Journal of Data Science,18(1). P. 78 – 100,2020
DOI:10.6339/JDS.202001_18(1).0004

On The Estimation Of The Shape Parameter Of A Symmetric Distribution

Jennifer S.K. Chan [1*], S.T. Boris Choy[2] and Stephen G. Walker[3]

[1]*Corresponding author. School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia.*

[2]*Discipline of Business Analytics, University of Sydney, NSW 2006, Australia.*

[3]*Department of Statistics and Data Science, University of Texas at Austin.*

ABSTRACT

The shape parameter of a symmetric probability distribution is often more difficult to estimate accurately than the location and scale parameters. In this paper, we suggest an intuitive but innovative matching quantile estimation method for this parameter. The proposed shape parameter estimate is obtained by setting its value to a level such that the central 1-1/n portion of the distribution will just cover all n observations, while the location and scale parameters are estimated using existing methods such as maximum likelihood (ML). This hybrid estimator is proved to be consistent and is illustrated by two distributions, namely Student-t and Exponential Power. Simulation studies show that the hybrid method provides reasonably accurate estimates. In the presence of extreme observations, this method provides thicker tails than the full ML method and protect inference on the location and scale parameters. This feature offered by the hybrid method is also demonstrated in the empirical study using two real data sets.

Keywords: Student-t; Exponential Power; Matching quantile; Tails; Consistency

1. Introduction

The paralogistic distribution is a sub-model of the generalized beta family which was introduced by McDonald (1984). The probability density function (PDF) and the cumulative distribution function (CDF) of the paralogistic distribution are given respectively by: Fitting a parametric distribution to data sometimes results in a model that agrees well with the data in high density regions, but not so in low density regions. For unimodal distributions such as the normal and Student-t, these low density regions are known as the “tails” of the distribution. One reason why a model might fit poorly in the tails is that there are fewer data points in the regions, and therefore models are often chosen based on their ability to fit data near the mode. Another reason might be that the underlying distribution is usually more complicated than the parametric model selected. It is well known that fitting the data well in the tails can significantly improve the overall model fit. This can be intuitively explained by observing most QQ-plots in which discrepancies between observed and fitted lie mostly at the tails of a distribution. Figure 5 clearly demonstrates this idea. Therefore, heavy-tailed distributions are often used because they have flexible tails whose thicknesses are determined by their shape parameters. By adjusting the shape parameter, one fits a distribution which can accommodate extreme observations and protect inferences from the distorting effects of these extreme observation on the location and scale parameters.

Parameters of a symmetric parametric model, including the shape parameter, can be estimated through the maximum likelihood (ML) method, moment method (MM), Bayesian method, etc. However, the MM may not be robust and the Bayesian method can be computational intensive. Moreover, in a full ML approach, the shape parameter is of- ten more difficult to estimate accurately than the location and scale parameters because the Newton-Raphson (NR) or Fisher scoring procedures may sometimes fail due to non-differentiable loglikelihoods, arisen when the probability density functions (PDFs) have sharp peaks. This occurs for distributions such as exponential power (EP) when the shape parameter falls inside a certain range. Even if the loglikelihoods are differentiable, the functional part involving the shape parameter can be very complicated and hence the differentiation with respect to the shape parameter may require tedious working. These difficulties may be avoided using some flexible optimization algorithms such as the R Package optim but it may be computational intensive if a full search is required. On the other hand, the Expectation Maximization (EM) type algorithms may have a slow convergence rate.

Many researches apply the idea of matching moments to matching quantiles (MQ) in inference (see, for example, Karian and Dudewicz, 1999; Small and McLeish, 1994). Matching the extreme quantiles could be attractive in mimicking the behavior of a target distribution at its ends (tails) which is controlled by its shape parameter. As these tails play an important role in describing the extreme observations, it is natural to estimate the shape parameter by matching the sample quantile of the most extreme observation to its theoretical quantile. We propose a hybrid method combining the MQ method to estimate the shape parameter, with the convenient ML method to estimate the location and scale parameters. Alternatively, Hill (1975) proposes measuring the tail thickness by approximating the tail cumulative distribution function (CDF) with a power function using extreme order statistics. This semi-parametric approach does not assume any global form of a distribution. Our approach is a full parametric approach and we estimate the tail index of a distribution not the tail CDF. We illustrate our methodology through two families of distributions with varying kurtosis; namely, the Student-t and EP distributions.

The rest of the paper is organized as follows. Section 2 introduces our proposed MQ method, provides a formal proof for its consistency property and reviews existing methods of inference under the two distributions: Student-t and EP distributions. Section 3 assesses the performance and robustness property of the proposed hybrid estimator through two simulation studies. Then in Section 4, the estimator is used to analyze two data sets and comparisons are made with some existing estimators. Finally, we conclude he paper by discussing its potential extension to the class of asymmetric distributions in Section 5.

2 Methodology

2.1 Matching quantile estimator

Suppose we have a set of independent observations $y_i, i = 1,\dots,n$ and they follow a certain distribution with location parameter μ, scale parameter σ and shape parameter θ. We propose a hybrid method to estimate $\mu, \sigma|\theta$ using some convenient methods such as the ML and estimate θ|μ, σ using our proposed MQ method. These two estimation procedures are then iterated until convergence is attained. In the MQ step, we estimate the shape parameter θ to a certain level such that the upper $1/2n$ quantile, Q of the standardized distribution for $z_i = (y_i - \mu^2)/\sigma^2$ matches the maximum absolute standardized residual $z^* = max_i |z_i|$. Note that z^* corresponds to the most extreme observation. Equivalently, we have

$$\hat{\theta} = \{\theta : Q_{\frac{1}{2n}} = S^{-1}\left(\frac{1}{2n}; 0,1,\theta\right) = z^* \}$$

where S(·) is the survival function. The idea is illustrated in Figure 1. Empirically, each data point zi accounts for 1/n portion of the distribution and so the portion from the most extreme data to the end of that side of the distribution is 1/2n. This is clearly demonstrated in Figure 1 by the proxy regularly spaced data denoted by solid circles. Essentially, θ is estimated such that the shaded area of the distribution will just cover all $\{z_i\}$ with $Q_{1/2n} = z^* = -z_{(1)}$ where $z_{(k)}$ is the k-th data value in ascending order.

2.2 Consistency

We show that our proposed MQ estimator is consistent. Assume that $Y_i, i = 1,\dots,n$, are independently and identically distributed from some probability distribution, which has a survival function

$$S(y) = P(Y > y).$$

We propose to use $Y^{(n)} = max(Y_1,\dots,Y_n)$ as the estimator for the upper quantile $Q_{1/n} = S^{-1}(1/n)$, where $S(\cdot)$ is determined by the shape parameter of the distribution under consideration. We show the consistency of the proposed estimator $\hat{Q}_{1/n} = Y$ in the following

Theorem:

For some monotone function $H(\cdot)$, the ratio $\frac{H(Y^{(n)})}{H(Q_{1/n})}$converges to 1 in probability.

Proof: Define $S_H(z) = P(H(Y) > z) = S(H(z))$ and recall we have

$$P(Y^{(n)} > y) = 1 - \{1 - S(y)\}^n.$$

Assume for now there exists a monotone function $H(\cdot)$ such that

$$nS_H((1 + \epsilon)H(Q1/n)) \rightarrow 0$$

and

$$nS_H((1 - \epsilon)H(Q1/n)) \rightarrow \infty$$

for all $\epsilon > 0$. Using the result

$$(1-\lambda_n)^n \rightarrow \begin{cases} 1 & \text{if } n\lambda_n \rightarrow 0 \\ 0 & \text{if } n\lambda_n \rightarrow \infty \end{cases}$$

and the assumption stated at the start of this proof, we have

$$P\left(H\left(Y^{(n)}\right) > (1+\epsilon)H\left(Q_{\frac{1}{n}}\right)\right) = 1 - \left\{1 - S_H\left((1+\epsilon)H\left(Q_{\frac{1}{n}}\right)\right)\right\}^n \rightarrow 0$$

and

$$P(H(Y^{(n)}) > (1-\epsilon)H(Q_{1/n})) = 1 - \{1 - S_H((1-\epsilon)H(Q_{1/n}))\}^n \rightarrow 1$$

for all $\epsilon > 0$. Putting these together, we have

$$\frac{H\left(Y^{(n)}\right)}{H(Q_{1/n})} \rightarrow 1$$

in probability. To consider the existence of $H(\cdot)$, we write

$$S(y) = e^{-H(y)},$$

where the monotone function H(y) is the cumulative hazard function. So

$$S^{-1}(t) = H^{-1}(-\ln t)$$

and hence $H(Q_{1/n}) = \ln n$. Now

$$nSH((1+\epsilon)H(Q1/n)) = nSH((1+\epsilon)\ln n) = n\exp\{-H(H^{-1}((1+\epsilon)\ln n))\}.$$

This is given by

$$n \times n^{-(1+\epsilon)} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Similarly,

$$nS_H((1-\epsilon)H(Q_{\frac{1}{n}})) = n \times n^{-1(1-\epsilon)} \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

Hence $H(\cdot)$ exists and this completes the proof.

