DEM Working Papers

N. 2018/11

**Estimating the wrapped stable distribution via
indirect inference**

Marco Bee

UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Economia e Management

Università degli Studi di Trento

Department of Economics and Management, University of Trento, Italy.

Editors

Luciano ANDREOZZI luciano.andreozzi@unitn.it

Roberto GABRIELE roberto.gabriele@unitn.it

Technical officer

Marco TECILLA marco.tecilla@unitn.it

Guidelines for authors

Papers may be written in Italian or in English. Faculty members of the Department must submit to one of the editors in pdf format. Management papers should be submitted to R. Gabriele. Economics Papers should be submitted to L. Andreozzi. External members should indicate an internal faculty member that acts as a referee of the paper.

Typesetting rules:

1. papers must contain a first page with title, authors with emails and affiliations, abstract, keywords and codes. Page numbering starts from the first page;
2. a template is available upon request from the managing editors.

Estimating the wrapped stable distribution via indirect inference

Marco Bee

Department of Economics and Management, University of Trento - Italy

December 7, 2018

Abstract. The wrapped stable distribution is a model for non-symmetric circular data. Maximum likelihood estimation is feasible, but computationally expensive and not exact, because the density does not exist in closed form. In light of these difficulties, we develop a constrained indirect inference approach based on a skewed- t auxiliary model. To improve the finite-sample properties of the estimators, we devise a bootstrap-based estimate of the weighting matrix employed in the indirect inference program. The simulation study suggests that the indirect inference estimators are definitely preferable to the maximum likelihood estimators as concerns computing time, and approximately equivalent in terms of root-mean-squared-error.

Keywords. Indirect inference; directional statistics; stable distribution; weighting matrix.

1 Introduction

Circular data are frequently encountered in applications. Because of their peculiar statistical features, they usually require a special treatment. Consider a point with Cartesian coordinates (x, y) : its polar coordinates are $(\rho \cos(\eta), \rho \sin(\eta))$, but the value of ρ is not important if we are only interested in the direction. Hence, in this case, it is convenient to set $\rho = 1$. Put it differently, since the distance from the origin is irrelevant when only the direction matters, observations are uniquely defined by angles with respect to a reference origin. Thus, an observation of a 2-dimensional circular distribution is an angle η , corresponding to a point on the unit circle in $\mathbb{R}^2$.

Circular distributions have received some attention in the literature. The pioneering contribution by Mardia (1972) has been the trigger of the research in directional statistics. More recent comprehensive treatments are Mardia and Jupp (2000) and Jammalamadaka and SenGupta (2001).

One of the typical ways of building probability distributions for directional data is based on the idea of “wrapping” on the unit circle a distribution defined on the real line (Jammalamadaka and SenGupta, 2001,

Sect. 2.1.1). However, most distributions obtained in this way, such as the wrapped normal, wrapped Cauchy and cardioid distribution, are symmetric, whereas directional data typically are not. This is one of the motivations for the introduction of the wrapped stable distribution, which is in general not symmetric.

Despite being a model with the potential of playing a key role in many fields of application of circular statistics, the wrapped stable distribution has received very little attention. The model is first mentioned in Mardia (1972, p. 57) and then studied extensively in Pewsey (2008). Not surprisingly, estimation inherits most of the difficulties encountered with the linear stable distribution, mainly related to the lack of an explicit density function. Pewsey (2008) tackles the problem of estimation via the method of moments and of maximum likelihood (ML), stressing that the latter is computationally heavy, because the estimates are obtained via a numerical maximization of an approximation of the density.

As usual when the evaluation of a model’s density is computationally demanding, simulation-based methods are an appealing alternative to maximum likelihood. In light of this remark, in this paper we develop a novel approach to the estimation of the wrapped stable distribution via constrained indirect inference. This method has been shown to work quite well by Garcia et al. (2011) for estimation of the linear stable distribution, and given the analogies between the linear and wrapped stable models, it seems sensible to apply constrained indirect inference to the wrapped stable distribution.

Indirect inference (II) has been proposed by Gouriéroux et al. (1993). Later on, Calzolari et al. (2004) extended it to the constrained case. Roughly speaking, the II estimators are obtained by minimizing a weighted squared distance between pseudo-maximum likelihood estimates obtained from observed and simulated data, where the term “pseudo-maximum likelihood” means that the estimates are based on an instrumental model. The “optimal” choice Ξ^* of the weighting matrix Ξ was found by Gouriéroux et al. (1993), who also suggest how to estimate it.

As far as we know, in all applications of (constrained) II, the weighting matrix Ξ is set equal to the identity, and not to Ξ^* . The use of the identity matrix is justified by asymptotic considerations: when the number of parameters of the estimated and instrumental models is the same, the limiting distribution of the II estimators does not depend on Ξ . However, according to a preliminary simulation study of the algorithm’s performance, in our setup the use of $\Xi = \mathbf{I}$ produces poor estimators with respect to the estimators obtained with $\Xi = \Xi^*$. Hence, we develop a bootstrap-based estimate of the optimal weighting matrix.

