On the Marshall-Olkin transformation as a skewing mechanism

F. J. Rubio and M. F. J. Steel[1]

Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK

Abstract

The use of the Marshall-Olkin transformation as a skewing mechanism is investigated. The distributions obtained when this transformation is applied to several classes of symmetric and unimodal distributions are analysed. It is shown that most of the resulting distributions are not flexible enough to model data presenting high or moderate skewness. The only case encountered where the Marshall-Olkin transformation can be considered a useful skewing mechanism is when applied to Student-t distributions with Cauchy or even heavier tails.

Keywords: moment; skewness measure; Student-t; tail behaviour

1. Introduction

The need for modeling data presenting departures from symmetry has fostered the development of more flexible classes of distributions. A popular approach is to modify a symmetric distribution by introducing a parameter that controls skewness (Azzalini, 1985; Fernández and Steel, 1998; Jones, 2004; Ferreira and Steel, 2006).

In the context of reliability and survival analysis, Marshall and Olkin (1997) proposed a transformation of a distribution $F(x;\theta)$ that introduces a new parameter $\gamma > 0$. This transformation is defined through the cumulative distribution function (cdf)

$$G(x;\theta,\gamma) = \frac{F(x;\theta)}{F(x;\theta) + \gamma(1 - F(x;\theta))}, \tag{1}$$

[1]Corresponding author: Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK. E-mail: M.F.Steel@stats.warwick.ac.uk. Tel.: +44 24 7652 3369, Fax: +44 24 7652 4532.

and assuming continuity of F throughout, the corresponding probability density function (pdf) is given by

$$g(x; \theta, \gamma) = \frac{\gamma f(x; \theta)}{[F(x; \theta) + \gamma(1 - F(x; \theta))]^2}. \tag{2}$$

The interpretation of the parameter γ is given in Marshall and Olkin (1997) in terms of the behavior of the ratio of hazard rates of F and G. This ratio is increasing in x for $\gamma \geq 1$ and decreasing in x for $0 < \gamma \leq 1$. This transformation is then proposed for the Exponential and Weibull distribution in Marshall and Olkin (1997) in order to generate more flexible models for lifetime data. Clearly, for $\gamma = 1$, G and F coincide.

Using the fact that the distribution in (1) describes a wider class than the original distribution F, García et al. (2010) define a generalised normal distribution (GN) by applying this transformation to a normal distribution F. They investigate the role of γ as a skewness parameter using the standardized third central moment EM $= \mu_3/\mu_2^{3/2}$ as a skewness measure (Edgeworth, 1904). In a similar search for families of skewed distributions, George and George (2011) apply the Marshall-Olkin transformation to the characteristic function of an Esscher transformed Laplace distribution (which, interestingly, leads to a very simple two-piece distribution with inverse scale factors, used later to generate data in Section 3.2). Maiti and Dey (2012) propose exactly the same distribution as the GN of García et al. (2010) and call it the tilted normal distribution. However, they focus mostly on its use for modelling survival data and less on the skewness properties.

We will focus here on the use of the Marshall-Olkin transformation in (1) as a mechanism for inducing skewness in symmetric and unimodal distributions F which are defined over the entire real line. It is immediate from (2) that $g(x; \theta, \gamma) = g(-x; \theta, 1/\gamma)$, which means that usual measures of skewness will change sign by inverting γ and that superficially suggests γ plays the part of a skewness parameter. Perhaps the most obvious choice for F is the normal, as explored by García et al. (2010) and Maiti and Dey (2012), and we will first investigate the wider class of Student-t distributions.

In Section 2 we study the tail behaviour induced by the Marshall-Olkin transformation and in the next section we define a generalised t distribution based on the transformation in (1). We explore the role of the parameter γ in the generalised t and the generalised normal distributions using different measures of skewness and we show that the standardized third central moment can lead to counterintuitive conclusions about the shape of the density. In fact, if we use a different measure of skewness based on the relative mass both sides of the mode, it becomes clear that the Marshall-Olkin transformation applied to normal and Student-t distributions with tails that are not extremely fat is unable to accommodate even moderate amounts of skewness. Section 3.2 illustrates this with some simulated data. Section 4 examines the use of the Marshall-Olkin transformation on other classes of distributions and Section 5 provides some intuitive explanation of the observed behaviour. Finally, we conclude that the Marshall-Olkin transformation can not generally be used as a skewing mechanism for unimodal symmetric distributions, and we find only one exception: the Student-t distribution with Cauchy or even heavier tails.