We can now establish conditions under which the above implies $Y^n / Q_{1/n} \rightarrow 1$ in probability. For this, define S(z) = S (e) , and so for some sequence $\epsilon_n \downarrow 0$ in probability, for which we assume $\epsilon_n \log n \rightarrow 0$ in probability, we have

$$\frac{H(Y^{(n)})}{H(Q_{1/n})} \to 1$$

in probability. To consider the existence of H(·), we write

$$S(y) = \mathrm{e}^{-H(y)}$$

where the monotone function H(y) is the cumulative hazard function. So

$$S^{-1}(t) = H^{-1}(-lnt)$$

and hence $H(Q_{1/n}) = lnn$. Now

$$nS_H((1+\epsilon)H(Q\,1/n)) = nS_H((1+\epsilon)lnn) = nexp\{-H(H^{-1}((1+\epsilon)lnn))\}.$$

This is given by

$$n \times n^{-(1+\epsilon)} \to 0\, as\, n \to \infty.$$

Similarly,

$$nS_H((1-\epsilon)H(Q_{\frac{1}{n}})) = n \times n^{-(1-\epsilon)} \to \infty\, as\, n \to \infty.$$

Hence H(·) exists and this completes the proof.

We can now establish conditions under which the above implies $Y^{(n)}/Q_{1/n} \to 1$ in probability. For this, define $\widetilde{S}(z) = S^{-1}(e^{-z})$, and so for some sequence $\epsilon_n \downarrow 0$ in probability, for which we assume $\epsilon_n$logn $\to$ 0 in probability, we have

$$\frac{Y^{(n)}}{Q_{1/n}} = \frac{\widetilde{S}((1+\epsilon_n)logn)}{\widetilde{S}(logn)}$$

That $Y^{(n)} = \tilde{S}((1+\epsilon_n)logn)$ for some sequence ϵ_n is derived as follows: Now for some sequence ϵ_n going to 0 we have

$$\frac{H(Y^{(n)})}{H(Q_{1/n})} = 1 + \epsilon_n$$

and $H(Q_{1/n}) = logn$. So

$$Y^{(n)} = H^{-1}((1+\epsilon_n)logn) = S((1+\epsilon_n)logn).$$

Therefore, under the weak condition that

$$\tilde{S}'(z)/\tilde{S}(z)$$

is bounded as $z \to \infty$, we have that $Y^{(n)}/Q_{1/n} \to 1$ in probability. This in turn implies, under continuity arguments, that the $\hat{\nu} \to \nu_0$, the true parameter value, in probability. This is based on the continuity assumption, which is that if ν_n is a sequence such that

$$\frac{S_{\nu_n}^{-1}\left(\frac{1}{n}\right)}{S_{\nu_0}^{-1}\left(\frac{1}{n}\right)} \to 1$$

then $\nu_n \to \nu_0$. For example, if $S_\nu(y) = exp(-\nu y)$ then

$$\frac{S_{\nu_n}^{-1}\left(\frac{1}{n}\right)}{S_{\nu_0}^{-1}\left(\frac{1}{n}\right)} \to \frac{\nu_0}{\nu_n}$$

Now $\tilde{S}$(z) is such that $\tilde{S}(0) = 0$ and $\tilde{S}(\infty) = \infty$ and S is monotone. So, for example, assume

$$\tilde{S}(z) = \sum_{j=1}^{M} a_j z^j$$

for some positive integer M, with the (a_j) such that $\tilde{S}$ is non-negative and increasing.
Then, it is easy to show that

$$z\tilde{S}'(z)/\tilde{S}(z) \rightarrow Ma_M \; as \; z \rightarrow \infty.$$

2.3 Two distributional examples

2.3.1 Student-t Distribution

The Student-t distribution is perhaps the most popular heavy-tailed distribution used in many statistical analyses. It contains normal distribution as the limiting distribution at one end and the Cauchy distribution at the other. Its PDF and CDF are given by

$$f_t(y;\mu,\sigma^2,\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma}[1+\frac{1}{\nu}(\frac{y-\mu}{\sigma})^2]^{-\frac{\nu+1}{2}}$$

and

$$F_t(y;\mu,\sigma^2,\nu) = 1-\frac{1}{2}I_{\frac{\nu}{y^2+\nu}}(\frac{\nu}{2},\frac{1}{2})$$

respectively, where μ, σ and ν are the location, scale and shape parameters, Γ(·) is the gamma function, and Ix(a, b) is the incomplete beta function. The variance of the distribution is $\nu\sigma^2/(\nu - 2)$, $\nu > 2$ while the kurtosis is $3 + \nu/(\nu - 4)$, $\nu > 4$.

Literature on parameter estimation for the Student-t distribution has been long and rich in history. In the context of ML estimation, the parameters (μ, σ, ν) can be estimated jointly using the full ML approach (Model 1 or M1) with the NR or FS iterative algorithms. However, the convergence to the solutions depends on the model complexity and hence it cannot be guaranteed. In case of numerical issues in using the NR or FS algorithms, we recommend a general purpose optimization algorithm optim in the R package to search for the ML estimates of μ and σ. The package optim is based on Nelder-Mead (Nelder and Mead, 1965; default method), quasi-Newton and conjugate-gradient algorithms. NelderMead algorithm works reasonably well for non-differentiable functions but is relatively slow. Alternatively, Liu and Rubin (1995) propose the Expectation Conditional Maximization (ECM) algorithm, multi-cycle version of the ECM (MCECM) algorithm, and Expectation Conditional Maximization Either (ECME) algorithm. Details of these estimators are given in Appendix A. We propose the hybrid method (M2) to estimate μ and σ conditional on ν using the ML method and ν conditional on μ and σ using our proposed MQ method. The ML estimation of (μ, σ|ν) is easier and more stable than the ML estimation of (μ, σ, ν) simultaneously. Figure 2 plotsν ˆ against z* for various sample sizes.

2.3.2 Exponential Power Distribution

The EP distribution, also known as the generalized Gaussian distribution and generalized error distribution, is a good model for signal and noise in many applications of science and engineering, because it allows both positive kurtosis (leptokurtosis) and negative kurtosis (platykurtosis). The distribution has a PDF and CDF given by

$$f_{EP}(y;\mu,\sigma^2,\beta) = \frac{1}{\sigma\beta\Gamma(\frac{\beta}{2})2^{\beta/2}} \exp\left(-\frac{1}{2}\left|\frac{y-\mu}{\sigma}\right|^{2/\beta}\right)$$

and

$$F_{EP}(y;\mu,\sigma^2,\beta) = \frac{1}{2} + \frac{\mathrm{sgn}\,(y-\mu)}{2\Gamma(\frac{\beta}{2})}\gamma\left(\frac{\beta}{2},\frac{1}{2}\left|\frac{y-\mu}{\sigma}\right|^{2/\beta}\right)$$

respectively, where μ, σ and β are the location, scale and shape parameters, respectively, and $\gamma(\cdot,\cdot)$ is the lower incomplete gamma function. The PDF has a sharp peak at μ when $\beta > 1$ and hence is not differentiable at μ. For the range of β, some set $\beta \in (0, 2]$ but others consider $\beta > 0$ (Mineo and Ruggieri 2005; Choy and Walker, 2003). As the EP distribution is symmetric, all odd central moments are zero while the even central moments are given by

$$E[(Y-\mu)^r] = (2^\beta\sigma^2)^{r/2}\frac{\Gamma(\frac{(r+1)\beta}{2})}{\Gamma(\frac{\beta}{2})}, r = 2,4,\ldots$$

The distribution is leptokurtic for $\beta > 1$ and platykurtic for $\beta < 1$. The Laplace ($\beta = 2$) and normal ($\beta = 1$) distributions are two special cases of the EP distribution. Moreover, the EP distribution with $\beta > 2$ has higher kurtosis than the conventional EP distribution with $\beta < 2$.

