (0.053)	-	-	-	2.481 (0.047)
(10)	4.068 (0.042)	0.957 (0.016)	0.007 (0.001)	2.326 (0.091)	0.972 (0.008)	0.018 (0.002)	-0.447 (0.061)	0.941 (0.015)	0.025 (0.005)	3.237 (0.033)	1.786 (0.200)	0.993 (0.004)	0.039 (0.007)	1.401 (0.170)	0.943 (0.023)	0.038 (0.006)	-0.479 (0.053)	0.756 (0.103)	0.021 (0.010)	0.021 (0.048)	2.559 (0.048)

Note: Models (9) and (10) are specified in the same way as (7) and (8) except that the switching probability is driven by the marginal distribution with respect to location.

Table VI. Box–Ljung test with $P = 5$, $Q(5)$, for residual correlation on the fitted scores from fitting, (5), the static mixture model, (6), the pure DAMM model, (7), the DAMM model with intra regime dynamic location and scale and (8), DAMM model with intra regime dynamic location, scale and concentration

Model	μ_1	λ_1	ζ_1	μ_2	λ_2	ζ_2	γ
(5)	767.826***	774.366***	2037.36***	550.021***	649.586***	1248.86***	2642.13***
(6)	753.425***	678.766***	1918.03***	587.870***	516.355***	1339.66***	26.827***
(7)	5.093	3.303	219.270***	25.646***	14.792**	159.494***	11.666**
(8)	3.479	3.737	38.265***	53.012***	6.806	41.579***	1.069
(9)	3.459	1.528	158.644***	25.686**	10.496*	141.566***	120.995***
(10)	10.298*	2.551	26.706***	26.684***	5.073	54.875***	43.294***

Note: Models (9) and (10) are specified in the same way as (7) and (8) except that the switching probability is driven by the marginal distribution with respect to location. (Nominal) significance levels: 0.01 '***', 0.05 '**', 0.1 '*'.

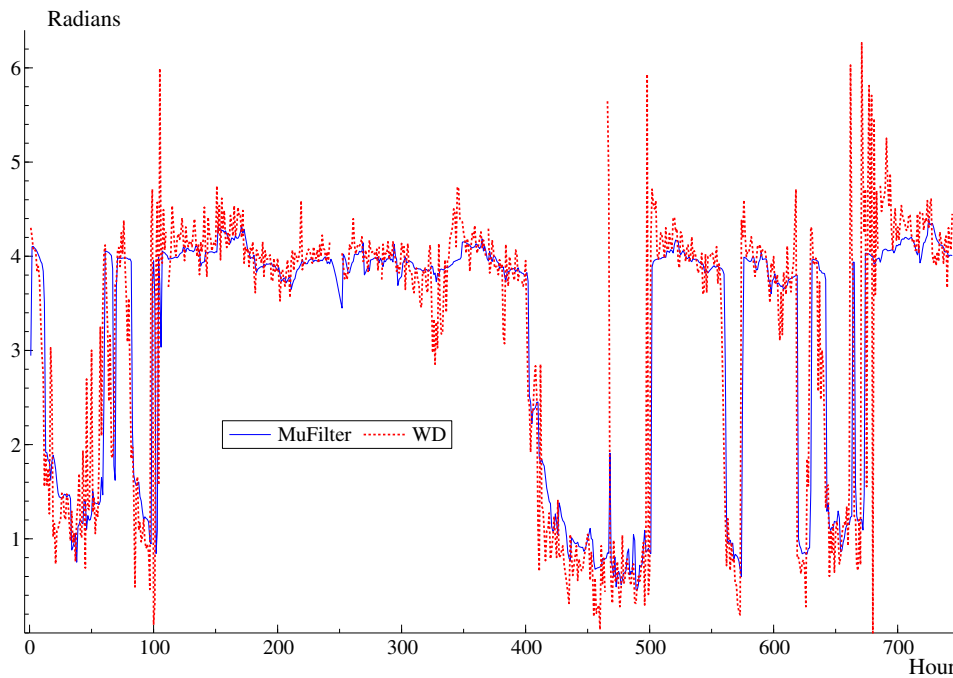


Figure 4. Wind direction (WD) and fitted wind direction (MuFilter) in Galicia with two regime heteroscedastic dynamic WeivM model

components, but, as noted earlier, this is not unusual with large sample sizes. The estimates of α in the Weibull parameter are well above one in almost all cases.

