UNIVERSITY OF CAMBRIDGE
Faculty of Economics

CAMBRIDGE WORKING PAPERS IN ECONOMICS

Regime switching models for directional and linear observations

Andrew Harvey
Dario Palumbo

Abstract

The score-driven approach to time series modeling provides a solution to the problem of modeling circular data and it can also be used to model switching regimes with intra-regime dynamics. Furthermore it enables a dynamic model to be fitted to a linear and a circular variable when the joint distribution is a cylinder. The viability of the new method is illustrated by estimating a model with dynamic switching and dynamic location and/or scale in each regime to hourly data on wind direction and speed in Galicia, north-west Spain.

Reference Details

CWPE 2123
Published 9 March 2021

Key Words Circular data, conditional score, cylinder, hidden Markov model, von Mises distribution, wind.
JEL Codes C22, C32

Website www.econ.cam.ac.uk/cwpe

Regime switching models for directional and linear observations

Andrew Harvey* and Dario Palumbo**
Faculty of Economics, Cambridge University, and University of Venice.

February 15, 2021

Abstract

The score-driven approach to time series modeling provides a solution to the problem of modeling circular data and it can also be used to model switching regimes with intra-regime dynamics. Furthermore it enables a dynamic model to be fitted to a linear and a circular variable when the joint distribution is a cylinder. The viability of the new method is illustrated by estimating a model with dynamic switching and dynamic location and/or scale in each regime to hourly data on wind direction and speed in Galicia, north-west Spain.

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

Email: *ach34@cam.ac.uk, **dp470@cam.ac.uk

Ethics: data is available. There are no ethical problems.

1 Introduction

A dynamic mixture distribution may be modeled using the score of the regime probabilities in the conditional distribution. Catania (2019) calls these dynamic adaptive mixture models (DAMMs). The DAMM model belongs to

the class of score-driven models developed in Harvey (2013) and Creal et al (2013), where they were called DCS and GAS models respectively. The locations and/or scales in each of the regimes contained in the mixture may also dynamic. Again the conditional scores are used so providing a unified approach based on well-established principles. The scores for location and scale, like the scores for the regime probabilities, have a natural and intuitive interpretation.

The DAMM model is designed for situations similar to those addressed by the regime switching model introduced by Hamilton (1989); see Fruhwirth-Schnatter (2006) for more recent references. These models now appear in standard textbooks in econometrics. They have been widely applied, for example with respect to modeling booms and recessions, albeit with limited forecasting success. The basic model 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. The regime is not observed: hence the term hidden Markov model (HMM) as in Zucchini, MacDonald and Langrock (2016). The probability of being in a given regime is given by a filter that depends on past observations. These probabilities yield a conditional distribution for the current observation. Hence a likelihood function may be constructed and maximum likelihood (ML) estimates of the transition probabilities obtained. The DAMM approach differs in that it by-passes the hidden Markov chain and derives dynamics directly from the probabilities in the conditionally distribution. In other words it sets up an observation driven model for the mixture distribution[1]. The intra-regime dynamics are similarly constructed and, as a result, past observations are weighted according to the probability that they are in a given regime. This not does happen with intra-day dynamics in a standard regime switching

[1]Bazzi et al (2017) add an extra dimension by modeling the time-varying transition probabilities using a score-driven approach. As such it combines a parameter-driven model - the Markov switching - with an observation-driven model. The DAMM, on the other hand, is fully observation driven.

model.

The structure of the DAMM is such that diagnostic tests based on the Lagrange multiplier (or score) test principle are easily constructed. When a static mixture model, that is with independent observations, 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. This is of great benefit for model specification. When dynamics have been estimated, diagnostics designed to assess the possibility of omitted dynamic effects can be constructed using the same principles.

Mixture distributions appear in the literature on the direction of wind, because there is often more than one direction for the prevailing wind. This suggests a switching model for moving between them because if they are a long way apart, perhaps in opposite directions, it may take time for a single regime model to adjust. Holtzmann et al (2006) propose a switching regime model for wind direction and the same approach is used by Zucchini, MacDonald and Langrock (2016, p 228-35) for modeling the change in direction of flight for fruit flies. However, there are no dynamics in individual regimes, only dynamics in moving from one state to another. The DAMM framework allows for a wide range of dynamics in the different regimes and recent work by Harvey et al (2019) has shown how the score-driven approach provides a solution to time series modeling of directional data. This paper shows how the DAMM model can be used for circular data in which the level of each regime is serially correlated. Modeling serial correlation in concentration or variance is an additional option.

The viability and effectiveness of our modeling approach 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 Garcia-Portugues 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 first observation in each hour. Garcia-Portugues 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. Because of circularity some of the measurements in the 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.

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. The recent paper by 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 paper by Fisher and Lee (1992). 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 resulting dynamic model may then be incorporated in a 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 a hidden Markov model for fitting a cylindrical distribution when there are three distinct regimes.

Section 2 describes the basic DAMM model, with constant parameters in the regimes, and compares it with Markov switching models. The extension to dynamic location and scale is then made. Section 3 lays out the diagnostic tests. The methods are then specialized to directional variables in Section 4 and applied to the Galicia data on wind direction in Section 5. Dynamic cylindrical distributions are described in Section 6 and extended to handle switching regimes in Section 7. These models are fitted to the Galicia data

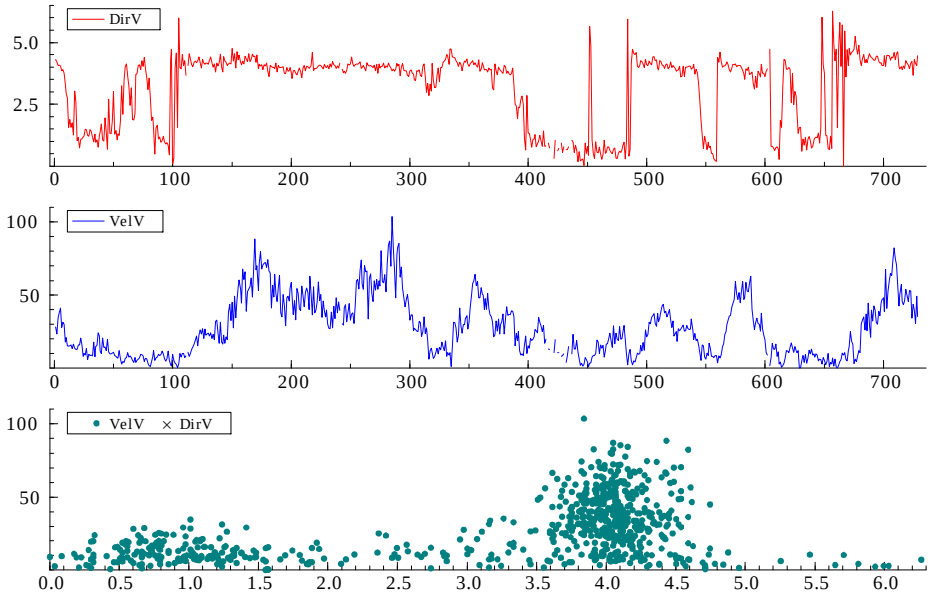

Figure 1: Hourly wind direction (DirV) and speed (VelV) in Galicia, with scatter plot.

on direction and speed in Section 8. Section 9 concludes.

2 Dynamic adaptive mixture model

2.1 Static mixture model

The PDF of a mixture of K distributions is

$$f(y_t) = \sum_{i=1}^{K} \xi_i f_i(y_t), \quad 0 \leq \xi_i \leq 1, \ \Sigma_i \xi_i = 1, \quad t = 1, ..., T,$$

where ξ_i is the probability of being in $i-th$ regime and $f_i(y_t) = f_i(y_t; \boldsymbol{\psi}_i)$, $i = 1, 2,..,K$ with $\boldsymbol{\psi}_i$ denoting the parameters in the $i-th$ regime. When the observations are independent, the probability of being in regime i given observation y_t is

$$\xi_i(y_t) = \xi_i f_i(y_t)/f(y_t), \quad i = 1, 2, ...K. \tag{1}$$

The ML estimates for the location μ_i and scale φ_i in each distribution, together with the unconditional probabilities, ξ_i, $i = 1, ..., K$, are obtained by solving

$$\sum_{t=1}^{T} \xi_i(y_t) \frac{\partial \ln f(y_t)}{\partial \mu_i} = 0 \ \text{ and } \sum_{t=1}^{T} \xi_i(y_t) \frac{\partial \ln f(y_t)}{\partial \varphi_i} = 0, \quad i = 1, 2, .., K, \tag{2}$$

and

$$\xi_i = \frac{\sum_{t=1}^{T} \xi_i(y_t)}{T}, \quad i = 1, 2, .., K. \tag{3}$$

These estimates can be computed iteratively by what turns out to be a special case of the EM algorithm; see Catania (2019, p9).

2.2 Score-driven dynamic mixtures

The two-state DAMM has the mixture weight at time t changing over time according to a filter, $\xi_{t|t-1}$, that is based on information at time $t-1$ and is driven by the score with respect to $\xi_{t|t-1}$. The conditional distribution is

$$f_{t|t-1}(y_t) = \xi_{t|t-1} f_{1,t|t-1}(y_t) + (1-\xi_{t|t-1}) f_{2,t|t-1}(y_t), \qquad t = 1, ..., T, \quad (4)$$

where $\xi_{t|t-1}$ is the probability of being in state one at time t, based on information available up to and including time $t-1$. The parameters in each regime may depend on past observations as well. The score is

$$\frac{\partial \ln f_{t|t-1}}{\partial \xi_{t|t-1}} = \frac{f_{1,t|t-1} - f_{2,t|t-1}}{f_{t|t-1}}. \quad (5)$$

Confirmation that it has zero expectation is straightforward.