The ML and moment estimators for the shape parameter β of the EP distribution have been studied in the literature. Since neither of them have closed form solutions, they are usually estimated numerically. Although the ML estimator is asymptotically more efficient than the moment estimator, the loglikelihood function for the ML estimator may not be differentiable. Thus, the NR algorithm does not always converge, especially when $\beta > 1$ and/or the number of observations is small (Agro, 1995). Nadarajah (2005) presents the FS algorithm to solve ML equations numerically. By taking expectation on the second order derivatives, this algorithm can solve some ML equations in which the NR algorithm fails. Another way to overcome the convergence problem is to consider a full MM method (M3) as described in Appendix B. The full ML method (M4) is also possible using some search algorithms in the R package optim. Again, we propose the hybrid method (M5) to estimate μ and σ conditional on β using the ML estimator and β conditional on μ and σ using the proposed MQ estimator. Figure 3 exhibits the relationship between $\hat{\beta}$ and z^* for various sample sizes.

3 Simulation studies

Two simulation experiments are performed to evaluate the performance and robustness properties of our proposed hybrid method and each experiment is repeated with the Student-t and EP distributions. The true values for the location and scale parameters are set to be $\mu = 5$ and $\sigma=2$ and their starting values are taken to be the moment estimates. For the shape parameter, we consider five levels for each of the Student-t and EP distributions. No starting values are needed for the shape parameters using the MQ method whereas the value of five and the moment estimate in (6) are adopted as starting values for the shape parameters in Student-t and EP distributions respectively using the full ML method.

3.1 Performance study

The first experiment studies the performance of the proposed hybrid method under two factors: the sample size and shape parameter. We simulate N = 500 independent data sets of size n = 20, 50, 100, 200, 500, 1000 observations, each at five levels of true shape parameter value using two scenarios. For each sample size and true shape parameter value, the parameter estimates and their standard error (SE) are the mean and standard deviation of estimates from the 500 simulated data sets. In scenario 1, we begin with a simplified situation using standard distributions with known $\mu = 0$ and $\sigma = 1$, to test if our proposed MQ method works well to estimate only the shape parameter θ. Then in scenario 2, non-standard distributions are adopted with unknown μ and σ to be estimated together with θ using our proposed hybrid method. Parameters are estimated iteratively using MQ method for $(\theta|\mu, \sigma)$ and ML method for $(\mu, \sigma|\theta)$. The whole procedure is implemented using the R package. These models using the hybrid method correspond to M2 and M5 for Student-t and EP distributions respectively.

3.1.1 Student-t distribution

In both scenarios, the true degrees of freedom (df) is set to $\nu = 2, 5, 10, 15$ and 20, respectively. The NR algorithm is adopted to estimate $(\mu, \sigma|\nu)$ and the first and second order derivatives for μ and σ as required can be obtained from the authors. Simulation results for the two scenarios are given in Tables 1 and 2, respectively. Note that the sample size of n = 20 is too small in scenario 2 to obtain convergent estimates.

As expected, the accuracy and precision of all parameter estimates increase with increasing *n* in general. The results also show that ν is more likely to be slightly overestimated but there are also a few underestimations. This shows two counter-acting effects for the ν estimate. Firstly, overestimation of ν is common in Student-t distribution with light tails since ν may tend to infinity, as shown by their higher SEs in Tables 1 and 2. On the other hand, our proposed hybrid method tends to underestimate ν to provide heavier tails for the distribution and the level of underestimation may be greater for a larger sample with more extreme observations. Hence we can observe underestimation of ν when n is large and ν is low but as ν increases, it will be dominated by overestimation. While ν is slightly overestimated in general, σ is slightly underestimated, particularly when n is small. This phenomenon is also common because both σ and ν simultaneously control the variance and kurtosis of the Student-t distribution and their estimated values tend to affect each other. Although σ is consistently underestimated, the bias is very small for moderate to large n. Lastly, μ is estimated with a high level of accuracy regardless of n and ν. To conclude, our proposed hybrid method provides a reasonably good estimate for *ν*.

3.1.2 EP distribution

The true shape parameter for the EP distribution is set to be β = 0.5, 1.0, 1.5, 2.0, 2.5 covering distributions with lower to higher kurtosis than normal distribution. Results for the two scenarios are reported in Tables 3 and 4 respectively. In Scenario 1, the β is likely to be slightly overestimated but the percentage error is less than 10%. In Scenario 2, β is again slightly overestimated, particularly when the sample size n is small and when μ and σ need to be estimated simultaneously. Accordingly, σ is consistently underestimated but its accuracy and precision increase with n. As with the Student-t distribution, μ is estimated accurately regardless of n and β. Generally speaking, our proposed hybrid method provides a satisfactory performance for estimating β of the EP distribution as the error percentages are at most 15% even at the sample size of n = 100 with one exception but their robustness advantage is demonstrated in the next simulation experiment.

3.2 Robustness study

The second experiment focuses on the performance of the proposed hybrid method and its comparison with the full ML method, M1 and M4 respectively for Student-t and EP distributions, in the presence of extreme observations. We generate contaminated data with two extreme observations by multiplying 2 and 2.5 respectively to the absolute values of the two most extreme standardized observations, that is $z^*_{(n-1)} = 2|z_{(n-1)}|$ and $z^*_{(n)} = 2.5|z_{(n)}|$. We only consider n = 20 and 100 which are the sample sizes most similar to the two data sets in Section 4. To compare the goodness-of-fit between the hybrid and ML methods, we report the Anderson-Darling (AD) statistic (Anderson and Darling, 1952) defined as:

$$A_n^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i - 1)[lnF(x_{(i)}) + ln(1 - F(x_{(n-i+1)}))].$$

We note that there is no test for significance when the distribution is Student-t and EP and AD statistic is used to compare the agreement between theoretical CDF F ($x_{(i)}$) with its empirical CDF $Fn(x_{(i)}) = i/n$ called model fit between the hybrid and full ML methods.

A smaller value of AD indicates a stronger agreement between F (x(i)) and Fn(x(i)). Tables 5 and 6 report the parameter estimates and AD statistics for the contaminated data of size n = 20 and 100 for the Student-t and EP distributions respectively. We remark that the true values denote various cases and should not be used to assess estimate accuracy as they have changed after contamination. Comparing AD measures between hybrid and full ML methods, results show close agreement between the two distributions. Obviously, the bias and SE of $\hat{\mu}$ and $\hat{\sigma}$ decrease with sample size n but $\hat{\mu}$ is not greatly inflated by the two right-sided extreme observations. To accommodate these observations, both methods diminish $\hat{\nu}$ or increase $\hat{\beta}$ giving thicker tails and estimate a lower $\hat{\sigma}$. These phenomena agree with the results of performance study. Comparatively, Table 5 shows that the hybrid method using Student-t distribution deflates ν and hence σ more and shifts $\hat{\mu}$ less towards the extreme observation than the full ML method. Similarly, Table 6 also shows that the hybrid method using EP distribution inflates β and deflates σ more than the full ML method. As a result, the hybrid method matches quantiles of the most extreme outlier to be the empirical 0.975 and 0.995 levels for n = 20 and 100 respectively which are lower that those of 0.990 and 0.999 respectively for Student-t and 0.992 and 0.9998 for EP distributions using the ML method. For model fit, we observe lower AD measures using the hybrid method. In conclusion, the hybrid method displays better model fit than the full ML method in the presence of extreme outliers as it gives heavier tailed distribution to downweigh the distorting effect of extreme observations and hence offers extra protection for statistical inference.