The instrumental distribution is probably the most important ingredient of the II method. We use the wrapped skewed- t distribution, i.e. the wrapped version of the skewed- t distribution of Fernández and Steel (1998), as auxiliary model in the constrained II procedure. This choice is motivated

by the fact that the wrapped skewed- t and the wrapped stable have the same number of parameters and the parameters have a similar interpretation in both distributions.

The outcomes of our Monte Carlo experiments suggest that the root-mean-squared-error (RMSE) efficiency of the constrained II estimators based on the optimal weighting matrix and of the numerical MLE estimators is similar on average, but with non negligible differences across setups and parameters: in particular, the II estimator of the skewness parameter is mostly preferable to the maximum likelihood estimator of the same parameter. Furthermore, the former is strongly preferable in terms of computing time.

The main novelties of the present paper are the development of the constrained II method and of the bootstrap approach to the estimation of the weighting matrix. In addition, the Monte Carlo study of the MLEs is by itself relevant, since, to the best of our knowledge, no simulation-based analysis of the properties of MLEs of the wrapped stable distribution is available in the literature.

The rest of the paper is organized as follows. Section 2 reviews the properties of the linear and wrapped stable distributions; Section 3 outlines the theory of II and develops the constrained II approach for estimation of the wrapped stable distribution as well as the bootstrap procedure for estimating the weighting matrix; Section 4 discusses some computational issues; Section 5 illustrates the results of the simulation experiments and of an application to headings of migrating birds; finally, Section 6 concludes.

2 Background: the wrapped stable distribution

The wrapped stable distribution shares many features with the linear stable distribution. Hence, we first briefly review the latter, referring the interested reader to Zolotarev (1986), Samorodnitsky and Taqqu (1994) and Nolan (2018) for further details. Throughout the paper, we will always use the $S0$ parametrization (Nolan, 2001).

The characteristic function is given by:

$$\phi_X(t) = \begin{cases} \exp \left\{ -\gamma^\alpha |t|^\alpha \left[1 + i\beta \operatorname{sign}(t) \tan\left(\frac{\pi\alpha}{2}\right) [(\gamma|t|)^{1-\alpha} - 1] \right] + i\delta t \right\}, & \alpha \neq 1; \\ \exp \left\{ -\gamma |t| \left[1 + i\beta \operatorname{sign}(t) \frac{2}{\pi} \log(\gamma|t|) \right] + i\delta t \right\}, & \alpha = 1. \end{cases}$$

Putting $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \delta)'$, the parameter space is $\boldsymbol{\Theta} = \{(0, 2] \times [-1, 1] \times (0, \infty) \times \mathbb{R}\}$. α is the index of stability, β , γ and δ are respectively the skewness, scale and location parameters.

In general, the density of a stable random variable $X \sim St(x; \alpha, \beta, \gamma, \delta)$ does not exist in closed form, with the exception of the three special cases corresponding to the normal, Cauchy and Lévy distribution. The Cauchy and Lévy distributions are respectively obtained when $\alpha = 1$, $\beta = 0$ and

when $\alpha = 1/2$, $\beta = 1$. The case $\alpha = 2$, $\beta = 0$ gives the normal distribution. The existence of the moments is related to the value of α : $E(|X|^p)$ is only finite for $0 < p < \alpha$, so that the normal ($\alpha = 2$) is the only member of the stable family with finite variance. As $\alpha \rightarrow 0$, the distribution becomes increasingly leptokurtic.

Since the likelihood function cannot be written explicitly, MLE is doable but requires numerical evaluation of the density. When the true parameter vector is not on the boundary of the parameter space, MLEs are consistent and asymptotically normal (Nolan, 2001). Other estimation methods based on the empirical characteristic function and on the quantile function are reviewed by Nolan (2001).

Given a linear random variable X with density $f(x; \boldsymbol{\theta})$, the wrapped version of X is defined as

$$Y = X(\bmod \pi), \quad (1)$$

so that $Y \in [-\pi, \pi]$. Note that, if Y has to have support in $[-\pi, \pi]$, the modulo operation must have the sign of the dividend. Intuitively, the definition in (1) wraps $\mathbb{R}$ around the circle of unit radius, accumulating probability over all the points $x = y, y + 2\pi, y + 4\pi, \dots$. The density of Y is given by

$$g(y) = \sum_{k=-\infty}^{\infty} f(y + 2\pi k), \quad y \in [-\pi, \pi]. \quad (2)$$

In practice, the series must be truncated, so that one has to deal with an approximation error. We will discuss this issue in Section 4.1.

In the following we need two versions of (2), the wrapped skewed- t (obtained when f is the linear skewed- t) and the wrapped stable (resulting when f is the linear stable density). In the latter case we use the numerical evaluation of the linear stable density provided by the `stabledist` R package (Wuertz et al., 2016). Therefore, the computation of the wrapped stable density is subject to two types of approximation errors: one deriving from the numerical evaluation of the linear stable density, and one caused by the truncation of the series.