With a logistic link function

$$\xi_{t|t-1} = \frac{\exp \gamma_{t|t-1}}{1 + \exp \gamma_{t|t-1}}, \qquad -\infty < \gamma_{t|t-1} < \infty$$

the score with respect to $\gamma_{t|t-1}$ is

$$\begin{aligned} w_t &= \frac{\partial \ln f_{t|t-1}}{\partial \gamma_{t|t-1}} = \frac{f_{1,t|t-1} - f_{2,t|t-1}}{f_{t|t-1}} \frac{\exp \gamma_{t|t-1}}{(1 + \exp \gamma_{t|t-1})^2} \quad (6) \\ &= \frac{f_{1,t|t-1} - f_{2,t|t-1}}{f_{t|t-1}} \xi_{t|t-1}(1 - \xi_{t|t-1}). \quad (7) \end{aligned}$$

There is no compelling reason to divide the score by the information quantity matrix; see Catania (2019, p 7).

The basic first-order dynamic equation is

$$\gamma_{t+1|t} = (1-\phi)\omega_\gamma + \phi \gamma_{t|t-1} + \kappa w_t, \quad t = 1, ..., T, \quad (8)$$

where the condition $|\phi| < 1$ 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; as $\kappa \to \infty$ there will be an abrupt change in regime when w_t changes sign, whereas when κ is close to zero any change will be gradual.

Catania (2019) discusses methods for extending the logistic transformation to more than two regimes.

2.3 Dynamics within regimes

Changing location in the i-th regime is driven by the score

$$\begin{aligned} \frac{\partial \ln f_{t|t-1}}{\partial \mu_{i,t|t-1}} &= \frac{\partial \ln f_{t|t-1}}{\partial f_{t|t-1}} \frac{\partial f_{t|t-1}}{\partial f_{i,t|t-1}} \frac{\partial f_{i,t|t-1}}{\partial \ln f_{i,t|t-1}} \frac{\partial \ln f_{i,t|t-1}}{\partial \mu_{i,t|t-1}} \\ &= \xi_{i,t|t-1}(y_t) \frac{\partial \ln f_{i,t|t-1}}{\partial \mu_{i,t|t-1}} = u_{it}, \qquad i = 1, ...K. \end{aligned} \tag{9}$$

where, following on from (1),

$$\xi_{i,t|t} = \xi_{i,t|t-1} f_{it} / f_{t|t-1}, \quad i = 1, 2, ..., K, \tag{10}$$

is the estimate of $\xi_{i,t}$ given y_t and $\xi_{i,t|t-1}$. When $\xi_{i,t|t}$, the estimated probability of being in the $i-th$ regime, is small, the contribution of y_t to the corresponding score is downweighted.

In the basic first-order case, the dynamic equations are

$$\mu_{i,t+1|t} = \omega_{\mu i}(1 - \phi_i) + \phi_i \mu_{i,t|t-1} + \kappa_i u_{it}, \quad |\phi_i| < 1, \quad i = 1, ..., K, \tag{11}$$

and the overall level is

$$\mu_{t|t-1} = \sum_{i=1}^{K} \xi_{i,t|t-1} \mu_{i,t|t-1}.$$

ML estimates are obtained by maximizing the log-likelihood function

$$\ln L = \sum_{t=1}^{T} \ln f_{t|t-1} = \sum_{t=1}^{T} \ln \left[\sum_{i=1}^{K} \xi_{i,t|t-1} f_{i,t|t-1} \right].$$

Starting values for $\omega_{\mu i}$ and φ_i, $i = 1, 2, .., K$ and ξ_i, $i = 1, 2, .., K$, can be obtained from the static model as described in sub-section 2.1. Similar formulae are obtained when other parameters, such as scale, are dynamic.

Remark 1 *In a Markov switching model, dynamics can be introduced into the location and/or scale of each regime by letting them depend on past observations. For example the conditional mean in the $i-th$ regime may be given by $\mu_i + \phi_i(y_{t-1} - \mu_i)$, $t = 1, ..., T$, $i = 1, 2$; see Hamilton (1994, p 691). This is very different from the filter in (11) which is driven by a forcing variable that is weighted by the probability of being in the $i-th$ regime.*

Remark 2 *Intra-regime dynamics feature very little in the HMM literature. For example, Zucchini, MacDonald and Langrock (2016, p 150-2) devote only a few pages under the title 'HMMs with additional dependencies'.*

3 Model selection and diagnostics

When a static mixture has been fitted, Lagrange multiplier (LM) tests are equivalent to portmanteau tests and they enable the researcher to separate out transition dynamics from location and/or scale dynamics. The pattern of the correlogram may be informative as to possible models and so this initial step has the potential for playing an important role in model specification.

Hamilton (1996) sets out tests for Markov switching models. Smith (2008) finds the LM test to have the best size and power properties. Here we show that in a score-driven model, LM tests can be set up very easily to test for serial correlation in switching probabilities, locations and scales.

3.1 Diagnostic tests when a (static) mixture has been fitted.

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

$$\gamma_{t|t-1} = \omega_\gamma + \kappa_1 w_{t-1} + ... + \kappa_P w_{t-P}, \tag{12}$$

where w_t is as in (7), may be constructed. When the model is static

$$w_t = \frac{f_1(y_t) - f_2(y_t)}{f(y_t)}\xi(1-\xi) = \xi(y_t) - \xi \tag{13}$$

because, from (1), $f_1(y_t) = \xi(y_t)f(y_t)/\xi$, where, as before, the subscript is dropped from ξ_1. Thus the LM test of the null hypothesis that the model is static, that is $H_0 = \kappa_1 = ... = \kappa_P = 0$, is equivalent to a portmanteau Q-test based on the correlogram of the estimated probabilities, $\xi_1(y_t)$, $t = 1, .., T$; see Harvey (2013, section 2.5). The critical values are taken from a χ^2_P distribution.

An LM test against level dynamics in the i-th regime may be constructed from the equation

$$\mu_{i,t|t-1} = \omega_{\mu i} + \kappa_{i1} u_{i,t-1} + ... + \kappa_{iP} u_{i,t-P}, \quad i = 1, .., K.$$

Combining the derivatives

$$\frac{\partial \mu_{i,t+1|t}}{\partial \kappa_{ij}} = u_{i,t-j}, \quad j = 1, ..., P,$$

with the score for location in (9) gives

$$\frac{\partial \ln f_{t|t-1}}{\partial \kappa_{i,j}} = \frac{\partial \ln f_{t|t-1}}{\partial \mu_{i,t|t-1}} \frac{\partial \mu_{i,t+1|t}}{\partial \kappa_{ij}} = u_{i,t} u_{i,t-j}, \quad i = 1, .., K, \quad j = 1, .., P,$$

where

$$u_{i,t} = \xi_i(y_t) \frac{\partial \ln f_{i,t}}{\partial \mu_i}, \quad t = 1, ..., T, \quad i = 1, 2, .. \tag{14}$$

When the test is against dynamics in the $i-th$ location only, and $\mu_j, j \neq i$ 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 $c_i(\tau) = \sum_{t=\tau+1}^{T} u_{i,t} u_{i,t-\tau}/T$, $i = 1, ...K, \tau = 1, ..., P$, that is

$$Q_i(P) = T \sum_{\tau=1}^{P} r_i^2(\tau), \quad i = 1, 2, .. \tag{15}$$

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

Remark 3 *The $Q_i(P)$ tests assume that the transition probabilities are fixed. When the transition probabilities are dynamic, but not modeled as such, the $u'_{i,t}s$ will be conditionally heteroscedastic martingale differences (MDs). The portmanteau statistic may be adjusted to deal with this situation. Lobato et al (2001) show that*

$$Q_i^*(P) = T \sum_{\tau=1}^{P} \frac{c_i^2(\tau)}{w_i(\tau)} = T \sum_{\tau=1}^{P} \frac{(\sum_{t=\tau+1}^{T} u_{i,t} u_{i,t-\tau})^2}{\sum_{t=\tau+1}^{T} u_{i,t}^2 u_{i,t-\tau}^2}, \quad i = 1, 2, ..., \tag{16}$$

where $w_i(\tau) = (T-\tau)^{-1} \sum_{t=\tau+1}^{T} u_{i,t}^2 u_{i,t-\tau}^2$, is asymptotically distributed as χ^2_P under the MD hypothesis. The structure of the test statistic is consistent with that of the LM statistic: as in the static case, the off diagonal terms in the information matrix have expectation zero but the conditional expectations of

the diagonal terms are no longer the same.

3.2 Diagnostic tests when a dynamic model has been fitted

Let $\boldsymbol{\psi}_h$ denote the parameters in the dynamic equation for the $\xi'_{i,t|t-1}s$. Suppose a basic DAMM has been fitted. The LM test statistic for location dynamics in the $h-th$ regime can be constructed as T times the uncentred R^2 from regressing the constant unity on $u_{i,t}u_{i,t-j},\ j=1,..,P$, together with $\partial \ln f_{t|t-1}/\partial\mu_h$ and $\partial \ln f_{t|t-1}/\partial\boldsymbol{\psi}_h,\ h=1,...,K$; see Davidson and MacKinnon (1990). The asymptotic distribution is χ^2_P under the null hypothesis. Calvori et al (2017) make use of this result for DCS/GAS but develop it in a different direction.