4 Data analysis

We analyze two real data sets which have different characteristics. They have minor skewness but different kurtosis. The Darwin data have a small sample size and two outliers in the lower side, and the Rivers data have mild excess kurtosis. We include the Shapiro-Wilk normality test and the p-values are 0.10 and 0.12 respectively showing that skewness is not significant for both data. For each data set, we begin with fitting a location-scale model with normal data distribution using the full ML approach for parameter estimation (Model 0 or M0). As it is well-known that the normal distribution is sensitive to outliers and in reality, the data generating distribution is always unknown, we further adopt the flexible Student-t and EP distributions which offer different distribution shapes in the central portion and accommodate heavy tails in the two ends to capture different levels of kurtosis. Model parameters are estimated using different methods. For the Student-t distribution, we apply either the full ML method (M1) to estimate all parameters simultaneously or the proposed hybrid method (M2) using ML method to estimate μ and σ given $\hat{\nu}$ and MQ method to estimate ν given $(\hat{\mu}, \hat{\sigma})$. For the EP distribution, we use the MM (M3) of Gonz´alez-Far´ıas et al. (2009), the full ML method (M4) using R package optim, and the proposed hybrid method (M5). Simulation studies confirm good performance of our proposed method in estimating the shape parameter when the data follow either a Student-t or EP distribution.

Table 7 reports the parameter estimates and the AD for each data set. Kernel density estimate and empirical CDF for each data set are displayed in Figure 4, together with their fitted values and the quantile-quantile (Q-Q) plots of the fitted models are given in Figure 5. The performance of each model in fitting the data can be revealed from these two graphs. Note that the peak of the smoothed kernel density estimate in Figure 4 depends on the bandwidth chosen and hence it cannot serve as a reference for comparison across models. Note also that our proposed hybrid method (M2 and M5) gives perfect quantile matches for the most extreme observations in Figure 5.

4.1 Darwin Data

Fisher (1960) analyzed the difference in height of 15 pairs of self- and cross-fertilized plants and the data are well-known as Darwin's data. It is obvious that this data set contains two extreme observations. Therefore, the residual sum of squares, SSE is substantially inflated to 19945, compared with a value of only 5568 when these two observations are discarded. These two observations distort statistical inference and the normal data distribution fails to provide sensible estimates for the location parameter μ and scale parameter σ. On the contrary, the Student-t and EP distributions provide robust inferences and more sensible μ estimates in general.

For the Student-t distribution, $\widehat{\nu}$ are 2.5 and 1.5 for M1 and M2 respectively. As pointed out in the simulation study, our proposed MQ estimate tends to slightly underestimate ν, suggesting a slightly heavier-tailed Student-t distribution. This is the result of accommodating the most extreme observation by thickening the tails of the distribution. For the EP distribution, $\widehat{\beta}$ are 1.85, 2.67, and 3.28 for M3, M4, and M5, respectively. The MM method (M3) is problematic as $\hat{\mu}$ is estimated by the sample mean which is sensitive to extreme observations. Hence its AD statistic is the second worst. The Q-Q plots in Figure 5 show that the normal distribution attempts to fit the two extreme observations and thus it fits poorly to the central portion of the data. This is confirmed by the higher CDF in this portion of data in Figure 4(c). Comparatively, the Student-t and EP distributions fit the central portion of the data much better than the normal distribution. Our proposed hybrid method in M2 and M5 fits well in both the central portion and the most extreme observation by matching quantiles. According to AD statistic, the hybrid method outperforms the ML method for both distributions and M5 using the hybrid method and EP distribution slightly outperforms M2 using Student-t distribution. Hence M5 provides the best model fit among all models. This agrees with Choy and Walker (2003), who used the Bayesian inference and the finding from the second simulation experiment that EP distribution captures extreme observations better than Student-t distribution.

Using the Student-t distribution, we compare the computational efficiency of the full ML method (M1), the hybrid method (M2), MCECM algorithm and ECME algorithm. The number of iterations required for all parameters to converge within a tolerance of 10^{-4} are 144, 9, 83, and 44 and the computing times (in seconds) are 1.09, 0.62, 0.54, and 0.40 respectively. Obviously, the hybrid method has a faster convergence rate than its competitors. Note that the MCECM and ECME algorithms under the full ML method give the same parameter estimates while the hybrid method gives a smaller $\hat{\nu}$ and hence different $\hat{\mu}$ and $\hat{\sigma}$.

4.2 Wind Data

The Wind data set contains 153 average wind speeds (in miles per hour) recorded between 7:00am and 10:00am from May to September in 1973 at LaGuardia Airport, New York, USA. The data are available in Chambers et al. (1983) and are saved in the data set called airquality of The R Datasets Package. The sample mean, standard deviation, skewness, and kurtosis are 9.96, 3.52, 0.35, and 3.11 respectively. The data are slightly skewed but have low level excess kurtosis. Obviously, the data contain no potential outlier as shown in Figure 4. Again, M2 using hybrid method and Student-t distribution provides the best model fit according to the AD statistic. The estimates of μ are very similar across models. From Figures 4 and 5, we see that all five models have similar peaks to describe the central portion of the data distribution and similar CDFs and QQ plots to reveal their close agreement for the overall distribution. Hence, they all provide similar and reasonably good model fit according to the AD statistic although the kernel density estimate shows a possible mixture model for the Wind data. Our proposed hybrid method estimates ν to be 11 for the Student-t distribution and β to be 1.35 for the EP distribution, again suggesting more heavy-tailed distributions than their ML counterparts of 83 and 1.03 respectively. Lastly, we check the adequacy of M2 using the Chi-square test with six symmetric bins in which the four middle bins have unit length on standardised scale. The p-value is 0.189 showing consistency with the null hypothesis of Student-t distribution.

5 Conclusion

The shape parameter of a symmetric probability distribution is often more difficult to estimate than the location and scale parameters. This paper proposes a simple way to estimate the shape parameter using a MQ method applied to the most extreme observation and the location and scale parameters using the ML method under a hybrid approach. This approach avoids the potential difficulties in estimating all model parameters simultaneously in a full ML approach and is particularly applicable for the EP distribution when the full ML method is infeasible and the MM fails to provide robust inference in the presence of outliers. The simulation and empirical studies confirm that the hybrid method is simple and feasible with satisfactory performance. Specifically, simulation studies show that the hybrid method provides parameter estimates reasonably close to the true values in a general situation whereas in the presence of very extreme observations, it estimates thicker tails than the full ML method to protect inference. This feature results in better model fit than the ML method with heavy tailed distribution.

This phenomenon is also demonstrated in the empirical studies of two real data with different characteristics. In the studies using Student-t distribution, the full ML method (M1) and our proposed hybrid method (M2) produce similar μ estimates but M2 has Smaller $\hat{\nu}$ than M1. With thicker tails, $\hat{\sigma}$ is also smaller in M2. Despite of this, both M1 and M2 give very similar estimates to the standard deviation of the distribution, which is $\sqrt{\frac{\nu}{\nu-2}}\sigma$ if it exists. When the EP distribution is adopted, the MM (M3), full ML method (M4) and hybrid method (M5) have different estimates for μ but M5 again estimates distribution with sharper peak and heavier tails to accommodate extreme observations. However, under M3, μ is estimated using the sample mean and hence, Figures 4(a)- (b) show that the centers of the fitted distributions using M3, as was M0, are shifted relatively closer to the extreme observations. Therefore, MM should not be recommended in parameter estimation if other methods are available. On the other hand, Table 8 shows that the fitted PDFs at the most extreme observation for Wind data are highest using M2, showing heavier tailed distributions but for Darwin data, it is M0 when the normal PDF is lifted up as a result of the left-shifted center by the two lower-sided extreme observations. This is again illustrated in Figure 4(a). The CDFs using M2 and M5 as reported in Table 8 and the QQ plot in Figure 5 show that the quantiles of most extreme observations are matched precisely and the AD statistic selects an estimation method that provides the best overall model fit.

Lastly, we conclude this paper with three remarks. Firstly, our proposed hybrid method suggests a slightly heavier-tailed distribution when the data is skewed and/or contains potential outliers. Perhaps the method can be extended to matching the quantile of the second or other less extreme points and the level of extremeness for the observation to be matched may be tuned to provide better model fit. This requires further investigation. Secondly, the method can be generalized to the class of asymmetric distributions and distributions with more than one shape parameters by matching several extreme observations but such generalization may not be straightforward and future research in this direction is required. Thirdly, apart from the class of distributions with real support, one can consider the distributions with positive real support. Some recently proposed distributions of this type include the three parameter paralogistic distribution (Idemudia and Ekhosuehi, 2019) and the exponentiated generalized extended Gompertz distribution (De Andrade et al., 2019). However, this extension can be very challenging as there are no distinct scale and shape parameters.