The II estimation approach bypasses the evaluation of the wrapped stable density, but requires the wrapped skewed- t . Hence, it is only affected by the error caused by the truncation of (2). On the other hand, maximum likelihood is based on the wrapped stable, so that it is influenced by both sources of error.

Random number generation of Y exploits the representation (1). Provided one knows how to sample X , Y is simulated by means of the following two steps:

1. sample a realization x of X ;
2. set $y = x(\bmod \pi)$.

3 Indirect inference

Indirect inference is a simulation-based estimation method introduced by Gourieroux et al. (1993). Similarly to other computationally-intensive techniques, its main strength is that it only requires the ability to sample the distribution of interest. This makes the method particularly valuable when the density and/or the moments are not explicitly available. With respect to other computer intensive methods, its desirable asymptotic properties are another advantage.

Although II has originally been devised for estimation problems with dependent observations, in the following we give a formal description restricted to the iid case, as it is the relevant setup in the present paper.

3.1 Background

Consider a random variable Y with cumulative distribution function F^θ , $\theta \in \Theta \subseteq \mathbb{R}^p$. Let M be an auxiliary random variable with density $f_M(m; \psi)$, where ψ is a vector of parameters. Given a random sample $\mathbf{y}_n = (y_1, \dots, y_n)'$ iid $\sim F^\theta$, the log pseudo-likelihood function built using the density of M and the sample $\mathbf{y}_n$ from F^θ is $\ell_n(\psi|F^\theta) = \sum_{i=1}^n \log f_M(y_i; \psi)$.

The auxiliary parameter vector $\psi(\theta) \in \Psi \subseteq \mathbb{R}^q$, $q \geq p$, is defined by the relationship $\psi(\theta) = \arg \max_{\psi} \ell_n(\psi|F^\theta)$. The function $\theta \rightarrow \psi(\theta)$ is called binding function and pseudo-true values $\psi(\theta)$ usually have a large informational content about θ . The binding function is typically unknown and must be estimated via simulation. Clearly, the auxiliary model chosen (and, consequently, the functional form of the binding map) has an impact on the properties of the estimators.

From the preceding discussion it follows that the II estimation procedure is carried out in two steps. First, it is necessary to compute the auxiliary estimate (also called *pseudo-maximum likelihood* estimate) $\hat{\psi}(\theta)$, obtained by solving the problem $\arg \max_{\psi} \ell_n(\psi|F^\theta)$ using the observed data $\mathbf{y}_n$.

Second, one simulates n_s observations $y_1, \dots, y_{n_s}$ from the true model F^θ and uses them to compute $\hat{\psi}^s(\theta) = \arg \max_{\psi} \ell_{n_s}(\psi|F^\theta)$. The II estimate $\hat{\theta}(\hat{\psi}(\theta)) \stackrel{\text{def}}{=} \hat{\theta}$ is finally obtained as solution of the problem $\min_{\theta \in \Theta} (\hat{\psi}^s(\theta) - \hat{\psi}(\theta))' \Xi^{-1} (\hat{\psi}^s(\theta) - \hat{\psi}(\theta))$, where Ξ is a weighting matrix.

In a just-identified setup, i.e. when $p = q$, the limiting distribution of $\hat{\theta}$ does not depend on the weighting matrix Ξ (Gourieroux et al., 1993, Proposition 5), and this result is typically used to justify the use of the identity matrix as weighting matrix. However, in finite samples, Ξ does affect the properties of the estimator. Gourieroux et al. (1993) show that there is an optimal choice of this matrix and provide a consistent estimator. We will discuss more thoroughly this issue in the next section.

3.2 Constrained indirect inference for the wrapped stable distribution

Building on the classical II setup, Calzolari et al. (2004) develop a theory of constrained II, which we now exploit to estimate the parameters of the wrapped stable distribution. The problem is tackled using a strategy based on the approach employed by Garcia et al. (2011) for II estimation of the linear stable distribution. In particular, we use the wrapped version of the same auxiliary model, namely the skewed- t distribution of Fernández and Steel (1998).

The density of the linear skewed- t distribution is given by

$$f_{st}(x; \omega, \lambda, \kappa, \nu) = \frac{2}{\kappa + \frac{1}{\kappa}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) (\pi\nu)^{1/2}} \lambda^{-1} \times \left[1 + \frac{(x - \omega)^2}{\nu\lambda^2} \left\{ \frac{1}{\kappa^2} \mathbb{1}_{[0, \infty)}(x - \omega) + \kappa^2 \mathbb{1}_{(-\infty, 0]}(x - \omega) \right\} \right]^{-(\nu+1)/2}, \quad (3)$$

where ν is the number of degrees of freedom and determines tail heaviness, κ is a skewness parameter and λ and ω are respectively a scale and a location parameter. Putting (3) into (2), one gets the wrapped skewed- t density $f_{WSt}(y; \omega, \kappa, \lambda, \nu) = \sum_{k=-\infty}^{\infty} f_{st}(y + 2\pi k; \omega, \kappa, \lambda, \nu)$, $y \in [-\pi, \pi)$.