When dynamics have been estimated within regimes, tests for omitted dynamics can be constructed by adding lagged scores to the filtering equation. Thus when (11) has been fitted to location it is augmented to give

$$\mu_{i,t+1|t} = (1-\phi_i)\omega_{\mu i} + \phi_i\mu_{t|t-1} + \kappa_{i0}u_{i,t} + \kappa_{i1}u_{i,t-1} + ... + \kappa_{iP}u_{i,t-P}. \quad (17)$$

Then

$$\frac{\partial\mu_{i,t+1|t}}{\partial\kappa_j} = (\phi_i + \kappa_{i0}u_{i,t}\frac{\partial\mu_{i,t|t-1}}{\partial\kappa_{ij}}) + u_{i,t-j}, \quad j=1,...,P.$$

This derivative can be computed recursively and in the LM test, the regressor $u_{i,t}u_{i,t-j}$ is replaced by $u_{i,t}(\partial\mu_{i,t|t-1}/\partial\kappa_j)$. The derivatives for ϕ_i and $\kappa_{i0},\ i=1,..,K$, must be included as well as those detailed in the opening paragraph.

Test for heteroscedasticity can be constructed in a similar way, as can tests of dynamics in location/scale models for non-negative variables.

If the effect of fitting dynamics to the $\xi'_i s$ and to location and/or scale, is ignored, simple $Q_i(P)$ tests may be used against serial correlation in each regime[2]. Harvey and Thiele (2016) show that this can often be a good strat-

[2]Some account could be taken of switching dynamics by using the $Q^*_i(P)'s$ of Remark

egy, particularly when the Escanciano and Lobato (2009) automatic data-driven procedure for selecting the number of lags, P, is adopted.

3.3 Residuals and PITs

The residuals

$$\begin{aligned} y_t - \mu_{t|t-1} &= y_t - \xi_{1,t|t-1}\mu_{1,t|t-1} - \xi_{2,t|t-1}\mu_{2,t|t-1} \\ &= \xi_{1,t|t-1}(y_t - \mu_{1,t|t-1}) + \xi_{2,t|t-1}(y_t - \mu_{2,t|t-1}) \end{aligned}$$

are MDs but they are not identically distributed. A better way forward is to note that the CDF and hence the PITs are obtained as

$$F(y_t) = \xi_{t|t-1}F_1(y_t) + (1 - \xi_{t|t-1})F_2(y_t).$$

If these PITs are transformed to observations with a normal distribution, a check on residual serial correlation can be made using their correlogram.

4 Score-driven models for directional data

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

$$f(y) = \frac{1}{2\pi I_0(\upsilon)} \exp\{\upsilon \cos(y - \mu)\}, \qquad -\pi \leq y, \mu < \pi, \quad \upsilon \geq 0, \tag{18}$$

where $I_k(\upsilon)$ denotes a modified Bessel function of order k, μ is the directional mean and υ is a non-negative concentration parameter that is inversely related to scale. When $\upsilon = 0$ the distribution is uniform whereas y is approximately $N(\mu, 1/\upsilon)$ for large υ. The ML estimator of location, μ, is the directional mean, $\overline{y}_d$; see Mardia and Jupp (1999). A class of general circu-

3.

lar distributions is described in Jones and Pewsey (2005). The cardioid and wrapped Cauchy are special cases.

4.1 Dynamic direction

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

$$y_t = z_t \bmod(2\pi) - \pi, \qquad t = 1, ..., T; \tag{19}$$

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

$$z_t = \mu_{t|t-1} + \varepsilon_t, \qquad t = 1, ..., T, \tag{20}$$

where the $\varepsilon_t's$ are independent and identically distributed (IID) random variables from a standardized circular distribution with location zero. The basic filter is

$$\mu_{t+1|t} = (1 - \phi)\omega + \phi\mu_{t|t-1} + \kappa u_t, \tag{21}$$

where the forcing variable, u_t, is defined as being (proportional to) the conditional score for location. A defining property of a (continuous) circular distribution is that it satisfies the periodicity condition $f(y \pm 2\pi k; \theta) = f(y; \theta)$, where k is an integer and θ denotes parameters. Provided the derivatives of the log-density with respect to the elements of θ are continuous, they too are circular in that the periodicity condition is satisfied. The distribution of $y_t - \mu_{t|t-1}$ in a model defined by (20), (21) and (19) is therefore the same as that of $z_t - \mu_{t|t-1}$ 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 $\varepsilon_t \sim vM(0, \upsilon)$ in (20),

the score is

$$u_t = \upsilon \sin(z_t - \mu_{t|t-1}) = \upsilon \sin(y_t - \mu_{t|t-1}), \qquad u_t \sim IID(0, A(\upsilon)/\upsilon). \quad (22)$$

The general continuous circular distribution proposed by Jones and Pewsey (2005) has a score that is equal to $\sin(y_t - \mu_{t|t-1})$ multiplied by a factor that depends on $\cos(y_t - \mu_{t|t-1})$, so, like (22), it is clearly invariant to wrapping as well as being IID.

The model is strictly stationary when $|\phi| < 1$ in (21). Harvey et al (2019) derive the asymptotic distribution of the ML estimators of ϕ, κ and μ for the stationary vM model. When ϕ is known to be one, the asymptotic distribution of the ML estimators of κ and δ may be similarly obtained.

Remark 4 *Note that wrapping $\mu_{t|t-1}$ changes the model unless $\phi = 1$. However, it may be useful to wrap the $\mu'_{t|t-1}s$ for the purpose of plotting on a graph with the observations.*

4.2 Tests

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 as in (15). For a vM distribution with $\upsilon > 0$, the scores under the null hypothesis of constant location are proportional to the sines of the angular observations measured as deviations from their directional mean, $\overline{y}_d$. Hence the sample autocorrelations are

$$r_c(\tau) = \frac{\sum \sin(y_t - \overline{y}_d) \sin(y_{t-\tau} - \overline{y}_d)}{\sum \sin^2(y_t - \overline{y}_d)}, \qquad \tau = 1, 2, ..., \qquad (23)$$

which correspond to the circular autocorrelations in Jammalamadaka and SenGupta (2001, p176-9). The limiting distribution when the observations are IID is standard normal, that is$\sqrt{T} r_c(\tau) \rightarrow N(0, 1)$; see Brockwell and Davis (1991, Theorem 7.7.2). When the Q-statistic in the portmanteau test is

based on the first P sample autocorrelations, it is asymptotically distributed as χ^2_P under the null hypothesis of serial independence.

4.3 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 (20) is distributed as $vM(0, \upsilon_{t|t-1})$ with the dynamics dependent on the score wrt $\upsilon_{t|t-1}$, that is

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

The scores are a MD with mean zero and variance $1-A(\upsilon_{t|t-1})^2-A(\upsilon_{t|t-1})/\upsilon_{t|t-1}$. An exponential link function can be used to ensure the concentration remains positive. Thus

$$\upsilon_{t|t-1} = \exp(\zeta_{t|t-1}).$$

The first-order dynamic model for $\zeta_{t|t-1}$ is then

$$\zeta_{t+1|t} = (1-\phi)\omega + \phi\zeta_{t|t-1} + \kappa u_t^{\zeta}, \tag{24}$$

where

$$u_t^{\zeta} = \upsilon_{t|t-1}u_t^{\upsilon} = \exp(\zeta_{t|t-1})[\cos(y_t - \mu_{t|t-1}) - A(\upsilon_{t|t-1})]. \tag{25}$$

In the results for heteroscedasticity, the constant term is reported as an estimate of $\exp(\omega) = \upsilon$.

For small deviations, $\cos(y_t - \mu_{t|t-1}) \simeq 1 - (y_t - \mu_{t|t-1})^2/2$ so u_t^{υ} is essentially quadratic as in the GARCH model that is widely used in financial econometrics. It works in the opposite direction from GARCH in that large deviations have a negative effect because they reduce concentration. When $\upsilon_{t|t-1}$ is large, so that the vM is close to being Gaussian, $\cos(y_t - \mu_{t|t-1}) - A(\upsilon_{t|t-1}) \simeq -(y_t - \mu_{t|t-1})^2/2$ because $A(\upsilon_{t|t-1}) \simeq 1$. With the exponential link function, the model is similar in structure to the exponential GARCH (EGARCH) model; see Harvey (2013, ch 4). When $\upsilon_{t|t-1}$ is

large, $u_t^{\zeta} \simeq -(y_t - \mu_{t|t-1})^2 \exp(\zeta_{t|t-1})/2$. Defining $\upsilon_{t|t-1} = \exp(-\vartheta_{t|t-1})$ gives a closer link to EGARCH because

$$u_t^{\vartheta} = \exp(-\vartheta_{t|t-1})[A(\upsilon_{t|t-1}) - \cos(y_t - \mu_{t|t-1})]. \tag{26}$$

The score-driven heteroscedastic model is still observation-driven; its likelihood function is

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

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

4.4 Static mixture model

In a static directional mixture model the conditions in (2) for location and concentration are

$$\widehat{\mu}_i = \arctan \left[\frac{\sum_{t=1}^{T} \xi_i(y_t) \sin y_t}{\sum_{t=1}^{T} \xi_i(y_t) \cos y_t} \right], \quad i = 1, 2, ... \tag{27}$$

where $\widehat{\mu}_i$ is the estimated directional mean in regime i and

$$\widehat{\upsilon}_i = A^{-1} \left[(1/T\xi_i) \sum_{t=1}^{T} \xi_i(y_t) \cos(y_t - \widehat{\mu}_i) \right] = A^{-1}(R_i), \quad i = 1, 2, ..., \tag{28}$$

where R_i is the term in square brackets and ξ_i is given by (3). Approximations for $A^{-1}(R)$ can be found in Mardia and Jupp (2001, p 85-6). The estimates obtained by iterating these equations may be fed into the EM algorithm, but numerical optimization of the log-likelihood with respect to the $\upsilon_i's$ and other parameters will yield ML estimates directly.