Appendix:

A. Suppose that $Y_i, i = 1,\dots,n$, are identically and independently distributed as a Student-t distribution with ν degrees of freedom (df), that is, $Y_i \sim t\nu(\mu, \sigma)$, Liu and Rubin (1995) express Yi as following a normal scale mixtures distribution with mixing variables λi. That is,

$$Y_i \sim N(\mu, \frac{\sigma^2}{\lambda_i}) \text{ and } \lambda_i \sim G(\frac{\nu}{2}, \frac{\nu}{2})$$

where G(a, b) is the gamma distribution with mean . At iteration k + 1 in the MCECM cycle, the first CM-step updates the parameters (μ, σ) to

$$\mu^{(k+1)} = \frac{\sum_{i=1}^{n} \lambda_i^{(k)} y_i}{\sum_{i=1}^{n} \lambda_i^{(k)}} \text{ and } \sigma^{(k+1)} = n^{-1/2}(\sum_{i=1}^{n} \lambda_i^{(k)} y_i^2 - \frac{1}{\sum_{i=1}^{n} \lambda_i^{(k)}}(\sum_{i=1}^{n} \lambda_i^{(k)} y_i)^2)^{1/2}$$

Then, the first E-step gives

$$\lambda_i^{(k+0.5)} = \frac{\nu^{(k)} + 1}{\nu^{(k)} + (\frac{y_i - \mu^{(k+1)}}{\sigma^{(k+1)}})^2}$$

and ν(k+1) is obtained from the second CM-step by solving

$$-\psi(\frac{\nu^{(k+1)}}{2}) + \ln(\frac{\nu^{(k+1)}}{2}) + \frac{1}{n}\sum_{i=1}^{n}(\ln(\lambda_i^{(k+0.5)}) - \lambda_i^{(k+0.5)}) + 1$$

$$+\frac{1}{n}\sum_{i=1}^{n}[\psi(\frac{\nu^{(k)}+1}{2}) - \ln(\frac{\nu^{(k)}+1}{2})] = 0$$

which maximizes the expected log-likelihood. Here, ψ(x) is the digamma function. Then, the second E-step gives $\lambda_i^{(k+1)}$ from (4) with $\nu^{(k+1)}$ replacing $\nu^{(k)}$. The ECM algorithm contains only the second E-step and hence the CM-step in (5) is evaluated using λ_i . The ECME algorithm has the same procedures as MCECM, except that ν(k+1) is evaluated from maximizing the log of Student-t densities in (1) directly. Although these ECM-type algorithms are easy to implement, they require numerical searches for the df estimate and hence the convergence rate is rather slow.

B. To estimate the shape parameter β of EP distribution, Varanasi and Aazhang (1989) proposed an entropy matching estimator which satisfies

$$\frac{(\Gamma(\frac{\beta}{2}))^{r/2-1}\Gamma(\frac{(r+1)\beta}{2})}{(\Gamma(\frac{3\beta}{2}))^{r/2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}|y_i - \bar{y}|^r}{(\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2)^{r/2}}$$

for any moments of order r≥ 4 Later, Rodr´ıguez-Dagnino and Leon-Garcia (1998) presented a closed form estimator for β based on an approximation using Gurlands inequality but this approximation is only well-behaved for 0.3 < p < 3 where p = 2/β . Later on, Gonz´alez-Far´ıas et al. (2009) proposed an alternative MM estimate for β which satisfies

$$\frac{\Gamma^2(\beta)}{\Gamma(\frac{\beta}{2})\Gamma(\frac{3\beta}{2})} = \frac{(\frac{1}{n}\sum_{i=1}^{n}|y_i - \overline{y}|)^2}{\frac{1}{n}\sum_{i=1}^{n}|y_i - \overline{y}|^2}$$

for 0.18 < p < 2 where p = β and they derived the following closed form solution for p:

$$\hat{p} = \begin{cases} \frac{ln(27/16)}{ln(3/(4m^2\ ))} & if\ m \in\ (0,0.131246), \\ \frac{1}{2a_1}\left(-a_2\ +\ a_4\ (m)\right) & if\ m \in\ [0.131246,0.448994) \\ \frac{1}{2b_3 m}\left(b_1\ -\ b_2\, m\ -\ b_4\ (m)\right) & if\ m \in\ [0.448994,0.671256), \\ \frac{1}{2c_3}\left(c_2\ -\ c_4\ (m)\right) & if\ m \in\ [0.671256,0.75) \end{cases}$$

where m denotes the right hand side of (6), $(6), a_1 = -0.535707, a_2 = 1.168940, a_3 = -0.151619, b_1 = 0.969443, b_2 = 0.872753, b_3 = 0.073508, c_1 = 0.365516, c_2 = 0.672353, c_3 = 0.033834, a_4(m) = (a_2^2 - 4a_1a_3 + 4a_1m)^{\frac{1}{2}}, b_4(m) = ((b_1 - b_2m)^2 - 4b_3m^2)^{\frac{1}{2}}, c_4(m) = (c_2^2 + 4c_3\ln(\frac{3-4m}{4c_1}))^{1/2}$

To estimate other parameters, one estimator adopts a hybrid approach and consists of the MM estimate of μ using the sample mean $\overline{y}$, the ML estimate of σ given by:

$$\hat{\sigma} = (\frac{1}{n\beta}\sum_{i=1}^{n}|y_i - \overline{y}|^{2/\beta})^{\beta/2}$$

and the ML estimate β by solving the following equation:

$$-\frac{n}{2}\psi(\frac{\beta}{2}) - \frac{n\ln\ 2}{2} - \frac{n}{\beta} + \frac{1}{\beta^2}\sum_{i=1}^{n}\{\ln\ |\frac{y_i - \overline{y}}{\hat{\sigma}}| \times |\frac{y_i - \overline{y}}{\hat{\sigma}}|^{2/\beta}\} = 0$$

Instead, we consider a full MM approach (M3) which estimates μ using the sample mean $\overline{y}, \sigma^2$ using

$$\hat{\sigma}^2 = \frac{\Gamma(\frac{\beta}{2})}{2^{\beta}\Gamma(\frac{3\beta}{2})}\frac{1}{n}\sum_{i=1}^{n}|y_i - \overline{y}|^2$$

and β using (7) as proposed in Gonz´alez-Far´ıas et al. (2009).

References

[1] Agro, G. (1995). Maximum likelihood estimation for the exponential power function parameters, Communications in Statistics - Simulation and Computation, 24, 523-536.

[2] Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain “goodness-of-fit” criteria based on stochastic processes. Annals of Mathematical Statistics, 23, 193-212.

[3] Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983). Graphical Methods for Data Analysis. Belmont, CA: Wadsworth.

[4] Choy, S.T.B. and Walker, S.G. (2003). The extended exponential power distribution and Bayesian robustness, Statistics and Probability Letters, 65, 227-232.

[5] De Andrade, T. A., Chakraborty, S., Handique, L., and Gomes-Silva, F. (2019). The exponentiated generalized extended Gompertz distribution. Journal of Data Science, 17(2), 299-330.

[6] Fisher, R.A. (1960). The Design of Experiments (2nd ed.), New York: Hafner.

[7] Gonz´alez-Far´ıas,G., Dom´ınguez-Molina,J. A., and Rodr´ıguez-Dagnino,R.M. (2009). Efficiency of the approximated shape parameter estimator in the generalized Gaussian distribution, IEEE Transactions on Vehicular Technology, 58, 4214-4223.

[8] Hill, B. (1975). A simple general approach to inference about the tail of a distribution, Ann. Statist., 3, 1163-1173.

[9] Idemudia, R., and Ekhosuehi, N. (2019). A new three-parameter paralogistic distribution: its properties and application. Journal of Data Science, 17(2), 239-257.

[10] Karian, Z. and Dudewicz, E. (1999). Fittin the generalized lambda distribution to data: a method based on percentiles. Communications in Statistics - Simulation and Computation, 28, 793-819.

[11] Liu, C., and Rubin, D.B. (1995). ML estimation of the t distribution using EM and its extension, ECM and ECME, Statistica Sinica, 5, 19-39.

[12] McNeil, D.R. (1977). Interactive Data Analysis, New York: Wiley.