2. Tail behaviour

Marshall and Olkin (1997) proved existence of moments of (1) for the cases when F is Exponential or Weibull. The next Theorem shows that this transformation preserves moment existence for general F.

Theorem 1. *The moments of (1) exist for exactly the same order as in the original distribution F.*
Proof. *Note that if $\gamma < 1$, then*

$$\gamma < \frac{\gamma}{\left[F(x;\theta) + \gamma(1 - F(x;\theta))\right]^2} < \frac{1}{\gamma}.$$

If $\gamma > 1$, then

$$\frac{1}{\gamma} < \frac{\gamma}{\left[F(x;\theta) + \gamma(1 - F(x;\theta))\right]^2} < \gamma.$$

Therefore

$$g(x;\theta,\gamma) = K(x,\theta,\gamma) f(x;\theta),$$

where $K(x,\theta,\gamma)$ takes values in between $\min\{\gamma, 1/\gamma\}$ *and* $\max\{\gamma, 1/\gamma\}$. *The result follows.*

Theorem 1 shows that transformation (1) produces a distribution with exactly the same tail behaviour as the original.

3. Generalised *t*

We now define a generalised t (Gt) distribution by applying the Marshall-Olkin transformation to the Student-t distribution.

Definition 2. *A random variable X is distributed according to the generalised t distribution if its cdf and pdf are given by*

$$Gt(x;\mu,\sigma,\nu,\gamma) = \frac{F(x;\mu,\sigma,\nu)}{F(x;\mu,\sigma,\nu)+\gamma(1-F(x;\mu,\sigma,\nu))}, \tag{3}$$

$$gt(x;\mu,\sigma,\nu,\gamma) = \frac{\gamma f(x;\mu,\sigma,\nu)}{[F(x;\mu,\sigma,\nu)+\gamma(1-F(x;\mu,\sigma,\nu))]^2}, \tag{4}$$

where F and f are the cdf and pdf of a Student-t distribution with location μ, scale σ and ν degrees of freedom.

Figure 1 shows some examples of density (4) for different choices of the parameters. Of course, panel (a) is just the Student-t, whereas panel (b) corresponds to $\gamma = 0.5$ and (c) is for $\gamma = 2$. Visually, two things are worth noting

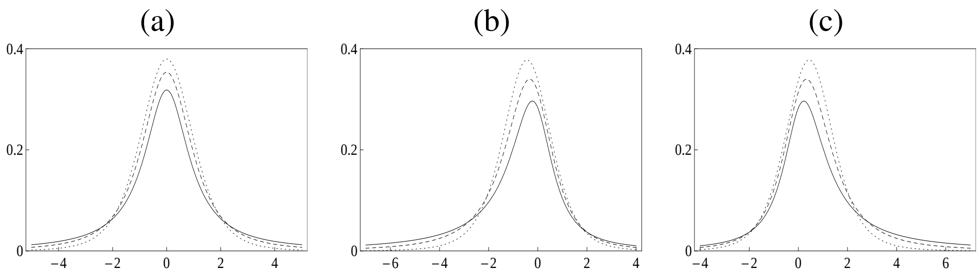

Figure 1: Examples of the density (4) with $\mu = 0$, $\sigma = 1$ and $\nu = 1$ (solid line), $\nu = 2$ (dashed line), $\nu = 5$ (dotted line): (a) $\gamma = 1$; (b) $\gamma = 0.5$; (c) $\gamma = 2$.

about Figure 1: the densities generated do not seem highly skewed (even though γ is rather far from one), especially for larger values of ν, and the amount of skewness seems to depend on the value of ν. This would suggest that ν and γ can not straightforwardly be assigned roles as tail and skewness parameters, respectively.

Just as in the symmetric case, the generalised normal distribution (GN) (García et al., 2010) is a limiting case of the Gt distribution, since $\lim_{\nu\to\infty} \mathrm{Gt}(x;\mu,\sigma,\nu,\gamma) = \mathrm{GN}(x;\mu,\sigma,\gamma)$.