A test that the probability of being in the $i-th$ regime is dynamic is

constructed from the correlogram of the estimates of the probabilities $\xi_i(y_t)$. A test against dynamics in the level of the $i-th$ regime is based on the correlogram of $\xi_i(y_t)\sin(y_t-\widetilde{\mu}_i)$, $i=1,...,K$, that is

$$r_{ic}(\tau)=\frac{\sum\xi_i(y_t)\sin(y_t-\widetilde{\mu}_i)\xi_i(y_{t-\tau})\sin(y_{t-\tau}-\widetilde{\mu}_i)}{\sum\xi_i(y_t)\sin^2(y_t-\widetilde{\mu}_i)},\qquad \tau=1,2,.... \quad (29)$$

Note that although the score is $\upsilon_i\xi_i(y_t)\sin(y_t-\overline{y}_i)$, the concentration parameter, υ_i, can be dropped as it cancels out.

Remark 5 *The arithmetic mean of $\xi_i(y_t)\sin(y_t-\widetilde{\mu}_i)$ is zero since if it were not, its directional mean would not be zero; see also Remark 10 in Harvey (2013, p 53).*

4.5 Dynamic mixture models

Dynamic mixture models with dynamics in the regimes may be set up as outlined in Section 2. The variable w_t driving the switching equation, (8), is obviously circular so invariance to wrapping is retained. The one-step ahead forecast in a dynamic mixture model with dynamic regimes is

$$E_T(y_{T+1})=\mu_{T+1|T}=\xi_{1,T+1|T}\mu_{1,T+1|T}+\xi_{2,T+1|T}\mu_{2,T+1|T}.$$

When a forecast of an observation, that is $\mu_{T+1|T}$, falls outside the range it can be reset, so that $\widetilde{y}_{T+1|T}=\mu_{T+1|T}\bmod 2\pi-\pi$ gives $\widetilde{y}_{T+1|T}$ in the range $[-\pi,\pi)$. Multi-step forecast can be computed recursively and the distributions of future observations can be simulated.

5 Application to directional data from Galicia

The Galicia data was described in the Introduction. Here we report the results from fitting models to hourly[3] wind direction. The parameter estimates are shown in Table 1. The estimates of the asymptotic standard errors, obtained from the numerical Hessian, are shown in brackets. When heteroscedastic models are fitted, the constant term for speed is reported (in the second row of the first and last groups) as an estimate of the level of the logarithm of concentration.

Diagnostic test statistics for assessing residual serial correlation in different components are shown in Table 2. However, the tests should not be treated formally because, as indicated earlier, the distribution is affected by the estimation of parameters. Furthermore, because the sample size is large, with $T = 744$, Q-statistics based on relatively small sample autocorrelations may be statistically significant. Having said that, some of the differences observed between models convey a very strong message about which are most effective.

5.1 One regime

The sample circular autocorrelation functions (CACFs) for the series shows very strong autocorrelation. If the possibility of more than one regime is ignored and a basic first-order model fitted, the result is $\widetilde{\phi} = 1.0$, $\widetilde{\kappa} = 0.19$ and $\widetilde{\upsilon} = 4.68$. The maximized log-likelihood is $\ln L = -520.4$. However, the residual CACF shows there is considerable serial correlation remaining. Furthermore the fit, as measured by dispersion (circular variance), is no

[3]Fitting regime switching models is best done over the full range of the sample and the minute data set is very large. It is difficult to see the bigger picture but once suitable models have been identified with hourly data estimating them with minute data is a possibility.

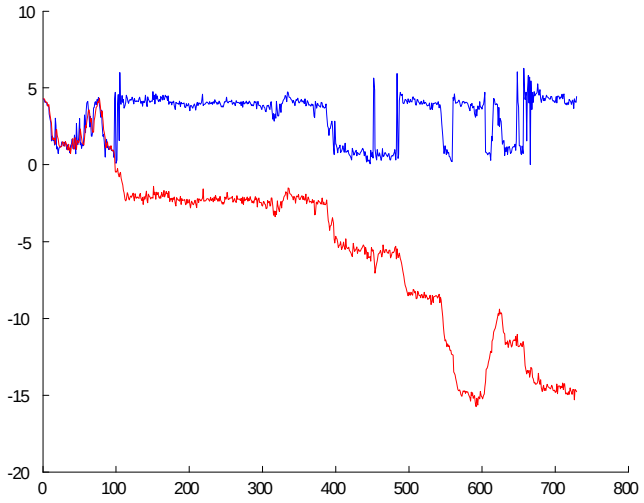

Figure 2: Filtered wind direction in Galicia from one regime heteroscedastic model

better than that of a random walk in which $\mu_{t|t-1} = y_{t-1}$; see Mardia and Jupp (2000, pp 18-19, 30) and HHT(2019). Adding heteroscedasticity gives a better fit, but strong serial correlation remains.

Figure 2 plots the evolution of the filtered location from the heteroscedastic model (the plot from the static model is similar). The fact that ϕ is estimated to be unity allows the location to travel round the circle several times. Figure 3 shows the corresponding wrapped filter; it is clear that the model is slow to adapt to a change in regime.

5.2 Static mixture

Although the static mixture model can be ruled out from the correlogram of the raw data, it is nevertheless informative about the presence of regimes. The parameter estimates obtained by the EM iterative procedure described in sub-section 2.1 were $\mu_1 = 4.06$, $\mu_2 = 1.08$, $\upsilon_1 = 12.78$, $\upsilon_2 = 2.00$ and

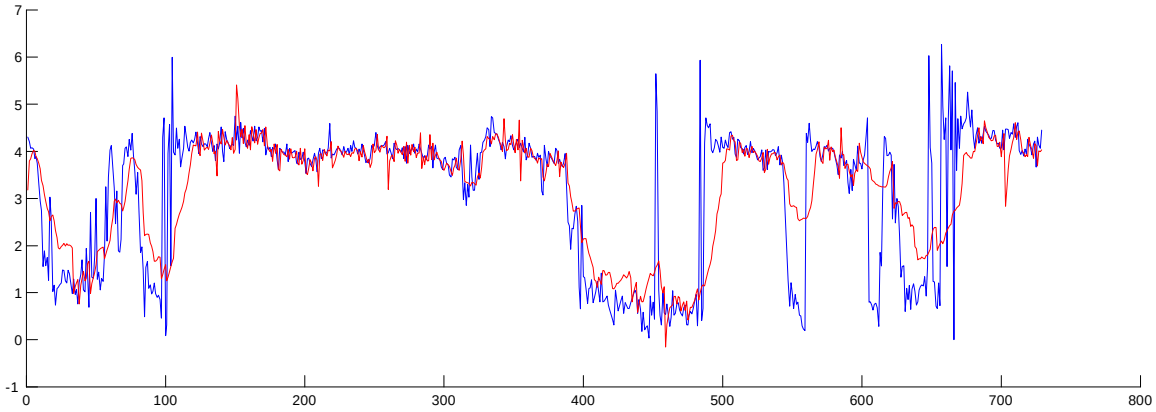

Figure 3: Wrapped filtered wind direction in Galicia from one regime heteroscedastic model. Note that some observations would appear less extreme if 2π were added or subtracted.

$\xi = 0.67$. Taking these as initial values for numerically optimizing the log-likelihood function gave $\widetilde{\mu}_1 = 4.05$, $\widetilde{\mu}_2 = 1.06$, $\widetilde{\upsilon}_1 = 12.69$, $\widetilde{\upsilon}_2 = 2.09$ and $\widetilde{\xi} = 0.66$, with $\ln L = -820.7$, which is far lower than the for the one regime dynamic model. The plot of $\xi(y_t)$ in Figure 4 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 $\xi_i(y_t)'s$ indicates that a fairly persistent first-order filter, as in (8), is likely to give a good fit. The 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.

5.3 Dynamic mixture: pure DAMM

Although the tests indicate dynamics within each regime, it is useful to begin by fitting a pure DAMM, that is one without dynamics within regimes. In the directional DAMM, the probability of being in the first regime, $\xi_{t|t-1}$,

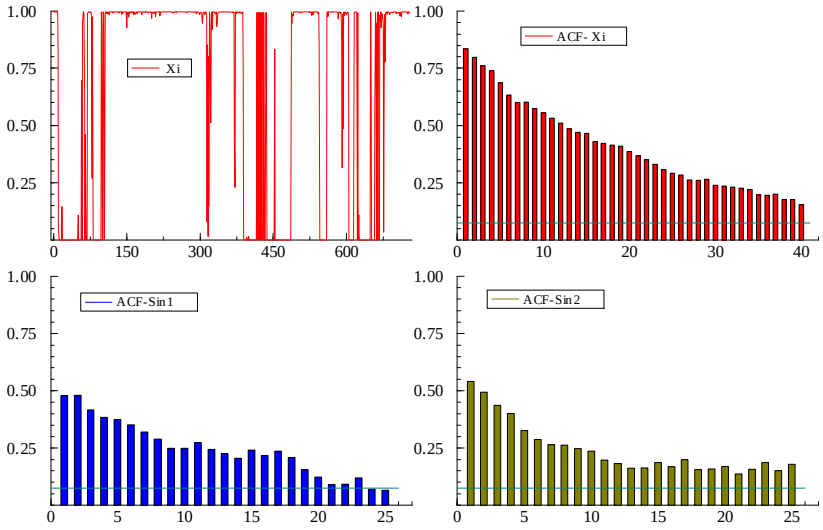

Figure 4: Regime probabilities and ACFs (from initial EM estimates) for static mixture model of wind direction in Galicia

comes from a dynamic equation for $\gamma_{t|t-1}$ as in (8).