[13] Mineo, A.M. and Ruggieri, M. (2005). A software tool of the exponential power distribution: the normalp package, Journal of Statistical Software, 12, 1-23.

[14] Nadarajah, S. (2005). A generalized normal distribution, Journal of Applied Statistics, 32, 685-694.

[15] Nelder, J. A. and Mead, R. (1965). A simplex algorithm for function minimization, Computer Jour-

nal, 7, 308-313.

[16] Rodr´ıguez-Dagnino,R. M. and Leon-Garcia, A. (1998). An explicit estimator for the shape parameter of the generalized Gaussian distribution, presented at the XII Foro Nacional de Estad´ıstica,Monterrey, Mexico.

[17] Small, C. and McLeish, D. (1994). Hilbert Space Methods in Probability and Statistical Inference. Wiley-Interscience.

[18] Varanasi, M.K., and Aazhang, B. (1989). Parametric generalized Gaussian density estimation, Journal of the Acoustical Society of America, 86, 1404-1415.

Table 1: Proposed MQ estimate for the degrees of freedom ν of the standard Student-t distribution.

True ν	2	SE	5	SE	10	SE	15	SE	20	SE
$n = 20$	3.03	*7.03*	9.92	*15.14*	17.08	*19.26*	19.36	*19.89*	18.26	*18.95*
$n = 50$	2.11	*2.28*	6.88	*9.84*	13.99	*15.60*	17.69	*17.34*	21.96	*18.93*
$n = 100$	1.89	*0.63*	5.98	*6.24*	13.42	*13.63*	18.87	*16.85*	22.44	*18.00*
$n = 200$	1.89	*0.50*	5.01	*2.87*	11.20	*9.58*	17.76	*14.78*	22.64	*16.71*
$n = 500$	1.92	*0.41*	4.82	*1.51*	10.70	*7.02*	16.47	*12.15*	21.90	*14.27*
$n = 1000$	1.88	*0.35*	4.86	*1.41*	10.10	*5.30*	16.78	*11.21*	21.24	*13.18*

Table 2: Proposed MQ estimate for ν and ML estimates for μ and σ of the non-standard Student-t distribution.

True ν	2	SE	5	SE	10	SE	15	SE	20	SE
$n = 50$	2.24	*2.37*	8.17	*12.43*	15.36	*18.14*	21.58	*20.17*	16.64	*18.43*
$n = 100$	2.05	*0.82*	7.38	*11.29*	13.98	*15.10*	20.57	*19.36*	21.33	*19.86*
$n = 200$	1.91	*0.57*	5.44	*5.86*	12.88	*13.17*	19.58	*17.89*	19.15	*17.86*
$n = 500$	1.85	*0.46*	4.80	*2.12*	11.96	*11.54*	19.85	*15.65*	23.45	*16.77*
$n = 1000$	1.84	*0.41*	4.76	*1.97*	12.08	*8.95*	17.02	*13.40*	22.69	*15.76*
True $\mu = 5$										
$n = 50$	4.98	*0.34*	4.99	*0.34*	4.98	*0.28*	5.01	*0.32*	4.97	*0.28*
$n = 100$	5.01	*0.28*	5.00	*0.23*	4.99	*0.20*	5.03	*0.21*	5.01	*0.22*
$n = 200$	5.00	*0.19*	5.00	*0.16*	5.02	*0.15*	5.02	*0.15*	5.00	*0.14*
$n = 500$	5.01	*0.12*	5.01	*0.11*	4.99	*0.11*	5.00	*0.10*	4.98	*0.09*
$n = 1000$	5.01	*0.08*	4.99	*0.07*	5.00	*0.06*	5.00	*0.06*	4.99	*0.07*
True $\sigma = 2$										
$n = 50$	1.89	*0.67*	1.89	*0.34*	1.84	*0.27*	1.92	*0.26*	1.74	*0.38*
$n = 100$	1.92	*0.34*	1.90	*0.27*	1.90	*0.23*	1.92	*0.19*	1.87	*0.21*
$n = 200$	1.92	*0.24*	1.93	*0.20*	1.94	*0.18*	1.94	*0.15*	1.90	*0.15*
$n = 500$	1.93	*0.18*	1.96	*0.14*	1.96	*0.13*	1.96	*0.09*	1.96	*0.11*
$n = 1000$	1.93	*0.15*	1.95	*0.11*	1.98	*0.11*	1.96	*0.09*	1.97	*0.08*

Table 3: Proposed MQ estimate for β of the standard EP distribution.

True β	0.5	SE	1	SE	1.5	SE	2	SE	2.5	SE
$n = 20$	0.55	*0.17*	1.07	*0.21*	1.59	*0.25*	2.09	*0.28*	2.62	*0.30*
$n = 50$	0.53	*0.11*	1.05	*0.14*	1.56	*0.18*	2.09	*0.21*	2.60	*0.23*
$n = 100$	0.53	*0.08*	1.04	*0.13*	1.56	*0.14*	2.07	*0.17*	2.58	*0.18*
$n = 200$	0.53	*0.07*	1.03	*0.10*	1.55	*0.12*	2.07	*0.15*	2.58	*0.17*
$n = 500$	0.52	*0.05*	1.04	*0.08*	1.54	*0.10*	2.04	*0.12*	2.55	*0.14*
$n = 1000$	0.52	*0.04*	1.03	*0.07*	1.53	*0.09*	2.04	*0.11*	2.54	*0.12*

Table 4: Proposed MQ estimate for β and ML estimates for μ and σ of the non-standard EP distribution.

True β	0.5	SE	1.0	SE	1.5	SE	2	SE	2.5	SE
$n = 20$	0.83	*0.40*	1.43	*0.58*	2.03	*0.75*	2.50	*0.83*	3.06	*0.78*
$n = 50$	0.65	*0.27*	1.21	*0.41*	1.77	*0.54*	2.34	*0.67*	2.90	*0.70*
$n = 100$	0.60	*0.20*	1.14	*0.35*	1.72	*0.43*	2.25	*0.54*	2.79	*0.60*
$n = 200$	0.58	*0.16*	1.11	*0.27*	1.67	*0.36*	2.23	*0.45*	2.76	*0.53*
$n = 500$	0.55	*0.12*	1.11	*0.23*	1.62	*0.31*	2.14	*0.41*	2.63	*0.49*
$n = 1000$	0.55	*0.10*	1.07	*0.19*	1.59	*0.26*	2.10	*0.34*	2.60	*0.42*
True $\mu = 5$										
$n = 20$	4.98	*0.29*	5.04	*0.47*	5.06	*0.70*	5.14	*1.04*	5.01	*1.45*
$n = 50$	5.00	*0.18*	5.02	*0.28*	5.00	*0.43*	5.06	*0.64*	5.08	*0.95*
$n = 100$	5.00	*0.12*	5.01	*0.20*	4.99	*0.31*	4.97	*0.48*	5.06	*0.64*
$n = 200$	5.00	*0.08*	4.99	*0.15*	5.01	*0.21*	4.99	*0.31*	5.07	*0.48*
$n = 500$	5.00	*0.05*	5.01	*0.09*	5.00	*0.14*	4.99	*0.20*	5.01	*0.26*
$n = 1000$	5.00	*0.04*	5.00	*0.06*	5.01	*0.09*	5.00	*0.13*	5.00	*0.18*
True $\sigma = 2$										
$n = 20$	1.56	*0.41*	1.46	*0.62*	1.39	*0.85*	1.52	*1.13*	1.41	*1.26*
$n = 50$	1.78	*0.32*	1.72	*0.54*	1.68	*0.77*	1.69	*1.02*	1.63	*1.19*
$n = 100$	1.86	*0.26*	1.82	*0.49*	1.73	*0.69*	1.74	*0.88*	1.73	*1.05*
$n = 200$	1.90	*0.22*	1.86	*0.41*	1.78	*0.55*	1.73	*0.74*	1.74	*0.91*
$n = 500$	1.93	*0.18*	1.85	*0.36*	1.85	*0.54*	1.88	*0.75*	2.00	*0.99*
$n = 1000$	1.94	*0.14*	1.90	*0.31*	1.88	*0.47*	1.90	*0.61*	2.01	*0.86*

Table 5: Proposed hybrid and full ML estimates for Student-t distribution with contamination.