The observations and combined filters for direction are shown in Figure 4. The filtered estimates are smoother than the raw observations. The filtered estimates stay well within the range $[0, 2\pi)$ whereas the observations sometimes move rapidly between the top and bottom of the graph. Note that the combined filter for mean direction is computed using the $a \tan 2$ function, as in (17), while the corresponding filter for scale is $\xi_{1,t|t-1} \exp(\lambda_{1,t|t-1}) + \xi_{2,t|t-1} \exp(\lambda_{2,t|t-1})$, $t = 1, \dots, T$.

Figures 5 and 6 show filtered wind direction and speed in individual regimes. It can be seen that when the probability of being in a given regime is small, the movements in the underlying filter change only gradually. For

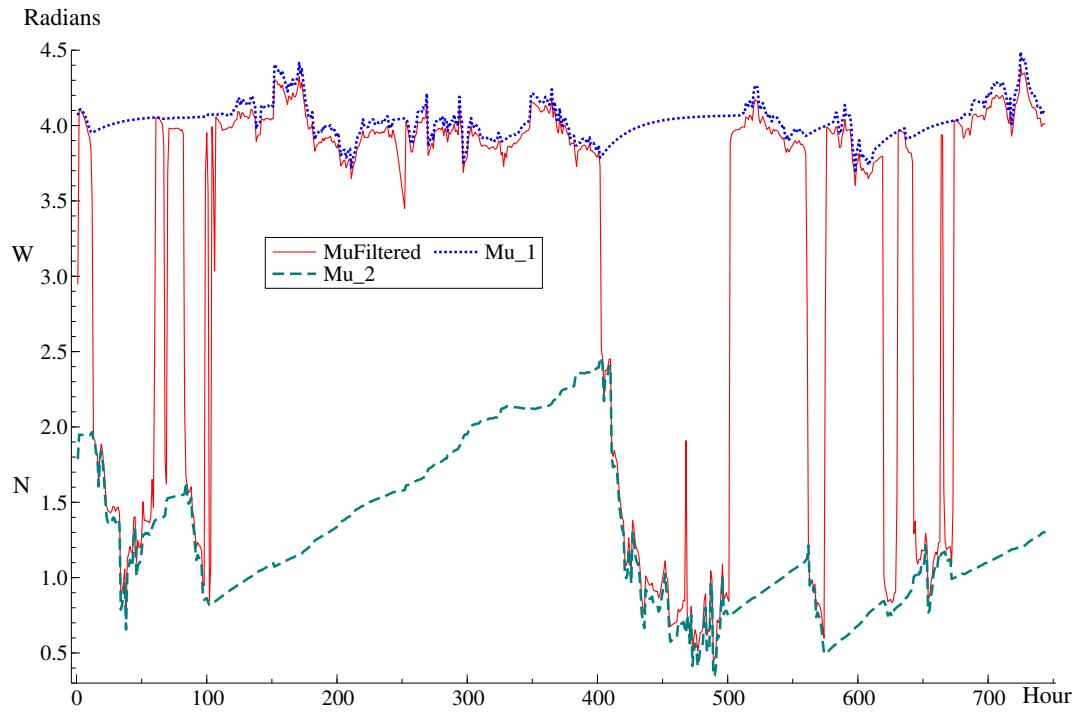


Figure 5. Filtered wind direction in individual regimes, mu1 and mu2, together with the combined filter