The estimates of μ_1, μ_2, ω_1 and ω_2 are similar to the estimates found for the static mixture model except that the estimate of ξ is, at 0.89, somewhat higher than the static estimate of 0.66. The log-likelihood is -487.6 so there is a clear improvement over both the static mixture model and the single regime dynamic model for location. However, it does not beat the single regime model when both location and scale are dynamic.

The diagnostic test based on the switching residuals indicates that there are no omitted dynamics in the regime-switching equation. However, the Q-statistics for location dynamics are still highly significant in both regimes; indeed the correlograms are not dissimilar from those in Figure 4.

The finding that the pure DAMM model is inadequate is important because most, if not all, of the research in this area has been restricted to pure Markov switching models, that is HMMs.

5.4 Dynamic mixture model with dynamic regimes

As regards the regimes, the first-order DCS models are as in (11) with $u_{i,t} = \xi_{i,t|t}\upsilon_{i,t|t-1}\sin(y_t - \mu_{i,t|t-1})$. The dynamics in locations are fairly persistent in both regimes, but the concentration in regime 1 is much higher than in regime 2. The mean of location in the second regime has increased to 3.16 but if it is constrained to its value for the static model, that is 1.03, the log-likelihood is much lower at -279.34 as opposed to -241.4 for the unconstrained model. It seems that the second regime is less clearly identified than the first. 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 the model is extended to allow for heteroscedasticity, the last line in the table shows there is a further improvement in goodness of fit and the level in regime 2 falls to 2.24. On the other hand the underlying probability of being in regime 1, ξ, rises from 0.77 to 0.88. The filtered location shown in Figure 5 tracks the observations quite well, although there are some discrepancies near the beginning and end.

6 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 next section makes the extension to a bivariate regime switching model.

6.1 Weibull-von Mises (Abe-Ley) distribution

The joint distribution proposed by Abe and Ley (2017) combines a von Mises directional distribution with a Weibull distribution. The latter is a special

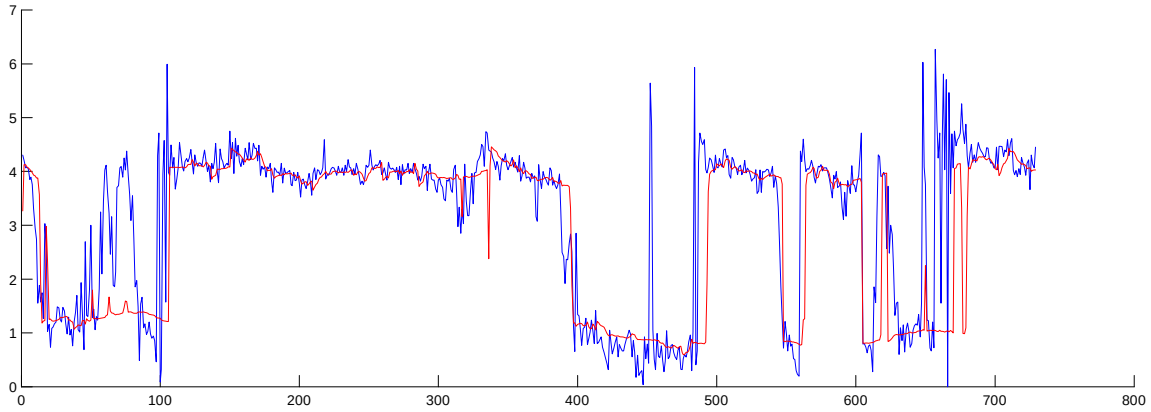

Figure 5: Filtered wind direction from the two-regime heteroscedastic vM model

case of the generalized gamma (GG) distribution, where the PDF is

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

with $\gamma, \alpha > 0$ and $-\infty < \lambda < \infty$. The gamma distribution is obtained by setting $\alpha = 1$, whereas the Weibull has $\gamma = 1$. The dynamics in location/scale are best modeled in terms of the logarithm of scale parameter so

$$x_t = \varepsilon_t \exp(\lambda_{t|t-1}),$$

where ε_t has unit scale and the forcing variable is the score

$$\frac{\partial \ln f(x_t)}{\partial \lambda_{t|t-1}} = u_t^{\lambda} = \alpha({y_t}^{\alpha} e^{-\lambda_{t|t-1}\alpha} - \gamma).$$

Remark 6 *Fat tailed distributions can be modeled with a GB2 distribution as discussed in Harvey (2013, ch 5).*

The Weibull-Sine Skewed-von Mises (WeiSSVM) proposed[4] by Abe and Ley (2017) is for a non-negative linear variable x and and circular z. The PDF is

$$\begin{aligned} f(z,x) &= \frac{\alpha \exp(-\alpha\lambda)}{2\pi \cosh(\upsilon)} S(z_t) x^{\alpha-1} \exp\{-(x/\exp\lambda)^{\alpha}(1-\tanh(\upsilon).\cos(z-\mu))\}, \quad (30) \\ -\pi &\leq z, \mu < \pi, \quad x \geq 0, \ \alpha > 0, \upsilon \geq 0, \end{aligned}$$

where $\exp(\lambda)$ is the scale, φ, for the linear variable and υ is a parameter that determines concentration for the circular variable. As in Jones and Pewsey (2005), $S(z_t) = 1 + \varsigma \sin(z_t - \mu)$, $|\varsigma| \leq 1$, skews the distribution. This skewing term will be dropped here to simplify the exposition. The result is the Weibull-von Mises (WeiVM) distribution. The distribution proposed by Johnson and Wehrly (1978), and used in Garcia-Portugues et al (2013), is a special case of the WeiVM with $\alpha = 1$.

6.2 Dynamic model

In the dynamic score-driven cylinder model the parameters μ and λ change over time. When z_t is defined over the whole real line, it may be wrapped, as in (19), to give $-\pi \leq y_t < \pi$. When the dynamic equations are driven by scores, as in the univariate model of sub-section 4.1, the log density of the conditional WeiVM distribution is

$$\begin{aligned} \ln f(y_t, x_t) &= \ln(\alpha/2\pi) - \ln\cosh(\upsilon) \ - \alpha\lambda_{t|t-1} \ + (\alpha-1)\ln x_t \quad (31) \\ &\quad -(x_t/\exp\lambda_{t|t-1})^{\alpha}(1-\tanh(\upsilon)\cos(y_t - \mu_{t|t-1})), \end{aligned}$$

with $\mu_{t|t-1}$ still defined over the whole real line. The conditional scores are

$$\frac{\partial \ln f}{\partial \mu_{t|t-1}} = u_t^{\mu} = \tanh(\upsilon)(x_t/\exp\lambda_{t|t-1})^{\alpha}\sin(y_t - \mu_{t|t-1}) \quad (32)$$

[4] Abe and Ley (2017) have $\beta = \exp(-\lambda)$ and $\kappa = \upsilon$.

and

$$\frac{\partial \ln f}{\partial \lambda_{t|t-1}} = u_t^{\lambda} = \alpha(x_t/\exp\lambda_{t|t-1})^{\alpha}(1-\tanh(\upsilon)\cos(y_t-\mu_{t|t-1}))-\alpha. \quad (33)$$

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.

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

$$\upsilon(x_t) = \tanh(\upsilon)(x_t/\exp\lambda_{t|t-1})^{\alpha}, \quad (34)$$

so the more x_t exceeds its expected value, the higher the concentration. Thus (32) can be written as

$$u_t^{\mu} = \upsilon(x_t)\sin(y_t-\mu_{t|t-1}). \quad (35)$$

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

$$\varphi(y_t) = (1-\tanh(\upsilon)\cos(y_t-\mu_{t|t-1}))^{-1/\alpha}\varphi_{t|t-1}, \quad (36)$$

where $\varphi_{t|t-1} = \exp(\lambda_{t|t-1})$ is the scale conditioned only on past observations. Substituting in (33) gives

$$u_t^{\lambda} = \alpha[(x_t/\varphi(y_t))^{\alpha}-1] \quad (37)$$

When y_t is close to $\mu_{t|t-1}$ it will boost the effect of x_t.