Given a random sample $y_1, \dots, y_n$ from the wrapped stable distribution $Y \sim WSt(y; \alpha, \beta, \gamma, \delta)$, the pseudo log-likelihood corresponding to (3) is:

$$\ell_n = \log L_n = \sum_{i=1}^n \log f_{WSt}(y_i; \omega, \kappa, \lambda, \nu). \quad (4)$$

The wrapped skewed- t is a well-suited auxiliary model because the location, scale and skewness parameters ω , λ and κ of the wrapped skewed- t are directly related to (and have the same interpretation of) the corresponding parameters δ , γ and β of the wrapped stable distribution. On the other hand, α is linked to ν in a less direct way, and this is the reason why we need the constrained II methodology.

It is indeed true that the kurtosis is a decreasing function of α (for the wrapped stable) and of ν (for the skewed- t). Similarly, the former converges to the normal when α increases, and the latter when ν gets larger. However, the convergence to the normal takes place when $\alpha = 2$ for the former distribution, where 2 is an upper bound for the parameter value, and when $\nu \rightarrow \infty$ for the latter. Thus, the binding function $\nu = \psi(\alpha)$ goes to ∞ when $\alpha \rightarrow 2$ and therefore the inverse problem of recovering α from ν becomes more and more ill-posed as the true value of α approaches 2.

We now apply constrained II to this problem; see Calzolari et al. (2004) for details in the general case. Under the inequality constraint $\nu \leq k$, $k \in \mathbb{R}^+$, the function to be maximized is the Lagrangian objective function

$$Q_n(\mathbf{Y}) = L_n(\boldsymbol{\psi}) + \rho(k - \nu), \quad (5)$$

where $\mathbf{Y} = (\boldsymbol{\psi}', \rho)'$ and ρ is the Kuhn-Tucker multiplier associated with the constraint. The choice of k is somewhat arbitrary: given that α has an upper bound at 2, here we set $k = 2$, analogously to Garcia et al. (2011). See Lombardi and Calzolari (2008) for a thorough discussion of this issue.

Constrained II estimation of the wrapped stable parameters $\boldsymbol{\theta}$ is summarized by the following pseudo-code.

Algorithm 1. (*Indirect inference for the wrapped stable distribution*)

1. Given an iid sample $\mathbf{y}_n = (y_1, \dots, y_n)'$ from $Y \sim WSt(y; \boldsymbol{\theta})$, compute the pseudo-MLE $\hat{\boldsymbol{\psi}} = \arg \max_{\boldsymbol{\psi}} \ell_n(\boldsymbol{\psi} | F^{\boldsymbol{\theta}}) = \sum_{i=1}^n \log f_{WSt}(y_i; \boldsymbol{\psi})$, where $f_{WSt}(\cdot; \boldsymbol{\psi})$ is the density of the skewed- t distribution with parameter vector $\boldsymbol{\psi}$;
2. Compute an estimate $\hat{\boldsymbol{\Xi}}$ of the weighting matrix $\boldsymbol{\Xi}$;
3. Simulate n_s observations $y_1, \dots, y_n$ from the $WSt(y; \boldsymbol{\theta})$ distribution and use them to compute $\hat{\boldsymbol{\psi}}^s(\boldsymbol{\theta}) = \arg \max_{\boldsymbol{\psi}} \ell_{n_s}(\boldsymbol{\psi} | F^{\boldsymbol{\theta}})$;
4. The II estimator of $\boldsymbol{\theta}$ is the solution $\hat{\boldsymbol{\theta}}$ of the problem

$$\min_{\boldsymbol{\theta}} (\hat{\boldsymbol{\psi}}^s(\boldsymbol{\theta}) - \hat{\boldsymbol{\psi}})' \hat{\boldsymbol{\Xi}}^{-1} (\hat{\boldsymbol{\psi}}^s(\boldsymbol{\theta}) - \hat{\boldsymbol{\psi}}). \quad (6)$$

Some comments are in order. First, Algorithm 1 is different according to whether $\hat{\nu}$ (the pseudo-MLE of ν) is smaller or larger than 2. In the former case, the constraint is not binding, the Kuhn-Tucker multiplier $\hat{\rho}$ is equal to zero and we just perform standard unconstrained II, i.e. we match the parameters ω , κ , λ and ν estimated via real and simulated data. On the other hand, when $\hat{\nu} \geq 2$, the constraint is binding, so that $\hat{\rho} > 0$, and $\hat{\nu}^r = 2$, where $\hat{\nu}^r$ is the restricted estimate. Since $\hat{\nu}^r$ is now uninformative, it is no longer used in step 4 above, where it is substituted by $\hat{\rho}$. Thus, we now match the parameters ω , κ , λ and ρ , because ρ contains all the information about the shape of the distribution (Calzolari et al., 2004).