3.1. The role of γ

Several measures of skewness have been proposed in the literature; see e.g. Groeneveld and Meeden (1984), Groeneveld (1991) and Arnold and Groeneveld (1995). We will assess the role of the parameter γ in the generalised t and the generalised normal by considering three different measures of skewness: the standardized third central moment EM= $\mu_3/\mu_2^{3/2}$, the Pearson measure of skewness (Pearson, 1895) defined as

$$\text{PM} = \frac{\text{Mean-Mode}}{\mu_2^{1/2}},$$

and the Arnold-Groeneveld measure of skewness (Arnold and Groeneveld, 1995), which is defined for any distribution S with unimodal density as AG $= 1 - 2S(\text{mode})$. The AG measure takes values in $(-1, 1)$, while negative values of AG are associated with left skewness and positive values of AG reflect right skewness. This skewness measure has a clear and intuitively appealing interpretation in terms of the allocation of mass both sides of the mode, and does not require the existence of any moment. We believe the three skewness measures considered here are representative of the most commonly used methods, but other measures of skewness could be used, such as the one proposed in Groeneveld and Meeden (2009) (where the direction of skewness is assumed known).

García et al. (2010) claim that the parameter γ in the generalised normal distribution plays the role of a skewness parameter as it "has a substantial effect on the skewness of the probability density function". This is shown using the standardized third central moment EM. Given that it is possible to cover a certain range of values of EM by varying the value of γ, García et al. (2010) conclude that this transformation can be used to introduce skewness. Here, however, we show that if we evaluate the role of γ using the AG and PM measures of skewness, then we have to conclude that the generalised normal and the generalised t models (except with small ν) are not flexible enough to model high or even moderate skewness.

Figure 2 shows the AG measure as a function of γ for several fixed values of ν for the generalised t. While for $\nu = 1$ the behaviour seems reasonable, for larger values of ν the AG measure as a function of γ is far from surjective and not even necessarily a one-to-one function. Surprisingly, in the practically relevant case with $\nu = 10$ the parameter γ has only a very small effect on AG and the direction of this effect changes with γ. If we consider instead the moment-based measure EM in Figure 3, we observe a similar worrying behaviour. The parameter γ has a relatively well-defined effect on EM for small ν (of course, we need $\nu > 3$ for EM to be defined), but for larger ν the effect is very small and not monotone. Figure 4 shows the AG, PM and EM measures as a function

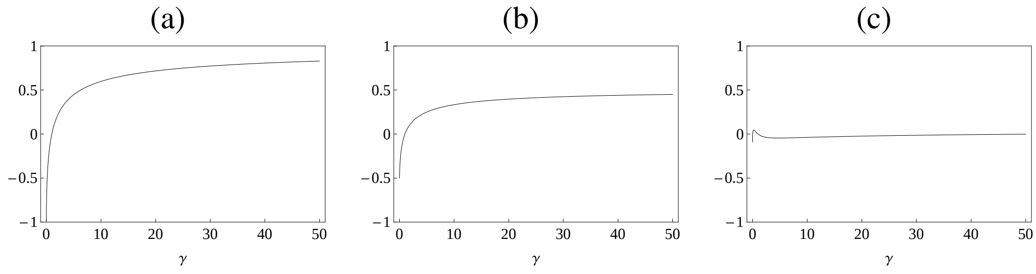

Figure 2: AG measure of skewness for the generalised t: (a) $\nu = 1$; (b) $\nu = 2$; (c) $\nu = 10$.

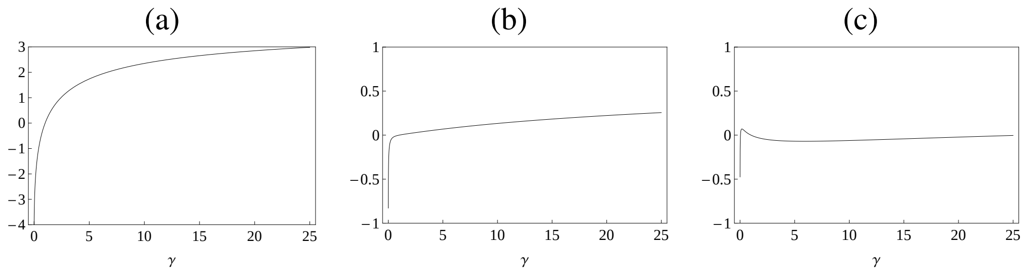

Figure 3: Standardized third central moment measure of skewness for the generalised t: (a) $\nu = 4$; (b) $\nu = 8$; (c) $\nu = 10$.