The filters for $\mu_{t|t-1}$ and $\lambda_{t|t-1}$ are driven by their scores, u_t^{μ} and u_t^{λ}

respectively, so for first-order dynamics

$$\begin{aligned} \mu_{t+1|t} &= (1-\phi_\mu)\mu + \phi_\mu \mu_{t|t-1} + \kappa_\mu u_t^\mu \\ \lambda_{t+1|t} &= (1-\phi_\lambda)\lambda + \phi_\lambda \lambda_{t|t-1} + \kappa_\lambda u_t^\lambda. \end{aligned} \tag{38}$$

Remark 7 *The asymptotic distribution can be derived for the model when μ and λ, but not υ, are dynamic.*

6.3 Heteroscedasticity

As the model stands, concentration, $\upsilon(x_t)$, changes only with x_t, depending on whether x_t it 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 $\upsilon(x_t)$ based on information at time $t-1$ is

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

Thus the prediction of $\upsilon(x_t)$ 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 υ be dynamic. The score with respect to this new dynamic parameter, denoted, $\upsilon_{t|t-1}$, is

$$u_t^\upsilon = (x_t/\exp\lambda_{t|t-1})^\alpha[1-\tanh^2\upsilon_{t|t-1}]\cos(y_t-\mu_{t|t-1}) - \tanh\upsilon_{t|t-1}. \tag{39}$$

The score u_t^υ is very close to that of $\lambda_{t|t-1}$, in (33), 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 $\tanh\upsilon_{t|t-1}$ replacing $\tanh\upsilon$. Using the filter

for $\upsilon_{t|t-1}$ now gives

$$E_{t-1}\upsilon(x_t) = E_{t-1}\tanh(\upsilon_{t|t-1})(x_t/\exp\lambda_{t|t-1})^\alpha = 0.5\sinh(2\upsilon_{t|t-1}). \tag{40}$$

The heteroscedastic dynamic model includes an equation for $\vartheta_{t|t-1} = -\ln\upsilon_{t|t-1}$. Thus the second equation in (38) is replaced by

$$\begin{aligned} \lambda_{t+1|t} &= (1-\phi_\lambda)\lambda + \phi_\lambda\lambda_{t|t-1} + \kappa_\lambda u_t^\lambda + \kappa_{\delta\lambda}u_t^\vartheta \\ \vartheta_{t+1|t} &= (1-\phi_\xi)\delta + \phi_\delta\delta_{t|t-1} + \kappa_\delta u_t^\vartheta + \kappa_{\lambda\delta}u_t^\lambda, \end{aligned} \tag{41}$$

where $u_t^\vartheta = -\upsilon_{t|t-1}u_t^\upsilon$ and u_t^λ and u_t^ϑ are included in both the last two equations for generality.

Remark 8 *An expression for the circular-linear correlation can be found in Abe and Ley (2017, p 96). Their Figure 2a shows how it varies with concentration.*

Remark 9 *As in univariate models the scores can be used to detect residual serial correlation.*

Remark 10 *The information matrix for μ, λ and υ is*

$$\mathbf{I}\begin{pmatrix}\mu\\ \lambda\\ \vartheta\end{pmatrix} = \begin{pmatrix}\cosh(2\upsilon) & 0 & 0\\ 0 & \alpha^2 & \alpha\upsilon\tanh\upsilon\\ 0 & \alpha\upsilon\tanh\upsilon & \upsilon^2(1+\tanh\upsilon^2)\end{pmatrix}; \tag{42}$$

see Appendix. This raises the issue of whether to pre-multiply the scores in the dynamic equations by the inverse[5] *of* $\mathbf{I}$.

Remark 11 *Abe and Ley (2017, p 96-7) give a generalization*[6]*, the GGSSVM, in which the generalized gamma distribution, denoted GG, replaces the Weibull;*

[5] Or even the inverse based on information at time t.

[6] Note that they have α replacing our $\alpha\gamma$ and γ replacing α.

the circular marginal distribution is the Jones-Pewsey distribution. A dynamic score model can again be formulated.

6.4 No observations on direction

When there is no wind, it has no direction. Similarly animals may be not be moving because they are eating or sleeping; see Zucchini et al (2016, pp 229-42). In such cases $x_t = 0$ and so the model gives $\upsilon(x_t) = 0$ which implies the (unobserved) wind direction is distributed uniformly[7]. It is evident from (35) 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. This is not the case for the scale of the linear variable because $u_t^{\lambda} = -\alpha$ and the concentration score where $u_t^{\upsilon} = -\tanh \upsilon_{t|t-1}$.

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.

6.5 Forecasts

Forecasts are based on information at T so for $T + 1$ we plug $\mu_{T+1|T}$, $\upsilon_{T+1|T}$ 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

$$f_T(y_{T+1}) = \frac{1}{2\pi} \frac{1 - \tanh^2(\upsilon_{T+1|T}/2)}{1 + \tanh^2(\upsilon_{T+1|T}/2) - 2\tanh(\upsilon_{T+1|T}/2)\cos(y_{T+1} - \mu_{T+1|T})}, \tag{43}$$

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

[7] When there is skewing, that is $\varsigma \neq 0$, the uniform is replaced by a circular cardioid with location at $\mu + \pi/2$ and concentration ς

variable is given in Abe and Ley (2017, p 94) as

$$f(x_{T+1}) = V_{T+1|T}(x_{T+1}) \frac{\alpha}{e^{\lambda_{T+1|T}}} \left(\frac{x_{T+1}}{e^{\lambda_{T+1|T}}} \right)^{\alpha-1} \exp\left(-(x_{T+1}/e^{\lambda_{T+1|T}})^{\alpha} \right), \tag{44}$$

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

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

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

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

where $P_\nu(.)$ in the normalizing term is the associated Legendre function of the first kind with degree ν and order zero. Except for the nomalizing term $V_{T+1|T}(x_{T+1})$, the form of (44) 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.

6.6 Switching cylinders

DAMMs can be applied to multivariate series as in Catania (2019, eq 3). In a bivariate model the switching filter for $\xi_{t|t-1}$ depends on the joint PDF $f(y_t, x_t)$. All parameters, including those that are fixed, such as α, are regime dependent.

In the Galicia example 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,

(5), is taken to be wrapped Cauchy, as in (43), and the same density is used in the contemporaneous probability equation (10).

Remark 12 *When switching depends only on direction, it raises the question of whether to model the dynamic scale parameters for speed in terms of one regime or two. If we decide on the former, the score for* $\lambda_{t|t-1} = \lambda_{i,t|t-1}$, $i = 1, 2$, *is the sum of the individual scores*

$$\begin{aligned} \frac{\partial \ln f_{t|t-1}}{\partial \lambda_{t|t-1}} &= \frac{\partial \ln f_{t|t-1}}{\partial f_{t|t-1}} \sum \frac{\partial f_{t|t-1}}{\partial f_{i,t|t-1}} \frac{\partial f_{i,t|t-1}}{\partial \ln f_{i,t|t-1}} \frac{\partial \ln f_{i,t|t-1}}{\partial \lambda_{t|t-1}} \\ &= \sum \xi_{i,t|t} \frac{\partial \ln f_{i,t|t-1}}{\partial \lambda_{t|t-1}} = \sum u_{it}^{\lambda}, \qquad i = 1, ...K, \end{aligned} \tag{45}$$

where, from (33),

$$\frac{\partial \ln f_{i,t|t-1}}{\partial \lambda_{t|t-1}} = \alpha(x_t / \exp \lambda_{t|t-1})^{\alpha}(1 - \tanh(\upsilon_{i,t|t-1}) \cos(y_t - \mu_{i,t|t-1})) - \alpha. \tag{46}$$

Although there is only one model for scale in each regime, the $u_{it}'s$ *depend on the changing parameters* $\upsilon_{i,t|t-1}$ *and* $\mu_{i,t|t-1}$ *as well as on the* $\xi_{i,t|t}{}'s$.

7 Wind direction and speed in Galicia

This section reports the experience of fitting bivariate WeiVM models, with and without heteroscedasticty in direction, to the Galicia data on wind direction and velocity.

7.1 Single regime

Table 3 first shows results for the single regime WeiVM bivariate model. The dynamics for both direction and speed show high persistence with the AR parameters greater than 0.9. The estimates of the $\kappa's$ for direction are similar to those found in the univariate models. The same is true for speed. (The univariate estimates are not reported here.)

The estimate of α in the speed Weibull distribution is about 2.6 so it is not heavy-tailed[8]. There is high serial correlation in the direction score residuals but not in those for speed.

7.2 Static and dynamic mixture models

As might be expected from the univariate results, a static mixture model, in line three, fares badly with $\ln L = -3871.8$ as opposed to $\ln L = -3424.3$ for the single regime model without heteroscedasticity and $\ln L = -3406.8$ with heteroscedasticity The pure DAMM model, shown in line 4, is much better, with $\ln L = -3.444.50$ but it too fails to beat the single regime models. The Q-statistics for residual serial correlation are huge.

7.3 Dynamic mixture model with dynamic regimes

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 direction only. The results favour the second possibility, especially when the dynamics include heteroscedasticity. Thus the estimates reported in the last line of Table 3 are for the preferred model.

Figure 6 shows the filtered locations in each regime, together with the path when they are combined. The combined filter is less inclined to go all the way round the circle than in the univariate model; compare Figure 2. Fitting the model with heteroscedasticity gives a higher likelihood and a slightly smoother filtered direction. However the overall picture remains the same.

Estimating the restricted regime switching model, in which the regime dynamics depend only on direction, gives an improvement in fit. As can be seen in Table 3, $\ln L = -2.987.89$ when heteroscedasticity is included. There

[8] The exponential distribution is obtained when $\alpha = 1$.