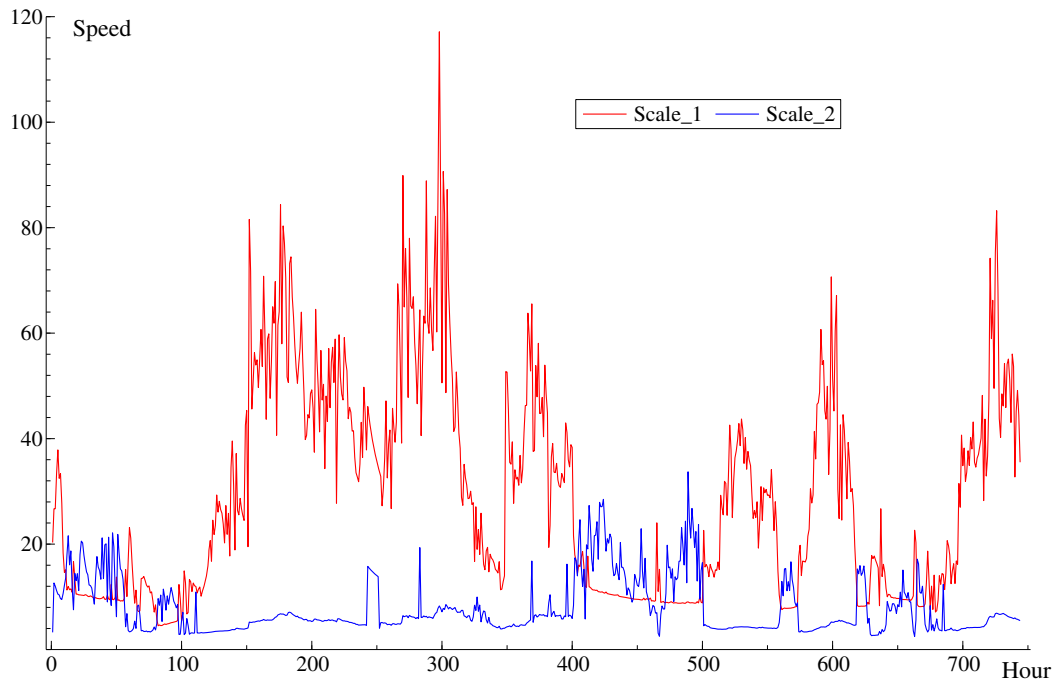


Figure 6. Filtered scale for wind speed in individual regimes

wind direction, the combined filter of Figure 4 is also shown. The combined filter for speed is not shown as it is just the sum of the individual filters.

Remark 3. When there is no wind, it has no direction. In such cases $x_t = 0$ and so the model gives $v(x_t) = 0$ which implies the (unobserved) wind direction is distributed uniformly. It is evident from (20) that the score for location, u_t^μ , is zero. Thus the observation is effectively ignored as in the naive solution for dealing with an observation that is missing.⁵ This is not the case for the scale of the linear variable because $u_t^\lambda = -\alpha$ and the concentration score where $u_t^\nu = -\tanh v_{t|t-1}$. As regards the likelihood, the difficulty is that $f(y_t, 0) = 0$ for $\alpha > 1$, indicating that $x_t = 0$ is impossible. For $\alpha < 1$, $f(0) = \infty$ which is also unhelpful. Only for $\alpha = 1$ is there a viable solution as in this case $f(y_t, 0) = 1/2\pi$. The simplest solution is to assume there is no contribution to the likelihood.

7. CONCLUSION

Score-driven regime switching models can be extended to handle circular observations and diagnostic tests can be constructed. The models allow for changing concentration as well as changing location. When fitted to hourly wind direction in a site in Galicia, pure regime switching models, without intra-regime dynamics, are unable to outperform the single regime model when both location and scale are dynamic. Although the diagnostic test based on the switching residuals indicates that there are no omitted dynamics in the regime-switching equation, the Q -statistics for location dynamics are still highly significant in both regimes. Fitting a score-driven switching model with location dynamics in each regime gives a big increase in the likelihood function.

The score-driven approach is then used to construct dynamic bivariate models for circular and linear variables with a conditional cylindrical distribution. The preferred specification for the Galicia data has dynamic location and concentration for wind direction and dynamic location/scale for its speed. Estimating a restricted regime switching model, in which the regime dynamics depend only on direction, gives a good fit when heteroscedasticity is included and the resulting filter for direction tracks the observations remarkably well. Again the modelling of intra-regime dynamics is crucial.

There is further scope for research extending the score-driven approach to bivariate dynamic cylindrical models based on copulas, as used by García-Portugués *et al.* (2013) and Lagona (2019), and to directional data on a sphere or torus.

ACKNOWLEDGEMENTS

We are grateful to Eduardo García-Portugués for providing the Galicia data used in García-Portugués *et al.* (2013). Earlier versions of some of the ideas in this article were presented at the Econometric Models of Climate Change conference in Milan in August 2019, at the 22nd Oxmetrics conference at Nuffield college, Oxford in September, 2019, and at workshops in Cambridge, Bologna, Ecole Polytechnique Fédérale de Lausanne, the QUT Centre for Data Sciences and the University of Konstanz. Later versions were given at a plenary (virtual) session of the 45th NBER-NSF conference in October 2021 at Rice University, Houston and at the ADISTA22 (Advances in Directional Statistics) conference in Santiago de Compostela in June 2022. We are grateful to Anthony Davison, Jurgen Doornik, David Hendry, Stan Hurn, Peter Jupp, John Kent, Francisco Lagona, Rutger-Jan Lange, Christophe Ley, Ken Lindsay, Oliver Linton, Alessandra Luati, Paul Myer, Alexiy Onatski, Richard Smith, Howell Tong and two referees for helpful comments.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author on reasonable request.