Method	True	n	$\nu = 2$	SE	$\nu = 5$	SE	$\nu = 10$	SE	$\nu = 15$	SE	$\nu = 20$	SE
Hybrid	ν	20	0.89	*0.22*	1.09	*0.23*	1.19	*0.25*	1.20	*0.23*	1.22	*0.27*
		100	1.17	*0.25*	1.68	*0.24*	1.88	*0.25*	1.98	*0.26*	2.02	*0.25*
	$\mu = 5$	20	5.08	*0.61*	5.07	*0.56*	5.12	*0.55*	5.08	*0.54*	5.13	*0.51*
		100	5.02	*0.27*	5.01	*0.25*	5.02	*0.23*	5.02	*0.23*	5.02	*0.21*
	$\sigma = 2$	20	1.41	*0.48*	1.33	*0.41*	1.31	*0.36*	1.31	*0.35*	1.28	*0.36*
		100	1.64	*0.24*	1.58	*0.17*	1.52	*0.15*	1.52	*0.16*	1.51	*0.16*
	AD	20	11.8		11.8		11.7		11.8		11.9	
		100	81.4		82.3		82.7		83.3		82.3	
Full ML	ν	20	1.34	*0.42*	1.76	*0.60*	1.99	*0.63*	2.00	*0.65*	2.06	*0.81*
		100	1.79	*0.39*	2.95	*0.56*	3.53	*0.70*	3.80	*0.77*	3.87	*0.76*
	$\mu = 5$	20	5.10	*0.59*	5.10	*0.53*	5.15	*0.52*	5.12	*0.49*	5.16	*0.48*
		100	5.03	*0.26*	5.02	*0.24*	5.04	*0.21*	5.03	*0.22*	5.04	*0.20*
	$\sigma = 2$	20	1.66	*0.56*	1.57	*0.48*	1.54	*0.42*	1.54	*0.42*	1.51	*0.42*
		100	1.90	*0.26*	1.81	*0.20*	1.74	*0.17*	1.74	*0.18*	1.72	*0.17*
	AD	20	12.4		12.3		12.0		12.2		12.2	
		100	87.3		86.7		86.7		87.0		85.9	

Table 6: Proposed hybrid and full ML estimates for EP distribution with contamination.

Method	True	n	$\beta = 0.5$	SE	$\beta = 1.0$	SE	$\beta = 1.5$	SE	$\beta = 2.0$	SE	$\beta = 2.5$	SE
Hybrid	β	20	2.97	*0.44*	3.52	*0.43*	3.82	*0.30*	3.93	*0.20*	3.99	*0.07*
		100	2.51	*0.20*	3.13	*0.34*	3.69	*0.31*	3.93	*0.16*	3.99	*0.05*
	$\mu = 5$	20	5.18	*0.49*	5.27	*0.58*	5.36	*0.68*	5.44	*0.96*	5.34	*1.22*
		100	5.05	*0.23*	5.07	*0.30*	5.10	*0.39*	5.16	*0.51*	5.18	*0.60*
	$\sigma = 2$	20	0.33	*0.16*	0.24	*0.15*	0.22	*0.13*	0.28	*0.13*	0.40	*0.11*
		100	0.42	*0.09*	0.31	*0.11*	0.24	*0.11*	0.25	*0.08*	0.35	*0.06*
	AD	20	10.5		10.6		11.3		11.7		11.9	
		100	77.5		74.9		76.9		82.0		84.6	
Full ML	β	20	2.38	*0.88*	2.99	*0.92*	3.58	*1.07*	3.81	*1.10*	4.26	*1.16*
		100	1.40	*0.16*	1.85	*0.24*	2.36	*0.32*	2.86	*0.39*	3.31	*0.46*
	$\mu = 5$	20	5.22	*0.42*	5.24	*0.55*	5.26	*0.75*	5.35	*1.01*	5.30	*1.36*
		100	5.08	*0.16*	5.07	*0.22*	5.04	*0.34*	5.02	*0.49*	5.11	*0.65*
	$\sigma = 2$	20	0.66	*0.49*	0.50	*0.49*	0.42	*0.47*	0.55	*0.70*	0.55	*0.67*
		100	1.11	*0.19*	1.03	*0.27*	0.94	*0.35*	0.91	*0.42*	0.94	*0.52*
	AD	20	11.8		12.0		12.1		12.1		12.1	
		100	82.8		82.6		84.5		85.5		85.5	

Table 7: Parameter estimates and model-fit measures under the six different approaches (M0-M5) for the two data sets. n, γ and κ are the sample size, sample skewness, and sample kurtosis respectively.

		Darwin: $n = 15$, $\kappa = 4.41$				**Wind:** $n = 153$, $\kappa = 3.11$			
Method	Dist.	μ	σ	ν or β	AD‡	μ	σ	ν or β	AD‡
M0 Full ML	N	20.93	36.46	-	0.618	9.96	3.51	-	0.742
M1 Full ML	t	26.84	22.70	2.50	0.293	9.94	3.47	83.45	0.721
M2 Hybrid	t	26.09	18.90	1.50	0.228	9.87	3.24	10.99	*0.709*
M3 Full MM	EP	20.93	15.40	1.85	0.446	9.96	3.43	1.03	0.741
M4 Full ML	EP	24.00	6.10	2.67	0.256	9.95	3.43	1.03	0.729
M5 Hybrid	EP	24.00	3.25†	3.28	*0.228*	9.82	2.57†	1.35	0.788

† The standard deviation of the EP distribution is a function of σ and β. Using Equation (2), their values are 40.02 and 0.5897 respectively for the Darwin and Wind data.

‡ AD in italic is the smallest across models for each data set, indicating the best overall model fit.

Table 8: PDFs ($\times 10^{-3}$) and CDFs at the most extreme observation for the two data sets.

	PDF♯						CDF♭						
Data	M0	M1	M2	M3	M4	M5	Obs†	M0	M1	M2	M3	M4	M5
Darwin	*21.7738*	9.8592	9.7323	0.6577	0.6186	0.7467	*0.0333*	0.0079	0.0181	*0.0333*	0.0160	0.0214	*0.0339*
Wind	3.6971	3.9789	*5.7857*	1.1716	1.1619	1.9903	*0.9967*	0.9989	0.9987	*0.9967*	0.9987	0.9987	*0.9967*

† The observed CDF is $1/(2n)$ for the observed minimum and $1 - 1/(2n)$ for the observed maximum.
♯ The PDFs in italic are the largest across models showing the heaviest tail density for the outlier.
♭ The CDFs in italic show that the observed and fitted quantiles at the outlier are matched for M2 and M5.

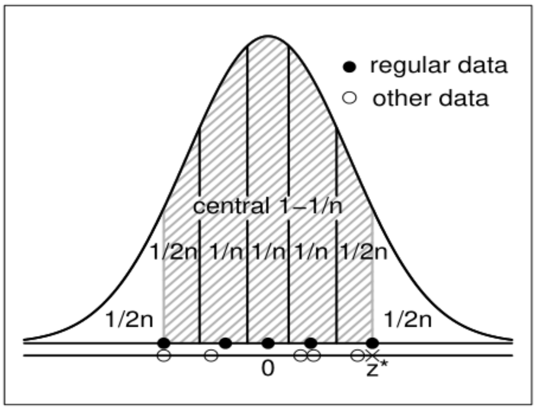

Figure 1: Illustration of the proposed matching quantile method when $n = 5$.

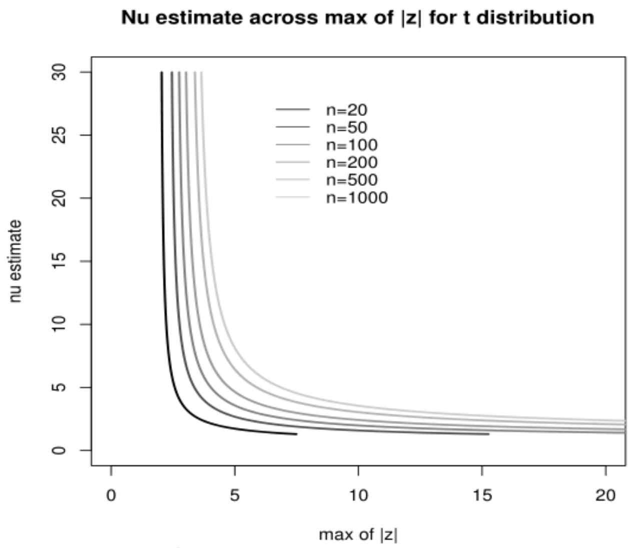

Figure 2: Plots $\widehat{\nu}$ against z^* for various sample sizes n.