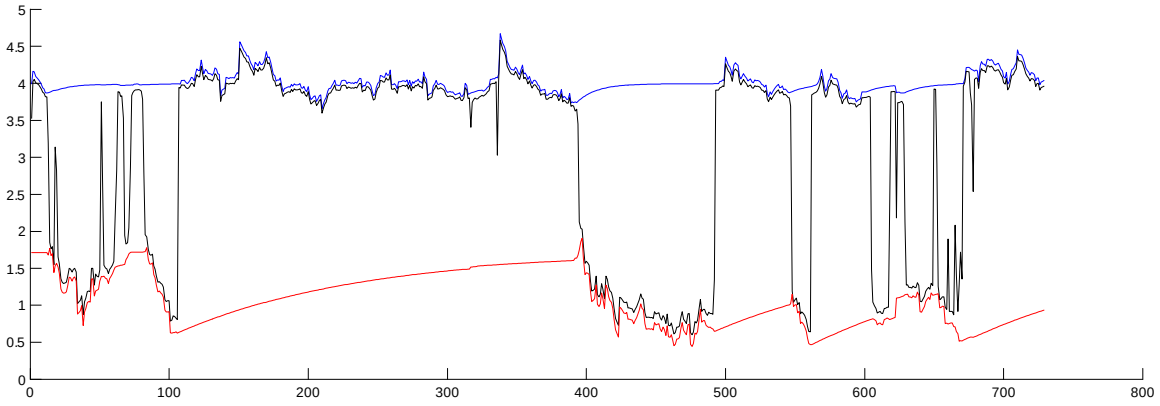

Figure 6: Filtered locations in each regime for the (unrestricted) regime switching model, without heteroscedasticity, together with the path when they are combined

is still some residual serial correlation in some of the 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 combined direction filter in Figure 7 tracks the data better than the corresponding filters for the unrestricted switching models. Although there appears to be some discrepancy towards the end, this can be explained by the direction moving round the circle; adding 2π to values of $\mu_{t|t-1}$ close to zero gives a path much closer to the observations.

8 Conclusion

Score-driven regime switching models can be extended to handle circular (directional) observations and diagnostic tests can be constructed. The models allow for changing scale (concentration) as well as changing location. The models are fitted to hourly wind direction in a site in Galicia. A hidden Markov model, without intra-regime dynamics is unable to outperform the

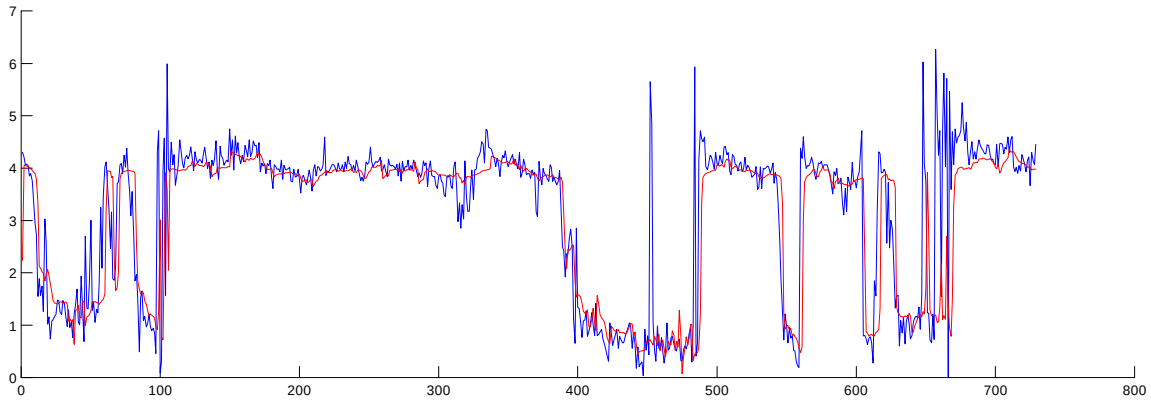

Figure 7: Filtered direction from the restricted switching regime WeiVM model with dynamic heteroscedasticity

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 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 circular and linear variables. The focus is on the Weibull-von Mises cylindrical distribution. The model can allow for changing scale as well as location. Furthermore it suggests a solution to the missing values for direction that arise when speed is zero, so there is no wind.

The bivariate models are estimated for the Galicia data. The preferred specification has dynamic location and scale 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.

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

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 paper 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. We are grateful to Anthony Davison, Jurgen Doornik, David Hendry, Stan Hurn, Peter Jupp, Rutger-Jan Lange, Christophe Ley, Ken Lindsay, Oliver Linton, Alessandra Luati, Paul Myer, Alexiy Onatski, Richard Smith, Howell Tong and others for helpful comments.

References

Abe, T. and C. Ley (2019). A tractable, parsimonious and flexible model for cylindrical data, with applications. *Econometrics and Statistics,* 4, 91-104.

Bazzi, M., F. Blasques, S. J. Koopman and Lucas, A. (2017). Time Varying Transition Probabilities for Markov Regime Switching Models, *Journal of Time Series Analysis, 38*, 458-78.

Bernardi, M. and Catania, L. (2018). Switching-GAS Copula Models With Application to Systemic Risk. *Journal of Applied Econometrics, 34*, 43-65.

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.

Calvori, F., Creal, D., Koopman, S.J. and A. Lucas (2017), Testing for

Parameter Instability in Competing Modeling Frameworks, *Journal of Financial Econometrics, 15*, 223-246.

Catania, L. (2019). Dynamic Adaptive Mixture Models with an Application to Volatility and Risk. *Journal of Financial Econometrics, 17, 1-34.*

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

Davidson, R. and J. G. MacKinnon (1990). Specification tests based on artificial regressions. *Journal of the American Statistical Association 85*, 220-227.

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

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

Frühwirth-Schnatter, S. (2006). Finite Mixture and Markov Switching Models. New York: Springer.

García-Portugués, E., Crujeiras, R.M., W. González-Manteiga (2013). Exploring wind direction and SO 2 concentration by circular–linear density estimation. *Stochastic Environmental Research and Risk Assessment, 27*, 1055–1067.

García-Portugués, E., Ana M. G. Barros, Rosa M. Crujeiras, Wenceslao González-Manteiga and J. M. C. Pereira (2014). A test for directional-linear independence, with applications to wildfire orientation and size. Stoch. Environ. Res. Risk Assess. 28, 1261-75.

Hamilton, J. D.(1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica, 57*, 357–384.

Hamilton, J. D. (1994). Time Series Analysis. Princeton.

Hamilton, J. D. (1996) Specification testing in Markov-switching time-series models. *Journal of Econometrics 70*, 127–57.

Harvey, A.C. (2013). *Dynamic Models for Volatility and Heavy Tails:*

with applications to financial and economic time series. Econometric Society Monograph, Cambridge University Press.

Harvey, A.C. and R-J. Lange (2017). Volatility Modelling with a Generalized t-distribution, *Journal of Time Series Analysis, 38*, 175-90.

Harvey, A.C., Hurn, S. and S. Thiele (2019). Modeling directional (circular) time series. *Cambridge Working Papers in Economics*,1971.

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. World Scientific Publishing Co., New York.

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

Lagona, F., Picone, M., and A. Maruotti (2015). A hidden Markov model for the analysis of cylindrical time series. *Environmetrics,* 26, 534–544.

Lobato, I., Nankervis, J. C. and N. E. Savin (2001). Testing for autocorrelation using a modified Box-Pierce Q test. *International Economic Review* **42**, 187–205.

Mardia, K. V. and P.E. Jupp (2000). Directional statistics. Wiley, Chichester.

Smith, D.R. (2008) Evaluating specification tests for Markov-switching time-series models. *Journal of Time Series Analysis, 29*, 629–52.

Zucchini,W., MacDonald, I. L. and R. Langrock (2016). Hidden Markov Models for Time Series: An Introduction using R. Monographs on Statistics and Applied Probability: Vol. 150. Boca Raton, FL: CRC Press.

1. Tables

AIC	BIC	Logl	$\omega_{\mu 1}$	$\phi_{\mu 1}$	$\kappa_{\mu 1}$	$\omega_{\vartheta 1}/\upsilon_1$	$\phi_{\vartheta 1}$	$\kappa_{\vartheta 1}$	$\omega_{\mu 2}$	$\phi_{\mu 2}$	$\kappa_{\mu 2}$	$\omega_{\vartheta 2}/\upsilon_2$	$\phi_{\vartheta 2}$	$\kappa_{\vartheta 2}$	ω_ξ	ϕ_ξ	κ_ξ	ξ_0
1,048.834	1,067.200	- 520.417	0.000	1.000	0.193	4.683	-	-										
			(0.273)	(0.001)	(0.014)	(0.049)												
858.205	885.755	- 423.103	4.294	1.000	0.152	-1.800	0.732	0.213										
			(0.349)	(0.001)	(0.009)	(0.071)	(0.041)	(0.032)										
1,651.475	1,674.434	- 820.738	4.051	-	-	12.688	-	-	1.061	-	-	2.094	-	-	0.672	-	-	0.662
			(0.014)			(0.101)			(0.055)			(0.128)			(0.086)			
989.239	1,021.381	- 487.620	4.041	-	-	11.435	-	-	1.032	-	-	2.334	-	-	2.112	0.960	4.737	0.892
			(0.014)			(0.066)			(0.049)			(0.090)			(0.825)	(0.012)	(0.723)	
504.785	555.294	- 241.393	4.139	0.915	0.013	27.606	-	-	3.163	0.988	0.264	3.365	-	-	1.205	0.958	5.821	0.769
			(0.036)	(0.025)	(0.002)	(0.066)			(0.411)	(0.005)	(0.027)	(0.073)			(0.835)	(0.012)	(0.700)	
476.137	545.013	- 223.069	4.089	0.908	0.017	2.932	0.813	0.336	2.241	0.989	0.130	1.436	0.669	0.519	1.980	0.932	6.131	0.879
			(0.040)	(0.025)	(0.003)	(0.108)	(0.068)	(0.083)	(0.188)	(0.005)	(0.023)	(0.147)	(0.088)	(0.099)	(0.674)	(0.017)	(0.908)	

Table 1: Estimated parameters from fitting the vM distribution to the Galicia Wind Direction data either with one (First and Second line) or two regimes in a DAMM model.