⁵ There is a run of missing observations on both wind and velocity around observation 250. In such cases we set all scores to zero: hence the slight dip in Figures 5 and 6. There are a few more missing observations around 455 and 680.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

REFERENCES

- Abe T, Ley C. 2017. A tractable, parsimonious and flexible model for cylindrical data, with applications. *Econometrics and Statistics* **4**:91–104.
- Breckling J. 1989. *The analysis of directional time series: Applications to wind speed and direction*.
- Calvori F, Creal D, Koopman S, Lucas A. 2017. Testing for parameter instability in competing modeling frameworks. *Journal of Financial Econometrics* **15**:223–246.
- Catania L. 2021. Dynamic adaptive mixture models with an application to volatility and risk. *Journal of Financial Econometrics* **19**:531–564.
- Creal D, Koopman S, Lucas A. 2013. Generalized autoregressive score models with applications. *Journal of Applied Econometrics* **28**:777–795.
- Fisher N, Lee A. 1992. Regression models for an angular response. *Biometrics* **48**:665–677.
- Fisher N, Lee A. 1994. Time series analysis of circular data. *Journal of the Royal Statistical Society B*(70):327–332.
- García-Portugués E, Barros A, Crujeiras R, González-Manteiga W, Pereira JC. 2014. A test for directional-linear independence, with applications to wildfire orientation and size. *Stochastic Environmental Research and Risk Assessment* **28**:1261–1275.
- García-Portugués E, Crujeiras R, González-Manteiga W. 2013. Exploring wind direction and so₂ concentration by circular-linear density estimation. *Stochastic Environmental Research and Risk Assessment* **27**:1055–1067.
- Haas M, Mittnik S, Paolella M. 2004. A new approach to Markov-switching garch models. *Journal of Financial Econometrics* **2**:493–530.
- Hamilton J. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57**(2):357–384.
- Hamilton J. 1994. *Time series analysis*.
- Hamilton J. 1996. Specification testing in Markov-switching time-series models. *Journal of Econometrics* **70**:127–157.
- Harvey A, Hurn S, Palumbo D, Thiele S. 2022. Modeling circular time series. mimeo. *Revised version of a 2019 Cambridge Working Papers in Economics*. Mimeo (1971).
- Harvey A, Thiele S. 2016. Testing against changing correlation. *Journal of Empirical Finance* **38**:575–589.
- Harvey AC. 2013. Dynamic models for volatility and heavy tails: with applications to financial and economic time series. *Econometric Society Monograph*.
- Holtzmann H, Munk A, Suster M, Zucchini W. 2006. Hidden Markov models for circular and linear-circular time series. *Environmental and Ecological Statistics* **13**:325–347.
- Imoto T, Shimizu K, Abe T. 2019. A cylindrical distribution with heavy-tailed linear part. *Japanese Journal of Statistics and Data Science* **2**:129–154.
- Jammalamadaka S, SenGupta A. 2001. *Topics in circular statistics*.
- Jones M, Pewsey A. 2005. A family of symmetric distributions on the circle. *Journal of the American Statistical Association* **100**:1422–1428.
- Lagona F. 2019. Copula-based segmentation of cylindrical time series. *Statistics and Probability Letters* **144**:16–22.
- Lagona F, Picone M, Maruotti A. 2015. A hidden Markov model for the analysis of cylindrical time series. *Environmetrics* **26**:534–544.
- Mardia KV, Jupp P. 2000. *Directional statistics*.
- Pewsey A, García-Portugués E. 2021. Recent advances in directional statistics. *TEST* **30**:1–58.
- Smith D. 2008. Evaluating specification tests for Markov-switching time-series models. *Journal of Time Series Analysis* **29**:629–652.
- Waltz R, Morales J, Nocedal J, Orban D. 2006. An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Mathematical Programming* **107**:391–408.
- Zucchini W, MacDonald IL, Langrock R. 2016. Hidden Markov models for time series: an introduction using r. *Monographs on Statistics and Applied Probability* **150**.