of γ for the generalised normal. This figure shows that, by varying the value of γ, the GN distribution can cover only a narrow range of values of the AG measure. For the PM and EM measures it is unclear what a reasonable range is, as they are not bounded. Thus, we will focus mostly on the AG measure in what follows. In addition, the effect of γ on AG and PM is not monotone. This clearly rules out any interpretation of γ as a skewness parameter in either the generalised normal or the generalised t models (for general ν). Interestingly, in contrast to the Student-t cases in Figure 3, skewness as measured by EM is positive for small γ, rather than negative. Visually, the density of the GN does not appear to have any substantial skewness even for very extreme values of γ. Clearly, EM is being driven mainly by the behaviour in the far tails.

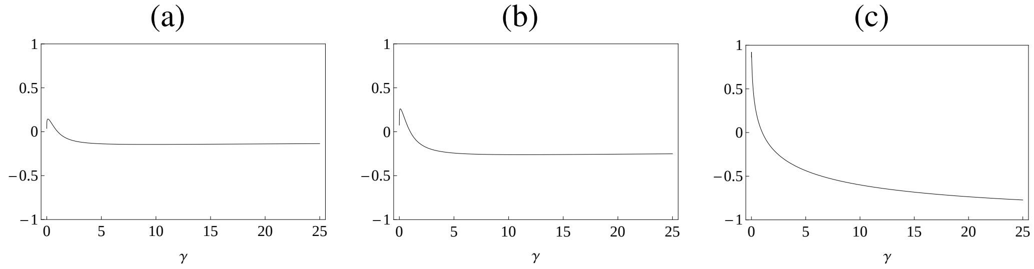

Figure 4: Measures of skewness of the generalised normal: (a) AG measure of skewness; (b) Pearson measure of skewness; (c) Standardized third central moment

3.2. Example

Here we consider a simulated data set of size 500 independently sampled from a two-piece t distribution with inverse scale factors (Fernández and Steel, 1998), which has density function

$$p(x; \mu, \sigma, \nu, \gamma) = \frac{2}{\sigma[\gamma + 1/\gamma]} \left[f(x; \mu, \sigma/\gamma, \nu) I_{(-\infty,\mu)}(x) + f(x; \mu, \sigma\gamma, \nu) I_{[\mu,\infty)}(x) \right],$$

where f is the pdf of a Student-t and we have chosen $\mu = 0$, $\sigma = 1$, $\nu = 10$ and $\gamma = 2$. The parameter $\gamma > 0$ has a clear interpretation as a skewness parameter in this model, and is linked to AG through AG$= (\gamma^2 - 1)/(\gamma^2 + 1)$. The theoretical AG measure of skewness for this example is thus 0.6.

Model	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\nu}$	$\hat{\gamma}$	$\widehat{AG}$	AIC
Two-piece t	-0.20	0.90	11.25	2.27	0.67	1725.5
Gt	-1.88	0.06	3.58	450122.2	0.28	1739.8
GN	4.30	1.80	–	0.03	0.13	1764.2

Table 1: Simulated data: maximum likelihood estimates. Values for the Akaike information criterion are shown in the last column.

The simulated data have somewhat heavy tails with significant right skewness. Neither the inverse scale factor transformation (see Fernández and Steel, 1998) nor the Marshall-Olkin transformation (see Theorem 1) affect the tail behaviour, so the degrees of freedom parameter ν has the same interpretation in terms of moment existence in both models. However, ν in the generalised t model affects both tail behaviour and the range of possible skewness and we have seen in Figure 2 that the generalised t model can not account for moderate AG skewness values with $\nu = 10$. This will affect the estimation of ν and lead to a compromise estimate which is too small for the tails, but allows for some of the skewness; this produces a poor fit in the right tail and underestimation of the AG. Table 1 shows the maximum likelihood estimates for the model from which we generated the data, as well as the generalised t and generalised normal models. The latter two models clearly perform worse than the "true" model in terms of the Akaike information criterion (AIC), although the generalised t does better than the generalised normal, which can only allow for a very small amount of AG skewness (see Figure 4). This is further illustrated in Figure 5, which presents the data histogram and the fit of the three models. Figure 6 shows the estimated