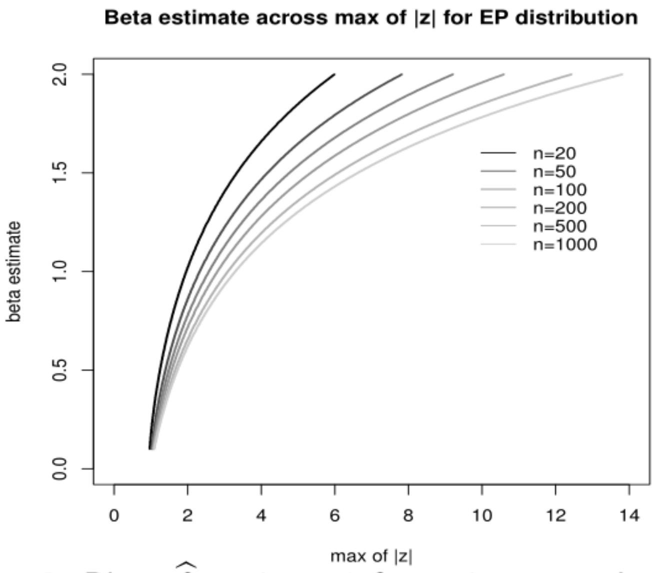

Figure 3: Plots $\widehat{\beta}$ against z^* for various sample sizes n.

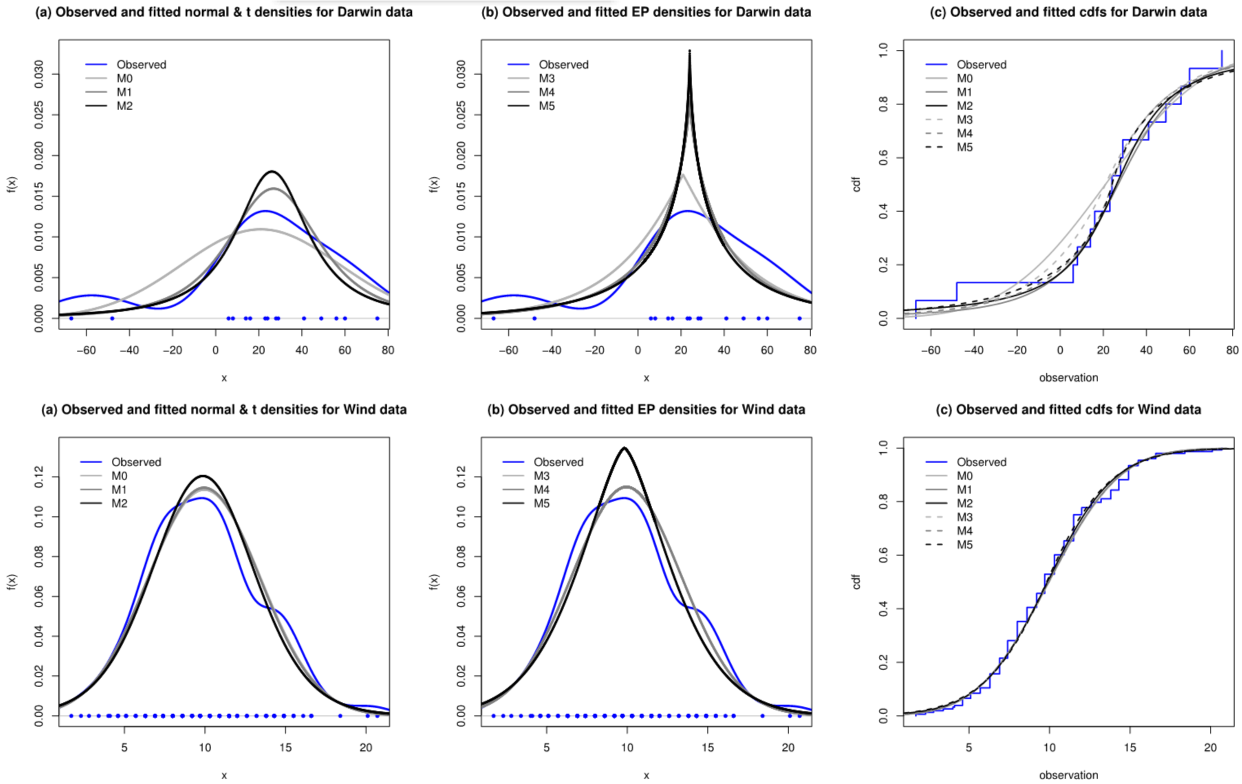

Figure 4: Kernel density estimates, fitted PDFs (a-b) and fitted CDFs (c) for Darwin and Wind data.

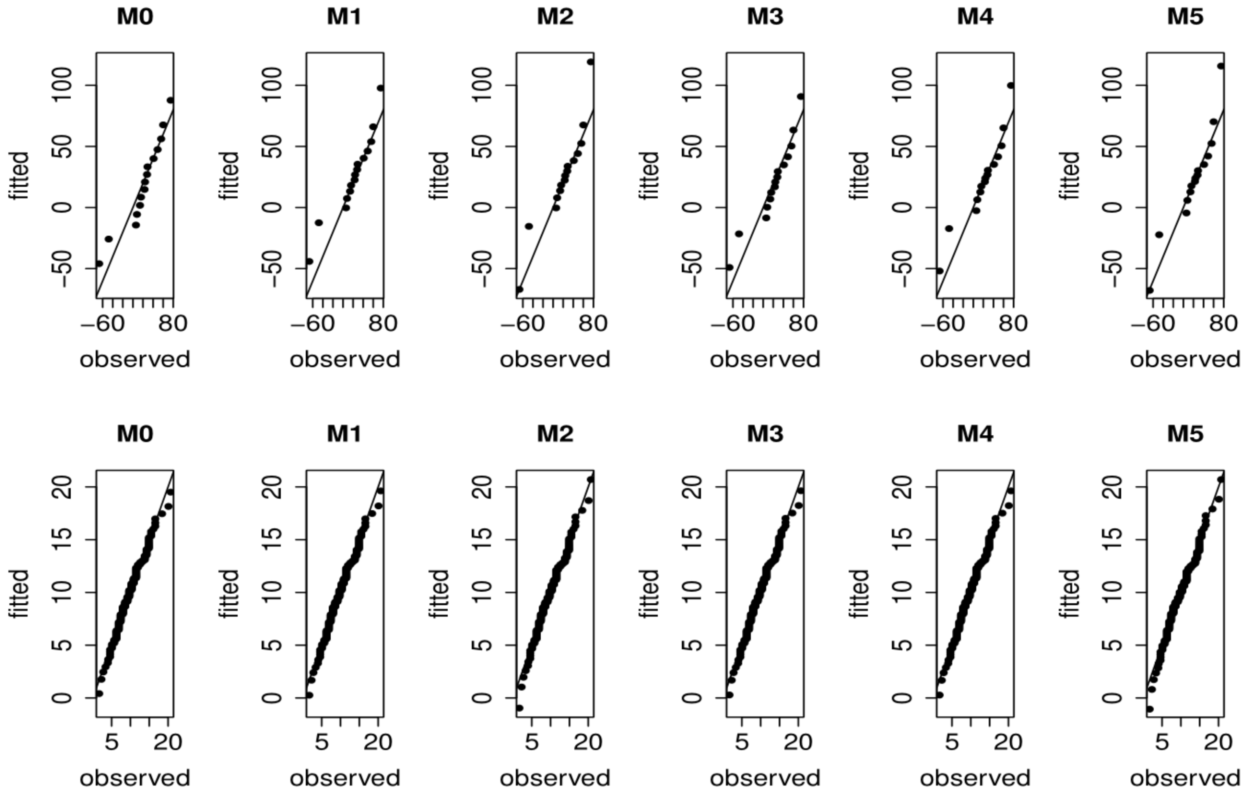

Figure 5: Q-Q plots of the Darwin and Wind data in row 1 and 2 respectively.