Second, one needs an estimate of $\boldsymbol{\Xi}$. Provided the setup is exactly identified, the optimal weighting matrix in classical (unrestricted) II is the asymptotic covariance matrix of $\hat{\boldsymbol{\Psi}}$ (Jiang and Turnbull, 2004). Gouriéroux et al. (1993, p. S112) show that it is given by the sandwich formula $\boldsymbol{\Xi}^* \stackrel{\text{def}}{=} \mathbf{J}_0 \mathbf{I}_0^{-1} \mathbf{J}_0$. In the restricted case, it is equal to $\boldsymbol{\Xi}^{*r} \stackrel{\text{def}}{=} \mathbf{J}_0^r (\mathbf{I}_0^r)^{-1} \mathbf{J}_0^r$ (Garcia et al., 2011, p. 330), where

$$\mathbf{J}_0^r = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \frac{\partial^2 L_n}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'}(\boldsymbol{\psi}_0)$$

is the Hessian of (4) and $\mathbf{I}_0^r = \sum_{\tau=-\infty}^{\infty} S_{\tau}(\boldsymbol{\theta}^0, \mathcal{Y}(\boldsymbol{\theta}^0))$, with $S_{\tau}(\boldsymbol{\theta}, \mathcal{Y}) = E(\mathbf{m}_i(\mathcal{Y}) \mathbf{m}_{i-\tau}(\mathcal{Y}) | \boldsymbol{\theta})$ and

$$\mathbf{m}_i(\mathcal{Y}) = \frac{\partial L_1}{\partial \boldsymbol{\psi}}(Y_i) - \rho \frac{\partial \nu}{\partial \boldsymbol{\psi}}.$$

In the case of independent draws, which is relevant to our setup, $\mathbf{I}_0^r$ simplifies to $\tilde{\mathbf{I}}_0^r = S_0(\boldsymbol{\theta}^0, \mathcal{Y}(\boldsymbol{\theta}^0))$. In practice, both $\tilde{\mathbf{I}}_0^r$ and $\mathbf{J}_0^r$ must be estimated using their sample counterparts, and a consistent estimator of $\boldsymbol{\psi}_0$ is required.

An appealing alternative which avoids the computation of the first two derivatives of the log pseudo-likelihood function at the expense of some computational effort is the use of non-parametric bootstrap. Given a sample $y_1, \dots, y_n$ from $Y \sim WSt(\boldsymbol{\theta})$, the bootstrap estimate of $\boldsymbol{\Xi}$ is obtained as follows:

Algorithm 2. *(Non-parametric bootstrap estimate of the weighting matrix)*

- (i) *Sample with replacement n observations $y_1^*, \dots, y_n^*$ from $y_1, \dots, y_n$;*
- (ii) *Use $y_1^*, \dots, y_n^*$ to maximize (4) and compute the quasi-maximum likelihood estimators $\hat{\boldsymbol{\psi}}(\boldsymbol{\theta}) = (\hat{\omega}, \hat{\kappa}, \hat{\lambda}, \hat{\nu})'$ of the skewed- t distribution;*
- (iii) *if $\hat{\nu} \leq k$, go to step (iv); otherwise, maximize (5), i.e. compute estimates $(\hat{\omega}, \hat{\kappa}, \hat{\lambda}, \hat{\rho})'$ under the constraint $\nu \leq k$;*
- (iv) *Repeat steps (i)-(iii) B times;*
- (v) *The bootstrap estimate of $\boldsymbol{\Xi}$ is the covariance matrix of the wrapped skewed- t parameter estimates obtained in the B replications.*

Since quasi-maximum likelihood estimation of the skewed- t distribution is reasonably fast, the computational cost of Algorithm 2 is low. Hence, we will employ this strategy in the rest of the paper.

4 Computational issues

4.1 Evaluation of wrapped densities

Actual computation of the wrapped density (2) requires the truncation of the series. By doing this we get the following approximation of (2):

$$\hat{g}(\theta) = \sum_{k=-K^*}^{K^*} f(\theta + 2\pi k), \quad \theta \in [-\pi, \pi).$$

In the present paper, f is either the stable or the skewed- t density. In the former case, Pewsey (2008) chooses K^* so as to exclude the terms whose contribution is less than 10^{-6} , linking the decision to the values of both γ and α . Nadarajah and Zhang (2017) carry out a massive sensitivity analysis for a wide range of wrapped distributions and of parameter values, finding that the relative errors of the computations of the probability and cumulative distribution functions are smaller than 10^{-10} when $K^* = 100$, leading to the conclusion that $K^* = 100$ should be used in the approximation. This

is a sort of lower bound that works well for all wrapped distributions and parameter values.