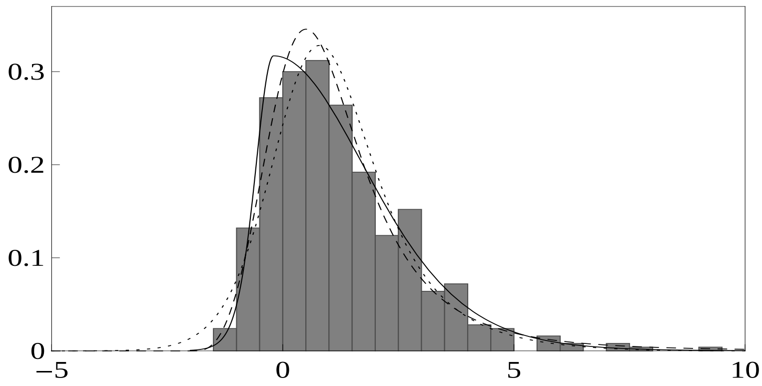

Figure 5: Simulated data: estimated two-piece-t density (continuous line); estimated generalised t density (dashed line); estimated generalised normal density (dotted line).

versus the empirical quantiles (QQ-plot), illustrating that the problem with the fit lies mainly in the right tail for the Gt and the left tail for the GN model. This example shows that the models obtained through the Marshall-Olkin transformation of normal and Student-t distributions are not flexible enough to deal with highly or moderate skewed data. Results with data simulated from a two-piece normal with the same theoretical AG skewness value similarly illustrate the lack of flexibility of the GN and Gt models.

4. Use with other distributions

Let us now investigate the use of the Marshall-Olkin transformation in the context of other classes of distributions. Figure 7 displays the AG measure as

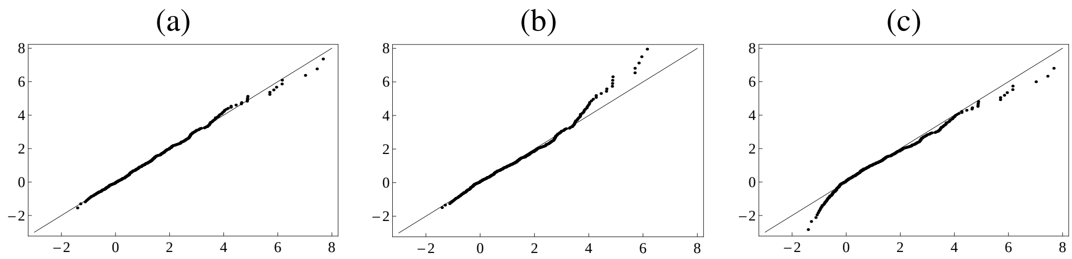

Figure 6: Simulated data: Estimated quantiles vs. empirical quantiles (a) Two-piece t; (b) generalised t; (c) generalised normal.

a function of the parameter γ in (1) for a variety of other underlying symmetric distributions.

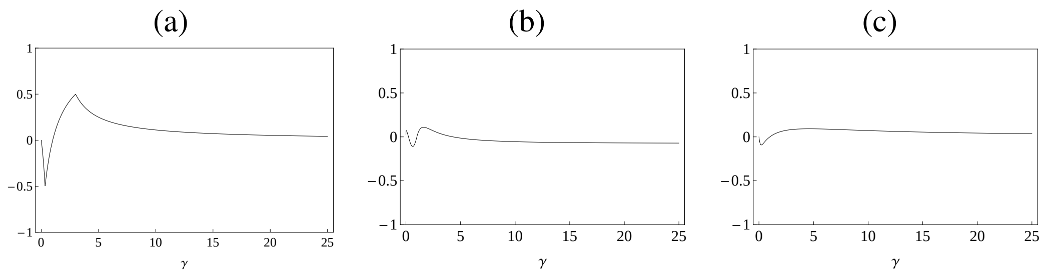

Figure 7: AG skewness measures as a function of γ for transformation of: (a) Laplace; (b) exponential power with $q = 3/2$; (c) hyperbolic secant distribution.