$\mu_{1t\|t-1}$			$\vartheta_{1t\|t-1}$			$\mu_{2t\|t-1}$			$\vartheta_{2t\|t-1}$			$\xi_{t\|t-1}$		
$Q(1)$	$Q(5)$	$Q(20)$	$Q(1)$	$Q(5)$	$Q(20)$	$Q(1)$	$Q(5)$	$Q(20)$	$Q(1)$	$Q(5)$	$Q(20)$	$Q(1)$	$Q(5)$	$Q(20)$
0.03	7.43	31.03*	128.15***	246.58***	308.81***									
19.09***	28.83***	43.04***	6.91***	37.87***	62.03***									
180.33***	699.79***	1371.29***	48.03***	182.28***	326.65***	222.24***	760.22***	1295.63***	137.07***	339.85***	413.64***	569.15***	2399.35***	5857.89***
249.05***	821.30***	1468.43***	128.63***	286.86***	377.86***	246.62***	775.45***	1389.42***	122.82***	229.51***	317.86***	0.02	18.64***	52.96***
0.18	2.14	28.60	14.64***	34.45***	60.59***	2.40	8.44	25.55	23.07***	65.50***	91.21***	1.93	44.32***	61.96***
0.72	2.81	30.62*	12.33***	18.55***	28.55*	4.08**	11.05*	31.78**	2.49	4.03	28.43	0.00	36.99***	68.08***

Table 2: Box-Ljung test for residual correlation on the fitted scores with respect to all the distribution parameters, either static or dynamic, in all the above model specifications. When the parameter is not modelled dynamically it reports instead the statistics for the simple LM test. "***" if the test results are significant with a confidence level of 0.01, "**" significant with a confidence level of 0.05, * significant with confidence level of 0.1.

AIC	BIC	Logl	$\omega_{\mu 1}$	$\phi_{\mu 1}$	$\kappa_{\mu 1}$	$\omega_{\lambda 1}/\varphi_1$	$\phi_{\lambda 1}$	$\kappa_{\lambda 1}$	$\omega_{\vartheta 1}/\upsilon_1$	$\phi_{\vartheta 1}$	$\kappa_{\vartheta 1}$	α_1	$\omega_{\mu 2}$	$\phi_{\mu 2}$	$\kappa_{\mu 2}$	$\omega_{\lambda 2}/\varphi_2$	$\phi_{\lambda 2}$	$\kappa_{\lambda 2}$	$\omega_{\vartheta 2}/\upsilon_2$	$\phi_{\vartheta 2}$	$\kappa_{\vartheta 2}$	α_2	ω_ξ	ϕ_ξ	κ_ξ	ξ_0
6,864.583	6,901.316	- 3,424.291	3.033	0.922	0.170	2.713	0.966	0.047	1.355	-	-	2.623														
			(0.072)	(0.009)	(0.014)	(0.105)	(0.007)	(0.004)	(0.028)			(0.028)														
6,833.643	6,879.560	- 3,406.822	3.179	0.939	0.158	2.702	0.974	0.049	-0.206	0.781	0.025	2.598														
			(0.132)	(0.010)	(0.012)	(0.118)	(0.007)	(0.005)	(0.038)	(0.079)	(0.008)	(0.028)														
7,761.762	7,803.087	- 3,871.881	4.054	-	-	11.402	-	-	2.001	-	-	2.136	0.884	-	-	3.883	-	-	1.705	-	-	1.732	0.995	-	-	0.730
			(0.012)			(0.051)			(0.024)			(0.033)	(0.028)			(0.104)			(0.052)			(0.054)	(0.096)			
6,910.994	6,961.502	- 3,444.497	4.048	-	-	11.315	-	-	2.084	-	-	2.175	0.924	-	-	5.392	-	-	1.378	-	-	1.792	1.035	0.851	11.574	0.738
			(0.011)			(0.045)			(0.022)			(0.029)	(0.036)			(0.025)			(0.054)			(0.052)	(0.597)	(0.019)	(1.591)	
6,112.906	6,200.147	- 3,037.453	3.994	0.930	0.010	2.218	0.978	0.021	2.104	-	-	1.272	1.712	0.992	0.042	1.490	0.772	0.110	1.620	-	-	0.730	1.355	0.926	7.226	0.795
			(0.033)	(0.025)	(0.002)	(0.079)	(0.005)	(0.002)	(0.021)			(0.029)	(0.120)	(0.003)	(0.007)	(0.025)	(0.066)	(0.018)	(0.041)			(0.037)	(0.670)	(0.016)	(0.991)	
6,064.883	6,170.491	- 3,009.441	4.079	0.942	0.008	2.325	0.975	0.019	0.664	0.906	0.011	3.619	1.106	0.995	0.017	1.871	0.770	0.088	0.193	0.954	0.114	2.083	0.978	0.927	16.219	0.727
			(0.041)	(0.019)	(0.001)	(0.079)	(0.008)	(0.002)	(0.045)	(0.031)	(0.003)	(0.028)	(0.120)	(0.006)	(0.007)	(0.033)	(0.062)	(0.015)	(0.132)	(0.023)	(0.024)	(0.044)	(0.969)	(0.020)	(6.950)	
6,099.026	6,186.268	- 3,030.513	3.992	0.918	0.010	2.264	0.972	0.028	2.102	-	-	3.511	1.489	0.970	0.059	1.348	0.921	0.058	1.572	-	-	2.138	0.922	0.970	5.401	0.715
			(0.029)	(0.026)	(0.002)	(0.102)	(0.006)	(0.003)	(0.020)			(0.031)	(0.174)	(0.015)	(0.013)	(0.187)	(0.044)	(0.015)	(0.057)			(0.051)	(0.987)	(0.010)	(1.209)	
6,021.775	6,127.384	- 2,987.888	4.034	0.946	0.007	2.276	0.980	0.017	0.564	0.930	0.017	3.407	1.754	0.989	0.063	1.345	0.954	0.037	0.326	0.910	0.064	2.062	-1.312	0.946	9.484	0.212
			(0.028)	(0.018)	(0.001)	(0.110)	(0.006)	(0.002)	(0.051)	(0.019)	(0.004)	(0.032)	(0.142)	(0.003)	(0.010)	(0.143)	(0.016)	(0.009)	(0.105)	(0.057)	(0.017)	(0.045)	(0.492)	(0.008)	(0.934)	

Table 3: Estimated parameters from fitting the WeibVM distribution to the Galicia Wind Direction and Velocity data either with one (First and Second line) or two regimes in a DAMM model. The last two lines assumes that the switching probabilty of the DAMM model is driven just by the marginal Wrapped Cauch distribution of the Wind Direction data.

	$\mu_{1t\|t-1}$			$\lambda_{1t\|t-1}$			$\vartheta_{1t\|t-1}$			$\mu_{2t\|t-1}$			$\lambda_{2t\|t-1}$			$\vartheta_{2t\|t-1}$			$\xi_{t\|t-1}$		
$Q(1)$	$Q(5)$	$Q(20)$	$Q(1)$	$Q(5)$	$Q(20)$	$Q(1)$	$Q(5)$	$Q(20)$	$Q(1)$	$Q(5)$	$Q(20)$	$Q(1)$	$Q(5)$	$Q(20)$	$Q(1)$	$Q(5)$	$Q(20)$	$Q(1)$	$Q(5)$	$Q(20)$
8.71***	55.36***	145.85***	2.70	5.84	32.20**	144.44***	506.19***	968.20***												
15.99***	44.21***	99.15***	1.13	5.72	32.53**	72.35***	298.13***	589.91***												
208.17***	734.12***	1,110.11***	205.37***	753.47***	1.042.69***	478.05***	1,968.95***	3,884.16***	101.16***	566.21***	901.82***	165.57***	577.84***	756.38***	358.68***	1,229.61***	1,582.69***	594.25***	2,561.45***	5,918.20***
207.15***	723.76***	1,085.00***	183.18***	656.69***	865.35***	475.69***	1,864.53***	3,380.22***	126.21***	605.67***	1,065.69***	162.79***	451.97***	562.27***	377.03***	1,313.55***	1,674.72***	15.29***	23.57***	84.56***
0.00	4.35	24.33	1.55	3.83	20.43	88.80***	212.02***	292.75***	0.27	27.75***	76.32***	0.03	14.99**	46.16***	46.23***	213.37***	237.71***	14.32***	15.42***	36.66**
0.63	3.72	18.59	0.08	3.05	21.02	24.56***	54.15***	79.81***	0.83	39.92***	138.90***	1.36	7.90	36.81**	7.66***	43.06***	91.45***	2.19	5.02	24.26
0.29	2.64	20.82	0.02	1.37	21.79	52.84***	130.60***	208.45***	0.88	29.90***	75.21***	0.83	15.42***	64.48***	68.54***	241.10***	257.85***	76.45***	91.24***	131.67***
1.42	6.31	19.11	0.04	2.33	18.16	15.62***	37.60***	73.32***	7.81***	54.71***	107.88***	1.89	3.97	17.90	0.09	23.74***	57.81***	67.35***	80.46***	111.98***

Table 4: Box-Ljung test for residual correlation on the fitted scores with respect to all the distribution parameters. either static or dynamic. in all the above model specifications. When the parameter is not modelled dynamically it reports instead the statistics for the simple LM test. "***" if the test results are significant with a confidence level of 0.01. "**" significant with a confidence level of 0.05. * significant with confidence level of 0.1.