It is worth noticing that $K^* = 100$ is much larger than the values found by Pewsey (2008), at least for not too small values of α . Moreover, the evaluation of the wrapped stable density is computationally very expensive when $K^* = 100$. Hence, in our simulation experiments, we privilege Pewsey (2008)'s guidelines, which can be applied directly to the wrapped stable case. On the other hand, for the wrapped skewed- t distribution we do not know the true values of the parameters, even in the simulation experiments. To overcome this issue, we first use the wrapped stable simulated observations to compute the quasi-MLEs $\hat{\psi}$ of the skewed- t with $K = 100$ and then find the value of K^* treating $\hat{\psi}$ as the true values. The values of K^* for various parameters configurations are shown in Table 1.

Table 1: Values K^* such that the terms $f(\theta + 2\pi k)$ with $k > K^*$ in $g(\theta) = \sum_{k=-\infty}^{\infty} f(\theta + 2\pi k)$ are smaller than 10^{-6} , where $g(\theta)$ is the density of the wrapped stable or of the wrapped skewed- t distribution. In both cases the observations are simulated from a wrapped stable distribution with parameters $\alpha, \beta, \gamma = 0.5$ and $\delta = 0$.

	wrapped stable	wrapped skewed- t
Parameters values	K^*	K^*
$\alpha = 1.9, \beta = 0$	6	9
$\alpha = 1.9, \beta = 0.5$	6	9
$\alpha = 1.5, \beta = 0,$	17	24
$\alpha = 1.5, \beta = 0.5$	20	33
$\alpha = 1.2, \beta = 0$	36	66
$\alpha = 1.2, \beta = 0.5$	43	92
$\alpha = 0.8, \beta = 0$	126	185
$\alpha = 0.8, \beta = 0.5$	135	186

4.2 Numerical computation of the estimates

To compute the II estimates, the constrained maximization of (4) is carried out using the `solnp` R function of the `Rsolnp` R package (Ghalanos and Theussl, 2015), because it also estimates the Kuhn-Tucker multiplier, which is essential in algorithms 1 and 2.

The solution of (6) has to be found via constrained numerical minimization. We have found that the `bobyqa` (Bound Optimization BY Quadratic Approximation; Powell, 2009) algorithm of the `minqa` R package (Bates et al., 2014) works well. As detailed in Algorithm 2 of Section 3.2, the optimal weighting matrix Ξ^* is estimated via non-parametric bootstrap.

As for maximum likelihood, the maximization of the stable log-likelihood function is carried out by means of the `optim` R function. For both approaches we use the method of moments (Pewsey, 2008, Sect. 3.1) to initialize the algorithms, even though it may produce starting values of β and

α outside the admissible ranges. When this is the case, we use values close to the bound of the parameter space.

5 Simulation and application

5.1 Simulation experiments

The choice of the simulation design is motivated by the remark that the wrapped stable distribution is mostly used when data are spiked and/or not symmetric. If not, other circular distributions are likely to give a similar fit with fewer computational difficulties.

Hence, we include one experiment in the asymmetric case, using $\beta = 0.5$. As for the index of stability, we use $\alpha \in \{1.2, 1.5, 1.9\}$: values of α smaller than 1.2 are not considered because K^* gets very large (see Table 1) and the computational burden of the simulation experiment becomes huge. Moreover, the distribution is very spiked for $\alpha < 1.2$, more than necessary in the majority of applications. All in all, we employ the following four pairs of numerical values for α and β (in all setups we set $\delta = 0$ and $\gamma = 0.5$):

1. $\alpha = 1.9, \beta = 0$;
2. $\alpha = 1.9, \beta = 0.5$;
3. $\alpha = 1.5, \beta = 0$;
4. $\alpha = 1.2, \beta = 0$;

At each replication we simulate $n = 1000$ observations from the $WSt(\alpha, \beta, \gamma, \delta)$ distribution and estimate the parameters via II and MLE. The experiments are repeated for $n = 100$, and in both cases we use $n_s = 1000$. The number of replications is $B = 50$, which is not very large if the purpose is to obtain a precise estimate of the standard error, but is sufficient to approximately assess the relative performance of the two methods. Given the long computing times of setups 3 and 4, we do not use a larger B in any setup. The number of bootstrap replications used for estimating the matrix $\hat{\Sigma}$ is equal to 50.

Panels (a) to (d) of Figure 1 show the kernel densities of the simulated distributions of the II and ML estimators of δ, γ, β and α in Setup 1. Figure 2 displays the same graphs for Setup 2.

Looking at Figure 1 first, the II estimator of β is much better than the corresponding MLE, whose performance is quite bad: in 8 cases out of 50, $\hat{\beta}^{ML}$ is equal to ± 1 , i.e. hits one of the boundaries of the admissible interval. As for the remaining parameters, the performances are similar, with a slight advantage of $\hat{\alpha}^{ML}$ over $\hat{\alpha}^{II}$ in terms of RMSE, but not of bias. Figure 2 conveys a similar message: in particular, the simulated distribution of $\hat{\beta}^{II}$ has much more desirable features than that of $\hat{\beta}^{ML}$.