Among the choices for F we use various members of the exponential power class, which has pdf

$$f(x; \mu, \sigma, q) = \frac{1}{2^{1+(1/q)}\Gamma[1 + (1/q)]\sigma} \exp\left[-\left(\frac{|x-\mu|}{2\sigma}\right)^q\right],$$

for $q > 0$. Within this exponential power class, the Marshall-Olkin transformation produces bimodal distributions for $q < 1$ and γ sufficiently far from one, and we do not consider these distributions of practical interest for modelling. For the Laplace, which corresponds to $q = 1$, there is a single mode, which remains at zero whenever $1/3 < \gamma < 3$ and shifts to $\ln[(\gamma - 1)/2]$ for $\gamma > 3$ and to $\ln[2\gamma/(1 - \gamma)]$ for $\gamma < 1/3$. From Figure 7 we deduce that γ does not operate as a skewness parameter for the Laplace or the case with $q = 3/2$. Other distributions used for F are the logistic and the hyperbolic secant (Johnson et al., 1995) distributions. For the transformed logistic distribution, the AG measure is exactly zero for any value of γ. In fact, the resulting distribution is symmetric around the mode, given by $\ln(\gamma)$. The hyperbolic secant distribution is another example where we can clearly not interpret γ as a skewness parameter, as shown in Figure 7(c). Finally, we consider the symmetric sinh-arcsinh distribution of Jones and Pewsey (2009), which is obtained by setting their skewness parameter ϵ to zero. This distribution contains an additional parameter δ which controls the tail weight. Values of $\delta < 1$ indicate heavier tails than the normal. For small values of δ the Marshall-Olkin transformation does manage to generate substantial amounts of skewness, but for $\delta < 0.5$ the transformed density is bimodal for certain values of γ. For $\delta \geq 0.5$ the range of possible AG skewness is already quite limited and the latter is not a monotone function of γ (see Figure 8). Thus, none of the distributions tried in this section leads to a practically useful class of skewed distributions by using the Marshall-Olkin transformation.

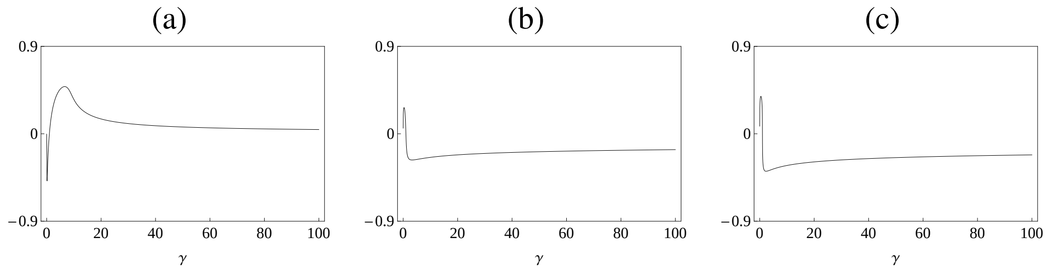

Figure 8: AG skewness measures as a function of γ for transformation of symmetric sinh-arcsinh distribution with: (a) $\delta = 0.5$; (b) $\delta = 1.5$; (c) $\delta = 4$.

5. Intuitive explanation

Let us try to understand the effect of the Marshall-Olkin transformation on the AG measure of skewness. There are two effects going on, which can cancel each other out (and they do so exactly for the logistic). Firstly, from (1) it is immediate that $G(x; \theta, \gamma)$ is a decreasing function of γ for fixed (x, θ). As a consequence, if the mode would not be affected by the transformation, the AG measure would be increasing with γ. This effect is illustrated by the transformed Laplace where the mode stays at zero for $1/3 < \gamma < 3$, so we see in Figure 7(a) that AG increases with γ within this range. Secondly, however, there is the effect of a possible shift of mode. The mode is the solution of

$$f'(x;\theta)[F(x;\theta) + \gamma(1 - F(x;\theta))] = 2(1 - \gamma)f^2(x;\theta),$$

where $f'(x;\theta)$ is the derivative with respect to x. This obviously leads to the mode of f for $\gamma = 1$ and for the Logistic leads to a mode equal to $\ln(\gamma)$. If the mode (as in the latter case) increases with γ, then this second effect will make AG decrease with γ. This is illustrated again by the Laplace in Figure 7(a), where for γ further from one the mode shifts away from zero which quickly counteracts the first effect, making the AG value a decreasing function of γ. As γ tends to very large or very small values, the AG value tends to zero.

Generally, the behaviour of the AG skewness measure as a function of γ depends on how these two effects interact. Changes in the relative strength of these two counteracting effects also explain the lack of monotonicity we have observed for most cases.

Another way to view the way the transformation works is through the ratio $K(x, \theta, \gamma)$ between $g(x;\theta, \gamma)$ in (2) and the symmetric $f(x;\theta)$. Viewed as a function of x for given (θ, γ), this ratio is always in between $1/\gamma$ and γ (see also Theorem 1), but what matters most for the skewness properties of the transformation is what happens around the mode of $f(x;\theta)$. If $F(x;\theta)$ increases very slowly with x in this region, the ratio $K(x, \theta, \gamma)$ will also change slowly (whenever $\gamma \neq 1$) and the Marshall-Olkin transformation will be able to accommodate a sensible amount of skewness. If $F(x;\theta)$ is more sharply increasing, $K(x, \theta, \gamma)$ will start to behave like a step function, with the main consequence being a shift in the mode, but the distributional shape will hardly be affected. Thus, we can expect that the Marshall-Olkin transformation can only be interpreted as a skewing mechanism if it transforms extremely leptokurtic distributions (with a very small amount of mass around the mode). A complication is that for some distributions the transformation can lead to bimodality, which seriously compromises the appeal for modelling. The only example we encountered where γ can be interpreted as a skewness parameter and which avoids bimodality is the Gt with $\nu \leq 1$ degrees of freedom, explored in detail in Section 3.

6. Conclusions

The use of the Marshall-Olkin transformation as a general mechanism for inducing skewness in unimodal symmetric densities can not be recommended. It can only accommodate substantial skewness when applied to very leptokurtic distributions and can easily lead to problems of bimodality. The only case we found where it can be used in practice is when applied to a Student-t distributions with Cauchy or heavier tails. Thus, we do not recommend its use in any other situation, including the generalised normal of García et al. (2010) or the equivalent tilted normal of Maiti and Dey (2012). The latter case also clearly illustrates the perils of the use of the common skewness measure EM, based on the standardized third central moment. EM can be seriously misleading in practice as it can be totally dominated by the behaviour in the far tails. In addition, it is not always defined and hard to interpret as it is not bounded and is not linked to a straightforward interpretation in terms of relative mass allocation. We recommend to use and compare a range of skewness measures. The mass-based skewness measure AG is an appealing option which we find quite intuitive and interpretable.

Acknowledgements:

We thank two reviewers and an Associate Editor for helpful comments. F. Javier Rubio gratefully acknowledges research support from Conacyt (Mexico) and Mark Steel benefitted from the kind hospitality of the Universidad Carlos III de Madrid.

References

Arnold, B. C. and Groeneveld, R. A. (1995). Measuring skewness with respect to the mode. *The American Statistician* 49: 34–38.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12: 171-178.

Edgeworth, F. Y. (1904). The law of error. *Transactions of the Cambridge Philosophical Society* 20: 36–65 and 113–141.

Fernández, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* 93: 359–371.

Ferreira, J.T.A.S. and Steel, M. F. J. (2006). A constructive representation of univariate skewed distributions. *Journal of the American Statistical Association* 100: 823–829.

García, V.J., Gómez-Déniz, E., Vázquez-Polo, F.J. (2010). A new skew generalization of the Normal distribution: Properties and applications. *Computational Statistics and Data Analysis* 54: 2021–2034.

George, D. and George, S. (2011). Marshall-Olkin Esscher transformed Laplace distribution and processes. *Brazilian Journal of Probability and Statistics* forthcoming.

Groeneveld, R. A. (1991). An influence function approach to describing the skewness of a distribution. *The American Statistician* 45: 97–102.

Groeneveld, R. A. and Meeden, G. (1984). Measuring skewness and kurtosis. *The Statistician*, 33: 391-399.

Groeneveld, R. A. and Meeden, G. (2009). An improved skewness measure. *Metron*, 67: 325–337.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Vol. 2 (2nd Ed. ed.). John Wiley & Sons, Inc., New York.

Jones, M. C. (2004). Families of distributions arising from distributions of order statistics. *Test* 13: 1–43.

Jones, M. C. and Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika* 96: 761–780.

Maiti, S. S. and Dey, M. (2012). Tilted normal distribution and its survival properties. *Journal of Data Science* 10: 225–240.

Marshall, A. W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika* 84: 641–652.

Pearson, K. (1895). Contributions to the mathematical theory of evolution, II: skew variation in homogeneous material. *Transactions of the Royal Philosophical Society*, Ser. A 186: 343-414.