Figure 1: Kernel densities of the II and ML estimators in Setup 1. The vertical lines are the averages of the simulated distributions of the two estimators.

Numerical results and computing times for all setups are displayed in Table 2 for $n = 1000$. Whereas II is definitely more efficient from the computational point of view, estimation results are mixed. $\hat{\beta}^{II}$ is much better than $\hat{\beta}^{ML}$ when $\alpha = 1.9$, $\hat{\alpha}^{ML}$ is better than $\hat{\alpha}^{II}$ in the first three setups, and the performance of the estimators of δ and γ is similar, with some advantage for MLE. Overall, for both methods there seems to be a sort of trade-off between the estimators of β and α : if a method is good at estimating α (β), it is bad at estimating β (α). An analogous outcome can be found in Nolan (2001, Figure 2) for MLE of the linear stable distribution: the upper-right panel shows that, when α increases, the estimator of β becomes much more volatile.

In the $n = 100$ case (see Table 3), II is always better for the estimation of β (more clearly when $\alpha = 1.9$, as in the $n = 1000$ case), ML is slightly preferable for the estimation of α , δ and γ .

Finally, unreported simulation experiments with $\Xi = \mathbf{I}$ give significantly

Table 2: Simulation results. In all cases, the sample size is equal to 1000 and the number of replications is equal to 50. Computing time is measured in seconds.

True pars	Method		δ	γ	β	α	time
Setup 1	II	bias	0.0037	0.0026	0.0220	0.0269	580
		RMSE	0.0013	0.0003	0.1253	0.0036	
	MLE	bias	0.0066	-0.0006	0.0488	0.0176	2544
		RMSE	0.0008	0.0002	0.3104	0.0022	
Setup 2	II	bias	0.0030	0.0021	0.0782	0.0306	1469
		RMSE	0.0013	0.0004	0.0354	0.0031	
	MLE	bias	0.0022	-0.0012	0.1807	0.0161	4407
		RMSE	0.0007	0.0002	0.1378	0.0011	
Setup 3	II	bias	-0.0025	-0.0003	0.0416	0.0534	1601
		RMSE	0.0015	0.0006	0.0270	0.0168	
	MLE	bias	-0.00176	-0.0012	0.0003	0.0723	4949
		RMSE	0.0010	0.0003	0.0168	0.0089	
Setup 4	II	bias	-0.0302	0.0279	0.0986	0.0593	3724
		RMSE	0.0021	0.0021	0.0268	0.0169	
	MLE	bias	-0.008	-0.0035	-0.0017	0.01089	10 119
		RMSE	0.0007	0.0003	0.0107	0.0174	

Table 3: Simulation results. In all cases, the sample size is equal to 100 and the number of replications is equal to 50. Computing time is measured in seconds.

True parameters	Method		δ	γ	β	α
Setup 1	II	bias	-0.0077	-0.0200	0.3173	-0.0788
		RMSE	0.0113	0.0045	0.2675	0.0319
	MLE	bias	-0.0165	-0.0162	0.0771	-0.0071
		RMSE	0.0054	0.0020	0.2981	0.0125
Setup 2	II	bias	0.0171	-0.0130	0.0415	-0.0250
		RMSE	0.0143	0.0021	0.1401	0.0176
	MLE	bias	0.0021	-0.0055	0.0530	-0.0145
		RMSE	0.0080	0.0015	0.2727	0.0128
Setup 3	II	bias	-0.0121	-0.0168	0.1100	0.1592
		RMSE	0.0090	0.0036	0.1084	0.0526
	MLE	bias	-0.0181	-0.0151	0.1139	0.0922
		RMSE	0.0090	0.0028	0.3352	0.0426
Setup 4	II	bias	0.0383	-0.0396	-0.0343	0.0407
		RMSE	0.0444	0.0091	0.0106	0.0190
	MLE	bias	-0.0082	-0.0035	-0.0017	0.1089
		RMSE	0.0007	0.0003	0.0107	0.0174

Figure 2: Kernel densities of the II and ML estimators in Setup 2. The vertical lines are the averages of the simulated distributions of the two estimators.

worse result. Hence, the bootstrap estimate of Ξ is of crucial importance in this framework.

5.2 Application

Finally, we use the II approach developed in the preceding sections to fit the wrapped stable distribution to the same data-set analyzed in Pewsey (2008). The data set, first used in Bruderer and Jenni (1990), consists of 1827 headings of migrating birds recorded at an observational post near Stuttgart in Germany during the autumnal migration period of 1987. “Heading” is the direction, measured clockwise from north, of a bird’s body during flight. Point estimates and standard errors are displayed in Table 4. For both methods, the standard errors have been computed via non-parametric bootstrap with 50 replications.

Figure 3 shows the histogram of the data and the wrapped stable den-

Table 4: Bird data estimation results. Standard errors have been computed via non-parametric bootstrap with $B = 50$ replications. The last column shows the p -value of the Kolmogorov-Smirnov (KS) non-parametric test.

Method		δ	γ	β	α	KS p -value
II	Estimate	4.0256	0.4516	-0.7782	1.5766	0.276
	Standard error	0.0371	0.0179	0.1609	0.0766	
MLE	Estimate	4.0100	0.4635	-0.6417	1.5434	0.742
	Standard error	0.0267	0.0100	0.1072	0.0368	

sity estimated with the two methods. The numerical values of the estimated parameters are similar, and consequently the densities are almost indistinguishable. Standard errors are also comparable, although MLEs are slightly less volatile.

We have tested the null hypothesis that the empirical and estimated distribution are identical. The p -values of the non-parametric Kolmogorov-Smirnov test reported in the last column of Table 4 confirm that the distance between the empirical and the estimated distribution is not significantly different from zero for both approaches.

Figure 3: Bird data and wrapped stable densities estimated via II and MLE.

6 Conclusion

The wrapped stable distribution is a flexible model for asymmetric directional data. Since maximum likelihood estimation is complicated by the lack of a closed-form density, in this paper we develop a constrained indirect inference approach based on a wrapped skewed- t auxiliary model. Besides analyzing the role of the input parameters of the algorithm, we propose a bootstrap-based method for estimating the weighting matrix, which turns out to have a significant impact on the statistical properties of the estimators.

In terms of comparison with numerical MLE, the results of our simulation experiments convey a twofold message. Indirect inference is significantly more efficient than MLE in terms of computational cost, because the computing time of indirect inference is from 60 to 80% smaller. As of statistical efficiency (measured in terms of RMSE), II is mostly better for β and ML for α .

Maximum likelihood estimation is not trivial also for various other circular distributions, so that it may be worth applying indirect inference to these cases. The choice of the instrumental model depends on the features of the distribution to be estimated, and must therefore be tackled on a case-by-case basis. This issue remains open for further research.

Acknowledgments. I would like to thank Bruno Bruderer, Dieter Peter and Arthur Pewsey for making available the “migrating birds” data employed in Section 5.2. The R codes used for II estimation are available at <http://marcobee.weebly.com>.

References

- Bates, D., Mullen, K. M., Nash, J. C., and Varadhan, R. (2014). *Minqa: Derivative-free optimization algorithms by quadratic approximation*. R package version 1.2.3.
- Bruderer, B. and Jenni, L. (1990). Migration across the Alps. In Gwinner, E., editor, *Bird Migration: Physiology and Ecophysiology*, pages 60–67. Springer, Berlin.
- Calzolari, G., Fiorentini, G., and Sentana, E. (2004). Constrained indirect estimation. *Review of Economic Studies*, 71:945–973.
- Fernández, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371.

- Garcia, R., Renault, E., and Veredas, D. (2011). Estimation of stable distributions by indirect inference. *Journal of Econometrics*, 161(2):325 – 337.
- Ghalanos, A. and Theussl, S. (2015). *Rsolnp: General Non-linear Optimization Using Augmented Lagrange Multiplier Method*. R package version 1.16.
- Gourieroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8, Supplement: Special Issue on Econometric Inference Using Simulation Techniques:S85–S118.
- Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*. World Scientific, Singapore.
- Jiang, W. and Turnbull, B. (2004). The Indirect Method: Inference based on Intermediate Statistics - A Synthesis and Examples. *Statistical Science*, 19:239–263.
- Lombardi, M. J. and Calzolari, G. (2008). Indirect estimation of α -stable distributions and processes. *Econometrics Journal*, 11:193–208.
- Mardia, K. V. (1972). *Statistics of Directional Data*. Academic Press, London.
- Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. Wiley, New York.
- Nadarajah, S. and Zhang, Y. (2017). Wrapped: An R package for circular data. *Plos One*, 12(12):1–26.
- Nolan, J. P. (2001). Maximum likelihood estimation and diagnostics for stable distributions. In Barndorff-Nielsen, O. E., Mikosch, T., and Resnick, S. I., editors, *Lévy Processes: Theory and Applications*, pages 379–400. Birkhäuser, Boston.
- Nolan, J. P. (2018). *Stable Distributions - Models for Heavy Tailed Data*. Birkhäuser, Boston. In progress, Chapter 1 online at <http://fs2.american.edu/jpnolan/www/stable/stable.html>.
- Pewsey, A. (2008). The wrapped stable family of distributions as a flexible model for circular data. *Computational Statistics and Data Analysis*, 52:1516–1523.
- Powell, M. J. D. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. Technical Report Report No. DAMTP 2009/NA06, Centre for Mathematical Sciences, University of Cambridge, UK.

- Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall, CRC, New York.
- Wuertz, D., Maechler, M., and core team members, R. (2016). *stabledist: Stable Distribution Functions*. R package version 0.7-1.
- Zolotarev, V. M. (1986). *One-dimensional Stable Distributions*. American Mathematical Society, Providence.